



Processing User Guide

October 7, 2026 | Version 12.1.xxx.x

For the most recent version of this document, visit our [documentation website](#).



Processing User Guide

October 7, 2026

For the most recent version of this document, visit our [documentation website](#).

Table of Contents

1 Processing	12
1.1 Basic processing workflow	13
1.2 Processing permissions	13
1.2.1 Viewing granular permissions	15
2 Supported file types for processing	18
2.1 Supported and unsupported file types for processing	18
2.2 Categories	18
2.3 Resources	32
3 Short message conversion for Cellebrite	32
3.1 Recommended workflow	33
3.2 Processing profile: import options	33
3.3 Generated RSMF fields	34
3.3.1 Device owner (custodian) labeling	34
3.3.2 Email data extracted from Cellebrite	35
3.3.3 Output format by data type	35
3.4 Deduplication mechanism in Relativity Processing (UFDR-specific)	37
3.5 Troubleshooting	37
3.6 Known issues and limitations	39
3.7 Resources	40
4 Short message conversion for Google Chat	40
4.1 Required file structure	40
4.2 Processing Google Chat conversations	41
4.3 Generated RSMF fields	41
4.4 Deduplication considerations for Google Chat	42
4.5 Troubleshooting	42
4.6 Known issues and limitations	43
4.7 Resources	44
5 Short message conversion for Slack	44
5.1 Required file structure	44
5.1.1 Understanding JSON files for Slack data exports	45
5.2 Generated RSMF fields	45
5.3 Processing organization-level exports	46

5.4 Deduplication considerations for Slack	47
5.5 Troubleshooting	47
5.6 Known issues and limitations	48
5.7 Resources	49
6 Short message conversion for Microsoft Teams	49
6.1 Microsoft eDiscovery (Standard) PST exports	49
6.2 Microsoft eDiscovery (Premium) HTML exports	50
6.3 Supported input format	50
6.4 Generated RSMF fields	50
6.5 Metadata used to group Teams conversations	51
6.6 Time zone for slice boundaries	51
6.7 Participant metadata	51
6.7.1 Participants are scoped to the slice, not the channel	51
6.8 Message counts	52
6.9 Deduplication considerations for Teams	52
6.10 Troubleshooting	52
6.11 Known issues and limitations	53
6.12 Resources	54
7 Microsoft	54
7.1 Outlook message item (.msg) to MIME encapsulation (.mht) conversion	55
7.1.1 Email image extraction support	55
7.1.2 Microsoft Office child extraction support	56
7.2 Excel files	57
8 Google Workspace	57
8.1 Overview	58
8.2 Required file structure	58
8.2.1 Collect	58
8.2.2 Google Vault	59
8.3 Google Workspace discovery	60
8.3.1 Sidecar file not found	61
8.4 Sidecar metadata in publishing	62
8.4.1 Google Drive sidecar fields	62
8.4.2 Gmail sidecar fields	64
8.4.3 Google Drive links sidecar fields	65

8.4.4 Google Classic Hangouts metadata fields	65
8.5 Other considerations	66
8.6 Resources	67
9 Other software support	67
9.1 Multi-part forensic file considerations	67
9.2 Password-protected .rar files	67
9.2.1 Lotus Notes	68
9.3 Multi-part container considerations	70
9.4 Calendar file, vCard file considerations	71
10 Supported short message sources for RSMF conversion	74
10.1 RSMF overview	74
10.2 Sources that can be converted to RSMF	74
10.3 Sources supported through file-based Processing	74
10.4 Relativity Collect sources	75
10.5 Sources not supported	75
10.5.1 Microsoft Teams — HTML exports (including Microsoft Purview)	75
10.5.2 Microsoft Teams — MSG files	75
10.5.3 Slack data embedded in a PST container	75
10.5.4 WhatsApp standalone exports	75
10.5.5 Facebook account downloads	75
10.5.6 Non-Cellebrite mobile forensic tools	75
10.5.7 AI conversation tool exports	75
10.6 Unlisted sources	76
11 RSMF slicing configuration reference	76
11.1 RSMF slicing overview	76
11.2 Where slicing happens	76
11.2.1 Slicing at the source	77
11.2.2 Slicing during short message conversion	77
11.2.3 Splitting to meet RSMF file limits	77
11.2.4 Slicing after publish	77
11.3 The Slice by processing profile setting	77
11.3.1 Supported Slice by intervals	77
11.3.2 Time zone handling	78
11.3.3 Orphaned messages	79

11.4 Changing slice boundaries after processing	79
11.4.1 Reprocess with a different Slice by value	79
11.4.2 Regenerate converted data after publish	79
11.4.3 Slicing published documents	79
11.4.4 Slicing one document at a time	79
11.4.5 Troubleshooting slice errors	80
11.5 Resources	80
12 Working with RSMF files built outside Relativity	80
12.1 How Relativity renders an externally built RSMF (rendering contract)	80
12.2 Validating an externally built RSMF file	81
12.2.1 Running the validator	81
12.2.2 Reading attachment and participant errors	81
12.2.3 Saving an error report	82
12.3 Attachment linking requirements	82
12.4 Supported elements require explicit manifest population	83
12.5 Remediation options	83
12.5.1 Option 1: Re-route the source data through Relativity's short message conversion	83
12.5.2 Option 2: Correct and rebuild the RSMF	83
13 Password bank	84
13.1 Password bank in processing workflow	84
13.2 Password Bank in imaging workflow	85
13.3 Creating or deleting a Password Bank entry	85
13.3.1 Fields	86
13.3.2 Example password	88
13.4 Validations and exceptions	88
13.5 Viewing audits	88
14 Mapping processing fields	89
14.1 Mapping fields	89
14.1.1 Relativity system field considerations	90
14.1.2 Field mapping validations	91
14.2 System-mapped fields	91
14.3 Optional fields	93
14.4 Email Store Name field	118
14.5 Virtual path	119

14.6 Processing folder path	120
14.7 Email folder path	121
14.8 Source path	121
14.9 Message ID considerations	121
14.10 Comments considerations	122
14.11 De-duped custodian and path considerations	122
15 Processing profiles	123
15.1 Creating or editing a processing profile	123
15.1.1 Fields	124
15.1.2 Inventory / Discover settings	132
15.1.3 Extraction settings	135
15.1.4 Short message conversion settings	141
15.1.5 Deduplication settings	145
15.1.6 Publish settings	146
15.1.7 Other considerations	148
16 Deduplication considerations	153
16.1 Global deduplication	154
16.2 Custodial deduplication	155
16.3 No deduplication	156
16.4 Global deduplication with attachments	157
16.5 Global deduplication with document-level errors	158
16.6 Technical notes for deduplication	159
16.6.1 Fields used for comparing files in deduplication	159
16.6.2 Calculating MD5/SHA1/SHA256 hashes	160
16.6.3 Calculating deduplication hashes for emails	160
16.6.4 Calculating the Relativity deduplication hash	161
16.6.5 Deduplication for short message formats	162
17 Processing sets	165
17.1 Processing sets default view	165
17.2 Creating a processing set	167
17.3 Processing set fields	167
17.4 Adding a data source	169
17.5 Data source fields	171
17.5.1 Data	171

17.5.2 Custodian Options	174
17.5.3 Advanced Options	175
17.5.4 Order considerations	176
17.5.5 Edit considerations for data sources	176
17.5.6 Processing data source view	176
17.5.7 Job Errors view	178
17.6 Processing Data Sources tab	179
17.7 Deleting a processing set	182
17.7.1 Permissions	183
17.8 Avoiding data loss across sets	186
17.9 Copying natives during processing	186
18 Inventory overview	190
18.1 Inventory process	191
18.2 Monitoring inventory status	192
18.2.1 Inventory job details	192
18.3 Inventory filters	193
18.4 Discovering files from Inventory	195
19 Inventory processing	195
19.1 Running inventory	195
19.1.1 Canceling inventory	196
19.2 Filtering files	197
19.2.1 Applying a Date range filter	200
19.2.2 Applying a File Size filter	202
19.2.3 Applying a deNIST filter	203
19.2.4 Applying a Location filter	204
19.2.5 Applying a File Type filter	204
19.2.6 Applying a Sender Domain filter	205
19.3 Removing filters	208
19.4 Inventory errors	208
19.4.1 Inventory error scenarios	209
19.5 Re-inventory	209
20 Discovering files	210
20.1 Running file discovery	211
20.1.1 Discovery process	213

20.1.2 Container extraction	214
20.2 Special considerations - OCR and text extraction	214
20.3 Monitoring discovery status	215
20.3.1 Discover job details	216
20.4 Canceling discovery	217
21 Files tab	219
21.1 Views on the Files tab	220
21.1.1 All Files view	220
21.1.2 All Files view (MDS)	222
21.1.3 Deleted Documents view	225
21.1.4 Deleted files view	226
21.1.5 Current Files with Exceptions	227
21.1.6 Current files with exceptions view	229
21.1.7 All Files with Exceptions	230
21.1.8 All files with exceptions view	232
21.2 Details modal	233
21.3 Retrying delete exceptions	235
21.4 Republishing files from the Files tab	236
21.4.1 Republishing files where job errors are present	239
21.4.2 Common use cases for using the Republish mass action	239
21.5 Retrying exceptions	239
21.6 Saved filters	241
21.7 Saved filters (MDS)	245
21.7.1 Right-click options	248
21.8 Single Download / Replace	250
21.9 Mass download and replacement of errored files	252
21.9.1 Mass download of errored files	252
21.9.2 Mass replacement of errored files	254
22 Publishing files	256
22.1 Running file publish	257
22.1.1 Publish process	261
22.2 Monitoring publish status	262
22.2.1 Publish job details	262
22.3 Canceling publishing	263

22.4 Republishing a processing set	265
22.5 Retrying exceptions after publish	266
23 Post-publish delete	267
23.1 Overview	267
23.2 Post-publish delete and duplicate documents	267
23.3 Running post-publish delete	267
23.4 Interacting with deleted documents	269
23.5 Retrying delete errors	270
24 Exception and error overview	270
24.1 Types of processing exceptions	271
24.2 Exception levels	271
24.3 Processing phases where exceptions may occur	271
24.4 Processing exception tabs	272
24.4.1 Files tab	272
24.4.2 Job Errors tab	277
24.5 Viewing exceptions from a processing set	280
24.6 Retrying exceptions after making changes to the processing profile	281
24.7 Required security permissions	282
25 Exception and error resolution	282
25.1 Types of processing exceptions	283
25.2 Exception levels	283
25.3 Processing phases where exceptions may occur	284
25.4 File exceptions	284
25.4.1 File exception actions	284
25.5 Resolving file exceptions	287
25.5.1 Retrying an exception from the processing set	287
25.5.2 Retry file exceptions mass action	289
25.5.3 Replacing corrupted files	289
25.5.4 Resolving a password-protected file exception	292
25.5.5 Publish documents with field length exceptions	292
25.5.6 Email file processing timeout	293
25.5.7 Troubleshooting corrupted .msg files in a .pst container	294
25.6 Job errors	305
25.6.1 Job error workflow	305

25.6.2 Resolving job errors	306
26 Processing duplication workflow	308
26.1 Downloading and installing the solution to your instance	309
26.1.1 Supported versions	309
26.1.2 Components	309
26.1.3 Considerations	309
26.2 Deploying and configuring the solution	310
26.3 Adding the solution to a workspace	310
26.4 Before you run the scripts	311
26.4.1 Fields	311
26.4.2 Saved searches	314
26.4.3 Other considerations	315
26.5 Running the scripts	321
26.5.1 All Custodians script	321
26.5.2 All Source Locations script	325
26.5.3 Update Duplicate Status script	329
27 Reports	335
27.1 Generating a processing report	335
27.2 Data Migration	336
27.3 Master Document Replacement Summary	337
27.4 Discovery Exclusion Results	338
27.5 Discovered Files by Custodian	338
27.6 Discovered Files by File Type	339
27.7 Document Exception	340
27.8 File Size Summary	341
27.9 Inventory Details	341
27.10 Inventory Details by Custodian	342
27.11 Inventory Exclusion Results	343
27.12 Inventory Exclusion Results by Custodian	343
27.13 Inventory Summary	344
27.14 Job Exception	345
27.15 Text Extraction	345
28 Processing administration	346
28.1 Monitoring active jobs	346

28.2 Active jobs mass operations	348
28.3 Post-publish delete job performance	350
28.4 Permissions	350
29 Processing to Data Grid	353
29.1 Enabling processing to Data Grid	353
30 Factors that affect processing throughput	355
30.1 Calculating processing discover throughput	355
30.2 Discover job performance	356
30.2.1 Factors related to the processing discover job	356
30.2.2 Processing profile settings that may impact discover job performance	357
30.3 Publishing job performance	357
30.3.1 Factors related to the processing publish job	357
30.3.2 Processing profile settings that may impact publish job performance	358
31 Frequently asked questions	358
31.1 Calendar appointments	358
31.2 Date and time	358
31.3 DeNIST	359
31.4 Discover	359
31.5 Errors	359
31.6 Media processing	359
31.7 Other	360
31.8 Password-protected files	360

1 Processing

Use Relativity's processing feature to ingest raw data directly into your workspace for eventual search, review, and production without the need for an external tool. You can use the various processing objects to create custom processing jobs that handle a wide array of information.

Some of the primary goals of processing are to:

- Discern, at an item level, exactly what data is found in a certain source.
- Record all item-level metadata as it existed prior to processing.
- Enable defensible reduction of data by selecting only items that are appropriate to move forward to review.

Note: There are no specific security requirements, but if a user needs to be restricted from running processing, then permissions need to be revoked to all processing objects.

See these related sections:

- [Supported file types](#)
- [Password bank](#)
- [Processing profiles](#)
- [Processing sets](#)
- [Inventory](#)
- [Discovering files](#)
- [Publishing files](#)
- [Processing error overview](#)
- [Processing error resolution](#)
- [Reports](#)
- [Deduplication considerations](#)

For a list of changes made to processing per monthly product update, see the Relativity RelativityaiR Release Notes.

1.1 Basic processing workflow

The following steps depict a typical processing workflow that uses all available processing objects and phases. Note that each user's workflow may vary. You may not be required to follow all of these steps for every processing job you run.

1. [Create a Processing Set](#)
 - Entities can be created on the fly, in advance, or automatically through imports or connections to HR systems.
 - Processing profiles carry over from template workspaces.
 - If necessary, create new Password Bank Entries with passwords for any password-protected files to be processed.
2. (Optional) [Inventory files](#)
 - Inventoried files can be filtered down based on several metadata attributes prior to publish.
 - Reports can be run to understand and communicate the files culled from publish.
3. [Discover files](#) and [Publish](#)
 - Post publish – view, ignore, or retry any errors that occurred during any phase of the processing job. If needed, republish the files.

1.2 Processing permissions

Relativity feature permissions are assigned through groups, which are then applied to workspaces, object types, tabs, and individual items. Administrators have precise control over feature actions accessible at multiple levels. To learn more, see Permissions. This section provides an overview of navigating Processing feature permissions.

Permissions

The screenshot shows the 'Permissions' interface. On the left is a sidebar with navigation links: 'Workspace Security', 'Groups', 'Features' (marked with a 'NEW' badge), 'Object Security', 'Tab Visibility', and 'Other Settings'. The 'Features' tab is selected. The main content area has a 'Features' header. Below it is a 'Selections' section with 'Selected Workspace' set to 'Documentation' and 'Selected Group' set to 'Documentation Team' (highlighted with a red box). Below this is another 'Features' section with a search bar containing 'processing'. It shows 'Showing 1 of 36 features'. Under the 'Ingestion' category, the 'Processing' feature is checked (checkbox highlighted with a red box) and assigned to the 'Administrator' role. A list of actions for 'Processing' is shown: 'View Processing', 'View Processing Errors', 'Manage Processing', 'Manage Processing Errors', 'Manage Processing Reports', 'Manage Processing Profiles', and 'Delete Processing'. A 'See details' link with a document icon is highlighted with a red box. At the bottom right are three buttons: 'Review and Save' (blue), 'Revert All' (white), and 'Close' (white).

To access FOIA feature permissions:

1. Use the search bar to navigate to **Workspace Details**.
2. Click **Manage Workspace Permissions** (in the right-side column).
3. Select the **Features** tab.
4. Use the **Select Group** drop-down menu to select the group for which you want to view permissions. If you do not know which group to select, contact your Relativity administrator. Groups and group assignments are managed by an organization's Relativity administrator.
5. Scroll down and check the box next to **Processing**. You can also use the search bar to jump to the Processing settings.
6. Click **See details**.

A modal appears listing the actions associated with Processing. One item to note is the **Preset** drop-down menu. When you select a preset, Relativity selects the actions defined for the preset. You can then add or clear actions as needed. For more information, see Presets.

Processing

✕

Preset

←

Administrator

▼

Description

Full access to Processing, including delete and Processing Profile management.

Actions

☒ Select all

☒ View **Processing**

☒ View **Processing** Errors

☒ Manage **Processing**

☒ Manage **Processing** Errors

☒ Manage **Processing** Reports

☒ Manage **Processing** Profiles

☒ Delete **Processing**

>> [Granular permissions mapping](#)

Review and Save

Revert All

Close

To edit actions:

1. Check or clear actions as needed.
2. Click **Review and Save**. This button is only enabled if you changed action settings.
3. Click **Save**. This button is only enabled if you changed action settings.
4. Click **Close** to exit the screen if you did not change any settings.

1.2.1 Viewing granular permissions

You can view granular permissions for some actions. Granular permissions are the specific, individual access rights bundled within an action. This helps you see exactly what an action allows before you enable it. Not all actions support this level of detail.

Preset

Administrator



Description

Full access to Processing, including delete and Processing Profile management.

Actions

☒ Select all☒ View **Processing**☒ View **Processing** Errors☒ Manage **Processing**☒ Manage **Processing** Errors☒ Manage **Processing** Reports☒ Manage **Processing** Profiles☒ Delete **Processing**[» Granular permissions mapping](#)

Review and Save

Revert All

Close

To view granular permissions:

1. Click **Granular permissions mapping**. The reference modal opens.

Granular permissions mapping

Processing - read-only reference

> View Processing

< View Processing Errors

Object Security Items

Job Error

View 

Processing Error

View 

Tab Visibility Items

Job Errors 

> Manage Processing

> Manage Processing Errors

> Manage Processing Reports

> Manage Processing Profiles

> Delete Processing

Icon Guide

 Enabled permission

 Disabled permission

 Permission used by multiple Features

 Permission not found by name

Review and Save

Revert All

Close

2. Expand the sections for each action to view specific functionality. For example, the **View Processing Errors** action allows you to view job and processing errors, and determine if the Job Errors tab is visible. The **Icon Guide** at the bottom of the modal provides a definition for each icon.
3. Click **Revert All** to reset permissions to the default. This button is only enabled if you changed action settings on the Actions details screen.

4. Click **Review and Save** to continue. This button is only enabled if you changed action settings on the Actions details screen.
5. Click **Save** to commit the permission changes.
6. Click **Close** if you did not change any action selections.

2 Supported file types for processing

Relativity supports file types where the metadata and text can be extracted and processed for further review and publication. If the metadata or text cannot be extracted, the file type is not supported.

Note: This documentation contains references to third-party software, or technologies. While efforts are made to keep third-party references updated, the images, documentation, or guidance in this topic may not accurately represent the current behavior or user interfaces of the third-party software. For more considerations regarding third-party software, such as copyright and ownership, see [Terms of Use](#).

2.1 Supported and unsupported file types for processing

The table below lists supported and unsupported file types for processing. If the **Supported** column is empty, the file type is not supported.

Important! Additional notes about file type support

Notes:

- Renaming a file extension has little effect on how Relativity identifies the file type. When processing a file type, Relativity looks at the actual file properties, such as digital signature, regardless of the named extension. Relativity only uses the named extension as a tie-breaker if the actual file properties indicate multiple extensions.
 - Relativity may not support files created by unofficial third-party applications, such as a .pdf file created by an unknown third-party application.
 - Data pulled from supported versus unsupported file types: Relativity only pulls limited metadata from unsupported file types. Data pulled from supported file types includes metadata, text, and embedded items.
 - Supported file types differ slightly in processing, imaging, and the viewer. See *Supported file types for imaging* and *Viewer-supported file types* for details specific to those features.
 - If you use the Native text extraction method on the profile, Processing does not handle pre-2008 Microsoft Office files that have the Protected view enabled. You must use the Relativity text extraction method to process these files.
 - Relativity currently does not support the extraction of embedded images or objects from Visio, Project, or OpenOffice files. In addition, Relativity never extracts any embedded objects or images that were added to any files as links. For a detailed list of the Office file extensions from which Relativity does and does not extract embedded objects and images, see [Microsoft Office child extraction support](#).
-

2.2 Categories

[Container files](#) | [Database, spreadsheets](#) | [Document, text, word processors](#) | [Email, messaging, chat](#) | [Graphics, images](#) | [Other](#)

File type	File extension	Notes	Supported	Supported for password bank
AccessData Logical Image Container files	.ad1	Relativity supports processing both single and multi-part non-encrypted .ad1 files. For encrypted .ad1 files, only single part files are supported. For multi-part .ad1 files, you must decrypt the files prior to processing. See Multi-part container considerations for more information.	✓	
Bloomberg	.xml	Relativity supports Bloomberg's email format as a processed email file type. Relativity does not support Instant Bloomberg (chat) files directly. Use Collect to convert these files to RSMF first; you can then process them as short messages.	✓	
Cabinet	.cab	Relativity does not support multi-part .cab files. Relativity does not support password protected .cab files.	✓	
Compressed archives	.7z, .zip, .tar, .gz, .bz2, .rar, .z, .cab, .alzip	For compressed container files, keep in mind the following: • Relativity supports multi-part .rar files. You do not need to combine multi-part .rar files before processing them. • Relativity does not support multi-part .zip, .tar, or .7z files. Self-extracting .rar files are also not supported. • Depending on how .rar files are encrypted, extraction may be impacted. See Password-protected .rar files for more information. • Very large container files may impact processing speed. One workaround is to extract the files from the container and process them from a new processing set. • When working with archives, there is no limit to the number of layers deep processing goes to extract data. It extracts until there is no more data to be extracted.	✓	

File type	File extension	Notes	Supported	Supported for password bank
		Inventory, however, only extracts data from first-level documents. For example, if you have a .zip file within a .zip file that contains an email with an attached Word document, Inventory only extracts up to the email. Additionally, .zip files within .zip files may impact performance. • Zip files do not store time zone information for the CreatedOn, LastModified, and LastAccessed fields. When extracting files, time stamps are only meaningful if you know the time zone in which the .zip file container was created. Relativity extracts file metadata and updates the CreatedOn and LastModified fields if available. Otherwise, CreatedOn defaults to 1/1/1900, and LastModified reflects the worker's local time zone. The LastModified and LastAccessed fields will usually match. • Zip files do not store character encoding information, which can be problematic when such characters use non-Unicode encoding. In this case, you may see garbled characters in the names of extracted files and folders, especially when the originating locale differs from the processing region configured for the Relativity instance. For example, given a .zip file originating in a Russian locale, with filenames encoded using Windows-1251 and processed on an instance set to US/English, the .zip's extracted file and folder names may be unrecognizable. To maximize the potential for successful extraction, confirm your instance is configured with a processing region		

File type	File extension	Notes	Supported	Supported for password bank
		that is appropriate to your geographical location. Contact your Customer Success Manager for questions regarding this and other instance settings.		
	.rar	Supported for Password Bank		✓
Compressed files	.7z, .alzip, .zip, .z, .bz2, .gz	Container file types that are supported by Relativity for Password Bank in Inventory.		✓
iCalendar	.ics	For Outlook meeting invites, the email that is sent with the meeting invite (the .msg file) will have a sent date that reflects when the sender sent out the meeting request. The resulting calendar file that is then added to the user's Outlook calendar (the .ics file) will not include a sent date, as the date does not apply to the calendar file itself.	✓	
Lotus Notes Database	.nsf	Relativity supports Lotus Notes version 10 and earlier. See Lotus Notes for more information.	✓	✓
MBOX Email Store	.mbox	.mbox is a standard format, in which case it does not matter whether you're using a Mac folder format or a Unix file format.	✓	
Microsoft Outlook for Mac	.olm			
Outlook Express Mail Folder	.dbx		✓	
Outlook Mail Folder	.pst	- Relativity assigns duplicate hash values to calendar invites, as it does with email messages and other documents. - Relativity supports Teams data conversion to RSMF out of PST. See Short message conversion for Microsoft Teams for more information.	✓	✓
Outlook Offline Storage	.ost		✓	

File type	File extension	Notes	Supported	Supported for password bank
PDF Portfolio	.pdf			✓
RAR	.rar			✓
File type	File extension	Notes	Supported	Pswd bank
Apple iWorks Suite Database, spreadsheets	.numbers	Numbers. Support includes identification and extraction of modern iWork formats (2013+), including attached files and password bank-based decryption.	✓	✓
Database files	.dbf	• Relativity only supports .dbf 3 and .dbf 4 files. • Relativity does not support the following database formats: • VisualFoxPro • VisualFoxPro autoincrement enabled • Relativity uses Microsoft Excel to extract text from .dbf file types. For details on .dbf file type handling, see Excel files.	✓	
Microsoft Access				
Microsoft Excel	.xlsx, .xlsm, .xlsb, .xlam, .xltx, .xltm, .xls, .xlt, .xla, .xlm, .xlw, .uxdc	Excel version 2.0 through the current product version is supported. See Excel files .	✓	✓ *

File type	File extension	Notes	Supported	Supported for password bank
		Notes: • If you save a Powerpoint or Excel document in pre-2007 format, .ppt or .xls files for example, and the document is read-only, we use the default known password to decrypt the document, regardless of whether or not the password exists in the Password Bank. • If you use the Native text extraction method on the profile, Processing does not handle pre-2008 Microsoft Office files that have the Protected view enabled. You must use the Relativity text extraction method to process these files. *Except .drm file types		
Unsupported spreadsheets		• Enable Versions 3.0, 4.0, and 4.5 • First Choice Versions through 3.0 • Framework Version 3.0 Lotus 1-2-3 (DOS and Windows) Versions through 5.0 • Lotus 1-2-3 (OS/2) Versions through 2.0 • Lotus 1-2-3 Charts (DOS and Windows) Versions through 5.0 • Lotus 1-2-3 for SmartSuite Versions 97 and Millennium 9.6 • Lotus Symphony Versions 1.0, 1.1, and 2.0 • Microsoft MultiPlan Version 4.0 • Mosaic Twin Version 2.5 • Novell Perfect Works Version 2.0 • PFS: Professional Plan Version 1.0 • Quattro Pro (DOS) Versions through 5.0		

File type	File extension	Notes	Supported	Supported for password bank
		- Quattro Pro (Windows) Versions through 12.0, X3 - SmartWare II Version 1.02 - SuperCalc 5 Version 4.0 - VP Planner 3D Version 1.0		
File type	File extension	Notes	Supported	Pswd bank
Adobe files Document, text, word processors	pdf, fm, ps, eps	Processing supports extraction for XFA and web form PDF files. Note however, some downstream workflows may require specific workarounds: - Native redactions are not currently supported. Workaround: Image the file first, then apply redactions. - The download PDF action, using the Natives file type, is not currently supported. Workaround: After selecting the PDF mass action, use either the Original Images or Produced Images file type or export the files as .pdfs using Import/Export. - Portfolio PDFs, also known as package PDFs, are essentially containers of other files with a PDF wrapper. Processing supports extraction of all files within these packages, and further associates certain parent metadata to the extracted children. The portfolio PDF is suppressed from publish, similar to ZIP containers. - Processing provides partial support for PDF files secured with Microsoft Information Protection (MIP) or Adobe Experience Manager (AEM) document security.	✓	✓

File type	File extension	Notes	Supported	Supported for password bank
		These files require server-side, web-based authentication that Relativity is not designed to support. As a result, processing is limited to file identification and the extraction of certain metadata. During Discovery, the restricted PDF is presented as a container, with a secondary PDF extracted that represents the encrypted payload. In the case of MIP security, the extracted PDF is named <i>MicrosoftIRMServices Protected PDF.pdf</i>. In the case of AEM security, the extracted file retains the name of the originally authored PDF. Because Relativity cannot overcome this type of document security, the extracted PDF is flagged with an exception status of <i>Not Resolved</i> and is published in place of the original PDF. In these cases, Relativity recommends collecting an unrestricted version of the PDF and reprocessing it using either Single Download / Replace or Mass download and replacement of errored files.		
Apple iWorks Suite	.pages	Pages. Support includes identification and extraction of modern iWork formats (2013+), including attached files and password bank-based decryption.	✓	✓
DOS Word Processors		Processing does not support files created with the following programs and versions: • DEC WPS Plus (.dx) Through 4.0 • DEC WPS Plus (.wpl) Through 4.1 • DisplayWrite 2 and 3 (.txt) All versions • DisplayWrite 4 and 5 Through Release 2.0 • Enable 3.0, 4.0, and 4.5		

File type	File extension	Notes	Supported	Supported for password bank
		• First Choice Through 3.0 • Framework 3.0 • IBM Writing Assistant 1.01 • Lotus Manuscript Version 2.0 • MASS11 Versions through 8.0 • MultiMate Versions through 4.0 • Navy DIF All versions • Nota Bene Version 3.0 • Office Writer Versions 4.0 through 6.0 • PC-File Letter Versions through 5.0 • PC-File+ Letter Versions through 3.0 • PFS:Write Versions A, B, and C • Professional Write Versions through 2.1 • Q&A Version 2.0 Samna Word IV+ Versions through • Samna Word • SmartWare II Version 1.02 • Sprint Versions through 1.0 • Total Word Version 1.2 • Volkswriter 3 and 4 Versions through 1.0 • Wang PC (.iwp) Versions through 2.6 • WordMARC Plus Versions through Composer • WordStar Versions through 7.0 • WordStar 2000 Versions through 3.0 • XyWrite Versions through III Plus		
Hangul	.hwp	Hangul word processing files 1997 up to 2010 are supported.	✓	

File type	File extension	Notes	Supported	Supported for password bank
HTML	.html, .mht, .htm, .mhtml, .xhtm, .xhtml	Relativity extracts metadata and attachments from multipurpose internet mail (MIME) file formats such as .mht and .eml during processing.	✓	
JungUm Global	.gul		✓	
Mac Word Processors		MacWrite II Version 1.1		
Text files	.txt, .csv.	Note: Processing supports any text file whose bytes are ASCII or Unicode text. Files are assumed to be in UTF8 if a Unicode BOM is not found. Files not in a Unicode format with characters outside the ASCII range may experience issues with text extraction.	✓	
Windows Word Processors		Processing does not support files created with the following programs and versions: • Adobe FrameMaker (.mif) Version 6.0 • JustSystems Ichitaro Versions 5.0, 6.0, 8.0, 13.0, 2004 • JustWrite Versions through 3.0 • Legacy Versions through 1.1 • Lotus AMI/AMI Professional Versions through 3.1 • Lotus Word Pro Millenium Versions 96 through Edition 9.6, text only • Novell Perfect Works Version 2.0 • Professional Write Plus Version 1.0 • Q&A Write Version 3.0 • WordStar Version 1.0		
Word	.docx, .docm, .dotx, .dotm, .doc, .dot, .rtf	Word 2.0 thupportrough the current product version is sed, including templates.	✓	✓ *

File type	File extension	Notes	Supported	Supported for password bank
		Note: If you use the Native text extraction method on the profile, Processing does not handle pre-2008 Microsoft Office files that have the Protected view enabled. You must use the Relativity text extraction method to process these files. *Except .drm file types		
WordPerfect	.wpd, .wps		✓	
File type	File extension	Notes	Supported	Pswd bank
Apple Double		AppleDouble-encoded attachments in e-mails.	✓	
Apple Mail	.emlxpart, .partial.emlx	Note: The .emlxpart and .partial.emlx are distinct from the .emlx file extension, which is supported by processing.		
Email	.pst, .ost, .nsf, .msg, .p7m, .p7s, .ics, .vcf, .mbox, .eml, .emlx, .tnef, .dbx, Bloomberg Mail .xml	- Original electronic email data (.eml file types) are parsed and stored inside a personal storage table (.pst files.) If the email contains embedded electronic email data, the email data is also parsed and stored in the personal storage table. The processing engine reads tables, properties, and rows to construct an .msg (Outlook message item) file from a .pst file. The .msg file format supports all rich metadata inside an email in a personal storage table. The original electronic email data is not preserved. - S/MIME-encrypted and digitally-signed emails are supported. - Even though the .emlx file type is supported, the following partial .emlx file extensions are not supported: .emlxpart .partial.emlx	✓	✓ *

File type	File extension	Notes	Supported	Supported for password bank
		Note: There is a known file size limit of 200MB for Outlook file attachments. Relativity cannot process Outlook attachments that exceed the file size limit. *Except S/MIME, .p7m file types		
Private mail certificate	.pem			
Short message	.rsmf	For details about short message use, metadata, and mapping, see Relativity's short message format.	✓	
Slack workspace export		Exported .json files contained in a .zip container For more information, see Short message conversion for Slack.	✓	
File type	File extension	Notes	Supported	Pswd bank
CAD files Graphics, images	.dxf, .dwg, .3dxml, .prt, .prt, .asm, .asm, .drw, .drw, .eprt, .easm, .edrw, .eprt, .edrw, .easm	The OCR output for processed CAD files can vary significantly. Identification and extraction/OCR of SolidWorks drawing files (.slddrw, .sldasm, .sldprt) is not currently supported. Convert these files to PDF with the native application in order to yield graphics and searchable text.	✓	
High Efficiency Image Container (HEIC)	.heic	All metadata extracted from .heic files, including but not limited to Exchangeable Image File Format (EXIF) metadata where present, is available for mapping to workspace fields. These fields can be found in the field catalog under the Metadata field type and have a prefix of HEIC/.	✓	
Image files	.jpg, .jpeg, .ico, .bmp, .gif, .tiff, .tif, .jng, .koala, .lbm, .pbm, .iff, .pcd, .pcx, .pgm, .ppm, .ras, .targa, .tga, .wbmp, .psd, .cut, .xbm, .dds, .fax, .sgi, .png, .exif, .exif, .webp, .wdp	Note: Image formats such as .gif support animations with multiple frames. When an animated image uses delta encoding for intermediate frames, Relativity cannot render all frames, which reduces the overall quality of OCR during text extraction.	✓	
Vector files	.svg, .svgz, .wmf,	Relativity only supports the text within	✓	

File type	File extension	Notes	Supported	Supported for password bank
	.plt, .emf, .snp, .hpgl, .hpg, .plo, .prn, .emz, .wmz	.svg files. Any images within .svg files will not be displayed. An .svg file with only images will be completely blank.		
Visio	.vsd, .vdx, .vss, .vsx, .vst, .vsw, .vsdx, .vsdm		✓	
File type	File extension	Notes	Supported	Pswd bank
Other		Symantec Ghost		
Apple iWorks Suite	.key	Keynote. Support includes identification and extraction of modern iWork formats (2013+), including attached files and password bank-based decryption.	✓	✓
Audio/Video files	.mp4, .mov, .avi, .mkv, .webm, .mpg, .mpeg, .mxf, .flv, .wmv, .mp3, .wav, .aac, .m4a, .flac, .ogg, .opus, .wma, .ts, .3gp, .aiif	Relativity is able to identify most media formats. However, extraction of internal metadata properties, such as media duration, is not currently supported.		
EnCase	.e01, .ex01, .l01, .lx01	- Relativity supports extraction of files and folders from EnCase forensic disk images (.e01, .ex01) for the following operating systems and corresponding file systems: - Windows—NTFS, FAT, ExFAT - Mac—HFS+ - Linux (Ubuntu)—EXT2, EXT3, EXT4 - For details regarding proper staging of forensic image multi-part extents, see Multi-part forensic files. - Encrypted EnCase files are not supported. You must decrypt EnCase files prior to processing them. - When processing EnCase forensic disk images (.e01,	✓	

File type	File extension	Notes	Supported	Supported for password bank
		.ex01) or logical evidence images (.I01, .lx01) containing a user's Recycle Bin/Trash, Relativity is able to extract the deleted files contained within. File system metadata available for extraction is typically limited. Note: Deleted files removed from the user's Recycle Bin/Trash, which may exist within the unallocated space captured in an Encase physical disk image (.e01, .ex01), are not recovered during processing. You must use a forensic tool to perform unallocated space file carving, and process the resulting files separately.		
Encryption		Pretty Good Privacy (PGP)		
Google Workspace		For details, see Supported file types - Google Workspace .	✓	
iCloud backup files				
Microsoft Works				
OneNote	.one	- During ingestion, Relativity extracts embedded items from OneNote files. - All child objects are extracted as separate documents. - Text extraction takes place during discovery (and not OCR.) - The Password Bank supports OneNote files. - Relativity supports OneNote 2003+ and 2007+.	✓	
OpenOffice	.odc, .ods, .odt, .odp, .xps		✓	
PowerPoint	.pptx, .pptm, .ppsx, .ppsm, .potx, .potm, .ppt, .pps, .pot	PowerPoint 97 through the current product version is supported, including the dual-format 95/97 version.	✓	✓ *

File type	File extension	Notes	Supported	Supported for password bank
		Modern Comments are supported with Relativity Text Extraction. Note: If you save a Powerpoint or Excel document in pre-2007 format, .PPT or .XLS files for example, and the document is read-only, we use the default known password to decrypt the document, regardless of whether or not the password exists in the Password Bank. *Except .drm file types		
Project	.mpp, .mpt, .mpd, .mpx	Note: The text extracted from Project files is from the Gantt chart view and includes Task Notes.	✓	
Raw partition files		- ISO - NTFS - HFS		

2.3 Resources

- For information on supported short message sources, see Supported short message sources for RSMF conversion.
- For information on enabling short message conversion, see Short message conversion settings.
- For information on short message use, metadata, and mapping, see Relativity's short message format topic.

3 Short message conversion for Cellebrite

This topic covers information for Cellebrite conversion. To view the Relativity supported and unsupported file type tables, see [Supported file types for processing](#).

Note: Relativity supports short message conversion for Cellebrite in Relativity aiR Government.

During processing, Relativity automatically converts Cellebrite's Universal Forensic Data Report (UFDR) chat extract from mobile devices to Relativity's Short Message Format (RSMF). Chat data includes text messages (SMS, MMS), and chat applications data (WhatsApp, Facebook Messenger, WeChat, Telegram). Chat data is extracted independent of the device operating system (iOS or Android). Relativity extracts all the other data available in the UFDR and makes it available for review.

Note: Relativity supports UFDR files generated by Cellebrite tools. UFDR files generated by other tools are not supported.

3.1 Recommended workflow

(Optional) Before importing data into Relativity, we recommend opening the file in Cellebrite Reader and re-exporting the UFDR. This step ensures that the UFDR has the latest schema and version changes, preventing conflicts when processed in Relativity. Opening the UFDR in Cellebrite Reader is a recommended step, and not a required one. You can also cull and remove any unnecessary data in the UFDR using different filters with Cellebrite Reader.



1. Use Cellebrite to extract UFDR data from a mobile device.
2. (Optional) Open the file in Cellebrite Reader and remove any unnecessary data, with the exception of the three data types listed below. This ensures you have the latest UFDR version. During the generate report workflow (using Cellebrite tools), be sure to include the following data types:
 - Chats
 - Instant Messages
 - Device Info

Processing the files into Relativity:

1. Use Import/Export (or Express Transfer) to upload UFDR files into Relativity.
OR
Use ROSE to upload UFDR files. If you use ROSE, create a processing set with the uploaded data as the data source.
2. Run processing discovery. Relativity automatically converts the files to RSMF. You can monitor the progress through the processing set layout details screen.
 - The number of discovered documents equals the number of generated RSMF files plus any attachments.
 - If you encounter processing errors, see [Troubleshooting](#).
3. Publish the files and then navigate to the Documents tab to view the RSMF files.

3.2 Processing profile: import options

This table lists import options from the processing profile Cellebrite settings and the data that is included when enabled or skipped when disabled.

Import option	What it includes	What is silently skipped if disabled
SMS/MMS	Text messages and multimedia messages	All SMS and MMS records
Chat apps	WhatsApp, Facebook Messenger, and similar apps	All chat application conversations
Media files	Photos, videos, and audio files	All media content
Other	Logs, location data, and miscellaneous files	Logs and other non-message content

Note: Enable all import options unless you have a specific reason to exclude a data type. Disabling any option silently skips all matching content during processing — no error or warning is generated, and the output will be smaller than expected.

3.3 Generated RSMF fields

The following table lists the metadata fields found in RSMF files after Cellebrite conversion.

Relativity source field name	Field type	Description
Date	Date	The date of first event in conversation.
EmailFrom	Long Text	The display name and email of the sender of the first message in this RSMF file. For conversations split across multiple files, this value may differ per file.
EmailTo	Multiple Choice	A list of display names and emails of all participants in conversation,
File Name		The conversation name + date + slice.
Subject	Long Text	The name of the conversation. • For public/private channels: CHANNEL_NAME • For group chats/dms: LIST OF PARTICIPANTS
Rsmf/Application	Long Text	Specific to the chat data type. For example, SMS or WhatsApp.
Rsmf/AttachmentCount	Whole Number	The number of attachments in conversation.
Rsmf/BeginDate	Date	The date and time of first event in conversation.
Rsmf/EndDate	Date	The date and time of last event in conversation.
Rsmf/EventCollectionId	Long Text	The conversation ID that is persisted between file slices.
Rsmf/Generator	Long Text	Relativity Processing
Rsmf/MessageCount	Whole Number	The number of events in the conversation.
Rsmf/Participants	Multiple Choice Caution: Use the Long Text field type when mapping the Rsmf/Participants field. Selecting Multiple Choice can lead to hundreds of unique choices that may impact performance.	A list of display names of all participants in conversation.
Rsmf/Version	Long Text	2.0.0

3.3.1 Device owner (custodian) labeling

During conversion, Relativity identifies the device owner using the phone number (MSISDN) in the UFDR, and labels that participant in the RSMF viewer with the custodian assigned to the data source. The participant appears as

Custodian: <custodian name> <phone number>. This label is generated by Relativity during processing and is not part of the source UFDR data.

Sent : 03/12/2024 6:34:15 AM
To : [REDACTED]
Subject : WhatsApp Chat: [REDACTED]
Attachments : rsmf.zip

Custodian: B [REDACTED]; R [REDACTED] <165 [REDACTED] 2>;

Custodian: B [REDACTED], R [REDACTED] <165 [REDACTED] 2>

3.3.2 Email data extracted from Cellebrite

Relativity extracts emails from Cellebrite forensic data (the UFDR file) and converts them into Email Message (EML) files.

Fields that are converted to EML:

- Subject
- Body
- TimeStamp
- Sender
- Recipients

Fields that are not converted to EML:

- Attachments
- IsDeleted
- IsEdited
- Labels
- Platform
- ParentId
- Identifier
- Conversation metadata

3.3.3 Output format by data type

Processing UFDR files can produce multiple output formats from the same job because different data types are extracted and converted in different ways. For example, chat data may be converted to RSMF, while other content is generated in additional file formats. The table below provides a simple mapping of common UFDR data types to their corresponding output formats.

Data type	Output format	Notes
-----------	---------------	-------

Chats, SMS, MMS	RSMF	Fully supported. Includes most common communication apps like: WhatsApp, Facebook Messenger, WeChat, Telegram and more when enabled in profile.
Emails	.eml	Each email becomes a standard .eml file processed by Relativity normally.
Media files and recordings	Native files	Extracted as raw native files (images, audio, video). Relativity handles these through standard processing.
Other documents	Native files or .txt fallback	No dedicated handler. Depending on how Cellebrite includes them in the UFDR, they either come out as native files (if treated as media) or fall through to a plain .txt fallback. See note below.
Structured data (calls, contacts, calendar, locations, notes, passwords, web history, and so forth)	Excel (.xlsx)	Structured data types produce Excel files placed in an Other Data folder. These are searchable and reviewable but are not documents in the traditional sense. Data types include: • Calendar • Call Log • Cell Towers • Contacts • Cookies • Credit Cards • Devices • Installed Applications • Locations • Mobile Cards • Notes • Passwords • Searched Items • Social Media • Transfers • User Accounts • Web Bookmarks • Web History

Notes:

- If it is important to keep loose files exactly as they are, export them directly from Cellebrite Reader and process them in Relativity as a separate job using a standard (non-short-message) processing profile. This avoids the .txt fallback risk and gives you full control over text extraction settings for those files.
- Attachments on emails extracted from a UFDR are not embedded in the .eml output. If email attachments are relevant to your matter, verify whether they are accessible via a separate media or document export from the same Cellebrite extraction.

3.4 Deduplication mechanism in Relativity Processing (UFDR-specific)

Relativity applies message-level deduplication during UFDR processing, where messages with identical time stamps, sender, body, and attachments are considered duplicates. Only one is retained, and the others are excluded from the final processed count.

When referencing the Rsmf/MessageCount field, note that it reflects only the messages retained after deduplication. Additionally, Cellebrite Reader may extract and label extra messages as "Genie" messages, which are displayed separately but included in Relativity's total message count. To learn more about Cellebrite's App Genie tool, see the [App Genie](#) blog on the Cellebrite website.

Cellebrite/UFDR is the only format that reports duplicate counts, and it produces two distinct values:

- **Detected duplicates**—a per-event custom field recording the length of the current consecutive run, written on the primary event.
- **X-RSMF-DuplicatedEventCount**—a per-conversation total written to the RSMF header. This value is surfaced downstream as the field Rsmf/DuplicatedMessageCount.

See [Deduplication for short message formats](#) for expanded details.

3.5 Troubleshooting

Errors reported during advanced access include:

Discover error message	Details
Processing UFDR did not create any RSMF files.	Make sure you have selected the correct processing profile. It is also possible the UFDR did not contain any chat data that could be converted to RSMF.
Error message, "An error occurred while adding some attachments in UFDR-to-RSMF conversion."	The error is likely due to a corrupt attachment. You can ignore this warning as the missing attachments will appear as links in the generated RSMF files.
Error message, "An error occurred during parsing date time of some messages in UFDR-to-RSMF conversion."	The error is likely due to a missing date time value. You can ignore this warning as the missing date time will appear as default value (01/01/0001 00:00:00) in the generated RSMF files.
Error message, "Multiple source extraction is not supported in UFDR-to-RSMF conversion. You can only process UFDR files created from a single device."	This error occurs when a UFDR file contains data extracted from more than one physical device (identified by mismatched device IMEIs). Relativity supports UFDR files from a single device, including files that contain multiple extraction passes (for example, logical and file system)

Discover error message	Details
	from the same device. To resolve: If your UFDR was created from multiple devices, re-export each device as a separate UFDR file in Cellebrite Reader and process them individually. If you believe your file is from a single device and you are still seeing this error, verify that the IMEI is present and consistent across all extractions in the Cellebrite export settings.
Error message, "An unexpected internal error occurred during RSMF conversion. Please retry the dataset to proceed."	Relativity may have experienced internal issues while processing the UFDR file(s). Retry the data set. If the issue persists, do not hesitate to contact support.
Error message, "An error occurred during RSMF conversion. The error may be due to an unsupported package structure. Please consider reprocessing the data."	Relativity has identified an unsupported structure in the UFDR file. Verify the file's integrity using Cellebrite Reader and export it again prior to reprocessing the data.
Error message, "The input file appears to be corrupted or is not a complete UFDR file. Please consider opening the file in Cellebrite Reader and exporting again before reprocessing the data."	Relativity has encountered an unexpected format in the UFDR file. Verify that the UFDR file was correctly exported from Cellebrite Reader, or consider re-exporting it before reprocessing.
Error message, "Extraction partially completed: X of Y media files were not fully processed due to data corruption issues. To resolve this issue, repair the corrupted media files, and export a new UFDR file before reprocessing."	The error is likely due to a corrupted media file within the UFDR file. To effectively process these files, verify their integrity in Cellebrite Reader prior to reprocessing.
Error message, "The UFDR report is encrypted or corrupted."	Relativity has identified UFDR file(s) that are likely encrypted, and this file type is currently unsupported. Use Cellebrite Reader to decrypt these files prior to reprocessing them in Relativity.
Error message, "An exception occurred during creation of EML file from emails in UFDR. The exception is due to invalid email addresses saved in UFDR report. The EML file is generated anyways but without the problematic data. You can ignore the warning or try reexporting the UFDR without the problematic emails and retry processing."	The exception arises from the presence of invalid email addresses stored in the UFDR report. You may choose to either disregard the warning or re-export the UFDR, excluding the problematic email addresses, and then attempt the processing again.
Error message, "An error occurred while adding attachments during RSMF conversion. One or more files appear to contain a virus or potentially unwanted software. You can ignore this warning as the missing attachments will appear as links in the generated RSMF files."	The error is due to an attachment file or files that have been recognized as potentially harmful and thus skipped in processing. You can find the links to the files in the RSMF generated from the source UFDR file. The processing of the remaining files was not interrupted and, unless other errors occurred, was finished successfully.
Error message, "Relativity has encountered an unsupported structure in the UFDR XML report due to a missing header. Open the file in Cellebrite Reader, then export it again before reprocessing the data."	The file might have been manually altered prior to processing. Opening the file in Cellebrite Reader removes unnecessary data and ensures the file is in the latest UFDR version.

Discover error message	Details
Multiple source extraction is not supported in UFDR-to-RSMF conversion.	The UFDR file contains data from more than one device, or has multiple extraction nodes in its internal structure Workaround: 1. Place each UFDR file in a separate processing set (do not combine multiple UFDRs in one set). 2. If the error persists on a single UFDR file, contact your Cellebrite vendor and request a new extraction with a single device per output file.

3.6 Known issues and limitations

Unsupported data types and extractions:

- UFDR files containing extractions from multiple different physical devices are not supported. Multiple extraction passes from the same device (for example, logical and file system) are supported, provided the device IMEI is consistent across all extractions in the file.
- Attachments that contain only a URL attribute are not supported.
- Tags added to the UFDR during creation are not supported.
- Avatars for chat participants are not supported.
- Voice Mail data is not supported as a recording.

Non-chat data exclusions: Relativity Processing does not support the following non-chat data types found in UFDR as text files:

- Application Usage data
- Cell Tower data
- Powering Event data
- Credit Card data

Large UFDR files:

When processing UFDR files larger than approximately 100 GB, the job queue may show no progress during the initial file copy phase. This is expected behavior — the job is running. No action is required. Allow additional time before contacting Support.

Formatting and metadata limitations:

- **Timezones:** Slicing settings reflect the original timezone. Recordings do not include timezone information.
- **Excel files:** There is no parent-child relation for text extracted from Excel files when the content of the Excel cell is larger than the maximum field length.
- **Missing metadata:** Any missing metadata following conversion to RSMF is likely due to the data not being in the UFDR file or the metadata being associated with a media file, such as images or videos.

General processing limitations:

- **Inventory:** Inventory is not supported for Cellebrite UFDR. We recommend you cull the files in Cellebrite Reader prior to uploading to Relativity.

- **Encryption:** Encrypted UFDR files are not supported. Use the Cellebrite Reader to decrypt them, then import them into Relativity.

3.7 Resources

- For information on supported short message sources, see Supported short message sources for RSMF conversion.
- For information on enabling short message conversion, see Short message conversion settings.
- For information on short message use, metadata, and mapping, see Relativity's short message format topic.

4 Short message conversion for Google Chat

Relativity supports short message conversion for Google Chat in Relativity aiR. During processing, Relativity automatically converts Google Chat exports in MBOX format to Relativity's Short Message Format (RSMF).

4.1 Required file structure

Relativity Processing supports Google Chat exports from Google Vault. A Google Vault export for Google Chat produces two files:

- An MBOX file
- A metadata XML file

Both files are generated automatically by Google Vault as part of the same export. When downloading the export from Google Vault, make sure to include the MBOX file, as it is required. If the metadata XML file is not uploaded alongside the MBOX file, conversion continues but the following warning is generated and some metadata will be missing in the output RSMF file and in the Short Message viewer:

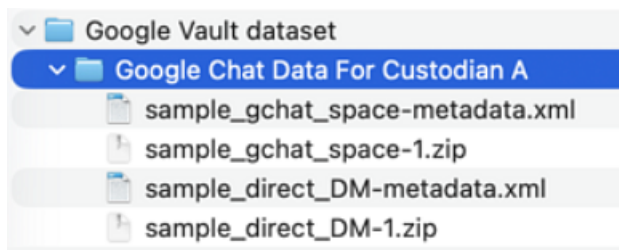
During conversion to RSMF, no metadata XML file was found. This will cause some of the metadata to be missing in the output RSMF file and in the Short Message viewer, but does not stop the full processing of the Google export file. You can continue.

You should import both the XML file and MBOX file(s) to the same level so that the XML file is matched to the MBOX file(s).

For example:

- **Correct**
 - Folder 1
 - name.mbox / name.zip (with MBOX file(s) inside)
 - name-metadata.xml
- **Incorrect**
 - Folder 1
 - name.mbox / name.zip (with MBOX file(s) inside)
 - Folder 2
 - name-metadata.xml

If you are using Express Transfer, you must use a two-level folder structure. The top-level folder contains sub-folders for each data source, for example, sub-folders for each custodian. If using the standard zip import option, both folder levels must be zipped together to maintain the two-level hierarchy.



See Create Processing Job in the Import/Export documentation for expanded details on importing data using Import/Export.

4.2 Processing Google Chat conversations

Relativity uses the conversation ID embedded in each MBOX message to group messages into conversations. Messages sharing the same conversation ID are grouped together and sorted chronologically before the conversion begins.

By default, conversations that span more than 24 hours, have more than 10,000 events, or are larger than 2 GB in size (including attachments) are split into separate RSMF files (slices). You can change the time slicing interval in the short message conversion settings of a processing profile.

Relativity processes the following message types from Google Chat exports:

- **Standard messages**—regular text messages.
- **Edited messages**—messages that were edited after being sent. Both the original body and the edited body are preserved in the RSMF file.
- **Deleted messages**—messages that were deleted by the sender. The deletion is recorded in the RSMF file.
- **Threaded replies**—messages that are quote replies to a specific message in a conversation.
- **Bot messages**—messages sent by Google Chat apps or bots within a conversation.

4.3 Generated RSMF fields

The following table lists the metadata fields found in RSMF files after Google Chat conversion.

Relativity source field name	Field type	Description
Subject	Long Text	The value of the RoomName field from the metadata XML file. The content depends on what Google Vault includes in that field for each conversation type. If no metadata XML file is provided, this field defaults to "Unknown".
EmailFrom	Long Text	The email address and name of the sender of the first event in the conversation. If no metadata XML file is provided, only the name extracted from the export content is available.
EmailTo	Multiple Choice	A list of email addresses and names of all participants in the conversation. If no metadata XML file is provided, only the names of participants who sent at least one message are available.
Date	Date	The date of first event in conversation.
Rsmf/Version	Long Text	2.0.0

Relativity source field name	Field type	Description
Rsmf/Participants	Multiple Choice	A list of email addresses of all participants in the conversation. If no metadata XML file is provided, only the names of participants who sent at least one message are available. Use the Long Text field type when mapping this field. Selecting Multiple Choice can lead to hundreds of unique choices that may impact performance. [relativity...assian.net] Caution: Use the Long Text field type when mapping the Rsmf/Participants field. Selecting Multiple Choice can lead to hundreds of unique choices that may impact performance.
Rsmf/Generator	Long Text	Relativity Processing
Rsmf/Application	Long Text	Google Chat
Rsmf/MessageCount	Whole Number	The number of events in the conversation.
Rsmf/AttachmentCount	Whole Number	The number of attachments in conversation.
Rsmf/EventCollectionId	Long Text	The conversation ID that is persisted between file slices.
Rsmf/BeginDate	Date	The date and time of first event in conversation.
Rsmf/EndDate	Date	The date and time of last event in conversation.

4.4 Deduplication considerations for Google Chat

During processing, no conversion-time deduplication is applied; edits/deletions are folded into the original event. File-level deduplication happens only at Publish.

See [Deduplication for short message formats](#) for expanded details.

4.5 Troubleshooting

Discover error message	Details
Error message, "Relativity encountered invalid MBOX file format. Export the source file again and retry processing. If the file was provided by a third party, contact the data source to request a new export. If the issue persists, contact Relativity Support."	Can't expand MBOX. The Google Chat export file isn't in the expected MBOX format. Continue by re-exporting the data from Google Vault without modifications and retry. If the issue persists, contact Support with Job ID: {jobId}.
Warning, "During conversion to RSMF, Relativity could not find the XML metadata file. Relativity processed the file, however, some metadata will be missing in the output RSMF file and the short message viewer. You can continue."	Metadata file missing. No XML metadata file was found, so some metadata may be missing in RSMF and Short Message Viewer. You can continue, or re-export including the XML file from Google Vault and reprocess for complete metadata.
Warning, "Relativity couldn't process Google Chat XML metadata. Verify if the file is correct and try again, if you'd like to process all metadata from the exported files. Processing continued anyway, but some metadata was not	Malformed sidecar XML. The XML metadata file could not be parsed (e.g. corrupt or invalid XML structure). Processing continued without sidecar metadata. Re-export the XML from Google Vault and reprocess for complete

included."	metadata.
Warning, "During conversion to RSMF, Relativity found empty messages in the input MBOX file. Relativity processed the file, however, these messages will appear as empty in the output RSMF file and the short message viewer. You can continue."	Some data types from Google Chat are not included in the export, which results in empty messages in MBOX file.
Error message, "An unexpected internal error occurred during RSMF conversion. Please retry the dataset to proceed."	Infrastructure-level failure on the last retry attempt. The job cannot continue. Retry the dataset.
Error message, "An error occurred during RSMF conversion. The error may be due to an unsupported package structure. Please consider reprocessing the data."	Catch-all for any unhandled conversion exception. May indicate an unsupported or corrupt package structure.
Warning, "Some attachments are likely corrupted. You can ignore this warning as the missing attachments will appear as names or links in the generated RSMF files."	Per-attachment: an occurred while adding an attachment to the RSMF file. The attachment is likely corrupted. It will appear as a name k instead of an embedded file.
Warning, "The total attachment size would exceed the maximum allowed limit for a single RSMF file. You can ignore this warning as the missing attachments will appear as names or links in the generated RSMF files."	Per-attachment: adding this attachment would push the RSMF past the 2 GB size limit. The attachment is skipped and appears as a name instead.
Warning, "One or more files appear to contain a virus or potentially unwanted software. You can ignore this warning as the missing attachments will appear as names or links in the generated RSMF files."	Per-attachment: the file was flagged as containing a virus or potentially unwanted software. The attachment is skipped and appears as a name instead.
Warning, "Relativity encountered invalid message format in {section_name} or an additional unexpected section. Export the source file again and retry processing. If the file was provided by a third party, contact the data source to request a new export. If the issue persists, contact Relativity Support."	Malformed message structure. A Google Chat message contained an invalid or unexpected section ({section_name}) is one of: data id, thread metadata, message metadata, quote metadata, or assets metadata), or an additional section that Relativity did not recognize. Some content from the affected message may be missing in the RSMF output and the Short Message Viewer. Processing continued. In case of any doubts or missing critical information, re-export the conversation from Google Vault without modifications and retry. If the issue persists, contact Relativity Support with the Job ID: {jobId}.

4.6 Known issues and limitations

- **Missing metadata file**—if the metadata XML file is not uploaded, some metadata will be missing, and the participant list will only include those who sent messages. This refers to the set of people who sent messages within a given slice of the conversation (that is, a single RSMF file).
- **Attachment size limit**—attachments are embedded in the RSMF file up to a combined total of 2 GB per conversation slice. Attachments exceeding this limit appear as links.
- **Corrupted attachments**—attachments that cannot be read are not embedded and appear as links.

- **Potentially unwanted files**—attachments flagged as containing a virus or potentially unwanted software are not embedded and appear as links.

4.7 Resources

- For information on supported short message sources, see Supported short message sources for RSMF conversion.
- For information on enabling short message conversion, see Short message conversion settings.
- For information on short message use, metadata, and mapping, see Relativity's short message format topic.

5 Short message conversion for Slack

This topic covers information for Slack conversion. To view the Relativity supported and unsupported file type tables, see [Supported file types for processing](#).

Note: Relativity supports short message conversion for Slack in Relativity aiR Government.

Slack exports are available across all Slack license plans. For information on Slack plans and their affect on documentation exports, see [Export your workspace data](#) on the Slack website.

You can choose whether or not to download short message attachments. If you do not download attachments, Relativity retains links to output RSMF files instead of downloading the actual files. See Short message conversion settings for information on enabling short message attachment downloads.

Note: If you do not download attachments, Relativity retains links to output RSMF files instead. You may see errors if Relativity cannot link to the source file. For example, an outage with Slack or network connectivity issues. See [Processing error resolution](#) for information on resolving processing errors.

Caution: Enabling attachment download can significantly increase total data volume. Slack exports commonly expand by 100 times or more when you download attachments. For example, a 1 GB export may result in 100–300 GB of processed data. Verify available storage and processing budget before enabling this option.

Slack AI-generated huddle summaries appear in the export as messages posted by a user named Slackbot. Each summary is represented as a standard RSMF message and includes the meeting summary as an attachment. In the short message viewer, these items display as a conversation with Slackbot, with the summary accessible through the attached file.

5.1 Required file structure

Processing supports original structure of zip containers coming from Slack.

Notes:

- Do not extract the Slack export zip before uploading. Upload and process the zip file directly. If you extract the archive before uploading, processing will complete without errors but will generate no RSMF output.
 - Do not change the contents of the zip file or you may see corrupt or incomplete results.
 - Slack exports greater than 10 GB are not supported and may be unable to process.
-

When processing Slack content, Relativity looks for information about the user and information about the messages. You must include the **users.json** file, and in the case of an [organization-level export](#), **org_users.json**, along with at least one of the following message files:

- channels.json
- dms.json
- groups.json
- mpims.json

5.1.1 Understanding JSON files for Slack data exports

Your JSON file structure will not match the images in this documentation exactly. You may see additional JSON files than what are depicted in the images. For an explanation of the JSON files and their use in Slack exports, see [How to read Slack data exports](#) on the Slack website.

5.2 Generated RSMF fields

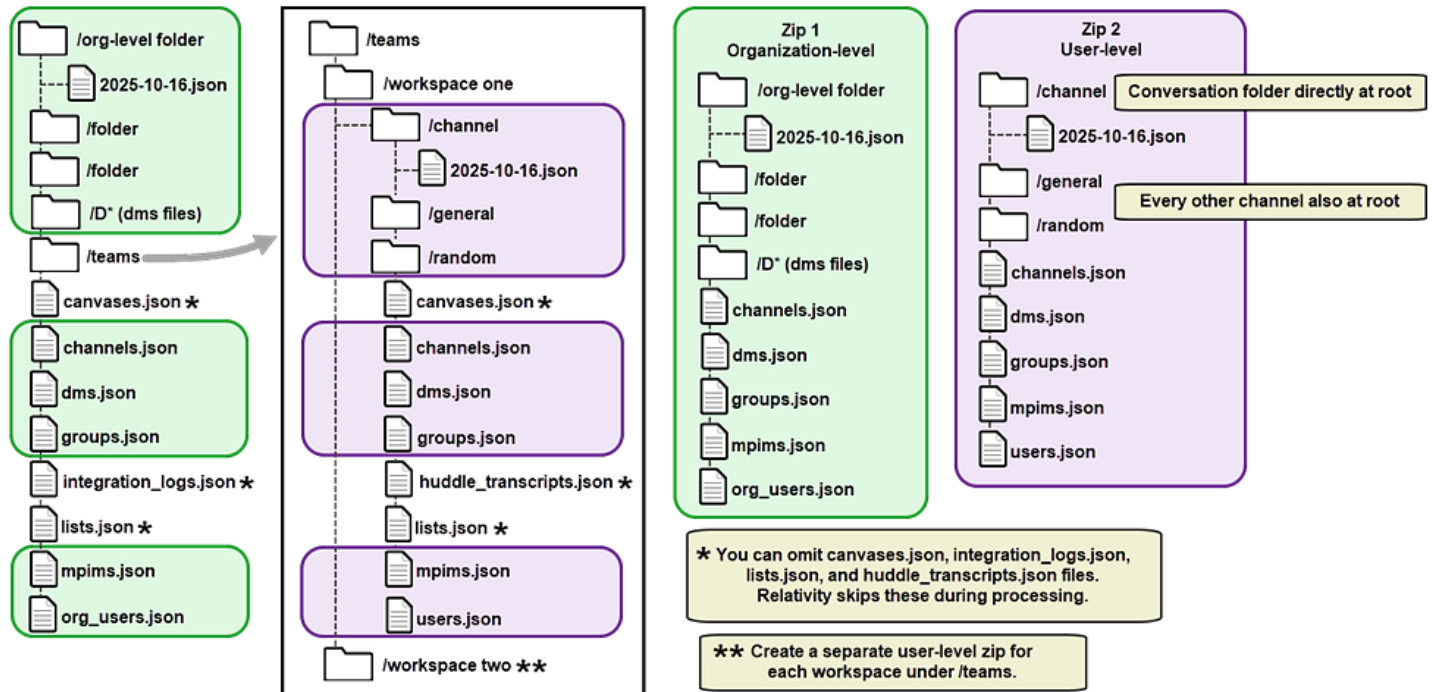
The following table lists the metadata fields found in RSMF files after Slack conversion.

Relativity source field name	Field type	Description
Subject	Long Text	The name of the conversation. • For public/private channels: CHANNEL_NAME • For group chats/dms: LIST OF PARTICIPANTS
EmailFrom	Long Text	The display name and email of the sender of the first message in this RSMF file. For conversations split across multiple files, this value may differ per file.
EmailTo	Multiple Choice	A list of display names and emails of all participants in conversation,
Date	Date	The date of first event in conversation.
Rsmf/Version	Long Text	2.0.0
Rsmf/Participants	Multiple Choice	A list of display names of all participants in conversation.
	Caution: Use the Long Text field type when mapping the Rsmf/Participants field. Selecting Multiple Choice can lead to hundreds of unique choices that may impact performance.	
Rsmf/Generator	Long Text	Relativity Processing
Rsmf/Application	Long Text	Slack
Rsmf/MessageCount	Whole Number	The number of events in the conversation.
Rsmf/AttachmentCount	Whole Number	The number of attachments in conversation.
Rsmf/EventCollectionId	Long Text	The conversation ID that is persisted between file slices.

Relativity source field name	Field type	Description
Rsmf/BeginDate	Date	The date and time of first event in conversation.
Rsmf/EndDate	Date	The date and time of last event in conversation.

5.3 Processing organization-level exports

Relativity supports organization-level exports from Slack, however you will need to manually create at least two zip files: the first containing the **organization-level** content, and then one for each **user-level** workspace (found in the /teams folder).



To export **organization-level** content from Slack:

1. Create a folder to contain the **organization-level** files.
2. Move the **/teams** folder out of the organization-level directory. You will work with the /teams folder in a later step.
3. Working in the organization-level directory, remove any JSON files that Relativity does not process (see the **Not processed** column in the table below).

Note: Relativity does not process all JSON files. It looks for information about the user and information about the messages. Therefore, the organization-level files must include the **org_users.json** file and the user-level files must include the **users.json** file, along with any related folders.

Relativity omits *Not processed* files during processing. That said, you do not need to add any *Processed* files to your folders if your environment does not have them. Just make sure you have either the **org_users.json** (for organization-level exports) or **users.json** file (for user-level or workspace exports), and at least one of the other *Processed* files (containing message information.)

Processed	Not processed
Channels (channels.json)	App activity logs (integration_logs.json)
Direct messages (dms.json)	Content flags (content_flags.json)
Private channels (groups.json)	A URL for every canvas (canvases.json)
Group direct messages (mpims.json)	Huddle transcripts (huddle_transcripts.json)
Members (org_users.json, users.json)	Task lists (lists.json)

4. Zip the organization-level folder; it is now ready for processing into Relativity.

To export **user-level workspace** content from Slack:

1. Create a folder to contain the **user-level** files.
2. Open the **/teams** folder.
3. Open the first **/workspace** folder.
4. Review the JSON files. Remove any **Not processed** files.
5. Move the remaining workspace files into the new folder.
6. Zip the contents of the folder (to match the diagram above); it is now ready for processing into Relativity.

Repeat the user-level steps for each workspace within the /teams folder, zipping each one individually.

Upload the zip files for processing with Import/Export or copy them to your processing source location with ROSE. If using ROSE, make sure the processing location points to the custodian folder containing the zip files.

5.4 Deduplication considerations for Slack

During processing, consecutive same-timestamp events are merged into the prior event (no duplicate count, no field/header).

See [Deduplication for short message formats](#) for expanded details.

5.5 Troubleshooting

Relativity tries to extract all available data it finds. In some cases, errors may occur due to the data in the Slack export package or availability of attachments. Errors reported include:

Discover error message	Details
An error occurred during downloading some attachments in Slack-to-RSMF conversion. The error is likely due to attachments no longer being available to download from Slack servers. Relativity downloaded all available attachments. You can ignore this warning as the missing attachments will appear as links in the generated RSMF files.	During downloading attachments Relativity takes the URL present in JSON. While attempting to download attachment the provided response was "404 not found". For most scenarios this means the file is no longer kept on Slack servers and won't be possible to obtain. This error serves informational purpose.
An error occurred during downloading attachment as part	During downloading attachments Relativity takes the

Discover error message	Details
of Slack-to-RSMF conversion. The error is likely due to a transient problem connecting to the Slack servers. You can ignore this or rediscover the data to proceed with attachments.	URL present in JSON. While attempting to download attachment the provided response was "500 server error". This means Slack had some transient problems and it's advised to retry this error to retry downloading attachments.
Relativity does not support organization-level exports for Slack-to-RSMF conversion. For instructions on how to proceed, see the Troubleshooting documentation for Slack conversion.	Provided Slack data was an organization-level export – it contains some additional file structure that is currently not supported. To process this data with Relativity Processing it needs to be repackaged into supported format. See "Processing org-level exports" section for more details.
An error occurred during the conversion of the Slack-to-RSMF conversion. The error may be due to an unsupported package structure. Repair the file and reprocess the data.	Relativity could not extract any RSMF data due to a corrupt Slack export file. Contact support for help in troubleshooting the file.
An unexpected internal error occurred during Slack-to-RSMF conversion. Rediscover the data to proceed.	Relativity encountered a transient issue. It's advised to retry the processing job. If the issue persist contact support for help.

5.6 Known issues and limitations

- To optimize processing speed, use a single Slack zip container within a single processing data source.
- Relativity does not support Inventory for Slack export containers.
- When using Web Import/Export, you must put your Slack zip container (let's call it the child container) inside another zip container (let's call it the parent container) and then import the parent container using the Import/Export tool. Relativity cannot convert Slack files to RSMF if you import a child container directly instead of placing it in a parent container. The data is processed as JSON.

Note: Orphaned messages

If a conversation extends across multiple time frames because of slicing, it will be divided into several files. As a result, messages from a single thread will be stored in different files that correspond to each sliced time frame. Messages that are saved outside of the parent conversation will become orphaned, meaning they will no longer be linked to the parent conversation. Relativity adds a note at the top of the message indicating it is an orphan.

Friday, July 30, 2021



Orphaned Replies from OrgWideTeamCreatedByDC_1800GroupConversation-AccessibleToEveryone

KA



That is one thing I do not miss

7:40 AM

5.7 Resources

For additional information, see:

- For information on supported short message sources, see [Supported short message sources for RSMF conversion](#).
- For information on enabling short message conversion, see [Short message conversion settings](#).
- For information on short message use, metadata, and mapping, see [Relativity's short message format](#).

6 Short message conversion for Microsoft Teams

This topic covers information for Microsoft Teams (Teams) conversion. To view the Relativity supported and unsupported file type tables, see [Supported file types](#) for processing.

Note: Relativity supports short message conversion for Microsoft Teams in Relativity aiR Government.

Processing can automatically convert discovered Teams conversations within Microsoft eDiscovery (Standard) export data files (PST). For information on enabling short message conversion, including how files should be converted (via slicing), see [Short message conversion settings](#).

Note: This documentation contains references to third-party software, or technologies. While efforts are made to keep third-party references updated, the images, documentation, or guidance in this topic may not accurately represent the current behavior or user interfaces of the third-party software. For more considerations regarding third-party software, such as copyright and ownership, see [Terms of Use](#).

6.1 Microsoft eDiscovery (Standard) PST exports

- To learn more about exporting Teams data with the Microsoft eDiscovery (Standard) solution, see [Feature reference for Content search](#) on the Microsoft website.
- Processing expects data to be provided as a PST container where a single PST might contain both email and Teams data.
- Teams messages are automatically identified based on the message class of emails inside the export. IPM.SkypeTeams.Message or IPM.SkypeTeams.SystemMessage is considered Teams messages.
- If you have enabled Teams conversion in the processing profile, the original MSG files that contain Teams information are not discovered. If you want to retain the original files, disable RSMF conversion.
- The email processing and Teams conversion jobs are independent within a single container processing. If the conversion fails with a document-level error, email files are still discovered.
- Any attachments embedded inside the PST are extracted and processed; any attachments not inside the PST are kept as links within RSMF files (for example, attachments shared via OneDrive or SharePoint.)

Note: Discovered file size will be smaller than the PST input size.

PST files exported from Microsoft 365 typically contain both email data and Teams conversation data. Relativity processes both types, but RSMF conversion applies only to the Teams content. The discovered file size reflects the Teams portion of the PST only — not the total PST size. This is expected behavior and does not indicate missing data.

Example: A 90 GB PST file containing Teams messages and email mailbox data will produce RSMF output sized around the Teams messages only. The remaining email data is processed separately as standard documents.

6.2 Microsoft eDiscovery (Premium) HTML exports

Microsoft eDiscovery (Premium) HTML exports are currently not supported for RSMF conversion in processing. See Relativity Collect documentation for collection capabilities for Microsoft 365.

6.3 Supported input format

You must export Teams data from Microsoft 365 as a PST file. Individual MSG files and HTML exports are not supported.

If your Teams data was exported using Microsoft Purview (Compliance portal), ensure the export format is set to PST, not HTML. The HTML export format introduced by Purview in May 2025 is not currently supported for RSMF conversion.

Note: Microsoft Purview users: Purview's default Teams export format changed to HTML in May 2025. If you are using Purview to collect Teams data, select PST as the output format before exporting.

6.4 Generated RSMF fields

The following table lists the metadata fields found in RSMF files after Teams conversion.

Relativity source field name	Field type	Description
Subject	Long Text	The name of the conversation. - For channels: Team: TEAM_NAME Channel: CHANNEL_NAME - For group chats/dms: Chat: list of participants/CHAT_NAME
EmailFrom	Long Text	The display name and email of the sender of the first message in this RSMF file. For conversations split across multiple files, this value may differ per file.
EmailTo	Multiple Choice	Display names and email addresses of all conversation participants who sent at least one message or reacted to a message. Users who were present in the conversation but did not send any messages are not included. This content is derived from the same data set as the Rsmf/Participants list and is rendered as the display name, along with the email address. Participants who have left the Teams workspace may appear with the display name only (and no email address).

Relativity source field name	Field type	Description
Date	Date	The date of first event in conversation.
Rsmf/Version	Long Text	2.0.0
Rsmf/Participants	Multiple Choice	Display names of all conversation participants who sent at least one message or reacted to a message. Users who were present in the conversation but did not send any messages are not included. This content is derived from the same data set as the EmailTo list and rendered as the deduped display name only. Caution: Use the Long Text field type when mapping the Rsmf/Participants field. Selecting Multiple Choice can lead to hundreds of unique choices that may impact performance.
Rsmf/Generator	Long Text	Relativity Processing
Rsmf/Application	Long Text	Teams
Rsmf/MessageCount	Whole Number	The number of events in the conversation.
Rsmf/AttachmentCount	Whole Number	The number of attachments in conversation.
Rsmf/EventCollectionId	Long Text	The conversation ID that is persisted between file slices.
Rsmf/BeginDate	Date	The date and time of first event in conversation.
Rsmf/EndDate	Date	The date and time of last event in conversation.
Rsmf/HasPlaceholders	Yes/No	Set if the generated RSMF contains any placeholders. See Troubleshooting for more information.

6.5 Metadata used to group Teams conversations

Relativity uses the ThreadID header to group Teams conversations. The header is found in every Teams conversation in a PST file. The full header string is *IOpenTypedFacet.SkypeSpaces_ConversationPost_Extension#ThreadId*.

6.6 Time zone for slice boundaries

Slice boundaries are always calculated in UTC, regardless of your processing profile time zone setting.

Example: A one-day slice will break at 00:00 UTC. If your matter requires slice boundaries aligned to a specific local time zone (such as US Eastern), apply the corresponding UTC offset when configuring your slice date range.

6.7 Participant metadata

When a Microsoft Teams export is converted to RSMF, Relativity records a list of participants for each RSMF file it produces. This section explains how that list is built, why it may look shorter than expected, and how to obtain a full channel membership list if you need one.

6.7.1 Participants are scoped to the slice, not the channel

Slices produced during conversion carry their own participant metadata. The metadata describes only the activity inside the one slice, not the channel as a whole.

A participant appears in a slice's participant metadata only if they sent at least one message within the slice's time window. Someone who is a member of the channel but did not send a message during that window is not listed in the slice. This includes members who:

- only read messages during the window
- only reacted to a message with an emoji
- joined the channel but never posted
- posted earlier or later, in a different slice

6.7.1.1 Guest users and external participants

Guest users and participants external to the organization may appear in participant metadata as a GUID (a1b2c3d4-5e6f-7890-abcd-ef1234567890) instead of a readable display name.

This happens when the source export does not include a resolved display name for that person, so Relativity records the underlying identifier that Microsoft 365 uses to represent them. The identifier is still a valid, stable reference to a single participant; only the friendly name is missing.

If you need to associate a GUID with a named individual, resolve it through the Microsoft 365 API, which can return the display name and account details for that identifier.

6.8 Message counts

The raw Teams message count (MSG files) may exceed the RSMF message count because edited and deleted messages are consolidated into a single log entry during RSMF conversion. This is expected behavior, not data loss.

6.9 Deduplication considerations for Teams

During processing, no duplicate filtering is applied. Unsupported types become placeholders, but same-ID edits/deletions are folded into the original event; file-level deduplication happens at Publish.

See [Deduplication for short message formats](#) for expanded details.

6.10 Troubleshooting

Relativity tries to extract all available data it finds. In some cases, errors may occur due to the data in the Microsoft eDiscovery (Standard) PST export files. Errors reported include:

Discover error message	Details
An error occurred while extracting some events during the Teams-to-RSMF conversion. The error is likely due to missing metadata in the PST file. Instead, Relativity extracted such events as standalone MSG files. You can ignore this warning or consider disabling Teams conversion and then rediscover the data.	Some messages are missing JSON metadata that Relativity uses to build RSMF conversations, even though they had the expected IPM.SkypeTeams.* message class. The missing data may be due to errors during the Microsoft eDiscovery export process or schema differences (such as an old export). Relativity provides these events as emails.
An error occurred extracting {0} type events found in PST during the Teams-to-RSMF conversion. Relativity currently does not support extracting events of that type. Instead, Relativity created placeholders for the events in the generated RSMF files. You can ignore this warning or consider disabling Teams conversion and then rediscover the data.	Relativity encountered unsupported events while converting Teams messages. Relativity created placeholders for these events in the RSMF files. {0} is the detected type, for example, <i>Media_CallRecording</i> . The event placeholder in the RSMF file will appear as fol-

Discover error message	Details
	lows:Unsupported type: Media_CallRecording. Placeholders contain all attachments provided in original event. RSMF metadata field Rsmf/HasPlaceholders can be used to identify files with placeholders. If this data is critical to the investigation, you can reprocess it as an email to retrieve the full body content. See Known issues and limitations for more information.
An error occurred extracting some events during the Teams-to-RSMF conversion. The error is likely due to missing metadata in the PST file. Instead, Relativity extracted the raw metadata for the events and included them in the RSMF files. You can ignore this warning or consider disabling Teams conversion and then rediscover the data.	Relativity encountered an unexpected error while extracting the metadata for a specific event. To avoid data loss, Relativity generates the raw metadata (JSON) for the event, which you can view and search. Where possible, contact support with examples of these errors, which will help improve future versions of Teams extraction.
An error occurred during Teams-to-RSMF conversion. The error may be due to an unsupported or corrupt PST file. Repair the file and reprocess the data.	Relativity could not extract any RSMF data due to a corrupt PST file. Contact support for help in troubleshooting the file.
An unexpected internal error occurred during Teams-to-RSMF conversion. Rediscover the data to proceed.	Relativity encountered a transient issue. You should retry the processing job. If the issue persists, contact support for help.
Warning: Media_xxx type events were found in the PST file during the Teams-to-RSMF conversion. Relativity currently does not support extracting events of this type. As a result, placeholders were created for these events in the generated RSMF files. You can safely ignore this warning, as it does not affect the conversion process itself.	This is a warning only. The conversion process is not affected.

6.11 Known issues and limitations

The Teams metadata available within the PST depends heavily on the Microsoft eDiscovery (Standard) search configuration. The following is a list of known issues and limitations:

- Teams metadata limitations in RSMF conversion capability:
 - Giphy app and stickers. GIFs are provided as links in the RSMF files. Only custom text from stickers is available within RSMF files. For more information, see [Send an emoji, GIF, or sticker in Microsoft Teams](#).
 - Loop components are downloaded as HTML attachments linked in the RSMF files. For more information, see [Send a loop component in a channel](#).
 - Media cards are provided by connectors and apps like Praise or Approvals. They are provided as place-

holders in RSMF files with HTML attached to the event.

- Meetings, Call Recordings, and Call Transcripts are provided as placeholders in RSMF files.
- Known issues with Microsoft eDiscovery (Standard) PST metadata:
 - Migrated Teams Chats do not contain the JSON metadata used by Relativity to build RSMF and are processed as MSG files.
 - Join/leave events are available only for chats exported with specific Microsoft eDiscovery content search. These are system events, so they are not retrieved if the export is filtered only to instant messages (itemclass:IPM.SkypeTeams.Message) or Teams (kind:microsoftteams).
 - Announcements do not include edit information. For more information, see [Send an announcement to a channel in Microsoft Teams](#).
 - Full participant information (full display name and email) becomes unavailable once the participant has left the chat.

Note: Orphaned messages

If a conversation extends across multiple time frames because of slicing, it will be divided into several files. As a result, messages from a single thread will be stored in different files that correspond to each sliced time frame. Messages that are saved outside of the parent conversation will become orphaned, meaning they will no longer be linked to the parent conversation. Relativity adds a note at the top of the message indicating it is an orphan.

Friday, July 30, 2021

 Orphaned Replies from OrgWideTeamCreatedByDC_1800GroupConversation-AccessibleToEveryone

KA



That is one thing I do not miss

7:40 AM

6.12 Resources

- For information on supported short message sources, see Supported short message sources for RSMF conversion.
- For information on enabling short message conversion, see Short message conversion settings.
- For information on short message use, metadata, and mapping, see Relativity's short message format topic.

7 Microsoft

This topic covers additional supported file type (for processing) information for Microsoft. To view information on Excel, PowerPoint, and Word, see [Supported file types for processing](#).

Note: This documentation contains references to third-party software, or technologies. While efforts are made to keep third-party references updated, the images, documentation, or guidance in this topic may not accurately represent the current behavior or user interfaces of the third-party software. For more considerations regarding third-party software, such as copyright and ownership, see [Terms of Use](#).

7.1 Outlook message item (.msg) to MIME encapsulation (.mht) conversion

The following table provides details on the differences between how Relativity handles .msg and .mht file types. This information may be especially useful if you plan on setting the Email Output field on the processing profile to MIME encapsulation.

Category	Field/attribute	Outlook message item (.msg)	MIME encapsulation (.mht)
Metadata fields	Show Time As	This field sometimes appears in the extracted text from MSG files when not explicitly stated in the message file itself. The default for a calendar invite is to show time as busy; the default for a cancellation is to show time as free.	Show Time As does not appear in the extracted text if the default value is populated.
Metadata fields	On behalf of	This field is sometimes present in text from a message item. In some cases, this field is populated with the same value as the From field.	On behalf of does not appear in the extracted text.
Interline spacing	N/A	The expected number of blank lines appears in the extracted text. Line wrapping for long paragraphs will also be present.	In some cases, the text in the .mht file format has fewer blank lines than the text from a message item. In addition, there is no built-in line wrapping for long paragraphs.
Intraline spacing	N/A	White-space characters are converted to standard space characters.	White-space characters may remain as non-breaking spaces.
Email addresses	Email	When a message file is converted to .mht, the text is extracted from the .mht file using Outlook. This can lead to a loss of data.	If joe.smith@acme.com renders as Joe Smith in the .mht file, the email address is not captured in the extracted text.

7.1.1 Email image extraction support

It is helpful to understand when Relativity treats an image that is attached to an email as an inline, or embedded, image and not as an actual attachment. The following table breaks down when this occurs based on email format and image characteristics:

Email format	Attachments that are inline, embedded, images
Plain text	None
Rich text	IPicture-based OLE embedded images
HTML	- Images with content ID referenced in the HTML body - Local, non-internet image references in the HTML that Relativity can match to an attachment .pst/.ost/.msg files containing metadata hints as to whether or not the image is marked hidden or is referenced in the HTML body

You can arrange for the discovery of inline images when creating [Processing profiles](#), specifically through the field called **When extracting children, do not extract**.

7.1.1.1 Inline image identification

Processing defines inline images within emails through the **HiddenAttachment** field. This field is not mapped by default. See [Mapping processing fields](#) for more information.

7.1.2 Microsoft Office child extraction support

See a breakdown of Relativity's support of Microsoft Office child extraction

The following table displays which Office file extensions will have their embedded objects and images extracted by Relativity and which will not.

- **√**—Relativity fully extracts the embedded object and image.
- **√***—Relativity partially extracts the embedded object or image.
- **Empty**—Relativity does not extract the embedded object or image.

Office program	File extension	Embedded object extraction	Embedded image extraction
Excel	.xlsx	√	√
Excel	.xlsm	√	√
Excel	.xlsb	√	√
Excel	.xlam	√	√
Excel	.ltx	√	√
Excel	.xltn	√	√
Excel	.xls	√	√*
Excel	.xlt	√	√*
Excel	.xla	√	√*
Excel	.xlm	√	√*
Excel	.xlw	√	√*
Excel	.uxdc		
Outlook	.msg	√	√
Word	.docx	√	√
Word	.docm	√	√
Word	.dotx	√	√
Word	.dotm	√	√
Word	.doc	√	√*
Word	.dot	√	√*

Office program	File extension	Embedded object extraction	Embedded image extraction
Word	.rtf	✓	✓
Visio	.vsd		
Visio	.vdx		
Visio	.vss		
Visio	.vsx		
Visio	.vst		
Visio	.vsw		
Visio	.vsdx	✓	✓
Visio	.vsdm	✓	✓
Project	.mpp		
PowerPoint	.pptx	✓	✓
PowerPoint	.pptm	✓	✓
PowerPoint	.ppsx	✓	✓
PowerPoint	.ppsm	✓	✓
PowerPoint	.potx	✓	✓
PowerPoint	.ppt	✓	✓
PowerPoint	.pps	✓	✓
PowerPoint	.pot	✓	✓
OneNote	.one	✓	✓

7.2 Excel files

Due to Excel specifications and limits, when processing a database file with the native text extraction method, the database file may miss data in extracted text. For example, if a database file contains more than 1,048,576 rows and 16,384 columns, the extracted text of these files will not contain text on row 1,048,577 and onward and on column 16,385 and onward. For more information, see [Excel specifications and limits](#) on the Microsoft website.

8 Google Workspace

This topic covers Relativity Processing support and considerations for Google Workspace data. To view a comprehensive list of supported file types, see [Supported file types for processing](#).

Note: This documentation contains references to third-party software, or technologies. While efforts are made to keep third-party references updated, the images, documentation, or guidance in this topic may not accurately represent the current behavior or user interfaces of the third-party software. For more considerations regarding third-party software, such as copyright and ownership, see [Terms of Use](#).

8.1 Overview

When you gather Google Workspace data directly using Google Vault or indirectly using Relativity Collect (Collect), the collection export includes zipped Google Gmail, Google Drive, or Google Chat data, along with supplemental metadata files. When generating a collection that includes both Google Gmail and linked Google Drive items, the export also includes a cross-reference file with linking data.

Processing automatically identifies these sidecar files when present in your data source, and captures this information alongside the naturally extracted metadata from the processed Gmail, Google Drive, and Google Chat items. This supplemental metadata is made available within the Field Catalog for publishing to your workspace.

Note: During processing, Relativity checks for the sidecar metadata files described below. If a sidecar file can't be located, Processing reports a file exception but does not stop or fail. Native files, text, and family relationships are unaffected, and the data can still be published. See [Google Workspace discovery](#) for details on when this occurs and how to resolve it.

8.2 Required file structure

There are two methods for collecting Google Workspace data:

- Collect—install this Relativity application to collect data from a variety of data sources. For detailed information on Collect, see the Google Workspace data source documentation.
- Google Vault—use this option to manually download data.

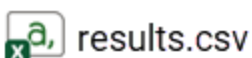
While each method exports collection data differently, Processing supports both. The following sections explain these structures and provide guidance for submitting Google Workspace data for processing.

Notes:

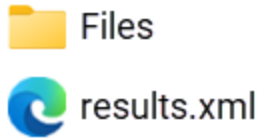
- Both Collect and Google Vault exports can contain additional files beyond the minimum requirements. You can remove these files before processing. Otherwise, they are processed as loose individual files without any special handling.
 - When collecting data manually from Google Vault, you must export Gmail or GChat data in MBOX format. Relativity does not capture sidecar file metadata if exported in other formats.
 - Do not unzip any containers included in the collection export. Doing so will result in improper handling during processing.
 - Do not edit the content or the names of the native files or supplemental metadata files. Doing so will result in improper handling during processing.
 - Relativity Import/Export may require additional preparation of the data, depending on the transfer mode. See [Importing raw unprocessed data for Processing via Import/Export](#) for details.
-

8.2.1 Collect

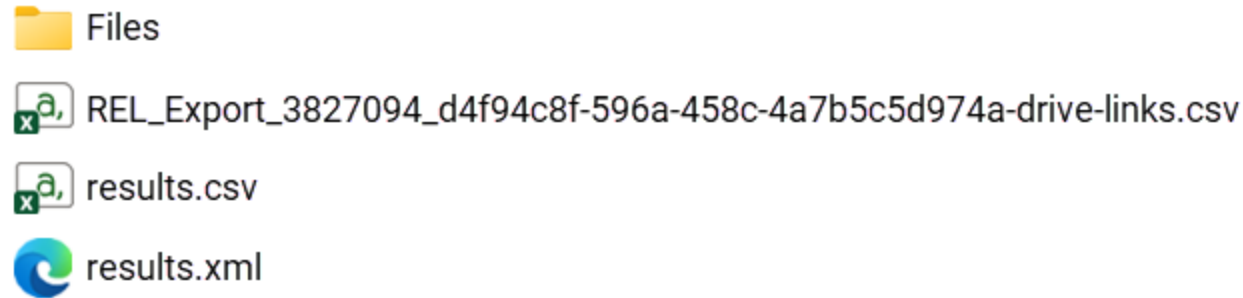
When you use Collect to gather Gmail data, messages are exported to one or more MBOX files, and zipped within the Files sub-directory. The Gmail sidecar metadata is provided in a *results.csv* file at the root of the export directory.



When you use Collect to gather Google Drive data, the items are exported to one or more zip files in the Files sub-directory. The sidecar metadata for Google Drive items is provided in a *results.xml* file at the root of the export directory.

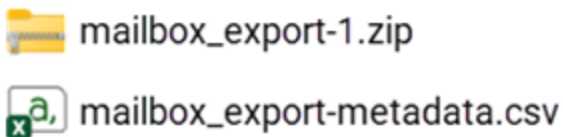


If you enable the Collect Linked Files option for Gmail collections, the export directory will contain all of the above content, as well as a file named *[COLLECTION_EXPORT_ID]-drive-links.csv*. This file captures the relationships between Gmail and linked Google Drive items. To learn more about enabling linked Google Drive item collection, see Google Workspace data source - Gmail.



8.2.2 Google Vault

When using Google Vault to gather Gmail data, messages are exported to one or more MBOX files in a single zip container. The Gmail sidecar metadata is provided in a file named *[EXPORT_NAME]-metadata.csv*. Place these files in the same folder.



When using Google Vault to gather Google Drive data, items are exported to one or more zip files. The Google Drive items sidecar metadata is provided in a file named *[EXPORT_NAME]-metadata.xml*. Place these files within the same folder.

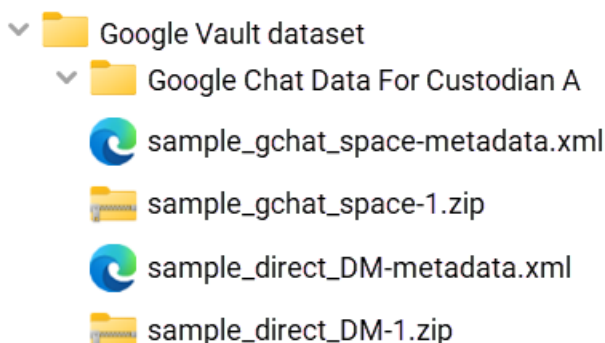


If you enable the Export linked Drive files option for a Gmail search export, the export will contain all of the above contents, with the Google Drive sidecar file named *[EXPORT_NAME]-Drive_Link_Export-metadata.xml*. The export will also contain a file named *[EXPORT_NAME]-drive-links.csv*. This file captures the relationships between Gmail and linked Google Drive items. To learn more about enabling linked Google Drive item collection, see the Google support article, [Export data from Vault](#).



8.2.2.1 Google Chat

When using Google Vault to gather Google Chat data, messages are exported to one or more MBOX files in a single zip container. The Gmail sidecar metadata is provided in a file named [EXPORT_NAME]-metadata.xml. Place these files in the same folder.



Note: Google Chat exports are recognized only if the first line of the MBOX file contains MBI-FLAT or MBI-THREADED. Exports created before Google's format change lack this marker, and use an older message structure. They are processed as individual EML files instead of RSMF. To fix, re-export the data from Google Vault using the current export process. Since Google Vault does not expose a cutoff date or version indicator, checking the first line of the MBOX is the only reliable test.

8.3 Google Workspace discovery

The following list outlines Relativity's handling of Google Workspace data and sidecar files during discovery.

- Relativity performs standard identification, expansion, text and metadata extraction on the zipped Gmail MBOX and Google Drive files. In the case of MBOX discovery, Relativity uses the RFC 4155 and RFC 5322 standards to properly identify and parse individual messages and their properties from the MBOX files. These messages are stored in EML format, with a file type designation of SMTP E-Mail Message.
- In addition to performing discovery of the Google Workspace data, Processing also identifies the related sidecar files. These files receive a special file type designation, and the Google Workspace metadata is captured and applied to the corresponding Gmail, Google Drive, and Google Chat files from the same data source. These sidecar files are ignored during publish, but can still be viewed from the Processing Files tab.
- Where possible, Processing attempts to identify data sources containing partial Google Workspace collections, for example, a missing sidecar metadata file. In these cases, a file exception is reported on either the Gmail or

Google Chat MBOX file or the Google Drive items zip file. Relativity still processes the data source. However, based on this warning, you may choose to prepare and process a more complete collection.

- For Gmail collections containing linked Google Drive files, Relativity processes both the Gmail and Google Drive data, but does not capture any formal relationship between them. Note however, the cross-referenced data found within the *drive-links.csv* file is captured and applied to their respective files. For example, processed Gmail messages include a metadata field that references a concatenated list of all linked Google Drive items by DriveItemid, also known as DocId. Refer to the tables on this topic for a complete list of metadata fields available for mapping.
- Google Gemini data is converted to Relativity Short Message Format (RSMF) for processing when using Collect. The sidecar metadata is also integrated. If using Google Vault directly, you should convert Google Gemini data to RSMF before processing. Otherwise, this data is processed as standard XML data, and will not include sidecar metadata.

Notes:

- All processing profile settings apply to Google Vault collection data, as with any other data source.
 - The Gmail sidecar file only includes metadata for top parent message items. Embedded attachments are extracted, but will not contain any supplemental metadata.
 - Similarly, neither embedded objects extracted from Google Drive items, nor files extracted from Google Drive-hosted compressed archives, are referenced in the Google Drive sidecar file. Relativity does not attempt to propagate Google Drive item metadata to extracted descendants. As such, these items will not contain any supplemental metadata.
 - Sidecar metadata is not used for publish deduplication. In the case of Google Drive files, Relativity relies on the standard hash it calculates from the native. In the case of Gmail messages, Relativity creates the standard composite hash using only the properties extracted directly from the MBOX file.
-

8.3.1 Sidecar file not found

If Processing cannot locate the expected sidecar file for a Google Drive or Google Gmail export, you will see the following exception:

A Google Drive sidecar file was not found... The Google Drive files have been processed without this supplemental metadata, and can be published..." or similarly for Google Mail, "A Google Mail sidecar file was not found..."

This exception means only the supplemental Google-provided metadata is missing, not the file content. You may publish as-is, or correct the issue and reprocess.

Check the filename separator. The ZIP filename and its sidecar use different separators, which are easy to mix up:

- Drive export ZIP—uses an underscore before the segment number (for example, *export_0.zip*).
- Drive sidecar metadata—uses a hyphen before *metadata.xml* (for example, *export-metadata.xml*).

Confirm your data is actually Google Workspace data. This exception can also appear on ZIP files that are not Google exports, if the ZIP's filename happens to end in an underscore followed by a number (for example, *Custodian_Export_003.zip*). If you did not collect this data from Google Drive or Google Vault, this exception can be safely ignored as no Google metadata was ever expected for these files.

If your sidecar file uses an underscore (for example, *export_metadata.xml* instead of *export-metadata.xml*), Processing will not recognize it as a match even though the file is present. Rename the file to use a hyphen and reprocess if you want the metadata captured.

8.4 Sidecar metadata in publishing

Google Vault sidecar files contain the metadata fields listed below, all of which are extracted and made available within the Field Catalog for publishing to your workspace. To learn more about using the Field Catalog, see Mapping Processing Fields.

Notes:

- Google Drive metadata fields available for publish may be found within the list of **Other Fields** in the Field Catalog, and have a prefix of GoogleDrive\. Gmail metadata fields may be found within the list of **Metadata Fields**, and have a prefix of Google\.
- As with other optional metadata fields, these fields are not made available in the Field Catalog until at least one item containing such metadata is discovered within your workspace.
- Google may add, remove, or edit fields at any time. If a metadata field is not listed here, it is not officially supported by Relativity. For the most recent list of exported fields, refer to the Google support article, [Vault export contents](#).

8.4.1 Google Drive sidecar fields

Google Drive .xml field	Relativity source field name	Field type	Description
#Author	GoogleDrive/Author	Long text	The email address of the person who owns the file in Google Drive. For a shared drive file, it shows the shared drive name.
Collaborators	GoogleDrive/Collaborators	Multiple Choice	The accounts and groups that have direct permission to edit the file or add comments. This field also includes users with indirect access to the file if you chose this option during export.
ContentVersionDate	GoogleDrive/ContentVersionDate	Date	If a version date was included in the export query, the date when a specific version of a Drive document was saved.
#DateCreated	GoogleDrive/DateCreated	Date	The date a Google file was created in Google Drive. For non-Google files, this is the date the file was uploaded to Google Drive.
#DateModified	GoogleDrive/DateModified	Date	The date the file was last modified.
DocID	GoogleDrive/DocID	Long Text	A unique identifier for the file. Also known as DriveItemID. For site exports, the value is the page ID.
DocParentID	GoogleDrive/DocParentID	Long Text	For Google sites, the site ID.
DocumentType	GoogleDrive/DocumentType	Long Text	The file type for Google files. Possible values are: DOCUMENT—a document cre-

Google Drive .xml field	Relativity source field name	Field type	Description
			ated in Google Docs. • SPREADSHEET—a spreadsheet created in Google Sheets. • PRESENTATION—a presentation created in Google Slides. • FORM—a form created in Google Forms. • DRAWING—a drawing created in Google Drawings. • SITES_PAGE—a page from a site created in new Google Sites.
Duration	GoogleDrive/Duration	Long Text	For Google Vids, the duration of the video recording in milliseconds.
FileName	GoogleDrive/FileName	Long Text	The file name. Use this value to correlate the metadata with the file in the export ZIP file.
FileSize	GoogleDrive/FileSize	Whole Number	The size of the file in bytes.
Hash	GoogleDrive/FileHash	Long Text	The MD5 hash of the file.
Others	GoogleDrive/Others	Multiple Choice	The accounts from your query that have indirect access to the file if you opted to exclude access level information during export. This field may also include users for whom Google Vault could not determine permission levels at the time of export.
PublishedURL	GoogleDrive/PublishedURL	Long Text	For Google sites, the web address of the published page. Value is empty for unpublished sites.
SharedDriveID	GoogleDrive/SharedDriveID	Long Text	The identifier of the shared drive that contains the file, if applicable.
SitesTitle	GoogleDrive/SitesTitle	Long Text	For Google sites, the name of the page.
SlideRecording	GoogleDrive/SlideRecording	Long Text	Metadata about recordings, if the exported item is a Slides presentation with associated recordings.
SourceHash	GoogleDrive/SourceHash	Long Text	A unique hash value for each version of a file. This field can be used to deduplicate file exports and verify the exported file is an exact copy of the source file. This field is supported by Google Docs, Google Sheets, and Google

Google Drive .xml field	Relativity source field name	Field type	Description
			Slides files only.
#Title	GoogleDrive/Title	Long Text	The file name as assigned by the user. Because some operating systems cannot expand zip files with extremely long file names, Google Vault truncates the file name at 128 characters during export. The value shown by the #Title tag is not truncated.
Viewers	GoogleDrive/Viewers	Multiple Choice	The accounts and groups that have direct permission to view the file. The field also includes users with indirect access to the file if you chose this option during export.

8.4.2 Gmail sidecar fields

Google mail .csv field	Relativity source field name	Field type	Description
Rfc822MessageId	Google/Rfc822MessageId	Long Text	A message ID that is the same for receiver and sender messages. Use this value to correlate metadata with the message in an MBOX export. For classic Hangouts, the value is for the first message in the thread.
GmailMessageId	Google/GmailMessageId	Long Text	A unique message ID. Use this value to manage specific messages with the Gmail API. For classic Hangouts, the value is for the first message in the thread.
Account	Google/Account	Long Text	The account that had the message in their inbox. For example, user1@company.com received a message sent to groupA@company.com because they are a member of the group. If a search returns that message because it was in user1's inbox, then the value of To is groupA@company.com while the value of Account is user1@company.com.
From	Google/From	Long Text	The sender account.
To	Google/To	Long Text	The recipient account. Multiple recipients are comma-separated and the list is in double quotes.
CC	Google/CC	Multiple Choice	Accounts in the cc: field.
BCC	Google/BCC	Multiple Choice	Accounts in the bcc: field.
Subject	Google/Subject	Long Text	The message subject.
Labels	Google/Labels	Multiple Choice	Labels applied to the message by Gmail or the

Google mail .csv field	Relativity source field name	Field type	Description
			user.
DateSent	Google/DateSent	Date	The message send date in UTC, yyyy-MM-dd 'T'HH:mm:ssZZZZ.
DateReceived	Google/DateReceived	Date	The message received date, yyyy-MM-dd 'T'HH:mm:ssZZZZ.

8.4.3 Google Drive links sidecar fields

Google Drive link-s.csv field	Relativity source field name	Field type	Description
Account	N/A	N/A	This field is not used for drive links overlay.
Rfc822MessageId	N/A	N/A	This fields is not used for drive links overlay.
GmailMessageId	GoogleDrive/LinkingGmailMessageIds	Long Text	Populated for Google Drive Items. This fields contains a semicolon-delimited list of messages within the same collection that contained a link to the Google Drive item.
DriveItemId	Google/LinkedDriveItemIds	Long Text	Populated for Gmail messages. This field contains a semicolon-delimited list of Google Drive items within the same collection that was linked to the Gmail message.
DriveUrl	Google/LinkedDriveUrls	Long Text	Populated for Gmail messages. This field contains a pipe-delimited list of URLs for Google Drive items within the same collection that was linked to the Gmail message.

8.4.4 Google Classic Hangouts metadata fields

The following table lists the Google Class Hangouts metadata fields found in the Google Gmail .csv file. Use these fields for mapping Google Classic Hangouts data.

Google mail .csv field	Relativity source field name	Field type	Description
SubjectAtStart	Google/SubjectAtStart	Long Text	The subject of the conversation when the first message was sent.
SubjectAtEnd	Google/SubjectAtEnd	Long Text	The subject of the conversation when the last message was sent.

Google mail .csv field	Relativity source field name	Field type	Description
DateFirstMessageSent	Google/DateFirstMessageSent	Date	The time stamp for when the first message in a conversation was sent.
DateLastMessageSent	Google/DateLastMessageSent	Date	The time stamp for when the last message in a conversation was sent.
DateFirstMessageReceived	Google/DateFirstMessageReceived	Date	The time stamp for when the first message in a conversation was received.
DateLastMessageReceived	Google/DateLastMessageReceived	Date	The time stamp for when the last message in a conversation was received.
ThreadedMessageCount	Google/ThreadedMessageCount	Decimal	The number of messages in the conversation.

8.5 Other considerations

- Upon export, Google Vault appends a DocID to the original name of Google Drive files so that it can be matched with Google Workspace metadata found in the .xml file. Google Vault may also truncate the original name portion to avoid file system errors when unzipping the Google Drive file package. The Google Vault exported file name is captured in the metadata .xml FileName tag, along with the original file name in the #Title tag. Relativity records the actual file name using the #Title value, but also captures the Google Drive FileName and Google Drive Title values.
- Google Vault may export Google Drive files that do not contain the original internal created or modified dates. This is especially true of Google Docs, Google Sheets, and Google Slides file exports. Relativity captures the actual created date and time, as well as the last modified date and time, using the #DateCreated and #DateModified values provided in the Google Workspace metadata .xml file. These dates may also appear in publish-calculated fields such as SortDate.
- In some scenarios, Google Vault may generate unique binary files for the same Google Drive file when it is exported multiple times. In other scenarios, Google Vault may generate duplicate binary files for the same Google Drive file when it is exported multiple times, but with different Google Workspace metadata, such as Title, Collaborator, and so forth. For deduplication purposes, Relativity always generates a standard SHA256 hash from the binary of non-email native files, including Google Drive files. As such, use caution when considering whether to perform deduplication of Google Vault Drive data during processing.
- Google Vault Drive exports contain a Google-calculated hash, exclusive to Google Docs, Google Sheets, and Google Slides, which is referenced in the metadata .xml file as SourceHash. This field is captured during processing and can be used post-processing, in conjunction with the Processing duplication workflow, to identify duplicate Google Docs, Google Sheets, and Google Slides files within the Relativity workspace.
- Google Drive and Gmail sidecar xml includes information about data classification labels applied to the source data. Relativity does not capture this information.
- Google Drive and Gmail sidecar xml includes query parameters used to generate the Google Vault export. Relativity does not capture this information.

- Google Vault exports of Google Drive data into a file named `[EXPORT_NAME]-custodian-docid.csv`. Relativity does not capture any information in this file; it is processed as a loose individual file unless removed before processing.
- Google Vault exports also typically include a `-results-count.csv`, `-errors.xml`, and `.md5` file. Relativity does not capture any information in these files; they are processed as loose individual files unless removed before processing.
- Beyond those mentioned here, Google Vault collections can contain other data types, some of which may be included within Google Drive or Gmail exports. For a complete list of these data types, see the Google support article [Supported services & data types](#).

8.6 Resources

To learn more about using Google Vault to collect Google Workspace data, see Google's resources on [Google Vault Help](#).

9 Other software support

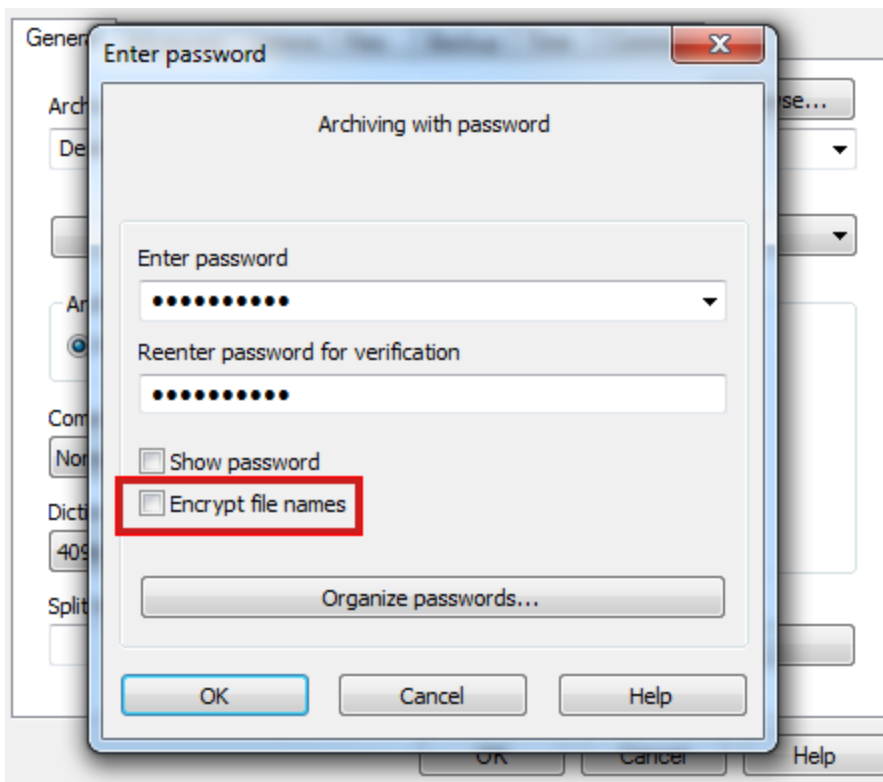
This topic covers additional supported file type (for processing) information for other software considerations. To view the Relativity supported and unsupported file type tables, see [Supported file types for processing](#).

9.1 Multi-part forensic file considerations

When processing a forensic image with multi-part extents, be sure to include all parts within the same directory. Not doing so will cause inventory and discovery to fail.

9.2 Password-protected .rar files

Processing does not decrypt a file that gets its encryption directly from the .rar file that contains it. This means that if you attempt to process a password-protected .rar file where the **Encrypt file names** property is checked, Processing is unable to extract the files inside that archive.



In addition, Processing can extract a single password-protected file from a .rar file, but not multiple password-protected files in the same archive.

The following table breaks down Processing's support of password-protected .rar files.

- ✓—Processing will decrypt the file.
- Empty—Processing will not decrypt the file.

Archive type	Single password-protected file	Multiple password-protected files	Encrypt file names property
.rar	✓		
Multi-part .rar	✓		

9.2.1 Lotus Notes

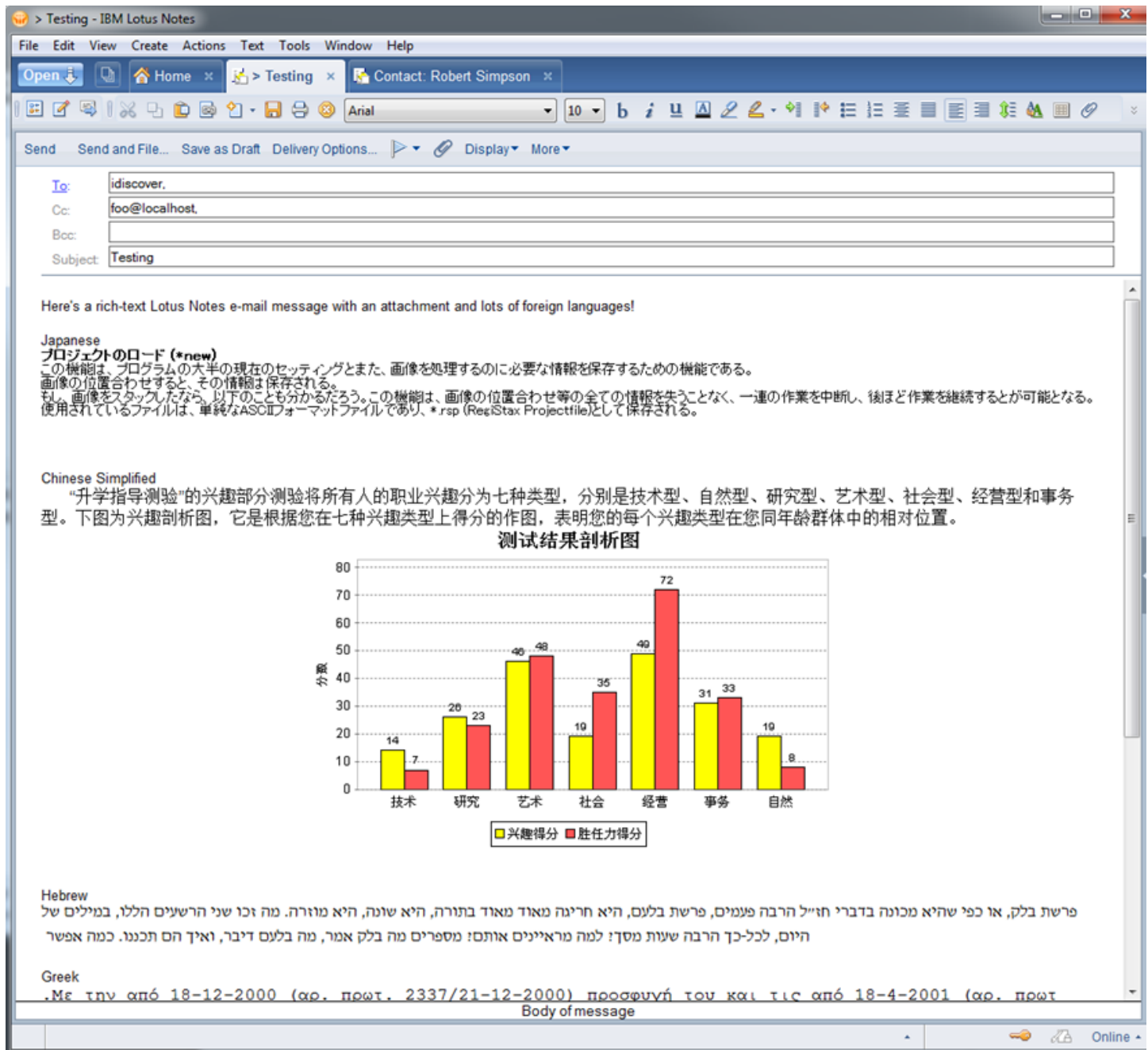
Note the following about how Processing handles note storage facility files:

- Processing does not perform intermediate conversion on .nsf files, meaning that they are not converted to .pst or .dxf files before discovering them. This ensures that document metadata is not missed during processing.
- Processing preserves the original formatting and attachments of the .nsf file. In addition, forms are not applied, since they are designed to hide information.
- Processing extracts the contents of .nsf files and puts them into individual message files using the Lotus Notes C/C++ API directly. This is because .nsf files do not have their own individual document entry file format. All of the original Lotus Notes metadata is embedded in the message, meaning if you look at the document metadata in an .nsf file within Lotus, all of the metadata listed is embedded in the message. In addition, the original Rich Text Format/HTML/Plaintext document body is written to the message. Relativity handles the conversion from .nsf to .msg files itself, and any errors regarding metadata or the inability to translate content are logged to the

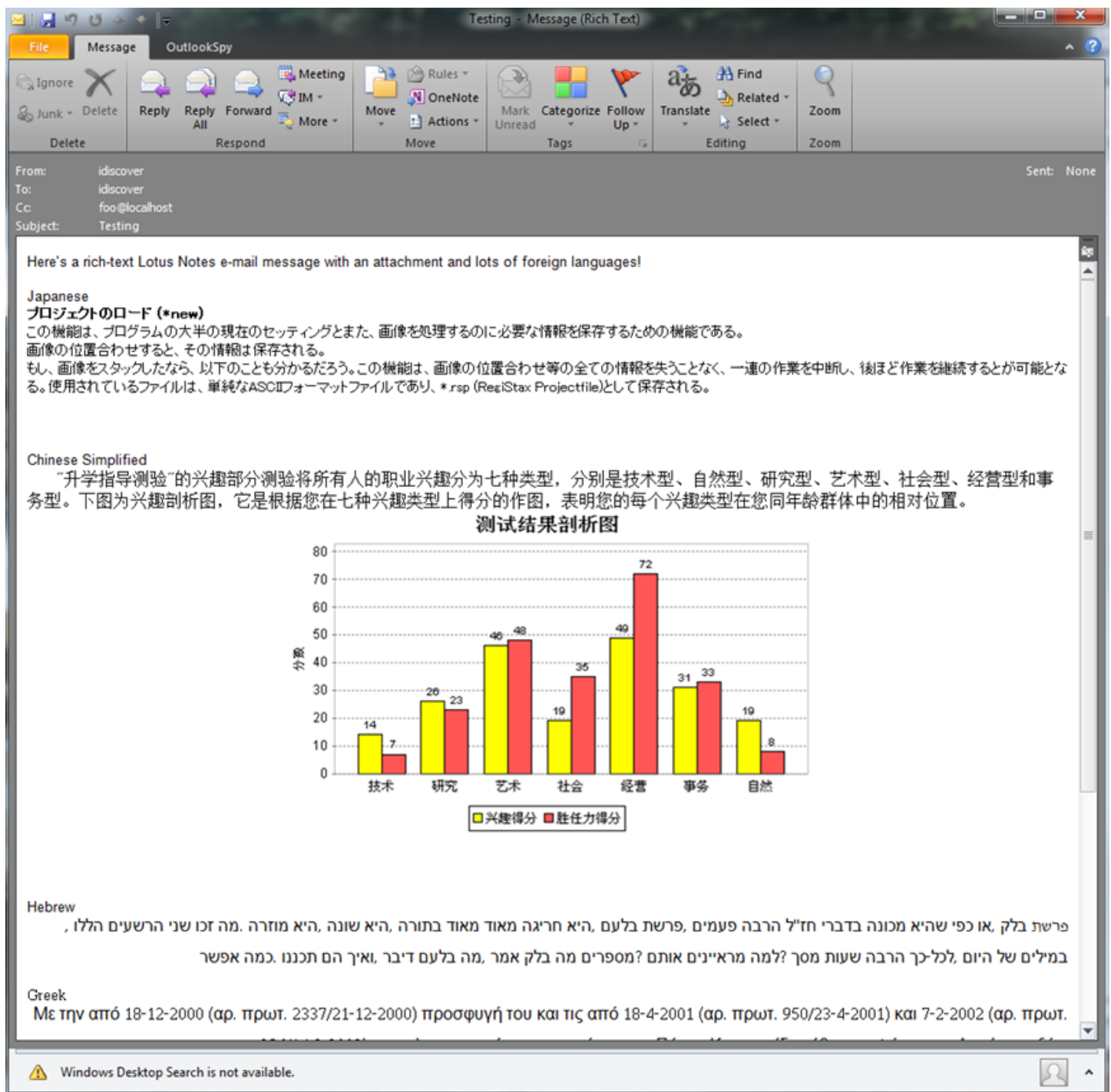
processing Errors tab. Relativity can process the following .nsf items as messages:

- Contacts
- Distribution lists
- Calendar items
- Emails and non-emails

This is an example of an original .nsf file before being submitted to the processing engine:



This is an example of an .nsf file that has been converted to a message:



9.2.1.1 Lotus Notes supported versions

Lotus Notes are supported through Version 10.

9.3 Multi-part container considerations

When processing a multi-part container, the first part of the container must be included. If the first part of the container is not included, the Processing engine ignores the file.

9.4 Calendar file, vCard file considerations

Calendar files (.ics) and vCard files (.vcf) are de-duplicated not as emails but as loose files based on the SHA256 hash. Since the system now considers these loose files, Relativity is no longer capturing the email-specific metadata that it used to get as a result of .ics or .vcf files going through the system's email handler.

The following table breaks down which metadata values the system populates for .ics files:

Processing engine property name	Relativity property name
Author	Author
DocTitle	Title
Email/AllDayEvent	[other metadata]
Email/AllowNewTimeProposal	[other metadata]
Email/BusyStatus	[other metadata]
Email/CommonEnd	[other metadata]
Email/CommonStart	[other metadata]
Email/ConversationTopic	[other metadata]
Email/CreatedOn	Email Created Date/Time
Email/DisplayTo	[other metadata]
Email/DomainParsedBCC	Recipient Domains (BCC)
Email/DomainParsedCC	Recipient Domains (CC)
Email/DomainParsedFrom	Sender Domain
Email/DomainParsedTo	Recipient Domains (To)
Email/Duration	[other metadata]
Email/EndDate	Meeting End Date/Time
Email/IntendedBusyStatus	[other metadata]
Email/IsRecurring	[other metadata]
Email/LastModified	Email Last Modified Date/Time
Email/Location	[other metadata]
Email/MessageClass	Message Class
Email/MessageType	Message Type
Email/NetMeetingAutoStart	[other metadata]
Email/ReminderMinutesBeforeStart	[other metadata]
Email/SentRepresentingEmail	[other metadata]
Email/SentRepresentingName	[other metadata]
Email/StartDate	Meeting Start Date/Time
EmailBCC	[other metadata]
EmailBCCName	[other metadata]

Processing engine property name	Relativity property name
EmailBCCSmtp	BCC (SMTP Address)
EmailCC	[other metadata]
EmailCCName	[other metadata]
EmailCCSmtp	CC (SMTP Address)
EmailConversation	Conversation
EmailFrom	[other metadata]
EmailImportance	Importance
EmailSenderName	Sender Name
EmailSenderSmtp	From (SMTP Address)
EmailSensitivity	Email Sensitivity
EmailSubject	Subject
EmailTo	[other metadata]
EmailToName	Recipient Name (To)
EmailToSmtp	To (SMTP Address)
SortDate	Sort Date/Time
Subject	[other metadata]

The following table breaks down which metadata values the system populates for .vcf files:

Processing engine property name	Relativity property name
DocTitle	Title
Email/BusinessAddress	[other metadata]
Email/BusinessAddressCity	[other metadata]
Email/BusinessAddressCountry	[other metadata]
Email/BusinessAddressPostalCode	[other metadata]
Email/BusinessAddressState	[other metadata]
Email/BusinessAddressStreet	[other metadata]
Email/BusinessPostOfficeBox	[other metadata]
Email/BusinessTitle	[other metadata]
Email/CellNumber	[other metadata]
Email/CompanyName	[other metadata]
Email/ConversationTopic	[other metadata]
Email/Country	[other metadata]
Email/Department	[other metadata]
Email/DisplayName	[other metadata]
Email/DisplayNamePrefix	[other metadata]

Processing engine property name	Relativity property name
Email/Email2AddrType	[other metadata]
Email/Email2EmailAddress	[other metadata]
Email/Email2OriginalDisplayName	[other metadata]
Email/Email3AddrType	[other metadata]
Email/Email3EmailAddress	[other metadata]
Email/Email3OriginalDisplayName	[other metadata]
Email/EmailAddrType	[other metadata]
Email/EmailEmailAddress	[other metadata]
Email/EmailOriginalDisplayName	[other metadata]
Email/FileUnder	[other metadata]
Email/Generation	[other metadata]
Email/GivenName	[other metadata]
Email/HomeAddress	[other metadata]
Email/HomeAddressCity	[other metadata]
Email/HomeAddressCountry	[other metadata]
Email/HomeAddressPostalCode	[other metadata]
Email/HomeAddressState	[other metadata]
Email/HomeAddressStreet	[other metadata]
Email/HomeNumber	[other metadata]
Email/HomePostOfficeBox	[other metadata]
Email/Locality	[other metadata]
Email/MessageClass	Message Class
Email/MessageType	Message Type
Email/MiddleName	[other metadata]
Email/OfficeNumber	[other metadata]
Email/OtherAddress	[other metadata]
Email/OtherAddressCity	[other metadata]
Email/OtherAddressCountry	[other metadata]
Email/OtherAddressPostalCode	[other metadata]
Email/OtherAddressState	[other metadata]
Email/OtherAddressStreet	[other metadata]
Email/OtherPostOfficeBox	[other metadata]
Email/PostOfficeBox	[other metadata]
Email/PostalAddress	[other metadata]

Processing engine property name	Relativity property name
Email/PostalCode	[other metadata]
Email/PrimaryFaxNumber	[other metadata]
Email/PrimaryNumber	[other metadata]
Email/State	[other metadata]
Email/StreetAddress	[other metadata]
Email/Surname	[other metadata]
EmailConversation	Conversation
EmailSubject	Subject
Subject	[other metadata]

10 Supported short message sources for RSMF conversion

This topic answers common questions about which data sources can be converted to Relativity Short Message Format (RSMF).

10.1 RSMF overview

RSMF is the file format Relativity Processing uses to represent conversational data. Supported data types include chat messages, SMS, and instant messages. Relativity converts the source data to RSMF during Processing.

10.2 Sources that can be converted to RSMF

There are two paths to RSMF in Relativity:

- File-based Processing workflows, where you export data from the source and upload it to Relativity for conversion.
- Relativity Collect integrations, where data flows directly into Relativity in RSMF format without a separate export step.

10.3 Sources supported through file-based Processing

Source types supported by Relativity through the file-upload processing workflow include the following (each has a dedicated documentation page covering required input formats, processing profile settings, and known limitations).

- **Microsoft Teams**—exported as PST from Microsoft 365. See [Short message conversion for Microsoft Teams](#).
- **Slack**—exported as JSON (organization-level or workspace-level). See [Short message conversion for Slack](#).
- **Cellebrite UFDR (Universal Forensic Data Recovery)**—produced by Cellebrite Physical Analyzer. See [Short message conversion for Cellebrite](#).
- **Google Chat**—produced by Google Vault. See [Short message conversion for Google Chat](#).

10.4 Relativity Collect sources

Relativity converts some sources to RSMF as part of a Relativity Collect integration. In this workflow, Relativity Collect connects directly to the source platform, retrieves the data, and produces RSMF output without requiring a separate export-and-upload step.

10.5 Sources not supported

The following sources cannot currently be converted directly to RSMF.

10.5.1 Microsoft Teams — HTML exports (including Microsoft Purview)

Relativity does not support Teams data exported as HTML for RSMF conversion. Microsoft Purview's default Teams export format changed to HTML in May 2025. If you are collecting Teams data through Microsoft Purview, configure the export to produce PST output instead of HTML. Then submit the data to Relativity.

10.5.2 Microsoft Teams — MSG files

Relativity does not support Individual MSG files as a Teams input format. Export Teams data as a complete PST.

10.5.3 Slack data embedded in a PST container

Slack data wrapped inside a PST or other email-container format is not supported. Provide Slack data as the direct JSON export from Slack's standard export workflow.

10.5.4 WhatsApp standalone exports

WhatsApp data exported directly from the application—such as a chat archive or device backup—is not supported for direct RSMF conversion.

WhatsApp data collected through a Cellebrite UFDR extraction is processed through the standard UFDR workflow. It may produce RSMF output as part of the chat apps category. See [Short message conversion for Cellebrite](#) for more details.

10.5.5 Facebook account downloads

Relativity does not support data downloaded from Facebook accounts (the "Download Your Information" export) for RSMF conversion.

10.5.6 Non-Cellebrite mobile forensic tools

Extractions produced by forensic tools other than Cellebrite Physical Analyzer — including Oxygen Forensic Detective and other vendor-specific formats — are not natively supported. The following alternative paths are available:

- **Export to RSMF directly from your tool.** If your forensic tool offers an RSMF export option, export the data as RSMF and import the file directly into Relativity. No short message conversion job is required.
- **Re-extract using Cellebrite Physical Analyzer.** If you have access to the original device or forensic image, re-extract the data to produce a supported UFDR file.
- **Submit a feature request.** This helps the product team in tracking demand for native support of your tool.

10.5.7 AI conversation tool exports

Conversation exports from AI tools — including Microsoft Copilot, ChatGPT, Claude AI, Fathom AI, and Perplexity — are not supported for direct RSMF conversion.

If AI conversation content appears inside a Cellebrite UFDR extraction (for example, captured during mobile forensic extraction of a browser session), the data is processed as part of the UFDR workflow but will not render as structured short messages in the output.

10.6 Unlisted sources

If your source is not listed in the preceding sections, the following options may apply.

- **If your source can export to RSMF directly**, import the RSMF file into Relativity. No short message conversion job is required.
- **If your data is collected through a Cellebrite UFDR**, follow the standard UFDR workflow regardless of the original chat application. RSMF support depends on the UFDR's content, not on the original app.
- **If neither applies**, submit a feature request so the product team can track demand for the source format.

11 RSMF slicing configuration reference

This topic explains where Relativity Short Message Format (RSMF) slicing happens, what the **Slice by** processing profile setting does, and how to slice a conversation again after processing.

11.1 RSMF slicing overview

Slicing splits one conversation into more than one RSMF document. Each resulting document is a slice. A slice behaves like any other document in your workspace, meaning you can review, image, and produce it.

Conversations are split for two different reasons:

- **Time-based slicing**—Relativity groups messages into time blocks during processing, so a conversation that runs for months becomes a set of documents sized for review.
- **Reviewer-driven slicing**—a reviewer selects specific events in the Short Message Viewer and saves them as a new document.

Slicing differs from filtering. Filtering changes which events the Short Message Viewer displays. Slicing creates a new document.

11.2 Where slicing happens

Four separate things can split a conversation. The processing profile controls only one of them. Identifying which one applies to your data answers most slicing questions.

Split point	What controls it	When it occurs
Source or collection tool	The export settings of the tool that generates the RSMF.	Before the file reaches Relativity.
Short message conversion	The Slice by setting on the processing profile.	During processing.
RSMF file limits	Fixed limits for file size and event count.	During processing.
Short Message Viewer	Reviewer event selection.	After publish.

11.2.1 Slicing at the source

Some tools slice a conversation as they generate the RSMF. When this happens, the file arrives in Relativity already sliced, and the slice boundaries reflect the settings of that tool rather than any Relativity setting.

Short message conversion settings apply only to processing jobs that convert source data into RSMF. They have no effect on data that is already in RSMF format when processing begins. A **Slice by** value therefore does not change files that arrive as RSMF.

This distinction matters most for Cellebrite data, which reaches Relativity by one of two routes:

- **You upload a Universal Forensic Extraction Device Report (UFDR) file.** Relativity converts the UFDR to RSMF during processing and applies the **Slice by** value from the profile.
- **You upload RSMF files exported from Cellebrite Physical Analyzer.** Cellebrite has already sliced the conversations. Relativity preserves those boundaries, and the **Slice by** value has no effect.

To apply Relativity slice boundaries to Cellebrite data, process the original UFDR file rather than exported RSMF files. For supported input formats, see [Short message conversion for Cellebrite](#).

11.2.2 Slicing during short message conversion

When Relativity converts Slack, Microsoft Teams, Cellebrite, or Google Chat source data into RSMF, it groups events into time blocks defined by the **Slice by** setting. See [Short message conversion settings](#) in the Processing profiles topic for details.

11.2.3 Splitting to meet RSMF file limits

Relativity splits a conversation into more files when it exceeds either of these limits, regardless of the **Slice by** value:

- A maximum of 10,000 events per file, which is a Short Message Viewer condition.
- A maximum size of 2 GB per file, which is a processing condition.

A conversation that is under the **Slice by** interval can still produce several documents for this reason.

11.2.4 Slicing after publish

Reviewers select individual events in the Short Message Viewer and save them as a new document. See RSMF slicing for more information.

11.3 The Slice by processing profile setting

Slice by determines how Relativity splits conversations into RSMF files by time. You configure it in the [Short Message Conversion Settings](#) section of the processing profile.

Slice by is required for RSMF conversion, and the profile layout marks it with an asterisk. Select a value whenever short message conversion is turned on.

11.3.1 Supported Slice by intervals

A slice begins with the first message and ends when the next interval boundary begins. You set the interval length with the **Slice by** value in the processing profile; the default is 24 hours. Slices do not extend past midnight.

Value	How Relativity groups events
4 hours	Messages are grouped into 4-hour intervals aligned to midnight. With a 4-hour interval, the boundaries are midnight, 4 AM, 8 AM, 12 PM, 4 PM, and 8 PM.

Value	How Relativity groups events
	For example: • If the first message occurs at 5 PM, the slice ends at 8 PM. • If the first message occurs at 11 PM, the slice ends at midnight.
8 hours	Messages are grouped into 8-hour intervals aligned to midnight. With a 8-hour interval, the boundaries are midnight, 8 AM, and 4 PM. For example: • If the first message occurs at 7 AM, the slice ends at 8 AM. • If the first message occurs at 5 PM, the slice ends at midnight.
12 hours	Messages are grouped into 12-hour intervals aligned to midnight. With a 12-hour interval, the boundaries are midnight and 12 PM. For example: • If the first message occurs at 10 AM, the slice ends at 12 PM. • If the first message occurs at 8 PM, the slice ends at midnight.
24 hours (default)	Messages are grouped into 24-hour slices, beginning at 00:00 on the calendar date that opens the slice. This is the default setting. If the Slice by value is empty, the time interval defaults to 24 hours.
1 week	Messages are grouped into 7-day slices, with Monday at 00:00 being the anchor. For example: • If the first message occurs on Monday at 7 AM, the slice ends at 00:00 the following Monday. • If the first message occurs on Friday at 5 PM, the slice ends at 00:00 the following Monday.

11.3.1.1 Time gaps in messages

Slices do not run continuously from the start of the conversation. Time gaps in message activity do not result in empty slices. A new slice begins only when the next message occurs.

For example:

A conversation contains messages on March 3, 2026 starting at 1:30 PM, further messages on the morning of March 4, 2026, then no activity until March 12, 2026. With **Slice by** set to 24 hours, Relativity produces three documents:

Slice	Window	Events included
1	March 3, 2026, 00:00 to 4 March, 2026, 00:00	Messages sent on March 3, 2026.
2	March 4, 2026, 00:00 to 5 March, 2026, 00:00	Messages sent on March 4, 2026.
3	March 12, 2026, 00:00 to 13 March, 2026, 00:00	Messages sent on March 12, 2026.

There are no slices for March 5, 2026 through March 11, 2026, because those dates do not contain messages. The 1:30 PM time stamp of the first message does not shift the ending slice boundary.

11.3.2 Time zone handling

Slicing uses the processing time zone and shifts slice boundaries to match it. Set **Default time zone** on the processing profile before you run the job.

The date range filter in the Short Message Viewer works differently. It displays timestamps in the time zone of the local machine, whichever time zone the processing profile uses. Account for the offset when you filter by date.

11.3.3 Orphaned messages

When slicing divides a conversation across time blocks, messages from one thread are stored in different documents. A message separated from its parent conversation becomes an orphan, and Relativity adds a note at the start of the message to identify it. Longer intervals produce fewer orphaned messages.

11.4 Changing slice boundaries after processing

It is possible to change slicing after processing by the following:

11.4.1 Reprocess with a different Slice by value

To change the **Slice by** value on processed files, update the value in the processing profile and run the job again.

11.4.2 Regenerate converted data after publish

To replace converted data that is already published, use the post-publish delete workflow to remove the published documents, then reprocess the source data with the interval you want. See [Post-publish delete](#) for more details.

Post-publish delete recalculates deduplication and evaluates documents individually rather than by processing set or data source. A document that belongs to a deduplication family shared with another data source is retained. Review the deduplication effects for your workspace before you start.

11.4.3 Slicing published documents

After documents are published, reviewers create slices in the Short Message Viewer by selecting events and saving them as a new document. The new document takes the control number of the original with **_sliceXXX** appended, and a banner in the Viewer links back to the original document.

For the full procedure, the permissions required, the fields a slice populates, and the audits slicing generates, see RSMF Slicing.

Two options on the **Save as new document** modal affect what the slice carries forward:

- **Copy Layout data**—copies coding decisions saved on the selected layout to the slice. A field that is absent from the layout is not populated on the slice.
- **Preserve all participants**—includes every participant from the original conversation in the metadata of the slice. When this option is turned off, the slice includes only the participants whose events you selected.

11.4.4 Slicing one document at a time

Reviewers slice documents individually. Relativity has no mass operation for slicing published documents, so a large volume of documents requires a separate operation for each one. For a conversation that spans a long date range, this takes considerable time.

Two options address this:

- **Set the interval during processing.** Choosing a **Slice by** value that matches how you intend to review the data is faster than slicing documents individually after publish. Use this approach whenever you control the processing job.
- **Automate slicing through the API.** The Short Message Viewer Manager services expose a slice method that accepts a list of event indexes, so a developer can script slicing across many documents. See Short Message Viewer Manager (REST).

11.4.5 Troubleshooting slice errors

The errors in this section are slice operation failures rather than configuration questions.

Error	Description
The slice exceeds the identifier field length	A slice control number consists of the original control number plus _sliceXXX, and attachments in a slice receive a further _XXX. When the resulting value is longer than the identifier field permits, the slice fails. Contact Relativity Support with the workspace name and the control number of the original document.
The slice fails on a layout field	Slicing copies fields from the layout you select in Copy Layout data. Slice the document without copying layout data to confirm whether a specific field causes the failure, then contact Relativity Support with the layout name.
A participant validation warning appears	When the participants in the RSMF email-style header differ from the participants in the manifest, Relativity records a validation warning. The warning does not stop processing or publishing.

11.5 Resources

- RSMF Slicing
- [Processing profiles](#)
- [Supported short message sources for RSMF conversion](#)
- Short Message Viewer
- [Post-publish delete](#)
- Processing an RSMF file

12 Working with RSMF files built outside Relativity

RSMF is an open container format that you can build with any tool outside of Relativity. Common sources include forensic suites, third-party collection applications, and in-house scripts. Relativity can ingest and render these externally built RSMF files, but the Short Message Viewer can only display the data that the external builder included in the RSMF manifest. If the manifest is missing elements, those elements do not appear in the viewer. This topic covers how Relativity renders externally built RSMF files, how to validate them, and how attachment linking works. It also describes the supported manifest elements and options for remediating missing content.

Related topics:

- Processing an RSMF file
- Short message conversion overview

12.1 How Relativity renders an externally built RSMF (rendering contract)

Relativity renders an RSMF file strictly according to its `rsmf_manifest.json`, validated against the published RSMF schema. The Short Message Viewer displays only the content the builder wrote into the manifest—no more

and no less.

The `rsmf_manifest.json` has four root properties: `version`, `participants`, `conversations`, and `events`. Each event can optionally carry:

- `reactions[]`—emoji reactions, including value, count, and participant list. See Reaction in the Relativity short message format topic for more information.
- `attachments[]`—files linked by ID to entries in `rsmf.zip`.
- `edits[]`—a sorted history of message edits (newest to oldest).
- Read receipts

All of these properties are optional in the RSMF schema. If an external builder did not emit a property, that property does not render in the Short Message Viewer.

Note: When an external builder omits or misreferences a supported element—such as reactions, attachments, edits, read receipts, or participant lists—that element silently fails to appear in the viewer, or renders as an error, even though Relativity fully supports that element. This is a source-construction gap, not a Relativity viewer or processing defect. Supported elements cannot be inferred, backfilled, or synthesized from other fields; the builder must populate them explicitly in the manifest.

12.2 Validating an externally built RSMF file

Validate an externally built RSMF file using the open-source RSMF validator before or immediately after ingestion. The validator checks `rsmf_manifest.json` against the published RSMF schema and verifies that attachments are correctly cross-referenced between the manifest and `rsmf.zip`.

The validator and schema are available at <https://github.com/relativitydev/rsmf-validator-samples>.

See Viewer interface for more information.

12.2.1 Running the validator

You can run the RSMF validator in two ways:

GUI/CLI application

The repository's `bin` directory contains a standalone application that accepts an RSMF file as input and returns a validation result. The result is one of the following:

Result	Meaning
Passed	The RSMF is valid.
PassedWithWarnings	The RSMF is structurally valid but has conditions that may affect rendering, such as participant list mismatches.
Failed	The RSMF has schema violations or attachment errors that will prevent correct rendering.

In-Viewer validator

You can also validate a single RSMF file from within the Relativity Short Message Viewer. This method validates the loaded file in place and surfaces the same error and warning categories without requiring a separate tool.

12.2.2 Reading attachment and participant errors

The validator produces two categories of attachment errors:

Error message	Meaning
The attachment with ID {id} was not found in the zip archive	The manifest references a file by ID, but that file is absent from <code>rsmf.zip</code> . The attachment will not render in the Viewer.
The file {name} in the zip archive does not have a reference	A file exists in <code>rsmf.zip</code> but is not linked from any event in the manifest. The file will not appear as a linked attachment in the Viewer; it remains accessible only through the <code>rsmf.zip</code> folder.

Participant warnings—for example, `X-RSMF-Participants` has different members than the JSON `participants list`—indicate a mismatch between the RSMF file header and the manifest participant list. This condition does not block rendering but can cause unexpected participant display behavior in the Viewer.

12.2.3 Saving an error report

The validator includes a **Save Error Report** feature that generates a report of all errors and warnings found during validation. Use this report when opening a support case with Relativity.

Important: Error reports may contain personally identifiable information (PII) extracted from the RSMF content. Scrub all PII from the report before sending it to Relativity Support.

12.3 Attachment linking requirements

For an attachment to appear in the Short Message Viewer as a family member of its parent message, both of the following conditions must be true:

1. The event entry in `rsmf_manifest.json` must include an `attachments[]` array where each entry's `id` field contains the exact filename—including extension—of the corresponding file in `rsmf.zip`.
2. The file named by that `id` must be present in `rsmf.zip`.

When either condition is not met:

- If the manifest references a file absent from `rsmf.zip`, the attachment link is broken and the file does not appear in the Viewer.
- If a file is present in `rsmf.zip` but is not referenced by any event, it is not promoted to a family member. It remains accessible through the `rsmf.zip` folder but does not render as a linked attachment.

The `id` value is case-sensitive and must exactly match the filename in the archive. Mismatches in case, path prefix, or extension between the manifest and the archive are a common source of attachment rendering failures in externally built RSMF files.

The following excerpt illustrates correct attachment linking in `rsmf_manifest.json`:

```
"events": [
  {
    "id": "msg-001",
    "type": "Message",
    "body": "See the attached document.",
    "participant": "user-a",
    "timestamp": "2024-03-15T10:22:00Z",
    "conversation": "conv-1",
```

```

      "attachments": [
        {
          "id": "report_final.pdf",
          "display": "report_final.pdf",
          "size": 204800
        }
      ]
    }
  ]
}
]

```

In this example, `rsmf.zip` must contain a file named exactly `report_final.pdf`. If the file in the archive is named `Report_Final.pdf` or `report_final.PDF`, the link will fail.

12.4 Supported elements require explicit manifest population

Relativity supports a rich set of message elements—reactions, edits, read receipts, participant metadata, and custom fields—but can display only elements the builder explicitly wrote into the manifest. Supported elements are not inferred from partial data or derived from source file properties at render time.

For a complete list of supported manifest fields, their behavior by source type, and known limitations, see the **RSMF Field Reference**. For authoritative field definitions and the full schema, see the published RSMF schema in the validator repository at <https://github.com/relativitydev/rsmf-validator-samples>.

If a supported element is absent from the manifest of an externally built RSMF, the builder must add it before the RSMF is loaded into Relativity. Missing manifest fields cannot be retroactively populated after ingestion.

12.5 Remediation options

When an externally built RSMF fails validation or renders incompletely, you have the following options.

12.5.1 Option 1: Re-route the source data through Relativity's short message conversion

If the source data is a forensic archive—for example, a Cellebrite UFDR file—routing it through Relativity's native short message conversion produces a schema-conformant RSMF. Relativity's conversion pipeline explicitly populates all supported elements for the source type and produces output that passes RSMF schema validation.

This option is available only when you have access to the original source archive. It is not applicable when the source data exists solely as an externally built RSMF with no access to the underlying forensic export.

For supported source types and processing instructions, see the short message conversion overview.

12.5.2 Option 2: Correct and rebuild the RSMF

If re-routing through Relativity's conversion is not possible, correct the RSMF at the source. Use the validator error report to identify specific schema violations and attachment mismatches, then follow these steps:

1. Fix the errors identified by the validator in `rsmf_manifest.json`.
2. Confirm that every attachment `id` value in the manifest has an exactly matching filename in `rsmf.zip`.

3. Add any missing supported elements—reactions, edits, read receipts, participant metadata—to the manifest.
4. Re-validate the corrected RSMF using the GUI/CLI validator before re-ingesting it into Relativity.

Contact the tool vendor—for example, Cellebrite, Oxygen, or ModeOne—or the team that wrote the in-house conversion script for assistance correcting the manifest output.

Note: Relativity Support can help you interpret validator output and identify the specific source of construction gaps in an externally built RSMF. Relativity Support cannot modify the manifest of an externally built RSMF on a customer's behalf.

13 Password bank

The Password Bank is a password repository used to decrypt certain password-protected files during inventory, discovery, and basic and native imaging. By creating a password bank, Relativity can run passwords against each encrypted document until it finds a match. Likewise, when you run an imaging job, mass image, or use image-on-the-fly, the list of passwords specified in the bank accompanies that job so that the encrypted files are imaged in that job. The password bank potentially reduces the number of errors in each job and eliminates the need to address password errors outside of Relativity.

Notes:

- You will find the Password Bank tab under both Imaging and Processing applications, if both are installed.
 - While the Password Bank documentation resides under the Processing heading, passwords stored in the bank are available to other Relativity features outside of Processing. For example, you upload a set of production files from another party that used Relativity. Some files are password-protected, and the passwords are stored in the Password Bank. Even though you are not processing the files, Relativity still opens the files using passwords from the bank.
-

13.1 Password bank in processing workflow

The following steps illustrate how the password bank typically fits into the processing cycle.

1. Create a password bank that includes a list of passwords that correspond with the files you intend to process.
2. Create a processing set and add data sources that contain the encrypted documents to the set.
3. Start inventory or discovery on the files in the data sources attached to the processing set.
4. All passwords supplied to the password bank become synced via an agent and accompany the job as it goes to the processing engine.
5. The processing engine discovers the files in the processing set. If the file is encrypted, Relativity checks the Password Bank to see if a password exists for the file. If a password exists, Relativity uses the password to open the file and extract the text, metadata, and native/basic imaging. The native file remains unmodified in its encrypted state, as the password is only used to open and extract content. The extracted text, metadata, and native/basic imaging are not encrypted.
6. Publish the discovered files in the processing set.
7. The published documents are now available for review in the workspace. To view or image any encrypted native files, the password must remain the Password Bank, otherwise you will see an exception.

The following scenario depicts the basic procedure by which you would address exceptions due to password-protected files in a processing set. In this scenario, you would:

1. Run publish on your discovered files.
2. Navigate to the Files tab after publish is complete and locate all exceptions resulting from password protection.
3. Outside of Relativity, locate the passwords designated to unlock those files.
4. Return to Relativity, go to the Password Bank, and create entries for every password that corresponds with the exceptioned files.
5. Run retry on the files that previously resulted in password-protection exceptions.
6. From the Files tab, use the Republish mass action to retry the job.

13.2 Password Bank in imaging workflow

The following steps depict how the Password Bank typically fits into the imaging cycle.

1. You create a password bank that includes a list of passwords that correspond with the files you intend to image.
2. You create an imaging set with the data source that contains the encrypted documents.
3. You start imaging the documents in the imaging set by clicking **Image Documents** in the Imaging Set console.
4. All passwords you supplied to the password bank become synced via an agent and accompany the job as it goes to the imaging engine.
5. The imaging engine images the files in the imaging set and refers to the passwords provided in the password bank. It then sends the imaged files back to Relativity.
6. Once the imaging status changes to Completed, you review and release images from QC review.
7. The imaged documents become available for review in the workspace, along with all the other previously-encrypted documents whose passwords you provided.

Note: The Password Bank works with both Native and Basic imaging.

To view and resolve password-protection errors:

1. After running an imaging set, click View Document Errors in the imaging set console.
2. In Relativity, go to the Password Bank and create entries for every password that corresponds with the files that resulted in errors.
3. Click Retry Errors in the imaging set console to re-image the files that previously resulted in password-protection errors.

The imaged documents, along with other documents that were previously encrypted, become available for review in the workspace.

13.3 Creating or deleting a Password Bank entry

Note: There is no limit on the number of passwords you can add to the password bank. However, having more than 100 passwords could degrade the performance of your processing and imaging jobs.

To create a new entry in the bank:

1. Navigate to **Processing**, and click the **Password Bank** tab.
2. Click **New** on the **Password Entry** category.

3. Complete the fields on the Password Entry Layout. See [Fields](#) for more information.
4. Click **Save**. The entry appears among the others under the Password Entry object.

To delete a password, select the check box next to its name and click **Delete** on the Password Entry object.

Note: When a password is successfully used to decrypt one or more files in a processing set, that password is remembered as the password for those files. Even if the password is subsequently deleted from the bank, it will continue to be used for the discovery of any hash duplicates submitted for processing.

13.3.1 Fields

The Password Bank layout has the following fields:

Password Information ?

Type*

Description*

Password(s)*

Other

Keywords

Notes

- **Type**—the type of password entry you are creating. The options are:
 - **Passwords**—any file that you want to decrypt that is not grouped with the three other types of Lotus Notes, Email encryption certificates, or AD1 encryption certificates.
 - Although you are able to process EnCase Logical Evidence files, the password bank does not support password-protected Encase files.
 - When you select this type, you must enter at least one password in the Passwords field to save.
 - The password bank does not support Microsoft OneNote files.
 - For imaging jobs, this is the only relevant option for a password entry.
 - For imaging and processing jobs, a slipsheet is not created automatically for documents that are password-protected. However, you can create an image outside of Relativity and use the password-protected document's control number as the image key to then load the image into Relativity through the Import/Export tool and have it display as the image for that encrypted document.
 - **Lotus Notes**—any file generated by Lotus Notes software.
 - Even though the Password(s) field does not display as required, you must enter passwords for all encrypted Lotus Notes files if you want to decrypt them during the processing job. This is because

Lotus Notes files require a matching password and file.

- When you select this type, you must upload a file with an extension of *User.ID* in the Upload file field.
- If you have Processing installed, you can associate a custodian with the Lotus files you upload. To do this, select a custodian from the **Custodians** field, which appears on the layout only when you select Lotus Notes as the type. Doing this syncs the password bank/custodian with the processing engine, which can then access partially encrypted Lotus Notes files. Passwords associated with a custodian have a higher priority.
- For encrypted Lotus documents, Relativity only supports user.id files whose public key size is 630 bits.
- **Email encryption certificate**—files protected by various encryption software certificates.
 - Even though the Password(s) field does not display as required, you must enter passwords for all email encryption certificates if you want to decrypt them during the processing job.
 - When you select this type, you must upload one .pfx or .p12 file in the Upload file field.
 - You can only upload one file per email encryption entry.
- **AD1 Encryption Certificate**—AD1 files protected by an encryption software certificate.
 - Even though the Password(s) field does not display as required, you must enter passwords for all AD1 encryption certificates if you want to decrypt them during the processing job.
 - When you select this type, you must upload one .pfx, .p12, .pem, or .key file in the Upload file field. You receive an exception if you try to upload any other file type.
 - You can only upload one file per AD1 encryption entry.
- **Description**—a description of the entry you are adding to the bank. This helps you differentiate between other entry types.
- **Password(s)**—the one or more passwords you are specifying for the type you selected. Only enter one password per line, and separate passwords with a carriage return. If you enter two passwords on the same line, the password bank interprets the value as a single password.
 - If you select Passwords as the file type, you must add at least one password here to save.
 - You can also add values here if you are uploading certificates that do not have passwords. See [Example password](#).
 - Unicode passwords for zip files are not supported.
 - Relativity bypasses passwords on .pst and .ost files automatically during file discovery. Thus, passwords are not required for these files to get discovered.
- **Upload file**—the accompanying file you are required to upload for Lotus Notes, Email encryption certificate, and AD1 encryption certificate types. If uploading for Lotus Notes, the file extension must be *User.ID* with no exceptions. The file types eligible for upload for the Email encryption certificate type are .pfx and .p12. The file types eligible for upload for the AD1 encryption certificate type are .pfx, .p12, .pem, and .key.

Note: If you save a Powerpoint or Excel document in pre-2007 format, .PPT or .XLS files for example. and the document is read-only, we use the default known password to decrypt the document, regardless of whether or not the password exists in the Password Bank.

13.3.2 Example password

When supplying passwords to the password bank, if you enter:

password@1

bookmark@56

123456

the password bank recognizes three passwords.

If you enter:

password@1

bookmark@56, 123456

the password bank only recognizes two passwords.

13.4 Validations and exceptions

Note the following:

- Including a password that does not belong to a document in your data set does not throw an exception or affect the process.
- A password can unlock multiple files. If you provide the password for a Lotus Notes file that also happens to correspond to a Word file, the password unlocks both files.
- If you delete a password bank entry after submitting a processing or imaging job, you can still complete those jobs.

You may see an exception called **Word template files** while using the password bank. In this case, the password bank cannot unlock an encrypted Word file that was created based on an encrypted Word template where the Word file password is different than the template password, regardless of whether both passwords are in the password bank.

You can resolve password bank exceptions by supplying the correct password to the bank and then retrying those exceptions in their respective processing or imaging jobs.

Note: When you supply a valid password to the password bank, the processing engine extracts metadata and extracted text from the document that the password unlocks. However, when you publish that document, its password security is not removed, in which case it technically remains in an encrypted state even after it is published to the workspace. However, you can view the still-encrypted document in the viewer, because the viewer recognizes that a valid password has been supplied. If the Password Protected field indicates that a document was decrypted, that designation only refers to the fact that you provided a valid password for it to the password bank for the purposes of processing.

13.5 Viewing audits

Every time you send a Password Bank to the processing engine, Relativity adds an audit. The Password Bank object's audit history includes the standard Relativity audit actions of update and run, and a list of all passwords associated with a discovery job at run time.

To view the passwords sent to the processing engine during a job:

1. Navigate to **Processing**, and then click **Password Bank**.
2. Click **View Audit** on the Password Bank layout.

- Click **Details** on the Password Bank history layout.

1 - 25 of 79 25 per page

Action	Field	Old Value	New Value	User Name	Timestamp	Details
Update				Service Account, Relativity	5/29/2025, 11:16 AM	↗
Run				Service Account, Relativity	5/29/2025, 11:16 AM	↗
Update				Service Account, Relativity	5/29/2025, 11:16 AM	↗
Run				Service Account, Relativity	5/29/2025, 11:16 AM	↗
Update				Service Account, Relativity	5/21/2025, 11:03 AM	↗
Run				Service Account, Relativity	5/21/2025, 11:03 AM	↗

- Refer to the **Value** field on the audit details window. Any properties not set on the password bank entry are not listed in the audit.

14 Mapping processing fields

To pull in all of your desired processing data, use the Field Catalog to map your document fields to Relativity's processing data.

System fields versus optional fields

When mapping processing fields, you will encounter Relativity system fields and optional fields. The following describes the differences between the two field types and how you might work with them.

- Relativity **system** fields refer to Relativity-specific fields that are automatically mapped and published by Relativity. Some examples include the Container ID, Custodian, and Extracted Text. You cannot modify Relativity system fields, and you will not see them in the Field Catalog.
- Optional** file fields include fields not considered a Relativity system field and vary from file to file. There are two types of optional file fields:
 - File** system metadata that comprises a file's properties, such as the file name, file size, date created, and file type.
 - Application** metadata stored inside the file itself, such as the author, message heading, and comments. You can only access this data if you have the ability to open the file and extract the content.

Sometimes, you will see a field stored in both Relativity system fields and optional fields, such as Date Created. In this case, Relativity determines which field to keep and which to ignore.

14.1 Mapping fields

To map processing fields, perform the following steps:

- Open the **Fields** tab.
- Click **New Field** or **Edit** on an existing field.
- Provide a name in the **Name** field. We recommend that you give the field an identical name to the one you are mapping to.
- In the Object Type field, select **Document**. Only Relativity document fields are eligible to map to a value in the Source field. Selecting any other object type disables the Source field.
- In the Field Type field, select the type of field to set what type of data can be entered into the field.
- When the Field Type is selected, you will see the menu for Field Settings and Advanced Settings appear. Click the **Advanced Settings** tab.

7. Click **Select** on the **Source** field to display the processing fields to which the Relativity field can be mapped.

Field Settings *

Advanced Settings

Unicode

☒

Open to Associations

☐

Relational

☐

Relational Properties

Propagation

☐

Propagate to

Select

Enable HTML

☐

Keyboard Shortcut

Select

Keywords

Notes

Relativity Application

Select

Source

Author

Select

Clear

List Column Width

Linked in List

☐

Wrapping

☐

8. From the available processing fields, select the one to which you want to map, and click **Set**.
9. Confirm that the field you just mapped appears in the Source field, complete the remaining required fields and click **Save**.

14.1.1 Relativity system field considerations

Note the following regarding Relativity system fields:

- Relativity system fields are mapped by default and cannot be modified.
- Relativity system fields are not listed in the Field Catalog.

A word on Field Catalog source fields

While processing data in an instance, Relativity discovers metadata fields and records them as source fields in the Field Catalog. You can map source fields in the Document object, where the field is then populated when a document is published.

Note: This occurs instance-wide. If one workspace processes a field with a unique metadata name, all other workspaces will see the source field as available for mapping, even if the workspace has never, and possibly will never, process a file with the same field name.

Example

Workspace 1 - processes a file with a unique metadata name, *UniqueData*. The field becomes part of the Field Catalog and is available to all other workspaces in the instance.

Workspace 2 - sees *UniqueData* in the Field Catalog, even though Workspace 2 has never processed a file with the metadata name.

14.1.2 Field mapping validations

When mapping fields, you will receive an error if:

- You attempt to map fields of mismatching types. For example, if you map a whole number field in Processing to a Relativity date field, you will receive an error upon saving the field.
- You attempt to map a fixed-length text field to a catalog field of a longer length.
- You do not have Edit permissions for the Field object. This is because mapping through the Source field is considered an edit to a field. If you only have Add permissions for the Field object and not Edit, and you attempt to map a field, you will receive an error stating, "Error saving field mapping."

14.2 System-mapped fields

The following system-created metadata fields are always populated when data is processed. These fields are automatically mapped and are not available for manual mapping through the Source field on the Field layout:

Note: Relativity system fields refer to Relativity-specific fields that are automatically mapped and published by Relativity. You cannot modify Relativity system fields, and they will not appear in the Field Catalog. As such, the Source field that exposes the Field Catalog will be blank for system fields. Care should be taken when renaming the system fields in your workspace.

Processing Field Name	Source view	Field Type	Description
Container Extension	Other Fields	Fixed-Length Text	Document extension of the container file in which the document originated.
Container ID	Other Fields	Fixed-Length Text	Unique identifier of the container file in which the document originated. This is used to identify or group files that came from the same container.
Container Name	Other Fields	Fixed-Length Text	Name of the container file in which the document originated.
Control Number		Fixed-Length Text	The identifier of the document.
Converted Email Format		Yes/No	Indicates whether an email was converted to MHT, based on the Email Output setting within your processing profile.
Custodian		Single Object	Custodian associated with, or assigned to, the processing set during processing.
Extracted Text	Other Fields	Long Text	Complete text extracted from content of electronic files or OCR data field. This field holds the hidden comments of Microsoft Office files. Extracted Text Size in KB is also an available mappable field outside of the standard fields. This field indicates the size of the extracted text field in kilobytes. To map this field, edit the corresponding Relativity field, open the Other Fields view via the Source control, and select Extracted Text Size in KB.
File Path		Long text	This field is auto-mapped to the native path.

Processing Field Name	Source view	Field Type	Description
			Note: Do not create a custom field with the same name, then map it to the source path (mappable) field. The results will reflect the native path and not the source path, as this field is auto-mapped.
Last Published On		Date	Date on which the document was last updated via re-publish.
Level	Other Fields	Whole Number	Numeric value indicating how deeply nested the document is within the family. The higher the number, the deeper the document is nested.
Originating Processing Set		Single Object	The processing set in which the document was processed.
Originating Processing Data Source		Single Object	A single object field that refers to the processing data source.
Processing Duplicate Hash		Fixed-Length Text	The hash value generated during processing duplication. For loose files, a SHA256 hash is generated from the physical file's SHA256 hash. For emails, a hash is generated from the email's metadata properties. See Calculating deduplication hashes for emails for more details.
Processing File ID	Other Fields	Fixed-Length Text	Unique identifier of the document in the processing engine database.
Processing Folder Path	Other Fields	Long Text	The folder structure and path to the file from the original location, which is used to generate the Relativity folder browser for your documents. This field is populated every time you process documents. See Processing folder path details for more information.
Relativity Attachment ID	Other Fields	Fixed-Length Text	A system field that the Short Message Viewer uses to provide enhanced support for attachments and avatars. See the Relativity Short Message Format guide for more information.
Relativity Native Time Zone Offset	Other Fields	Decimal	A numeric field that offsets how header dates and times appear in the Native Viewer for processed emails. Accepts whole numbers as well as fractional, negative or positive, in the form of 0.00. This field is populated with the UTC offset value of the time zone chosen in the processing profile, and further adjusted by the Daylight Savings Time (DST) offset when applicable. For example, documents processed to Central Time that fall within standard months (CST), are populated with a value of "-6.00" because CST is UTC-6. Documents processed to Central Time during Daylight Savings months (CDT) are populated with a value of "-5.00" because CDT is UTC-5. For more details on this field, see the Admin Guide.
Time Zone Field		Single Object	Indicates which time zone is used to display dates and times on a document image. Note, this does not reflect any applic-

Processing Field Name	Source view	Field Type	Description
			able DST adjustments.
Virtual Path	Other Fields	Long Text	Folder structure and path to file from the original location identified during processing. See Virtual path details for more information.

14.3 Optional fields

The following, optional, metadata fields can be mapped through the Field Catalog. The Field Catalog contains a list of all available fields to map regardless of discovered data.

If you are setting up Processing prior to Discovery and Publish, you have the following options available in the Source field modal:

- Standard Fields—contains a collection of fields from both the Metadata Fields and Other Fields options.
- Metadata Fields—contains fields extracted from the actual file or file system.
- Other Fields—contains static, or Relativity system fields such as control number, processing set name, custodian, and so forth.

Please note:

- The table below includes only the most commonly used optional fields and is not an exhaustive list.
- You can map one processing field to multiple Document object fields.
- Fixed length and long text fields must be Unicode-enabled before you can map to them.
- The following metadata fields can be mapped to similar field types in the Field Catalog. To map different field types outside of the 135 metadata fields to one another, select **Metadata Fields** from the drop-down menu in the Source field modal.
- Consider the following data compatible field types with valid mapping:
 - You can map long text document fields to fixed-length text processing fields. However, Relativity does not support mapping fixed-length text document fields to long text processing fields.
 - You can map single choice Catalog fields to destination fields of fixed-length text, long text, choice, or single object fields.
 - You can map a DateTime field to a Date field if the source field is DateTime and the type of destination field is Date.

Caution: Use caution when mapping to multiple-choice fields. Mapping multiple-choice fields creates a unique value for each choice option. Depending on the multiple-choice fields mapped, you could have hundreds of unique choices, impacting performance.

Processing field/ source name	Source view	Field type	Description	Example value
All Custodians	Standard Fields	Multiple Object	All custodians, de-duped and primary, associated with a file. The All Custodians field is mapped to a document	Lay, Kenneth; Doe, John

Processing field/ source name	Source view	Field type	Description	Example value
			and is updated only when Global or Custodial de-duplication is enabled on the set and the field has been mapped, even if no duplicates exist for the document that was published in the workspace.	
All Paths/Locations	Standard Fields	Long Text	This is the same as DeDuped Paths except that the virtual path of the current document is appended to the end of the list. The All Paths/Locations field is populated only when Global or Custodial de-duplication is enabled on the set and the field has been mapped, even if no duplicates exist for the document that was published in the workspace.	Lay, Kenneth \Lay, Kenneth\kenneth_lay_000_1_2_1.pst \lay-k\Kenneth_Lay_Dec2000\Notes Folder-s\Notes inbox; Doe, John \Doe, John\John_Doe_000_1_2_1.pst \Doe-J\John_Doe_Dec2000\Notes Folder-s\Discussion threads
Attachment Document IDs	Standard Fields	Long Text	Attachment document IDs of all child items in family group, delimited by semicolon, only present on parent items.	KL0000000031.0001;KL00000000031.0002
Attachment List	Standard Fields	Long Text	Attachment file names of all child items in a family group, delimited by semicolon, only present on parent items.	EC PRC Meeting Agenda.doc;Map to The St.Regis.doc
Author	Standard Fields	Fixed-Length Text (50)	Original composer of document or sender of email message.	Jane Doe
BCC	Standard Fields	Long Text	The names, when available, and email addresses of the Blind Carbon Copy recipients of an email message.	Capellas Michael D. [Michael.Capellas@COMPAQ.com]
BCC (SMTP Address)	Standard Fields	Long Text	The full SMTP value for the email address entered as a recipient of the Blind Carbon Copy of an email message.	Michael.Capellas@COMPAQ.com

Processing field/ source name	Source view	Field type	Description	Example value
CC	Standard Fields	Long Text	The names, when available, and email addresses of the Carbon Copy recipients of an email message.	Capellas Michael D. [Michael.Capellas@COMPAQ.com]
CC (SMTP Address)	Standard Fields	Long Text	The full SMTP value for the email address entered as a recipient of the Carbon Copy of an email message.	Michael.Capellas@COMPAQ.com
Child MD5 Hash Values	Standard Fields	Long Text	Attachment MD5 hash value of all child items in a family group, only present on parent items. Relativity cannot calculate this value if you have FIPS (Federal Information Processing Standards cryptography) enabled for the instance.	BA8F37866F59F269AE1D6 ↳2D962B887B65DE7474D1 ↳3679D9388B75C95EE7780FE
Child SHA1 Hash Values	Standard Fields	Long Text	Attachment SHA1 hash value of all child items in a family group, only present on parent items.	1989C1E539B5AE9818 ↳20648623954872BEE3E483; 58D9E4B4A3068DA6E9 ↳BCDD969523288CF38F9FB3
Child SHA256 Hash Values	Standard Fields	Long Text	Attachment SHA256 hash value of all child items in a family group, only present on parent items.	7848EEFC40C40F86892960 ↳0BF033617642E0D37C2F5 ↳FA444C7EF83350AE19883;628B6233D-D6E ↳0C89F32D6EFF2885F26917F14 ↳4B19F3678265BEBAC7E9ACAAAF5B
Comments	Standard Fields	Long Text	Comments extracted from the metadata of the native file. For more information, see Comments considerations .	Oracle 8i ODBC QueryFix Applied
Company	Standard Fields	Fixed- Length Text (255)	The internal value entered for the company associated with a Microsoft Office document. This field has a maximum length of 255 alpha-numeric characters.	Oracle Corporation
Contains Embedded Files	Standard Fields	Yes/No	The yes/no indicator of whether a file such as a Microsoft Word doc-	Yes

Processing field/ source name	Source view	Field type	Description	Example value
			ument has additional files embedded in it.	
Control Number Beg Attach	Standard Fields	Fixed-Length Text (50)	The identifier of the first document in a family group. This field is also populated for documents with no family members. This field has a maximum length of 50 alphanumeric characters.	KL0000000001
Control Number End Attach	Standard Fields	Fixed-Length Text (50)	The identifier of the last document in a family group. This field is also populated for documents with no family members. This field has a maximum length of 50 alphanumeric characters.	KL0000000001.0002
Conversation	Standard Fields	Long Text	Normalized subject of email messages. This is the subject line of the email after removing the RE and FW that are added by the system when emails are forwarded or replied to.	Sigaba Secure Internet Communication
Conversation Family	Standard Fields	Fixed-Length Text (44)	Relational field for conversation threads. This is a maximum 44-character string of numbers and letters that is created in the initial email.	01C9D1FD002240FB633CEC8 ↪94C1985845049B1886B67
Conversation Index	Standard Fields	Long Text	Email thread created by the email system. This is a maximum 44-character string of numbers and letters that is created in the initial email and has 10 characters added for each reply or forward of an email.	01C9D1FD002240FB633CEC8 ↪94C1985845049B1886B67
Created Date	Standard Fields	Long Text	The date on which a file was created. This field is derived from the Created Date/Time field.	12/24/2015

Processing field/ source name	Source view	Field type	Description	Example value
Created Date/Time	Standard Fields	Date	The date and time from the Date Created property extracted from the original file or email message. This field will display the filesystem date created for the document if that is the only date created value available. If a document has both a filesystem date created value and a document metadata date created value, this field will display the document metadata date created value.	"12/24/2015 11:59 PM"
Created Time	Standard Fields	Long Text	The time at which a file was created. This field is derived from the Created Date/Time field.	11:59 PM
DeDuped Count	Standard Fields	Whole Number	The number of duplicate files related to a primary file. This is present only when Global or Custodial Deduplication is enabled and duplicates are present. If you discovered and published your set before Relativity Foxglove, you cannot map this field and re-publish the set. This is populated on the primary document.	2
DeDuped Custodians	Standard Fields	Multiple Object	The custodians associated with the deduped records of a file. The DeDuped Custodians file is mapped to a document and is present only when Global or Custodial Deduplication is enabled and duplicates are present. This is populated on the primary document.	Lay, Kenneth;Doe, John

Processing field/ source name	Source view	Field type	Description	Example value
			The All Custodians field is mapped to a document and is updated only.	
DeDuped Paths	Standard Fields	Long Text	The virtual paths of duplicates of a file. This is present only when Global or Custodial Deduplication is enabled and duplicates are present. Each path contains the associated custodian. This is populated on the primary document.	Lay, Kenneth \Lay, Kenneth\kenneth_lay_000_1_2_1.pst \\lay-k\Kenneth_Lay_Dec2000\Notes Folders\Notes inbox Doe, John \Doe, John\John_Doe_000_1_2_1.pst\Doe-J \John_Doe_Dec2000\Notes Folders\Discussion threads
Delivery Receipt Requested	Standard Fields	Yes/No	Indicates whether a delivery receipt was requested for an email.	No
Document Subject	Standard Fields	Long Text	Subject of the document extracted from the properties of the native file.	RE: Our trip to Washington
Discover Errors on Child Containers	Standard Fields	Long Text	Contains the Processing File IDs and category of child containers in parent documents that have errors associated with them.	1 - Password Protected Container; 2 - Corrupt Container; 3 - Configuration Error
Document Title	Standard Fields	Long Text	The title of a non-email document. This is blank if there is no value available.	Manual of Standard Procedures
Email Categories	Standard Fields	Long Text	Categories assigned to an email message.	Personal
Email Created Date/Time	Standard Fields	Date	The date and time at which an email was created.	"12/24/2015 11:59 PM"
Email Entry ID	Standard Fields	Long Text	The unique Identifier of an email in an mail store.	000000005B77B2A7467F564 68D820375BC3DC58244002000
Email Folder Path	Standard Fields	Long Text	The folder path in which a custodian stored an email. See Email folder path details for more information.	Inbox\New Business
Email Format	Standard Fields	Single Choice	The indicator of whether an email is HTML, Rich	HTML

Processing field/ source name	Source view	Field type	Description	Example value
			Text, or Plain Text.	
Email Has Attachments	Standard Fields	Yes/No	The yes/no indicator of whether a parent document has attachments. This field applies to the parent document and not child (email) files or attachments.	Yes
Email In Reply To ID	Standard Fields	Long Text	The internal metadata value within an email for the reply-to ID.	<F9B1A278195DF640A4CC6E C973DFF0C85FBBEDED @Prod-EX-MB-01.company.corp>
Email Last Modified Date/Time	Standard Fields	Date	The date and time at which an email was last modified.	"12/24/2015 11:59 PM"
Email Modified Flag	Standard Fields	Yes/No	The yes/no indicator of whether an email was modified.	Yes
Email Sensitivity	Standard Fields	Single Choice	The indicator of the privacy level of an email.	Company Confidential
Email Sent Flag	Standard Fields	Yes/No	The yes/no indicator of whether an email was sent, versus saved as a draft.	Yes
Email Store Name	Standard Fields	Fixed-Length Text (255)	Any email, contact, appointment, or other data that is extracted from an email container, .pst, .ost, .nsf, .mbox, and other files, will have this field populated with the name of that email container. Any children of those extracted emails, contacts, and appointments will not have anything populated in this field. For more information on this field, see Email Store Name details on page . This field has a maximum length of 255 alphanumeric characters.	kenneth_lay_000_1_1_1_1.pst
Email Unread	Standard Fields	Yes/No	The yes/no indicator of whether an email was not read.	Yes

Processing field/ source name	Source view	Field type	Description	Example value
Email/EDRMMMessage IdentificationHash	Metadata Fields	Long Text	An EDRM Message ID Hash (EDRM MIH) is an MD5 hash of the Message ID value of Email Messages, following EDRM guidelines.	1de319c276884bd0c 9e2f1621ada26cc
Email/IsRecurring	Metadata Fields	Yes/No	Indicates if an appointment is recurring. See Processing FAQs for an explanation of how Relativity processes recurring appointments as attachments.	Yes
Email/RecurrencePattern	Metadata Fields	Long Text	Displays the details of an appointment recurrence.	Occurs every Tuesday from 5:00 PM to 5:25 PM effective 2/18/2025 until 8/12/2025
Error Category	Standard Fields	Single Choice	The category assigned by the system to a processing error.	Password Protected Container
Error Message	Standard Fields	Long Text	The message that details the error, cause, and suggested resolution of the error prioritized by processing phase—discovery, text extraction, publish, file deletion.	There was an error during extraction of an email from this Notes container. It may be password protected. Consider adding the User.ID file and password(s) to Password Bank and retrying.
Error Phase	Standard Fields	Single Choice	The phase of processing in which the error occurred—discovery, text extraction, publish, file deletion.	Discovery
Error Status	Standard Fields	Single Choice	The status of the error—undetermined, ready to retry, retrieved, submitted, unresolvable.	Ready to retry.
Excel Hidden Columns	Standard Fields	Yes/No	The yes/no indicator of whether an Excel file contains one or more hidden columns.	No
Excel Hidden Rows	Standard Fields	Yes/No	The yes/no indicator of whether an Excel file contains one or more hidden rows.	Yes
Excel Hidden Worksheets	Standard Fields	Yes/No	The yes/no indicator of whether an Excel file contains one or more hidden	No

Processing field/ source name	Source view	Field type	Description	Example value
			worksheets.	
Excel Pivot Tables	Standard Fields	Yes/No	The yes/no indicator of whether an Excel file contains pivot tables.	Yes
ExtractedTextLocation	Other Fields	Long Text	The full path to the file Relativity creates to store extracted text/OCR for the file.	\\files.SERVER.relativity.one\0000\Files\XYZ000\Pro- cessing\000\INV1111564\INTERMEDIATE\0\141_6b5cefd1-3abf-4b70-b619-f2d99faa2e9c.TXT
Extracted Text Size in KB	Other Fields	Decima- l	This field indicates the size of the extracted text field in kilobytes.	6.7
Family Group (formerly "Group Identifier")	Standard Fields	Fixed- Length Text (40)	Group the file belongs to, used to identify the group if attachment fields are not used. This field has a maximum length of 40 alpha-numeric characters.	KL0000000002
File Extension	Standard Fields	Fixed- Length Text (25)	The extension of the file, as assigned by the processing engine after it reads the header information from the original file. This may differ from the value for the Original File Extension field. If you publish processing sets without mapping the File Extension processing field, the Text Extraction report does not accurately report document counts by file type. This field has a maximum length of 25 alpha-numeric characters.	MSG Note: During processing, some file extensions may be trimmed if they exceed Relativity's metadata storage limit of eight characters. Relativity keeps the first five characters of the file extension name and then appends ZZZ to indicate that the name has been trimmed. For example, a file named SomeFile.fileslack will be trimmed to SomeFile.filesZZZ for metadata storage. The original file name and extension (SomeFile.fileslack) are maintained in the File Name metadata field and can be mapped as a source to a Document field for publishing.
File Name	Standard Fields	Fixed- Length Text (255)	The original name of the file. This field has a maximum length of 255 alpha-numeric characters.	enron corp budget.xls
File Size	Standard Fields	Decima- l	Generally a decimal number indicating the size in	15896

Processing field/ source name	Source view	Field type	Description	Example value
			bytes of a file.	
File Type	Standard Fields	Fixed- Length Text (255)	Description that represents the file type to the Windows Operating System. Examples are Adobe Portable Document Format, Microsoft Word 97 - 2003 Document, or Microsoft Office Word Open XML Format. This field has a maximum length of 255 alpha-numeric characters.	Microsoft Excel 97-2003 Worksheet
From	Standard Fields	Fixed- Length Text (255)	The name, when available, and email address of the sender of an email message. This field has a maximum length of 255 alpha-numeric characters.	Capellas Michael D. [Michael.Capellas@COMPAQ.com]
From (SMTP Address)	Standard Fields	Fixed- Length Text (255)	The full SMTP value for the sender of an email message. This field has a maximum length of 255 alpha-numeric characters.	Michael.Capellas@COMPAQ.com
Has Hidden Data	Standard Fields	Yes/No	Indication of the existence of hidden document data such as hidden text in a Word document, hidden columns, rows, or worksheets in Excel, or slide notes in PowerPoint. If a document contains hidden data that was found during processing, this field displays a value of Yes. If no hidden data was found, this field is blank. Note that this field does not display a value of No if no hidden data was found. This is because Relativity cannot	Yes

Processing field/ source name	Source view	Field type	Description	Example value
			definitively state that a document contained no hidden data just because the system could not detect it.	
Has OCR Text	Standard Fields	Yes/No	The yes/no indicator of whether the extracted text field contains OCR text.	Yes
HiddenAttachment	Other Fields	Yes/No	The yes/no indicator of whether or not the record is an Email inline image.	Yes
Image Taken Date/Time	Standard Fields	Date	The date and time at which an original image, for example a document scan or .jpg, was taken.	"12/24/2015 11:59 PM"
Importance	Standard Fields	Single Choice	Notation created for email messages to note a higher level of importance than other email messages added by the email originator.	Low
Is Embedded	Standard Fields	Yes/No	The yes/no indicator of whether a file is embedded in a Microsoft Office document.	No
Is Parent	Standard Fields	Yes/No	The yes/no indicator of whether a file is a parent with children or a child/loose record with no children. If this reads Yes, it is a top-level parent with children. If this reads No, it is an attachment or a loose record such as a standalone email or an Edoc.	No
IsLinkedCloudFile	Other Fields	Yes/No	The yes/no indicator of whether a file is a linked cloud file.	Yes
Keywords	Standard Fields	Long Text	The internal value entered for keywords associated with a Microsoft Office document.	Enron, Security Agreement

Processing field/ source name	Source view	Field type	Description	Example value
Last Accessed Date	Standard Fields	Long Text	The date on which a loose file was last accessed.	12/24/2015
Last Accessed Date/Time	Standard Fields	Date	The date and time at which the loose file was last accessed.	"12/24/2015 11:59 PM"
Last Accessed Time	Standard Fields	Long Text	The time at which the loose file was last accessed.	11:59 PM
Last Modified Date	Standard Fields	Long Text	The date on which a file was last modified. This field is derived from the Last Modified Date/Time field.	12/24/2015
Last Modified Date/Time	Standard Fields	Date	The date and time from the Date Modified property extracted from the original file or email message. This field displays the filesystem date modified for the document if that is the only date modified value available. If a document has both a filesystem date modified value and a document metadata date modified value, this field displays the document metadata date modified value.	"12/24/2015 11:59 PM"
Last Modified Time	Standard Fields	Long Text	The time at which changes to a file were last saved. This field is derived from the Last Modified Date/Time field.	11:59 PM
Last Printed Date	Standard Fields	Long Text	The date on which a file was last printed.	12/24/2015
Last Printed Date/Time	Standard Fields	Date	The date and time at which a file was last printed.	"12/24/2015 11:59 PM"
Last Printed Time	Standard Fields	Long Text	The time at which a file was last printed.	11:59 PM
Last Saved By	Standard Fields	Fixed- Length	The internal value indicating the last user to	ymendez

Processing field/ source name	Source view	Field type	Description	Example value
		Text (255)	save a document. This field has a maximum length of 255 alpha-numeric characters.	
Last Saved Date	Standard Fields	Long Text	The date on which a file was last saved.	12/24/2015
Last Saved Date/Time	Standard Fields	Date	The internal value entered for the date and time at which a document was last saved.	"12/24/2015 11:59 PM"
Last Saved Time	Standard Fields	Long Text	The time at which a file was last saved.	11:59 PM
Lotus Notes Other Folders	Standard Fields	Long Text	A semi-colon-delimited listing of all folders that a Lotus Notes message or document appeared in, except for the one indicated in the Email Folder Path. For example: (Mail Threads);\$All);\$Drafts)	(Mail Threads);\$All);\$Drafts)
MD5 Hash	Standard Fields	Fixed-Length Text (40)	Identifying value of an electronic record that can be used for de-duplication and authentication generated using the MD5 hash algorithm. MD5 hashes are generated for the physical file of all file types. Relativity cannot calculate this value if you have FIPS (Federal Information Processing Standards cryptography) enabled for the instance.	21A74B494A1BFC2FE217 ↪CC274980E915
MS Office Document Manager	Standard Fields	Fixed-Length Text (255)	The internal value entered for the manager of a document. This field has a maximum length of 255 alpha-numeric characters.	Fabienne Chanavat
MS Office Revision Number	Standard Fields	Fixed-Length Text (255)	The internal value for the revision number within a Microsoft Office file. This field has a maximum length of 255 alpha-	72

Processing field/ source name	Source view	Field type	Description	Example value
			numeric characters.	
Media Type	Standard Fields	Single Choice	A standard identifier used on the Internet to indicate the type of data that a file contains.	application/msword
Meeting End Date	Standard Fields	Long Text	The date on which a meeting item in Outlook or Lotus Notes ended.	12/24/2015
Meeting End Date/Time	Standard Fields	Date	The date and time at which a meeting item in Outlook or Lotus Notes ended.	"12/24/2015 11:59 PM"
Meeting End Time	Standard Fields	Long Text	The time at which a meeting item in Outlook or Lotus Notes ended.	11:59 PM
Meeting Start Date	Standard Fields	Long Text	The date on which a meeting item in Outlook or Lotus Notes started.	12/24/2015
Meeting Start Date/Time	Standard Fields	Date	The date and time at which a meeting item in Outlook or Lotus Notes began.	"12/24/2015 11:59 PM"
Meeting Start Time	Standard Fields	Long Text	The time at which a meeting item in Outlook or Lotus Notes started.	11:59 PM
Message Class	Standard Fields	Single Choice	The type of item from an email client. For example, email, contact, calendar, and others.	IPM.Note
Message Header	Standard Fields	Long Text	The full string of values contained in an email message header.	date: Wed, 4 Oct 2000 18:45:00 -0700 (PDT) Wed, 4 Oct 2000 18:45:00 -0700 (PDT) Message-ID: MIME-Version: 1.0 Content-Type: text/plain; charset="us-ascii" Content-Transfer-Encoding: 7bit from: "Rosalee Fleming" to: "Telle Michael S." subject: Re: Referendum Campaign filename: klay.nsf folder: \Kenneth_Lay_Dec2000\Notes Folders\sent
Message ID	Standard Fields	Fixed- Length	The message number created by an email	<PLSRGLMRNQWEDFYJL5 ↪ZJFF41USDEIQHB@zlsvr22>

Processing field/ source name	Source view	Field type	Description	Example value
		Text (255)	application and extracted from the email's metadata. For more information, see Message ID considerations on page . This field has a maximum length of 255 alpha-numeric characters.	
Message Type	Standard Fields	Single Choice	Indicates the email system message type. Possible values include Appointment, Contact, Distribution List, Delivery Report, Message, or Task. The value may be appended with '(Encrypted)' or 'Digitally Signed' where appropriate.	Message
Native File	Standard Fields	Long Text	The path to a copy of a file for loading into Relativity.	\\files2.T026.ctus014128.r1.company.com\↳T026\Files\EDDS2544753\Processing\↳1218799\INV2544753\SOURCE\0\982.M-SG
Number of Attachments	Standard Fields	Whole Number	Number of files attached to a parent document.	2
Original Author Name	Standard Fields	Fixed- Length Text (50)	The display name of the original author of an email. This field has a maximum length of 50 alpha-numeric characters.	Jane Doe
Original Email Author	Standard Fields	Fixed- Length Text (255)	The email address of the original author of an email. This field has a maximum length of 255 alpha-numeric characters.	Jane.Doe@COMPAQ.com
Original File Extension	Standard Fields	Fixed- Length Text (25)	The original extension of the file. This may differ from the value for the File Extension field, since that value is assigned based on the processing engine's reading of the file's header information. This	DOC

Processing field/ source name	Source view	Field type	Description	Example value
			field has a maximum length of 25 alpha-numeric characters.	
Other Metadata	Standard Fields	Long Text	Metadata extracted during processing for additional fields beyond the list of processing fields available for mapping. This includes TrackChanges, HiddenText, HasOCR, and dates of calendar items. Field names and their corresponding values are delimited by a semicolon.	Excel/HasHiddenColumns=True;Office↵/Application=Microsoft Excel;↵InternalCreatedOn=7/25/1997↵9:14:12 PM;Office/Security=2;↵Office/PROPID_23=528490;Office/↵Scale=0;Office/LinksDirty=0;Office/↵PROPID_19=0;Office/PROPID_22=0;↵Office/Parts=sum,ENRON;Office/↵Headings=Worksheets,2;Office/↵_PID_GUID=UnknownPROPVARIANT↵type 65;Excel/HasHiddenRows=True;↵LiteralFileExtension=XLS
Outlook Flag Status	Standard Fields	Single Choice	Indicates if an Outlook item is flagged. The field is blank if the item is not flagged.	Flagged
Parent Document ID	Standard Fields	Fixed-Length Text	Document ID of the parent document. This field is only available on child items.	EN11, 12345
Password	Other Fields	Long Text	When the Password Protected field indicates that a document was decrypted by Relativity, this field will contain the value from the Password Bank that was successfully applied. See Password bank in processing workflow for more information.	Test1234!
Password Protected	Standard Fields	Single Choice	The Password Protected field shows the decryption status of password-protected documents. The field displays Decrypted if a password is known, Encrypted if a password is not known, or no value if the file does not have a password. To track decrypted	Encrypted

Processing field/ source name	Source view	Field type	Description	Example value
			documents, map the Password field. To map this field, add or edit the corresponding Relativity field, then open the Source control. From the drop-down menu, select Other Fields . In the results list, select Password .	
PowerPoint Hidden Slides	Standard Fields	Yes/No	The yes/no indicator of whether a PowerPoint file contains hidden slides.	Yes
Primary Date/Time	Standard Fields	Date	Date taken from Sent Date, Received Date, or Last Modified Date in that order of precedence.	"12/24/2015 11:59 PM"
ProcessingPageCount	Other Fields	Whole Number	The total count of pages within PDF files calculated during processing.	3
Read Receipt Requested	Standard Fields	Yes/No	Indicates whether a read receipt was requested for an email.	Yes
Received Date	Standard Fields	Long Text	The date on which an email message was received.	12/24/2015
Received Date/Time	Standard Fields	Date	The date and time at which an email message was received.	"12/24/2015 11:59 PM"
Received Time	Standard Fields	Long Text	The time at which an email message was received.	11:59 PM
Recipient Count	Standard Fields	Whole Number	The total count of recipients in an email which includes the To, CC, and BCC fields.	1
Recipient Domains (BCC)	Standard Fields	Multiple Object	The domains of the 'Blind Carbon Copy' recipients of an email. For information on domains and steps to create the Domains object and associative multiple object fields, see Relativity Objects.	enron.com;bellatlantic.com

Processing field/ source name	Source view	Field type	Description	Example value
Recipient Domains (CC)	Standard Fields	Multiple Object	The domains of the 'Carbon Copy' recipients of an email. For information on domains and steps to create the Domains object and associative multiple object fields, see Relativity Objects.	enron.com;bellatlantic.com
Recipient Domains (To)	Standard Fields	Multiple Object	The domains of the 'To' recipients of an email. For information on domains and steps to create the Domains object and associative multiple object fields, see Relativity Objects.	enron.com;bellatlantic.com
Recipient Name (To)	Standard Fields	Long Text	The names of the (To) field recipients of an email message.	Jane Doe
Record Type	Standard Fields	Single Choice	The single choice field that indicates that the file is an Email, Edoc, or Attach.	Edoc
Replaced Extracted Text	Standard Fields	Yes/No	This Yes/No field indicates if the document had its extracted text replaced with a placeholder due to its size exceeding 2GB.	Yes
Rolled up image text	Standard Fields	Fixed- Length Text (255)	The comma (,) delimited list of child FileIDs that have had their text rolled into the parent document. This field is only set for parent files.	Roll Up Image Text
Note: *You will not see RSMF fields in the catalog until you discover them. Any discovered RSMF fields are then available for mapping.				
Caution: Use caution when mapping to multiple-choice fields. Mapping multiple-choice fields creates a unique value for each choice option. Depending on the multiple-choice fields mapped, you could have hundreds of unique choices, impacting performance.				
*RSMF Application	Metadat- a Fields	Long Text	This is used to identify source of the data, which is intended to be ambiguous. For example, it could be the application	Slack

Processing field/ source name	Source view	Field type	Description	Example value
			of the data contained in the RSMF file.	
*RSMF Attachment Count	Metadata Fields	Whole Number	This field should be a number that is a sum of all of the attachments present in the RSMF.	10
*RSMF Begin Date	Metadata Fields	Date	The time stamp (ISO8601) of the earliest short message event within the file.	11:59 PM
*RSMF Custodian	Metadata Fields	Long Text	This field is used to identify from whom the data was collected from.	John Doe
*RSMF End Date	Metadata Fields	Date	The time stamp (ISO8601) of the latest short message event within the file.	11:59 PM
*RSMF Event Collection Id	Metadata Fields	Long Text	This field should be a unique ID that is to be used to help keep many RSMFs from a single conversation together.	D4C4EB398980E82B4B3064
*RSMF Message Count	Metadata Fields	Whole Number	The number of short message events captured within the file.	5
*RSMF Generator	Metadata Fields	Long Text	Identifies the author of the RSMF file.	Relativity v2.4
*RSMF Participants	Metadata Fields	Long Text	This field can be used to choose from a string of names (comma delimited) that are present in the conversation in the RSMF file. Note: Relativity discovers the RSMF Participants field type as Multiple Choice. To maximize performance, map this field as Long Text.	John Doe <john.doe@relativity.com>, Jane Doe <jane.doe@relativity.com>
*RSMF Version	Metadata Fields	Long Text	The version of the RSMF specification that the file adheres to.	2.0.0

Processing field/ source name	Source view	Field type	Description	Example value
SHA1 Hash	Standard Fields	Fixed- Length Text (50)	Identifying value of an electronic record that can be used for de-duplication and authentication generated using the SHA1 hash algorithm. This field has a maximum length of 50 alpha-numeric characters. SHA1 hashes are generated for the physical file of all file types.	D4C4EB398980E82B4B3064 ↪CC2005F04D04BBAAE6
SHA256 Hash	Standard Fields	Fixed- Length Text (70)	Identifying value of an electronic record that can be used for de-duplication and authentication generated using the SHA256 hash algorithm. This field has a maximum length of 70 alpha-numeric characters. SHA256 hashes are generated for the physical file of all file types.	4F8CA841731A4A6F78B919 ↪806335C963EE039F33214A04 ↪1F0B403F3D156938BC
Sender Domain	Standard Fields	Multiple Object	The domain of the sender of an email.	enron.com
Sender Name	Standard Fields	Fixed- Length Text (255)	The name of the sender of an email message. This field has a maximum length of 255 alpha-numeric characters.	Kenneth Lay
Sent Date	Standard Fields	Long Text	The date on which an email message was sent.	12/24/2015
Sent Date/Time	Standard Fields	Date	The date and time at which an email message was sent.	"12/24/2015 11:59 PM"
Sent Time	Standard Fields	Long Text	The time at which an email message was sent.	11:59 PM
Sort Date/Time	Standard Fields	Date	For parent documents, the field is populated with the Primary Date/Time	"12/24/2015 11:59 PM"

Processing field/ source name	Source view	Field type	Description	Example value
			value. For child documents, the field is populated with the Sort Date/Time of the parent document. All documents in a family will therefore have the same Sort Date/Time value, keeping family members together when sorting on this field. Note: When you filter for dates in Inventory, you are filtering specifically on the Sort Date/Time field, which is taken from the file's Sent Date, Received Date, and Last Modified Date fields in that order of precedence. This happens on email messages repeated for the parent document and all child items to allow for date sorting. You have the following options for applying a date range filter:	

Processing field/ source name	Source view	Field type	Description	Example value
			Note: When processing documents without an actual date, Relativity provides a null value for the following fields: Created Date, Created Date/Time, Created Time, Last Accessed Date, Last Accessed Date/Time, Last Accessed Time, Last Modified Date, Last Modified Date/Time, Last Modified Time, Primary Date/Time, and Sort Date/Time. The null value is excluded and not represented in the filtered list.	
Source Path	Standard Fields	Long Text	The folder structure and path to the file from the original location identified during processing. For emails, this displays the subject rather than the email's entry ID. This provides you with better context of the origin of the email. Previously, the Virtual Path field displayed the entry ID with the email file name, and if you followed this virtual path, it was difficult to tell by that entry ID where the email came from. See Source path details for more information.	Reports\User\Sample.pst\Inbox\Requested February report
Speaker Notes	Standard Fields	Yes/No	The yes/no indicator of whether a PowerPoint file has speaker notes associated with its slides.	Yes
Subject	Standard Fields	Long Text	The subject of the email message.	Blackmore Report - August
Suspect File Extension	Standard Fields	Yes/No	The yes/no indicator if whether the extension of	Yes

Processing field/ source name	Source view	Field type	Description	Example value
			a file does not correspond to the actual type of the file. For example. XLS for a Word document.	
Text Extraction Method	Standard Fields	Single Choice	For text-based natives where text extraction was successful, this field indicates the extraction method, such as Excel, Word, PowerPoint, PDF Library, Outlook, Relativity, dtSearch, Native, or Native dtSearch Fail-over. For image-based natives where OCR was performed, the field indicates OCR, or in the case of partial text, may reflect PDF Library; OCR.	Excel
Title	Standard Fields	Long Text	The title of the file. For emails, this is the subject line. For non-emails, this is any available title.	June Scrum Notes
To	Standard Fields	Long Text	The names, when available, and email addresses of the recipients of an email message.	Capellas Michael D. [Michael.Capellas@COMPAQ.com]
To (SMTP Address)	Standard Fields	Long Text	The full SMTP value for the recipient of an email message, for example, "bob@example.com."	Michael.Capellas@COMPAQ.com
TrackChanges	Standard Fields	Yes/No	The yes/no indicator of whether the track changes metadata on an Office document is set to True. This does not necessarily indicate if tracked changes were made to the document or not. On Word documents, the track changes toggle may have been	Yes

Processing field/ source name	Source view	Field type	Description	Example value
			set to True, changes made to the document, then set back to False. In this situation, this field will still indicate 'No' because it is looking only at the setting and not for the actual existence of changes even though tracked changes still exist in the document. • If the same situation is applied to Excel documents, the result is slightly different. Microsoft deletes tracked changes on Excel documents when the toggle is set back to False. The returned value will also indicate 'No' but there is no concern about missed tracked changes as none exist. • For file types that cannot contain tracked changes, such as PDFs, email, and images, this field is blank.	
Track Changes	Standard Fields	Yes/No	The yes/no indicator of whether the track changes toggle is set to True and/or there are tracked changes in the document. This field maps to the TrackedChangesCombi	

Processing field/ source name	Source view	Field type	Description	Example value
			ned Invariant field. This will be Yes if either of the following are true: - The Track Changes button is enabled in the document. - There is actual Tracked Change content in the document.	
Unified Title	Standard Fields	Long Text	The subject of the file. For emails, this is the subject line. For non-emails, this is the file name. Note that in short message (RSMF) conversion, Relativity may use the conversation name as the subject, and therefore the Unified Title.	Company Memo
Unprocessable	Standard Fields	Yes/No	The yes/no value indicating if a file was able to be processed. If the file could not be processed, this field is set to Yes. - Even if a file is flagged as Unprocessable, it may still be visible in the native file viewer if Oracle is able to render the file. - The Unprocessable field is set to Yes on any file for which Relativity does not have an Invariant plugin that is capable of extracting text or ima-	No

Processing field/ source name	Source view	Field type	Description	Example value
			ging/OCRing that document type. For example, it is not set for a corrupt file for which we cannot extract text, such as a corrupt Word document that logs an error during data extraction. Unprocessable documents do not have errors associated with them because they never reach a point at which they can register a processing error.	
Virtual Path	Other Fields	Long Text	Contains the complete folder structure and internal path of the processed file, up to but not including this file. This path includes any containers that the file may be in and, in the case of attached or embedded items, includes the file name of the parent document. See Virtual path for more information.	\\small outlook mail sample\\REL0000428454.MSG\\Zone A G J 11-5.doc

Note: You can track which passwords successfully decrypted published documents by mapping the **Password** field found in the fields catalog. After creating or editing a field, click the **Advanced Settings** tab, then click **Source**. Select **Other Fields** from the drop-down control, then filter for the word **Password**. The field should be a long text type.

14.4 Email Store Name field

To better understanding how the Email Store Name field works, consider the following examples:

- When an email comes from .pst file, the .pst is listed in the Email Store Name field. When a child Word document comes from a .rar archive and is attached to the email, the Email Store Name field is blank for the Word document.

	Control Number	Email Store Name
	Filter	Filter
✉	-file-0000000001	PSTWithRarWithOfficeFiles.pst
W	-file-0000000001.0001	
W	-file-0000000001.0002	
W	-file-0000000001.0003	

- The RAR/ZIP information for the Word documents mentioned above is found in the Container Name field.

	Control Number	Email Store Name	Originating Processin...	Container name
	Filter	Filter	Filter	Filter
✉	-file-0000000001	PSTWithRarWithOfficeFiles.pst	Kolhe, Aneeta - 1048232	PSTWithRarWithOfficeFiles.pst
W	-file-0000000001.0001		Kolhe, Aneeta - 1048232	Word.rar
W	-file-0000000001.0002		Kolhe, Aneeta - 1048232	Word.rar
W	-file-0000000001.0003		Kolhe, Aneeta - 1048232	Word.rar

- In the following example, email 00011 comes from a .pst file named *PSTContainingEmbeddedPSTInFolders.pst*, which is the value for the Email Store Name field for that email. The other emails, 00011.001 and 00011.002, come from a .pst file attached to the 00011 email. This .PST file is named *PSTWithEmails.pst*. In this case, the Email Store Name field for those child messages is *PSTWithEmails.pst*, not the top-level .pst file named *PSTContainingEmbeddedPSTInFolders.pst*.

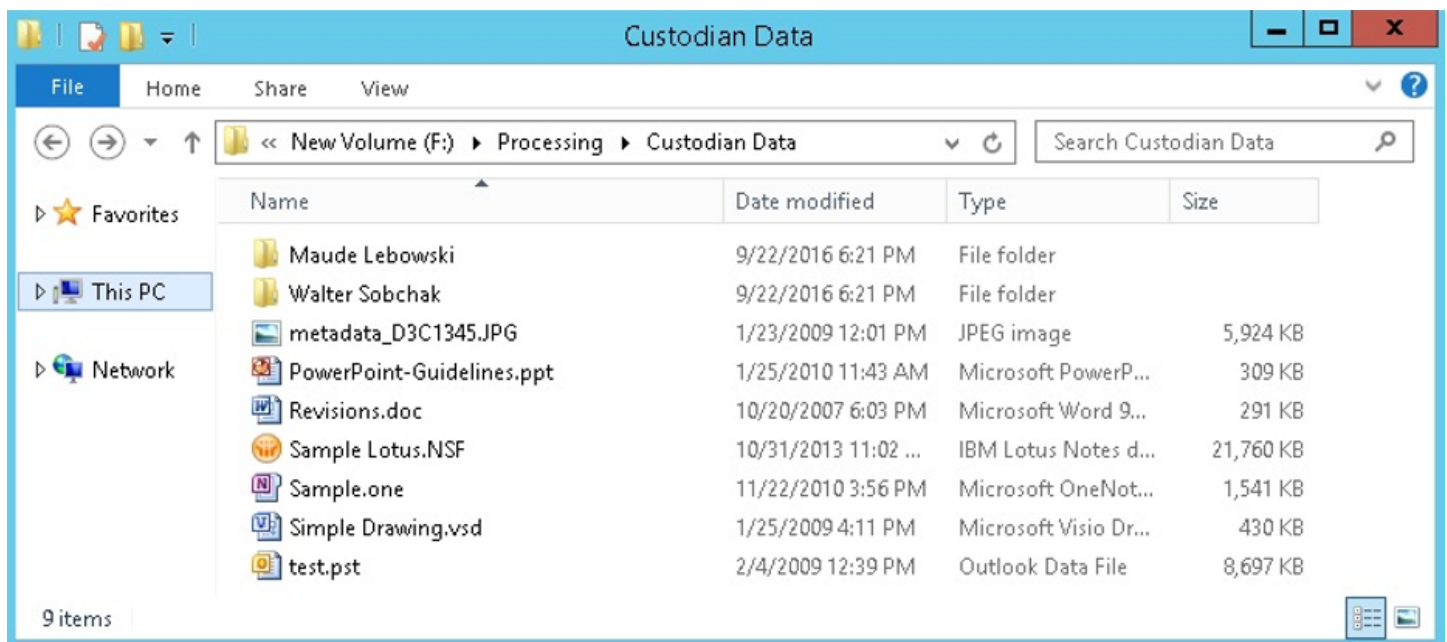
	Control Number	Email Store Name
	Filter	Filter
✉	-file-0000000011	PSTContainingEmbeddedPSTInFolders.pst
✉	-file-0000000011.0001	PSTWithEmails.pst
✉	-file-0000000011.0002	PSTWithEmails.pst

	Control Number	Email Store Name	Container name	Originating Processin...	Originating Processing Data Source
	Filter	Filter	Filter	Filter	Filter
✉	-file-0000000012		ZipWithEmailNoAttachments.zip	Zip with Email	Kolhe, Aneeta - 1049077

14.5 Virtual path

The virtual path is the complete folder structure and path from the original folder or file chosen for processing to the file. This path includes any containers that the file may be in and, in the case of attached or embedded items, includes the file name of the parent document.

This path does not include the name of the file itself. If a file is selected for import instead of a folder, the virtual path for that file is blank.



Processing Data Source						Items 1 - 4 (of 4)	
☐	Source path	Custodian	Document numbering prefix	Time zone	OCR language(s)		
☐ Edit	\\Relativity\Processing\Custodian Data\Maude Lebowski		REL	(UTC) Coordinated Universal Time	English		
☐ Edit	\\Relativity\Processing\Custodian Data\Walter Sobchak		REL	(UTC) Coordinated Universal Time	English		
☐ Edit	\\Relativity\Processing\Custodian Data\test.pst		REL	(UTC) Coordinated Universal Time	English		
☐ Edit	\\Relativity\Processing\Custodian Data\Revisions.doc		REL	(UTC) Coordinated Universal Time	English		

0 Selected Item(s) Select Page Size: 100

The following are examples of virtual paths created from the folders, per the above images:

- \Maude Lebowski\Loose Docs
- \Walter Sobchak\Walter.pst\Inbox\Unimportant\Fest Junk\Walter
- test.pst\My Test Box
 - In the case of a container or loose file being directly selected for processing, the virtual path does not have a leading backslash.
- test.pst\My Test Box\000000009B90A00DCC4229468A243C71810F71BC24002000.MSG
- Revisions.doc
 - This is the virtual path of a file embedded in the Revisions.doc file.

14.6 Processing folder path

The processing folder path is the folder structure created in the folder browser of the Documents tab. Relativity creates this path by keeping any folders or container names in the virtual path and discarding any file names that a file may be attached to or embedded in.

Files without a virtual path and items embedded within them do not have a processing folder path. If a container is embedded in a loose file, the items in that container have a processing folder path that matches the name of the container.

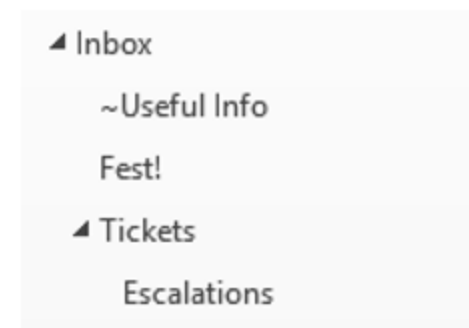
The following are examples of virtual paths and corresponding processing folder paths.

Virtual Path	Processing Folder Path
test.pst\Inbox	test.pst\Inbox
test.pst\Inbox\000000009B90A00DCC4229468A243C71810F71BC24002000.MSG	test.pst\Inbox
test.pst\Inbox\000000009B90A00DCC4229468A243C71810F71BC24002000.MSG\Pics.zip	test.pst\Inbox\Pics.zip

14.7 Email folder path

The email folder path is the folder path within the email container file in which an email was stored. All attachments to emails have no value for this field.

For example, an email stored in the *Escalations* folder in the following image below would have a value of "Inbox\Tickets\Escalations."



14.8 Source path

The source path is a modified display of the virtual path. In the case of attachments to emails, any entry IDs of emails appearing in the virtual path are replaced by the subject of that email instead. In all other cases the source path value is identical to the virtual path.

For example, an attachment to an email could have the following virtual path and source path values:

Virtual Path	Source Path
Sample.pst\Inbox\000000009B90A00DCC4229468A243C71810F71BC24002000.MSG	Sample.pst\Inbox\Requested February reports

Note: This source path field is not to be confused with the Source Path field found on the Processing Data Source layout on the saved processing set.

14.9 Message ID considerations

Note the following details regarding the Message ID field:

- Message ID is an identifier applied to an email by the program that created the email, such as Outlook, Eudora, or another.
- Email programs can use whatever they want for a message ID, or they can leave it off entirely. The mail server is free to assign an identifier even if an email client did not.
- There is no guarantee that every message ID is unique because every email client and mail server uses a different algorithm to create one.
- Message ID is unique only in the fact Relativity does not know what tool generated the identifier or what algorithm generated it. In addition, Relativity cannot assume that the identifier will even exist in an email.
- Relativity cannot validate the message ID because it is made up of opaque data associated with an email.
- It is possible that two entirely different emails might share the same message ID.
- Using the Message ID is not a reliable alternative to SHA256 de-duplication. For the purposes of de-duplication, we recommend that you use the Processing Duplicate Hash. If you processed the information in another tool, it is recommended that you use the Hash Algorithm you selected in that tool.

14.10 Comments considerations

There are two kinds of comments that are possible in all Office documents: metadata and inline. The following table breaks down which optional processing fields are populated by each type of comment.

Comment type	Location in file	Hidden Data value	Comments value
Metadata	Details tab of the Properties window (when you right-click on file name)	Null (blank)	Contents of comments property on the file
Inline	In the body of the document	"Yes"	Null (blank)
Both	Details tab of file and body of document	"Yes"	Contents of comments property on the file

Note: There are a number of reasons why a document could contain hidden text. A returned value of Yes for the Hidden Data field doesn't automatically mean that the document has inline comments.

14.11 De-duped custodian and path considerations

If you run de-duplication as part of your processing job, you may want to know where the documents that eventually get de-duplicated came from, the path, as well as which custodian those documents were associated with.

The DeDuped Custodians and DeDuped Paths optional fields allow you to track this information. When a document is de-duplicated, these fields are populated upon publish, or republish.

- **DeDuped Custodians**—a multiple object field with object type Document and associated object type Entity. You should only associate this field with the Entity object. If this field is associated with any other object type, you will not be able to publish documents to your workspace.
- **DeDuped Paths**—a long text document field that provides the location of the de-duplicated document.

To use these fields, simply add them to a document view and refer to that view after your publish job has completed. You can then export the results to an Excel file, if necessary.

	Doc ID Beg	DeDuplicated Custodians	DeDuplicated Paths
✉	IVAN00000000001	Two, David	Two, David \All mappable metadata\3 files inside
✉	IVAN00000000002	Two, David	Two, David \All mappable metadata\3 files inside
✉	IVAN00000000003	Two, David	Two, David \All mappable metadata\3 files inside
📄	IVAN00000000004	Two, David	Two, David \All mappable metadata
✉	IVAN00000000005	Two, David	Two, David \All mappable metadata
📄	IVAN00000000006	Two, David	Two, David \All mappable metadata
📄	IVAN00000000007	Two, David	Two, David \All mappable metadata
📄	IVAN00000000008	Two, David	Two, David \All mappable metadata

Note: When Relativity populates the Deduplicated Custodians and Deduplicated Paths fields during republish, it performs an overlay. Because of this, if you modify a document's identifier field in Relativity, your information could become out of sync. For this reason, we recommend that you do not modify the identifier field.

15 Processing profiles

A processing profile is an object that stores the numbering, deNIST, extraction, and deduplication settings that the processing engine refers to when publishing the documents in each data source that you attach to your processing set. You can create a profile specifically for one set or you can reuse the same profile for multiple sets.

Note: This documentation contains references to third-party software, or technologies. While efforts are made to keep third-party references updated, the images, documentation, or guidance in this topic may not accurately represent the current behavior or user interfaces of the third-party software. For more considerations regarding third-party software, such as copyright and ownership, see [Terms of Use](#).

Relativity provides a Default profile upon installation of processing.

15.1 Creating or editing a processing profile

To create or edit a processing profile:

1. Use the search bar to navigate to the **Processing Profile** tab.
2. Click **New Processing Profile** or select any profile in the list.
3. Complete or modify the fields on the Processing Profile layout. See [Fields](#).
4. Click **Save**. Once you save the processing profile, you can associate it with a processing set. For more information, see [Processing sets](#).

Note: You can't delete the Default processing profile. If you delete a profile that is associated with a processing set you've already started, the in-progress processing phase will continue with the original profile settings you applied when you submitted the job, but you won't be able to proceed to the next phase. For example, if you delete a profile during discovery, you won't be able to publish those discovered files until you add a new profile to the set. If you have an existing processing set that you haven't started that refers to a profile that you deleted after associating it to the set, you must associate a new profile with the set before you can start that processing job.

15.1.1 Fields

Note: Relativity doesn't re-extract text for a re-discovered file unless an extraction error occurred. This means that if you discover the same file twice and you change any settings on the profile, or select a different profile, between the two discovery jobs, Relativity will not re-extract the text from that file unless there was an extraction error. This is because processing always refers to the original/primary document and the original text stored in the database.

15.1.1.1 Processing profile information

The **Processing Profile Information** category of the profile layout provides the following fields:

- **Name**—the name you want to give the profile.

15.1.1.2 Numbering settings

The **Numbering Settings** category of the profile layout provides the following fields.

Numbering Settings

Numbering Type*

☒ Auto Number
☐ Define Start Number
☐ Level Numbering

Default document numbering*
prefix

Number of digits

Number of Digits*
(padded with zeros)

Parent/Child Numbering*

☒ Suffix Always
☐ Continuous Always
☐ Continuous, Suffix on Retry

Delimiter*

- **Numbering Type**—determines how the documents in each data source are numbered when published to the workspace. This field gives you the option of defining your document numbering schema. It is useful in keeping your document numbering consistent when importing documents from alternate sources. The choices for this field are:
 - **Auto Number**—determines that the next published document will be identified by the next available number of that prefix.
 - **Define Start Number**—sets the starting number of the documents you intend to publish to the workspace.
 - Relativity uses the next available number for that prefix if the number is already published to the workspace.

- To ensure continuity, Relativity will never assign a control number below the defined starting number in future processing sets. For example, if you define a starting number of 100, the numbers 0-99 become unavailable for future use for that prefix.
- This option is useful when you process from a third-party tool that does not provide a suffix for your documents and you want to define a new start number for the next set of documents to keep the numbering continuous.
- Selecting this choice makes the Default Start Number field available below and the Start Number field on the data source layout.
 - **Default Start Number**—the starting number for documents that are published from the processing set(s) that use this profile.
 - This field is only visible if you selected the Define Start Number choice for the Numbering Type field above.
 - If you use a different start number for the Start Number field on the data source that you attach the processing set, that number takes precedence over the value you enter here.
 - The maximum value you can enter here is 2,147,483,647. If you enter a higher value, you'll receive an Invalid Integer warning next to field value and you won't be able to save the profile.
- **Level numbering**—option to number documents with a control number that follows the format PPP.BBBB.FFFF.NNNN at a **document level**. For details on level numbering, see [Level numbering special considerations](#).
 - **Number of Digits**—determines how many digits each level of the document's control number contains.
 - **Level 2 (box number)**—corresponds to the BBBB level . Selecting 4 in the drop-down list will allow for the following range in this level: 0001-9999. By default, this field is set to 3.
 - **Level 3 (folder number)**—corresponds to the FFFF level . Selecting 4 in the drop-down list will allow for the following range in this level: 0001 - 9999. By default, this field is set to 3.
 - **Level 4 (document number)**—corresponds to the NNNN level at the document level . Selecting 4 in the drop-down list will allow for the following range in this level: 0001 - 9999. By default, this field is set to 4.
- **Default document numbering prefix**—the prefix applied to each file in a processing set once it is published to a workspace. The default value for this field is REL.
 - When applied to documents, this appears as the prefix, followed by the number of digits you specify. For example, <Prefix>xxxxxxxxxx.
 - If you use a different prefix for the Custodian field on the processing data source(s) that you add to your processing set, the custodian's prefix takes precedence over the profile's.
 - The character limit for this prefix is 75.

Note: When Level numbering is selected, the prefix corresponds to the PPP section in the PPP.BBBB.FFFF.NNNN format and it can be used to identify the source or owner of the documents also known as 'party code' or 'source'.

- **Number of Digits**—determines how many digits the document's control number contains. The range of available values is 1 to 10 when Define Start Number is selected. By default, this field is set to 10 characters.

- **Parent/Child Numbering**—determines how parent and child documents are numbered relative to each other when published to the workspace. The choices for this field are as follows. For examples of each type, see [Parent/child numbering type examples](#).
 - **Suffix Always**—child documents are numbered according to their parent, along with a suffix and delimiter of your choice.
 - **Continuous Always**—child documents receive the next sequential control number after their parent.
 - **Continuous, Suffix on Retry**—child documents are assigned control numbers in continuous format by default. However, new child documents discovered through retry go through the Suffix Always numbering process.

Note: It is possible for your workspace to contain a document family that has both suffixed and non-suffixed child documents. See [Suffix special considerations](#) for details.

- **Delimiter**—the delimiter you want to appear between the different fragments of the control number of your published child documents. The choices for this field are:
 - **- (hyphen)**—adds a hyphen as the delimiter to the control number of child documents. For example, REL0000000001-0001-0001.
 - **.** **(period)**—adds a period as the delimiter to the control number of child documents. For example, REL0000000001.0001.0001.
 - **_ (underscore)**—adds an underscore as the delimiter to the control number of child documents. For example, REL0000000001_0001_0001.
-

Notes:

- Level numbering cannot be used with Quick-Create Set(s).
 - Level numbering and data source cannot be changed upon publish, retry, or republish. Non-level numbering cannot be changed to level numbering on a published processing set and then republished. Once published, Numbering Type cannot be changed.
-

Level numbering special considerations

When Level numbering is selected as the Numbering Type in the Processing Profile, the prefix corresponds to the PPP section in the PPP.BBB.FFF.NNNN format. It can be used to identify the source or owner of the documents also known as 'party code' or 'source'.

In the Number of digits section, you can determine the number of digits to use in each level. For example, selecting 4 in the drop-down list will allow for the following range in that level: 0001 - 9999.

Level 2 (box number)—corresponds to the BBB level in the PPP.BBB.FFF.NNNN format. Default value is 3 digits.

Level 3 (folder number)—corresponds to the FFF level in the PPP.BBB.FFF.NNNN format. Default value is 3 digits.

Level 4 (document number)—corresponds to the NNNN level in the PPP.BBB.FFF.NNNN format. Default value is 4 digits.

Once published, Numbering Type cannot be changed. Thus, Level numbering and data source cannot be changed upon publish, retry, or republish. Non-level numbering cannot be changed to level numbering on a published processing set and then republished.

Create a new Processing Set and add the data sources that you need. If the profile used by the Processing Set is Level Numbering, you can define the start number for each Data Source when adding or modifying data sources to the Processing Set.

Start numbers

Level 2 (box number) ###

Level 3 (folder number) 1

Level 4 (document number) 0001

When you create a new data source, the system will use # to indicate how many digits were configured for that level in the Processing Profile used on the current Processing Set. If a level was configured to take up to 3 digits, you can enter a start number with no padding, (e.g., 1), or with padding, (e.g., 0001).

Level numbering and control numbers

By using Level Numbering, you can define a prefix text and three numbering levels as the control number to be used on documents that are published. For example:

- Prefix: REL.
- Level one numbering: 001
- Level two numbering: 001
- Level three numbering: 0001

When creating the control number, each level will be separated by a dot symbol, e.g., REL.001.001.0001.

Each level has a range of numbers that it can support. For example, 01 supports from 01 to 99. On the other hand, 001 supports from 001 to 999.

Fields like Family/Group Identifier, Attachments, and Parent ID are created based on the new control number.

Document level numbering vs page level numbering

The Level Numbering applies only at the document level. For example, if a data source is processing data using 01 for Level 1 numbering, 001 for Level 2 numbering, and 0001 for level 3 numbering, then the corresponding control numbers will be as follows:

Example List of Documents to Process	Resulting Control Number
Doc 1: document with 3 pages	PREFIX.001.001.0001
Doc 2: a one-page document	PREFIX.001.001.0002
Doc 3: a one-page document	PREFIX.001.001.0003
Doc 4: a 5 pages document	PREFIX.001.001.0004
Doc 5: a 2 pages document	PREFIX.001.001.0005
Doc 6: an email with no attachments	PREFIX.001.001.0006

15.1.1.3 Keeping families together

Families roll over to new level

When a family does not fit on the current level, the whole family rolls over to next level to keep the family together. See the example below:

When a family does not fit on the current level, the whole family rolls over to next level to keep the family together. See example below:

Number	Document
REL.001.0001.9999	Excel document

Number	Document
REL.001.0002.0001	Word document
REL.001.0002.0002	Word document

9,997 documents later

Number	Document
REL.001.0002.9997	email with no attachments
REL.001.0003.0001	email with 4 attachments
REL.001.0003.0002	attachment 1
REL.001.0003.0003	attachment 2
REL.001.0003.0004	attachment 3
REL.001.0003.0005	attachment 4

The email with 4 attachments couldn't use 9998 because the current level only had 2 values left (9998 - 9999), but families are required to stay together in the same level, so it roll overs to the next level.

Multi-level families must roll over

A family is every document that can be traced to the same parent. Grandchildren are in same family as children, thus, grandchildren stay in the same level as the rest of the family.

Publish scenario:

Number - Document
REL.001.0001.9999 – excel document
REL.001.0002.0001 – word document
REL.001.0002.0002 – word document

9,995 documents later

Number - Document
REL.001.0002.9995 – email with no attachments
REL.001.0003.0001 – email with 4 attachments
REL.001.0003.0002 – attachment 1 from REL.001.0003.0001
REL.001.0003.0003 – attachment 2 from REL.001.0003.0001
REL.001.0003.0004 – attachment 3 from REL.001.0003.0001
REL.001.0003.0005 – attachment 4 from REL.001.0003.0001

Family does not fit in one level

If there are more children documents than it can fit in a single level, then Relativity will suffix the children that overflow.

Publish scenario:

Number - Document
REL.001.0003.0001 – email with 10,000 attachments
REL.001.0003.0002 – attachment 1

Number - Document
REL.001.0003.0003 – attachment 2
REL.001.0003.0004 – attachment 3
REL.001.0003.0005 – attachment 4
[...]
REL.001.0003.9999 – attachment 9998
REL.001.0003.0001_0001 – attachment 9999
REL.001.0003.0001_0002 – attachment 10,000

15.1.1.4 Republish scenarios

New child found during republish

If during Retry-Discover, Relativity finds new children from a password-protected file, then Relativity will publish these children using the parent control number and a suffix appended to it. See the example below:

Initial Publish:

Number	Document
REL.001.001.0001	
REL.001.001.0002	
REL.001.001.0003	(Password-protected file)
REL.001.001.0004	

Republish:

Number	Document
REL.001.001.0001	
REL.001.001.0002	
REL.001.001.0003	(Password-protected file)
REL.001.001.0003_0001	new child found in REL.001.001.0003
REL.001.001.0003_0002	new child found in REL.001.001.0003

New child found during republish in a document with the highest possible control number at a specific level

If during Retry-Discover, Relativity finds new children in a document that holds the highest control number in the last level, then Relativity will publish these children with their parent's control number and a suffix appended to it. The family will not be moved to a new folder. See example below:

Initial Publish:

Number	Document
REL.001.001.9997	
REL.001.001.9998	
REL.001.001.9999	(Password-protected file)
REL.001.002.0001	
REL.001.002.0002	

Republish:

Number	Document
REL.001.001.9997	
REL.001.001.9998	
REL.001.001.9999	(Password-protected file)
REL.001.001.9999_0001	new child found in password-protected file
REL.001.002.0001	
REL.001.002.0002	

A child with multiple children is found during republish

If during Retry-Discover, Relativity finds new children in a document that holds the highest control number in a level, and those children also have children, then Relativity will publish these children with the ORIGINAL parent control number + a suffix appended to it. Family will not be moved to a new folder. See example below:

Initial publish:

Number	Document
REL.001.001.9997	
REL.001.001.9998	
REL.001.001.9999	(Password-protected file)
REL.001.002.0001	
REL.001.002.0002	

Republish:

Number	Document
REL.001.001.9997	
REL.001.001.9998	
REL.001.001.9999	(Password-protected file)
REL.001.001.9999_0001	new child found in REL.001.001.9999
REL.001.001.9999_0002	new child found in REL.001.001.9999_0001
REL.001.001.9999_0003	new child found in REL.001.001.9999_0001
REL.001.001.9999_0004	new child found in REL.001.001.9999
REL.001.002.0001	
REL.001.002.0002	

New documents from a container

When Relativity finds new root level documents, Relativity will not suffix them. Instead, Relativity will assign them to the next control number available.

Initial publish received error on ZIP container and can publish only two documents:

Number	Document
REL.001.001.9997	

Number	Document
REL.001.001.9998	
REL.001.001.9999	SourceFolder/containerFile.Zip
REL.001.002.0001	SourceFolder/containerFile.Zip/1.txt
REL.001.002.0002	SourceFolder/containerFile.Zip/2.txt
REL.001.002.0003	SourceFolder/flatDocument

When retry-discover yields two more documents from the ZIP container:

Number	Document
REL.001.001.9997	
REL.001.001.9998	
REL.001.001.9999	SourceFolder/containerFile.Zip
REL.001.002.0001	SourceFolder/containerFile.Zip/1.txt
REL.001.002.0002	SourceFolder/containerFile.Zip/2.txt
REL.001.002.0003	SourceFolder/flatDocument
REL.001.002.0004	SourceFolder/containerFile.Zip/3.txt (new doc)
REL.001.002.0005	SourceFolder/containerFile.Zip/3.txt (new doc)

Republish new root documents, each family is a single document

If new documents are published with a start number that is within a range that have unused numbers, new documents will be published in those gaps. See example below:

Initial Publish started at REL.001.001.001. First 998 documents are single documents with no families or attachment. Document 999 is family with 30 documents, so it is published on the next level:

Number	Document
REL.001.001.001	
REL.001.001.998	Next document is a family with 30 documents that rollovers.
REL.001.002.001	Family with 30 attachments.
REL.001.002.030	Last document published.

Republish finds 3 new root documents, each family is a single document. Thus, new documents are published using any numbering gaps.

Number	Document
REL.001.001.999	First document is published using 999.
REL.001.002.031	Second document is published in the next available number.
REL.001.002.032	Third document uses next available number and so on.

15.1.1.5 Collisions

Collisions with a new data source

Let's assume that documents were already published using numbers REL.001.001.001 to REL.001.001.010. If a new data source is created with a start number that was already used (e.g., REL.001.001.008), then the new data source start number is the next available number: REL.001.001.011.

Collisions among multiple data sources

If a user adds 3 data sources and each data source has 10 documents and the same start number, when published, each data source start number will be the next available number. For example:

Data source 1: REL.001.001.0001- REL.001.001.0010

Data source 2: REL.001.001.0011- REL.001.001.0020

Data source 3: REL.001.001.0021 - REL.001.001.0030

15.1.1.6 Overflow scenarios

Children overflow during republish

If the number of new children found during republish is higher than the maximum allowed by the suffix padding digits, then Relativity would use the next consecutive number without increasing the padding of the previous published children.

Initial Publish:

Number	Document
REL.001.001.9998	
REL.001.001.9999	(Password-protected file)
REL.001.002.0001	
REL.001.002.0002	

Republish:

Number	Document
REL.001.001.9998	
REL.001.001.9999	(Password-protected file)
REL.001.001.9999_0001	new child found in REL.001.001.9999
REL.001.001.9999_0002	new child found in REL.001.001.9999_0001
REL.001.001.9999_0003	new child found in REL.001.001.9999_0001
...	
REL.001.001.9999_9999	new child found in REL.001.001.9999 (uses 4 digits padding)
REL.001.001.9999_10000	new child found in REL.001.001.9999(uses 5 digits padding)
REL.001.002.0001	

15.1.2 Inventory / Discover settings

The **Inventory | Discover Settings** category of the profile layout provides the following fields.

DeNIST Settings

DeNIST* ☒ Yes
☐ No

DeNIST Mode ☒ DeNIST all files
☐ Do not break parent/child groups

Default Custodian Settings

Default OCR languages* English

Default time zone* (UTC) Coordinated Universal Time

Inclusion/Exclusion File List

Include/Exclude* ☒ Yes
☐ No

Mode* ☒ All Files
☐ Do not break parent/child groups

File Extensions*

(required)* ☐ Inclusion ☒ Exclusion

- **DeNIST**—if set to Yes, processing separates and removes files found on the [National Institute of Standards and Technology \(NIST\)](#) list from the data you plan to process so that they don't make it into Relativity when you publish a processing set. The NIST list contains file signatures—or hash values—for millions of files that hold little evidentiary value for litigation purposes because they are not user-generated. This list may not contain every known junk or system file, so deNISTing may not remove 100% of undesirable material. If you know that the data you intend to process contains no system files, you can select No. If the DeNIST field is set to Yes on the profile but the Invariant database table is empty for the DeNIST field, you can't publish files. If the DeNIST field is set to No on the processing profile, the DeNIST filter doesn't appear by default in Inventory, and you don't have the option to add it. Likewise, if the DeNIST field is set to Yes on the profile, the corresponding filter is enabled in Inventory, and you can't disable it for that processing set. The choices for this field are:
 - **Yes**—removes all files found on the NIST list. You can further define DeNIST options by specifying a value for the DeNIST Mode field.

Note: When DeNISTing, the processing engine takes into consideration everything about the file, including extension, header information and the content of the file itself. Even if header information is removed and the extension is changed, the engine is still able to identify and remove a NIST file. This is because it references the hashes of the system files that are found in the NIST database and matches up the hash of, for example, a Windows DLL to the hash of known DLL's in the database table.

- **No**—doesn't remove any files found on the NIST list. Files found on the NIST list are then published with the processing set.

Note: The same NIST list is used for all workspaces in the environment. Relativity makes new versions of the NIST list available shortly after the National Software Reference Library (NSRL) releases them quarterly. The NIST list is deployed to all instances by Relativity.

- **DeNIST Mode**—specify DeNIST options in your documents if DeNIST is set to Yes.
 - **DeNIST all files**—breaks any parent/child groups and removes any attached files found on the NIST list from your document set.
 - **Do not break parent/child groups**—doesn't break any parent/child groups, regardless if the files are on the NIST list. Any loose NIST files are removed.
- **Default OCR languages**—the language used to OCR files where text extraction isn't possible, such as for image files containing text. This selection determines the default language on the processing data sources that you create and then associate with a processing set. For more information, see [Adding a processing data source](#).
- **Default time zone**—the time zone used to display date and time on a processed document. This selection determines the default time zone on the processing data sources that you create and then associate with a processing set. The default time zone is applied from the processing profile during the discovery stage. For more information, see [Adding a processing data source](#).

The processing engine discovers all natives in UTC and then applies the time zone's UTC offset to all date and time metadata fields. When applicable, date and time values are further adjusted by the Daylight Savings Time (DST) offset. These adjusted values are also reflected in the text extracted during processing, for example the Sent Date/Time displayed in the header portion of email text.

Notes:

- Processing also uses the configured time zone to populate two system fields that are automatically mapped and published to the workspace: Relativity Native Time Zone Offset and Time Zone Field. These are used downstream for Native Viewer file conversion and Imaging activities. For more information, see [System-mapped fields](#).
 - This time zone setting does not affect the RSMF Filter date range in the Short Message Viewer, which uses the viewer's local machine time zone instead. See [RSMF file creation limitations](#) for details.
 - RSMF extracted text preserves source-file timestamps (typically UTC) regardless of the processing profile offset. Only the RSMF metadata fields are adjusted.
-
- **Include/Exclude**—enables the toggle for the inclusion/exclusion fields. The Inclusion/Exclusion File List allows you to manually enter custom lists of file extensions to either include or exclude. This gives greater flexibility to cull down data sets during Processing, resulting in faster Discovery, increased relevancy for review, and storage reduction. If DeNist and Include/Exclude are both selected, DeNist will run first.
 - **Yes**—reveals the additional associated inclusion/exclusion fields as required.
 - **No**—hides the additional associated inclusion/exclusion fields.
 - **Mode**—specifies Include/Exclude options in your documents if Include/Exclude is set to Yes.
 - **All files**—breaks any parent/child groups and removes any attached files found on the inclusion/exclusion list from your document set.
 - **Do not break parent/child groups**—doesn't break any parent/child groups, regardless if the files are on the inclusion/exclusion list. Any loose inclusion/exclusion files are removed.

- **File Extensions**—cross references the identified File Extension of the file, not its original extension. This long text field is used to enter the list of file extensions. The file extensions will be determined based on groupings of case insensitive alphanumeric characters. Hard returns are determined as delimiters and file a new extension. For example, the following list:

.dwg
.xml
.iso
.exe
.d

will create a list of .dwg, .xml, .iso, .exe, .d to exclude from Discovery

Note: File extensions must be separated with a hard return in order to be filed as a new extension. Extensions are case insensitive and should be entered as just the name of the extension (such as exe versus .exe).

- **Inclusion/Exclusion Selection**

- **Inclusion**—causes any File Extension within the list to be Discovered while all other to be filtered out.
- **Exclusion**—causes any File Extension within the list to be filtered out while all other File Extensions get included.

15.1.3 Extraction settings

The **Extraction Settings** category of the profile layout provides the following fields.

Note: For all text extraction methods described below, **Relativity** is recommended over both Native settings and dtSearch for performance and accuracy.

Extraction Settings ?

Extract children* ☒ Yes

☐ No

When extracting children, do not extract

☒ MS Office / Apple iWorks embedded images

☒ MS Office / Apple iWorks embedded objects

☒ Email inline images

Roll up image text

☐ Yes

☐ No

Include Linked Cloud Files With Email ?



Email Output*

Choose... ▼

Excel Text Extraction Method*

Relativity (Recommended) ▼

Excel Header/Footer* Extraction

☐ Do not extract

☒ Extract and place at end

☐ Extract and place inline (slows text extraction)

PowerPoint Text Extraction* Method

Relativity (Recommended) ▼

Word Text Extraction Method*

Relativity (Recommended) ▼

OCR*

☒ Enabled

☐ Disabled

OCR Accuracy

Medium (Average Speed) ▼

OCR Text Separator

☒ Enabled

☐ Disabled

- **Extract children**—Choose whether or not to extract child items during discovery, including message attachments, embedded objects and images, and other non-parent files. Select either:
 - **Yes**—to extract all child files during discovery so that both child and parent items are included in the processing job.

- **No**—to exclude child items, so that only parent items are included in the processing job. Selecting No hides the corresponding options for embedded image and object extraction, and image text roll up.

Note:

- The **Extract children** field must be set to Yes if you are working with short message data. If you set this field to No, attachments that are discovered and published are not viewable.
- You do not need to set the **Extract children** field to Yes to have files within PST, L01, ZIP, and other container files extracted. Relativity automatically expands these containers by default.

-
- **When extracting children, do not extract**—choose to include or exclude MS Office embedded images, MS Office embedded objects, and Email inline images. Options are:
 - **MS Office/Apple iWorks embedded images**—selecting this option excludes images of various file types found inside Microsoft Office files or Apple iWorks from discovery. As a result, these embedded images are not subject to text extraction and OCR, nor are they published as discrete items to the workspace.
 - **MS Office/Apple iWorks embedded objects**—selecting this option excludes objects of various file types found inside Microsoft Office files or Apple iWorks from discovery. As a result, these embedded objects are not subject to text extraction and OCR, nor are they published as discrete items to the workspace.
 - **Email inline images**—selecting this option excludes images of various files types found inside emails (such as .jpg, .bmp, or .png in an email) from discovery. As a result, these embedded in-line images are not subject to text extraction and OCR, nor are they published as discrete items to the workspace.

Notes:

- See [Microsoft Office child extraction support](#) for information on what MS Office documents can have embedded images and objects extracted.
- See [Email image extraction support](#) for information on what type of emails can have inline images extracted.
- Depending on how an object is embedded within an MS Office file, the content may display natively as either an icon, or as a static - and sometimes limited - view of the content itself. For example, in the case of a Word document with embedded Excel spreadsheet, Word creates and stores an .emf image file of the active sheet and no other sheets. This is in addition to storing the Excel file, unless the Microsoft option **Link to File** was chosen when embedding the object, in which case the .emf image file is the only representation of the embedded content. These .emf files are treated as embedded images by Relativity, and as such are subject to the **Extract children** and **Do not extract... MS Office embedded images** settings. Choosing not to suppress these MS Office embedded images results in the .emf file being OCR'd and published to the workspace. If text is needed for searching, but you want to reduce your workspace size, consider additionally enabling the **Roll Up Image Text** option described below.

-
- **Roll up image text**—selecting this option appends the image's text to the end of the parent document, following a text separator, and excludes the image from being published to the workspace. The benefit of this feature is cost savings due to reduced file count and size in the hosted workspace.

Notes:

- You must enable OCR to get any text for image text roll up.
- Images that have their text rolled up are marked as deleted. You can view a list of deleted items on the Files tab > Deleted files view.
- An image can be attached to a document as an MS Office embedded image, an Email inline image, or even an MS Office embedded object if the image is stored as an object. The **Roll up image text** option applies to any and all of these scenarios, as long as you have not chosen to suppress extraction using the **When extracting children, do not extract...** options.
- The exceptions to this general rule are images that are non-inline hidden attachments of MSG files, and images embedded in Microsoft OneNote files. The **Do not extract...** and **Roll up image text** options do not apply to these. Setting **Extract Children** to Yes results in the unconditional extraction, OCR, and roll up of these images, with only the parent file published to the workspace.

-
- **Include Linked Cloud Files With Email**—enabling this setting creates parent-child relationships between emails and any linked cloud files included in the data source. This option supports Google Vault exports of Google Mail with downloaded linked Google Drive items. See [Google Workspace](#) for more information on processing linked cloud files.

View a table summarizing the general behavior related to extract children settings

Extract children	When extracting children, do not extract embedded images, embedded objects, email inline images	Roll-up Image Text	Child Extracted?	Child text and OCR appended to parent?	Child published?
No	N/A	N/A	×	N/A	N/A
Yes	Checked - don't extract	No	×	N/A	N/A
Yes	Checked - don't extract	Yes	×	N/A	N/A
Yes	Unchecked - do extract	No	✓	×	✓
Yes	Unchecked - do extract	Yes	✓	✓	×

15.1.3.1 Other extraction settings

- **Email Output**—determines the file format in which emails will be published to the workspace. The options are:
 - **MSG**—publishes emails which are handled as MSGs during processing as **MSG**
 - **MHT**—converts and publishes emails which are handled as MSGs during processing as **MHT**

Notes:

- This option affects the following file types: Outlook files, Lotus Notes files, Bloomberg files.
 - Relativity does not convert EML formats to MHT.
 - Hashing for deduplication is performed on emails before conversion to MHT. The Processing Duplicate Hash value contains the Body, Header, Recipient, and Attachment hashes instead of the SHA256 hash used on native MHTs. After conversion, unique information from MSGs may render the same in the resulting MHT due to the files format. An example is two MSG's that contain "[www.test.com [http://www.test.com]]" and "www.test.com<http://www.test.com/>" in their respective text. During hash generation, these MSG's result in unique body hashes. When converted to an MHT, this text renders as "www.test.com<http://www.test.com/>". You can view or map individual Body, Header, Recipient, and Attachment hashes from the Files tab.
-
- This conversion happens during discovery.
 - MSG files take up unnecessary space because attachments to an MSG are stored twice, once with the MSG itself and again when they're extracted and saved as their own records. As a result, when you convert an MSG to an MHT, you significantly reduce your file storage because MHT files do not require duplicative storage of attachments.
 - If you need to produce a native email file while excluding all privileged or irrelevant files, convert the email native from MSG to MHT by using the Email Output field. After an email is converted from MSG to MHT, the MHT email is published to the workspace separately from any attachments, reducing the chance of accidentally producing privileged attachments.
 - Once you convert an MSG file to MHT, you cannot revert this conversion after the files have been published. For a list of differences between how Relativity handles MSG and MHT files, see [MSG to MHT conversion considerations](#).

Note: There is also a Yes/No Processing field called Converted Email Format that tracks whether an email was converted to MHT.

- **Excel Text Extraction Method**—determines whether the processing engine uses Excel, Relativity, or dtSearch to extract text from Excel files during publish. For a list of the differences between text extraction methods, see [Text extraction method considerations](#).

- **Relativity (Recommended)**—Relativity uses its built-in engine to extract text from Excel files.

Note: Using Relativity's built-in engine is the recommended method for performance and accuracy.

- **Native (Legacy)**—Relativity uses Excel to extract text from Excel files.
 - **Native - failover to dtSearch (Legacy)**—Relativity uses Excel to extract text from Excel files with dtSearch as a backup text extraction method if extraction fails.
 - **dtSearch- failover to Native (Legacy)**—Relativity uses dtSearch to extract text from Excel files with Native as a backup text extraction method if extraction fails. This typically results in faster extraction speeds; however, we recommend considering some differences between dtSearch and Native

extraction. For example, dtSearch doesn't support extracting the Track Changes text from Excel files. For more considerations like this, see [dtSearch special considerations](#). If you select dtSearch, you will see a warning that indicates this is a less optimal choice and to consider selecting Relativity (Recommended) instead.

- **Excel Header/Footer Extraction**—extract header and footer information from Excel files when you publish them. This is useful for instances in which the header and footer information in your Excel files is relevant to the case. Any selections for this field are nullified if you select dtSearch for the extraction method. Header and footer options are:
 - **Do not extract**—doesn't extract any of the header or footer information from the Excel files and publishes the files with the header and footer in their normal positions. This option is selected by default; however, if you change the value for the Excel Text Extraction Method field above from dtSearch, back to Native, this option will be de-selected and you'll have to select one of these options in order to save the profile.
 - **Extract and place at end**—extracts the header and footer information and stacks the header on top of the footer at the end of the text of each sheet of the Excel file. Note that the native file will still have its header and footer.
 - **Extract and place inline (slows text extraction)**—extracts the header and footer information and puts it inline into the file. The header appears inline directly above the text in each sheet of the file, while the footer appear directly below the text. Note that this could impact text extraction performance if your data set includes many Excel files with headers and footers. Note that the native file will still have its header and footer.
- **PowerPoint Text Extraction Method**—determines whether the processing engine uses PowerPoint, Relativity, or dtSearch to extract text from PowerPoint files during publish. For a list of the differences between text extraction methods, see [Text extraction method considerations](#).
 - **Relativity (Recommended)**—Relativity uses its built-in engine to extract text from PowerPoint files.

Note: Using Relativity's built-in engine is the recommended method for performance and accuracy.

 - **Native (Legacy)**—Relativity uses PowerPoint to extract text from PowerPoint files.
 - **Native - failover to dtSearch (Legacy)**—Relativity uses PowerPoint to extract text from PowerPoint files with dtSearch as a backup text extraction method if extraction fails.
 - **dtSearch - failover to Native (Legacy)**—Relativity uses dtSearch to extract text from PowerPoint files with Native as a backup text extraction method if extraction fails. This typically results in faster extraction speeds; however, we recommend considering some differences between dtSearch and Native extraction. For example, dtSearch doesn't support extracting watermarks from pre-2007 PowerPoint files, and also certain metadata fields aren't populated when using dtSearch. For more considerations like this, see [dtSearch special considerations](#).
- **Word Text Extraction Method**—determines whether the processing engine uses Word, Relativity, or dtSearch to extract text from Word files during publish. For a list of the differences between text extraction methods, see [Text extraction method considerations](#).
 - **Relativity (Recommended)**—Relativity uses its built-in engine to extract text from Word files.

Note: Using Relativity's built-in engine is the recommended method for performance and accuracy.

 - **Native (Legacy)**—Relativity uses Word to extract text from Word files.
 - **Native - failover to dtSearch (Legacy)**—Relativity uses Word to extract text from Word files with dtSearch as a backup text extraction method if extraction fails.

- **dtSearch - failover to Native (Legacy)**—Relativity uses dtSearch to extract text from Word files with Native as a backup text extraction method if extraction fails. This typically results in faster extraction speeds; however, we recommend considering some differences between dtSearch and Native extraction. For example, dtSearch doesn't support extracting watermarks from pre-2007 Word files, and also certain metadata fields aren't populated when using dtSearch. For more considerations like this, see [dtSearch special considerations](#).
- **OCR**—select Enable to run OCR during processing. If you select Disable, Relativity won't provide any OCR text in the Extracted Text view.

Notes:

- For information regarding how processing OCR works, see [Special considerations - OCR and text extraction](#).
 - OCR performed during processing is not intended to be a complete replacement for imaging and OCR features available post-publish in the workspace. For example, no options are provided here for configuring image dpi, dithering algorithm and threshold, image pre-processing such as deskewing and parallax distortion removal, and image auto-rotation.
 - If OCR is not essential to your processing job, it is recommended to disable OCR on your processing profile. Forgoing OCR at this stage can significantly reduce processing time and prevent irrelevant documents from having OCR performed on them. You can then perform OCR on only relevant documents post-publish within the workspace.
-
- **OCR Accuracy**—determines the desired accuracy of your OCR results and the speed with which you want the job completed. This drop-down menu contains three options:
 - **High (Slowest Speed)**—Runs the OCR job with the highest accuracy and the slowest speed.
 - Use the High setting for documents with dark backgrounds, or handwritten text.
 - **Medium (Average Speed)**—Runs the OCR job with medium accuracy and average speed.
 - **Low (Fastest Speed)**—Runs the OCR job with the lowest accuracy and fastest speed.
 - Use the Low setting for crisp, or electronically-generated text.
 - **OCR Text Separator**—select Enable to display a separator between extracted text at the top of a page and text derived from OCR at the bottom of the page in the Extracted Text view. The separator reads as, “--- OCR From Images ---”. With the separator disabled, the OCR text will still be on the page beneath the extracted text, but there will be nothing to indicate where one begins and the other ends. By default, this option is enabled.

Note: When you process files with both the OCR and the OCR Text Separator fields enabled, any section of a document that required OCR will include text that says **OCR from Image**. This can then pollute a dtSearch index because that index is typically built off of the extracted text field, and OCR from Image is text that was not originally in the document.

15.1.4 Short message conversion settings

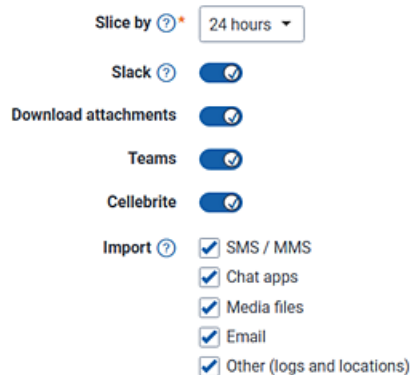
Import short message files in their native format directly into Relativity for processing. This feature eliminates having to convert short message files to RSMF (Relativity Short Message Format) before processing. You can define conversion settings in the processing profile's **Short Message Conversion Settings** section. The short message conversion settings you define only apply to processing jobs where RSMF conversion occurs during processing. The settings do not impact data already in RSMF format before processing takes place.

Note:

- To view more detailed information on processing an RSMF file, including the steps needed to prepare for processing, see Processing an RSMF file.
 - To view information on supported file types for short messages, see:
 - Short message conversion for Slack
 - Short message conversion for Cellebrite
 - Short message conversion for Microsoft Teams.
 - For short message mapping considerations, see Relativity's short message format.
-

Use the following settings to define short message conversion extraction parameters.

Short Message Conversion Settings



The screenshot shows the 'Short Message Conversion Settings' interface. At the top, there is a 'Slice by' dropdown menu set to '24 hours'. Below this are four toggle switches, all of which are turned on: 'Slack', 'Download attachments', 'Teams', and 'Cellebrite'. Under the 'Import' section, there are five checkboxes, all of which are checked: 'SMS / MMS', 'Chat apps', 'Media files', 'Email', and 'Other (logs and locations)'.

- **Slice by**—determines how Relativity splits conversations into RSMF files in terms of time.
 - **4 hours**—conversations are grouped in 4-hour blocks starting from 00:00 on any day where the message in the conversation exists.
 - **8 hours**—conversations are grouped in 8-hour blocks starting from 00:00 on any day where the message in the conversation exists.
 - **12 hours**—conversations are grouped in 12-hour blocks starting from 00:00 on any day where the message in the conversation exists.
 - **24 hours**—conversations are grouped in 24-hour blocks starting from 00:00 on any day where the message in the conversation exists.
 - **1 week**—conversations are grouped in 7-day blocks starting from Monday where the message in the conversation exists.
-

Notes:

- This setting determines how a conversation is split into RSMF documents in terms of time.
 - Slicing takes into consideration the processing time zone and shifts slices accordingly.
 - Slicing does not affect any files RSMF processed. It affects only the files generated with short message conversion.
 - See RSMF slicing configuration reference for complete details on slicing and slicing intervals.
-

- **Slack**—use this toggle to turn on or off the conversion of Slack export containers to RSMF.
 - **On**—select this option to convert Slack export containers to RSMF during processing. When toggled on, you will see additional fields for downloading attachments and for setting slicing time blocks.
 - **Off**—select this option to turn off conversion of Slack export containers to RSMF. When toggled off, you will not see the additional fields for downloading attachments or setting slicing time blocks.
- **Download Attachments**—use this toggle to download attachments from Slack servers (files.slack.com/) while converting conversations to RSMF.
 - **On**—select this option to download attachments and include them as standalone files during RSMF conversion.
 - **Off**—select this option to exclude attachments. Relativity retains links to output RSMF files instead of downloading the actual files.

Caution: Enabling Download Attachments can significantly increase total data volume. Slack exports commonly expand by 100 times or more when you download attachments. For example, a 1 GB export may result in 100–300 GB of processed data. Verify available storage and processing budget before enabling this option.

- **Teams**—use this toggle to turn on or off the conversion of Teams data to RSMF during the processing of PST files.
 - **On**—select this option to convert Teams data to RSMF during the processing of PST files.
 - **Off**—select this option to turn off conversion of Teams data to RSMF during the processing of PST files. This will process all Teams data found in PST as MSG files.
- **Cellebrite**—use this toggle to turn on or off the conversion of Cellebrite UFDR files to RSMF. See the table below the options to view what is included and what is silently skipped.
 - **SMS/MMS**—message data.
 - **Chat apps**—including but not limited to:
 - WhatsApp
 - Facebook Messenger
 - WeChat
 - Telegram
 - **Media files**—including but not limited to:
 - Images
 - Audio
 - Videos
 - Documents
 - **Email**—see the Generated RSMF fields table for a full list of extracted email metadata.
 - **Other (logs and locations)**—including but not limited to:
 - Call logs
 - Browser history logs
 - Browser search logs

- Notes
- Passwords
- Contacts
- Web bookmarks
- **Import**—choose whether to include data from UFDR.
 - SMS / MMS—includes text messages and multimedia messages.
 - Chat apps—covers messages from applications likes WhatsApp, Facebook Messenger, and similar apps.
 - Media files—encompasses photos, videos, and audio files.
 - Email—includes email files. See Email data extracted from Cellebrite for a list of fields included or not included during conversion.
 - Other (logs and locations)—contains logs, location data, and other files.

This table lists the data that is included when Cellebrite options are enabled or skipped when disabled.

Import option	What it includes	What is silently skipped if disabled
SMS/MMS	Text messages and multimedia messages	All SMS and MMS records
Chat apps	WhatsApp, Facebook Messenger, and similar apps	All chat application conversations
Media files	Photos, videos, and audio files	All media content
Other	Logs, location data, and miscellaneous files	Logs and other non-message content

Note: Enable all import options unless you have a specific reason to exclude a data type. Disabling any option silently skips all matching content during processing — no error or warning is generated, and the output will be smaller than expected.

15.1.4.1 Limitations when creating RSMF files

There are a few limitations when creating RSMF files.

- There is a viewer condition that restricts the number of events to 10,000.
- There is a processing condition that limits each RSMF file to a maximum size of 2 GB. During short message conversion of source data (such as Slack, Teams, or Cellebrite events) into an RSMF file, Relativity splits conversations into multiple files if they exceed either of these conditions.
- Each RSMF file is limited to a maximum size of 2 GB. Files exceeding this limit cannot be parsed, previewed, or downloaded within Relativity. There is no in-product path for processing oversized files. If an RSMF file exceeds 2 GB, split it into smaller files externally before uploading. When splitting, ensure that conversation threads are not divided across file boundaries, as this may result in incomplete conversation records in the output.

Note: If an oversized file has already been processed and is not accessible, contact Relativity Support for assistance.

- The date range filter in the Short Message Viewer displays timestamps in the viewer's local machine time zone, regardless of the time zone configured in the processing profile. This behavior affects both Relativity aiR and Relativity Server 2024. As a workaround, manually adjust your filter dates to account for the offset between your local time zone and the processing profile time zone.

Note: Orphaned messages

If a conversation extends across multiple time frames because of slicing, it will be divided into several files. As a result, messages from a single thread will be stored in different files that correspond to each sliced time frame. Messages that are saved outside of the parent conversation will become orphaned, meaning they will no longer be linked to the parent conversation. Relativity adds a note at the top of the message indicating it is an orphan.

Friday, July 30, 2021



Orphaned Replies from OrgWideTeamCreatedByDC_1800GroupConversation-AccessibleToEveryone

KA



That is one thing I do not miss

7:40 AM

15.1.5 Deduplication settings

The **Deduplication Settings** category of the profile layout provides the following fields:

Deduplication Settings ?

Deduplication method* ☐ None
☒ Global
☐ Custodial

Propagate deduplication data* ☐ Yes
☒ No

- **Deduplication method**—the method for separating duplicate files during discovery. During deduplication, the system compares documents based on certain characteristics and keeps just one instance of an item when two or more copies exist. The system performs deduplication against published files only. Deduplication doesn't occur during inventory or discovery. Deduplication only applies to parent files; it doesn't apply to children. If a parent is published, all of its children are also published. Select from the following options. For details on how these settings work, see [Deduplication considerations](#):

Note: Don't change the deduplication method in the middle of running a processing set, as doing so could result in blank DeDuped Custodians or DeDuped paths fields after publish, when those fields would otherwise display deduplication information.

- **None**—no deduplication occurs.
 - Even when you select None as the deduplication method, Relativity identifies duplicates by storing one copy of the native document on the file repository and using metadata markers for all duplicates of that document.
 - Relativity doesn't repopulate duplicate documents if you change the deduplication method from None after processing is complete. Changing the deduplication method only affects subsequent

processing sets. This means that if you select global deduplication for your processing settings, you can't then tell Relativity to include all duplicates when you go to run a production.

- **Global**—arranges for documents from each processing data source to be de-duplicated against all documents in all other data sources in your workspace. Selecting this makes the Propagate deduplication data field below visible and required.

Note: If you select Global, there should be no exact e-mail duplicates in the workspace after you publish. The only exception is a scenario in which two different e-mail systems are involved, and the e-mails are different enough that the processing engine can't exactly match them. In the rare case that this happens, you may see email duplicates in the workspace.

- **Custodial**—arranges for documents from each processing data source to be de-duplicated against only documents in data sources owned by that custodian. Selecting this makes the Propagate deduplication data field below visible and required.

Note: Deduplication is run on custodian ID's; there's no consequence to changing a custodian's name after their files have already been published.

- **Propagate deduplication data**—applies the deduplication fields you mapped out of deduped custodians, deduped paths, all custodians, and all paths field data to children documents, which allows you to meet production specifications and perform searches on those fields without having to include family or overlay those fields manually. This field is only available if you selected Global or Custodial for the deduplication method above. You have the following options:
 - Select **Yes** to have the metadata fields you mapped populated for parent and children documents out of the following: All Custodians, Deduped Custodians, All Paths/Locations, Deduped Paths, and Dedupe Count.
 - Select **No** to have the following metadata fields populated for parent documents only: All Custodians, Deduped Custodians, All Paths/Locations, and Deduped Paths.
 - If you republish a processing set that originally contained a password-protected error without first resolving that error, then the deduplication data won't be propagated correctly to the children of the document that received the error.
 - In certain cases, the Propagate deduplication data setting can override the extract children setting on your profile. For example, you have two processing sets that both contain an email message with an attachment of a Word document, Processing Set 1 and 2. You publish Processing Set 1 with the Extract children field set to Yes, which means that the Word attachment is published. You then publish Processing Set 2 with the Extract children field set to No but with the Deduplication method field set to Global and the Propagate deduplication date field set to Yes. When you do this, given that the emails are duplicates, the deduplication data is propagated to the Word attachment published in Processing Set 1, even though you didn't extract it in Processing Set 2.

15.1.6 Publish settings

The **Publish Settings** category of the profile layout provides the following fields.

Auto Publish

Auto-publish set* ☐ Yes
☒ No

Default Destination Folder Structure

Default destination folder* Sample Folder

Do you want to use source* ☒ Yes
folder structure ☐ No

- **Auto-publish set**—arranges for the processing engine to automatically kick off publish after the completion of discovery, with or without errors. By default, this is set to No. Leaving this at No means that you must manually start publish.
- **Default destination folder**—the folder in Relativity into which documents are placed once they're published to the workspace. This value determines the default value of the destination folder field on the processing data source. You have the option of overriding this value when you add or edit a data source on the processing set. Publish jobs read the destination folder field on the data source, not on the profile. You can select an existing folder or create a new one by right-clicking the base folder and selecting **Create**.
 - If the source path you selected is an individual file or a container, such as a zip, then the folder tree does not include the folder name that contains the individual file or container.
 - If the source path you selected is a folder, then the folder tree includes the name of the folder you selected.
- **Do you want to use source folder structure**—maintain the folder structure of the source of the files you process when you bring these files into Relativity.

Note: If you select **Yes** for **Use source folder structure**, subfolders matching the source folder structure are created under this folder. See the following examples:

Example 1 (recommended)

- **Select Source for files to process:** \\server.ourcompany.com\Fileshare\Processing Data\Jones, Bob\
- **Select Destination folder for published files:** Processing Workspace \ Custodians \

Results: A subfolder named Jones, Bob is created under the *Processing Workspace \ Custodians \ destination folder*, resulting in the following folder structure in Relativity: *Processing Workspace \ Custodians \ Jones, Bob *

Example 2 (not recommended)

- **Select Source for files to process:** \\server.ourcompany.com\Fileshare\Processing Data\Jones, Bob\
- **Select Destination folder for published files:** Processing Workspace \ Custodians \ Jones, Bob \

Results: A sub-folder named Jones, Bob is created under the *Processing Workspace \ Custodians \ Jones, Bob \ destination folder*, resulting in the following folder structure in Relativity: *Processing Workspace \ Custodians \ Jones, Bob \ Jones, Bob *. Any folder structure in the original source data is retained underneath.

If you select **No** for **Do you want to use source folder structure**, no sub-folders are created under the destination folder in Relativity. Any folder structure that may have existed in the original source data is lost.

15.1.7 Other considerations

The follow sections describe other considerations for numbering, prioritizing publishing speed, and dtSearch,

15.1.7.1 Parent/child numbering type examples

To better understand how each parent/child numbering option appears for published documents, consider the following scenario.

Your data source includes an MSG file containing three Word documents, one of which is password protected:

- MSG
 - Word Child 1
 - Word Child 2
 - Word Child 3 (password protected)
 - sub child 1
 - sub child 2

When you process the .msg file, three documents are discovered and published, and there's an error on the one password-protected child document. You then retry discovery, and an additional two sub-child documents are discovered. You then republish the processing set, and the new two documents are published to the workspace.

If you chose Suffix Always for the Parent/Child Numbering field on the profile, the identifiers of the published documents would appear as the following:

Numbering Settings		Control Number	File Name
Numbering Type *		Filter	Filter
☒ Auto Number			
☐ Define Start Number			
☐ Level Numbering			
Default document numbering * prefix			
REL			
Number of Digits * (padded with zeros)			
10			
Parent/Child Numbering *			
☒ Suffix Always			
☐ Continuous Always			
☐ Continuous, Suffix on Retry			
Delimiter *			
. (period)			
		REL0000000001	MSG1 .msg
		REL0000000001.0001	Word Child 1.docx
		REL0000000001.0002	Word Child 2.docx
		REL0000000001.0003	Word Child 3.docx
		REL0000000001.0003.0001	image1.emf
		REL0000000001.0003.0002	Microsoft_Word_Docume

If you chose Continuous Always for the Parent/Child Numbering field on the profile, the identifiers of the published documents would appear as the following:

Numbering Settings

Numbering Type* ☒ Auto Number
☐ Define Start Number
☐ Level Numbering

Default document numbering* prefix REL

Number of Digits* (padded with zeros) 10

Parent/Child Numbering* ☐ Suffix Always
☒ Continuous Always
☐ Continuous; Suffix on Retry

	Control Number	File Name
	Filter	Filter
 	REL0000000001	MSG1 .msg
 	REL0000000002	Word Child 1.docx
 	REL0000000003	Word Child 2.docx
 	REL0000000004	Word Child 3.docx
 	REL0000000005	Document 1.docx
 	REL0000000006	Document 2.docx

- In this case, the .msg file was the last document processed, and Word Child 3.docx was the first error reprocessed in a larger workspace. Thus, the sub child documents of Word Child 3.docx do not appear in the screen shot because they received sequence numbers after the last document in the set.

If you chose Continuous, Suffix on Retry for the Parent/Child Numbering field on the profile, the identifiers of the published documents would appear as the following:

Numbering Settings

Numbering Type* ☒ Auto Number
☐ Define Start Number
☐ Level Numbering

Default document numbering* prefix REL

Number of Digits* (padded with zeros) 10

Parent/Child Numbering* ☐ Suffix Always
☐ Continuous Always
☒ Continuous, Suffix on Retry

Delimiter* . (period)

	Control Number	File Name
	Filter	Filter
 	REL0000000001	MSG1 .msg
 	REL0000000002	Word Child 1.docx
 	REL0000000003	Word Child 2.docx
 	REL0000000004	Word Child 3.docx
 	REL0000000004.0001	image1.emf
 	REL0000000004.0002	Microsoft_Word_Docume

- Suffix on retry only applies to errors that have not been published to the workspace. If a document has an error and has been published, it will have a continuous number. If you resolve the error post-publish, the control number doesn't change.

15.1.7.2 Prioritizing publishing speed special considerations

Publishing speed can be prioritized by performing one of the following actions:

- setting the **Deduplication method** to None
- setting the **Create Source Folder Structure** to No

15.1.7.3 Suffix special considerations

Note the following details regarding how Relativity uses suffixes:

- For suffix child document numbering, Relativity indicates secondary levels of documents with a delimiter and another four digits appended for additional sub-levels. For example, a grandchild document with the assigned

prefix REL would be numbered *REL0000000001.0001.0001*.

- It is possible for your workspace to contain a document family that contains both suffixed and non-suffixed child documents. This can happen in the following example scenario:
 - You discover a primary (level 1) .msg file that contains child (level 2) documents and grandchild (level 3) documents, none of which contain suffixes.
 - One of the child documents yields an error.
 - You retry the error child document, and in the process you discover two grandchildren.
 - The newly discovered grandchildren are suffixed because they came from an error retry job, while the primary and non-error child documents remain without suffixes, based on the original discovery.

15.1.7.4 dtSearch special considerations

When you publish Word, Excel, and PowerPoint files with the text extraction method set to **dtSearch** on the profile, you'll typically see faster extractions speeds, but note that those file properties may or may not be populated in their corresponding metadata fields or included in the Extracted Text value.

The dtSearch text extraction method does not populate the following properties:

- In Excel, Track Changes in the extracted text.
- In Word, Has Hidden Data in the corresponding metadata field.
- In Word, Track Changes in the corresponding metadata field.
- In Powerpoint, Has Hidden Data in the corresponding metadata field.
- In Powerpoint, Speaker Notes in the corresponding metadata field.

Note: The dtSearch text extraction method will display track changes extracted text in-line, but changes may be poorly formatted. The type of change made is not indicated. The Native text extraction method will append track changes extracted text in a Tracked Change section.

The following table breaks down which file properties are populated in corresponding metadata fields and/or Extracted Text for the dtSearch text extraction method:

File type	Property	Included in dtSearch Corresponding metadata field	Included in dtSearch Extracted text
Excel (.xls, .xlsx)	Has Hidden Data	✓	✓
Excel (xls, .xlsx)	Track Changes (Inserted cell, moved cell, modified cell, clear cell, inserted column, deleted column, inserted row, deleted row, inserted sheet, renamed sheet)	✓	
Word (.doc, .docx)	Has Hidden Data		✓
Word (.doc, .docx)	Track Changes (Insertions, deletions, moves)		✓
Powerpoint (.ppt, .pptx)	Has Hidden Data		✓
Powerpoint (.ppt, .pptx)	Speaker Notes		✓

Note: Check marks do not apply to .xlsb files.

Note: Relativity does not possess a comprehensive list of all differences between the Native application and dtSearch text extraction methods. For additional information, see support.dtsearch.com.

15.1.7.5 Text extraction method considerations

As text extraction directly impacts search results, the following table lists which features are supported by the Relativity, Native, and dtSearch methods:

Features	Relativity			Native			dtSearch		
FEATURE DIFFERENCES	Excel Features Supported	Word Features Supported	Power Point Features Supported	Excel Features Supported	Word Features Supported	Power Point Features Supported	Excel Features Supported	Word Features Supported	Power Point Features Supported
Math equations. For more information, see Math equations .	Not Supported	Not Supported	Not Supported	✓	✓	✓	✓	✓	✓
Math formulas (sum, avg, etc.)	Not Supported	Not Supported	Not Supported	Not Supported	Not Supported	Not Supported	✓	✓	✓
SmartArt	✓ *	✓ *	✓ *	✓ *	✓ *	✓ *	✓ *	✓ *	✓ *
Speaker notes	N/A	N/A	✓ **	N/A	N/A	✓	N/A	N/A	✓ ***
Track changes	✓	✓	N/A	✓	✓	N/A	✓ ***	✓ ***	N/A
Hidden data	✓	✓	✓	✓	✓	✓	✓ ***	✓ ***	✓ ***
2016+ new chart styles	Not Supported	✓	✓	Not Supported	✓	✓	Not Supported	✓	✓
* Pre-2007 Office SmartArt are considered attachments and will be extracted and OCRd. ** When a header or footer is in the Speaker Notes section, field codes are not extracted. *** For more information, see dtSearch special Considerations .									
FULLY COMPATIBLE AND SUPPORTED FEATURES									
Bullet lists	✓	✓	✓	✓	✓	✓	✓	✓	✓
Chart box	✓	✓	✓	✓	✓	✓	✓	✓	✓
CJK and other foreign language characters	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clip art	✓	✓	✓	✓	✓	✓	✓	✓	✓
Comments and replies	✓	✓	✓	✓	✓	✓	✓	✓	✓

Currency format	✓	✓	N/A	✓	✓	N/A	✓	✓	N/A
Date / Time format	✓	✓	✓	✓	✓	✓	✓	✓	✓
Field codes	✓	✓	✓	✓	✓	✓	✓	✓	✓
Footer	✓	✓	✓	✓	✓	✓	✓	✓	✓
Header	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hidden slide	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A	✓
Macros	N/A	✓	N/A	N/A	✓	N/A	N/A	✓	N/A
Margins / Alignment Format	N/A	✓	N/A	N/A	✓	N/A	N/A	✓	N/A
Merged cell (horizontal)	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Merged cell (vertical)	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Number format (positive / negative)	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Number format (fraction)	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Number format (with comma)	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Number format (with decimal point)	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Password protected (cell level)	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Password protected (column level)	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Password protected (file level)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Password protected (row level)	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Password protected (sheet /	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A

page level)									
Phone number format	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Pivot table	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Right to left test format	N/A	✓	N/A	N/A	✓	N/A	N/A	✓	N/A
Slide numbers	N/A	N/A	✓	N/A	N/A	N/A	N/A	N/A	✓
Table	✓	✓	✓	✓	✓	✓	✓	✓	✓
Text box	✓	✓	✓	✓	✓	✓	✓	✓	✓
Transitions	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A	✓
WordArt	✓	✓	✓	✓	✓	✓	✓	✓	✓
Word wrapping format	N/A	✓	N/A	N/A	✓	N/A	N/A	✓	N/A

15.1.7.6 Math equations

The following table includes examples of what the extracted text would look like if Native or dtSearch are used rather than Relativity:

Original Document	Text Extraction Method		
	Relativity	Native	dtSearch
NO TEXT $(x+a)^n = \sum_{k=0}^n \binom{n}{k} x^k a^{n-k}$ $f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right)$		$(x+a)^n = \sum_{k=0}^n \binom{n}{k} x^k a^{n-k}$ $f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right)$	$(x+a)^n = \sum_{k=0}^n \binom{n}{k} x^k a^{n-k}$ $f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right)$

16 Deduplication considerations

The following scenarios depict what happens when you publish a processing set using each of the deduplication methods available on the processing profile.

Note the following special considerations regarding deduplication:

- Deduplication is applied only on Level 1 non-container parent files. If a child file (Level 2+) has the same processing duplicate hash as a parent file or another child file, then they will not be deduplicated, and they will be published to Relativity, regardless of whether the hash field has the same value. This is done to preserve family integrity. You can find out the level value for a file by mapping the Level metadata from the Field Catalog.
- In rare cases, it's possible for child documents to have different hashes when the same files are processed into different workspaces. For example, if the children are pre-Office 2007 files, then the hashes could differ based on the way the children are stored inside the parent file. In such cases, the child documents aren't identical (unlike in a ZIP container) inside the parent document. Instead, they're embedded as part of the document's

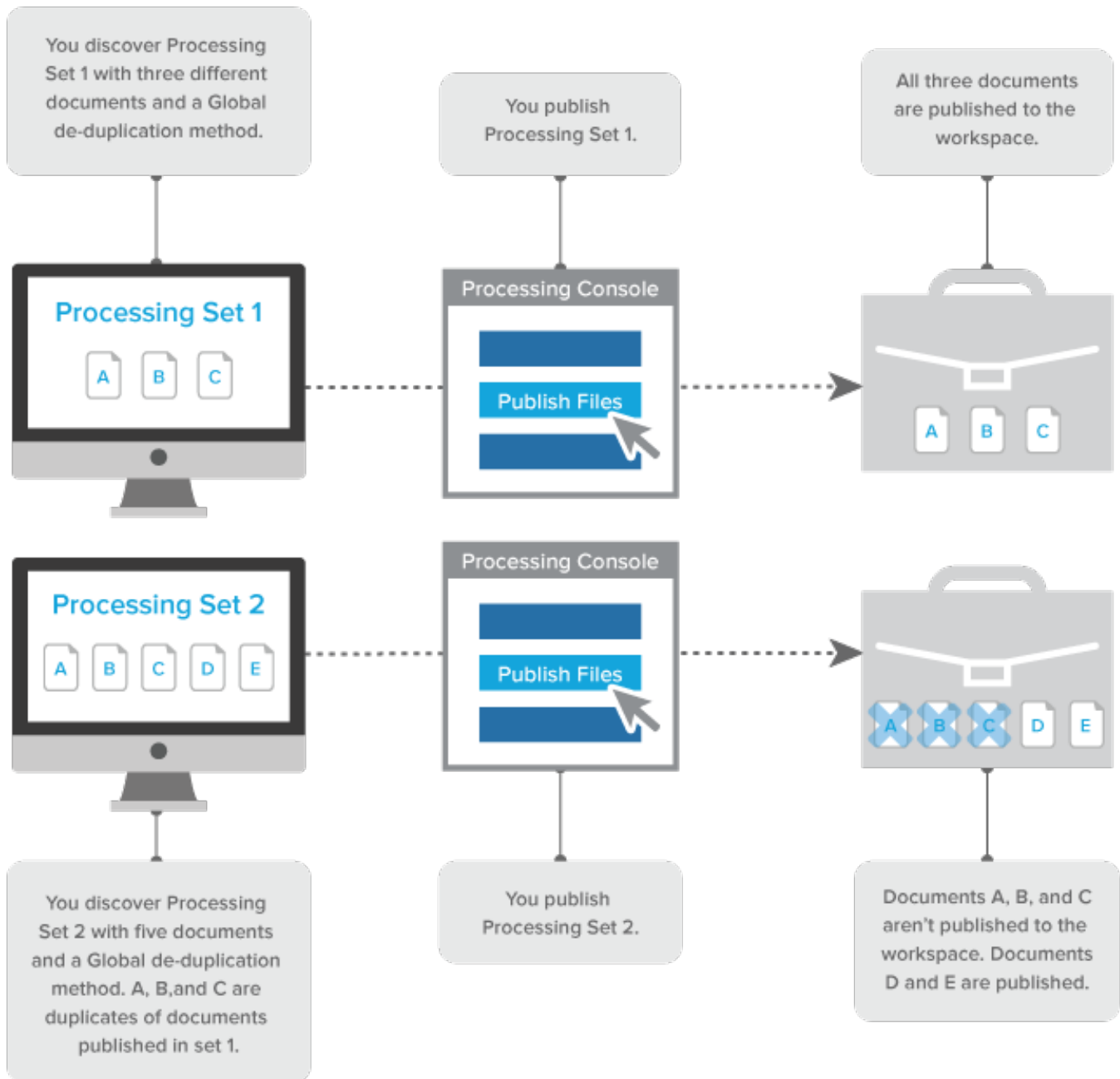
structure and have to be created as separate files, and the act of separating the files generates a new file entirely. While the content of the new file and the separated file are the same, the hashes don't match exactly because the OLE-structured file contains variations in the structure. This is not an issue during deduplication, since deduplication is applied only to the level-1 parent files, across which hashes are consistent throughout the workspace.

- At the time of publish, if two data sources have the same order, or if you don't specify an order, deduplication order is determined by Artifact ID.
- The deduplication method selected for a given Processing Set applies only to the data being published in that set. When a set is published with deduplication enabled, Relativity deduplicates it against all previously processed data in the workspace — regardless of what deduplication settings those prior sets used. See the following examples:
 - If the current set is configured with **None**, all documents will be published, regardless if past sets had deduplication enabled.
 - If the current set is configured with **Custodial** or **Global**, the current set will deduplicate against all applicable data (matching the Custodian or all previously Published sets), even data that was originally published with None.

16.1 Global deduplication

When you select Global as the deduplication method on the profile, documents that are duplicates of documents that were already published to the workspace in a previous processing set aren't published again.

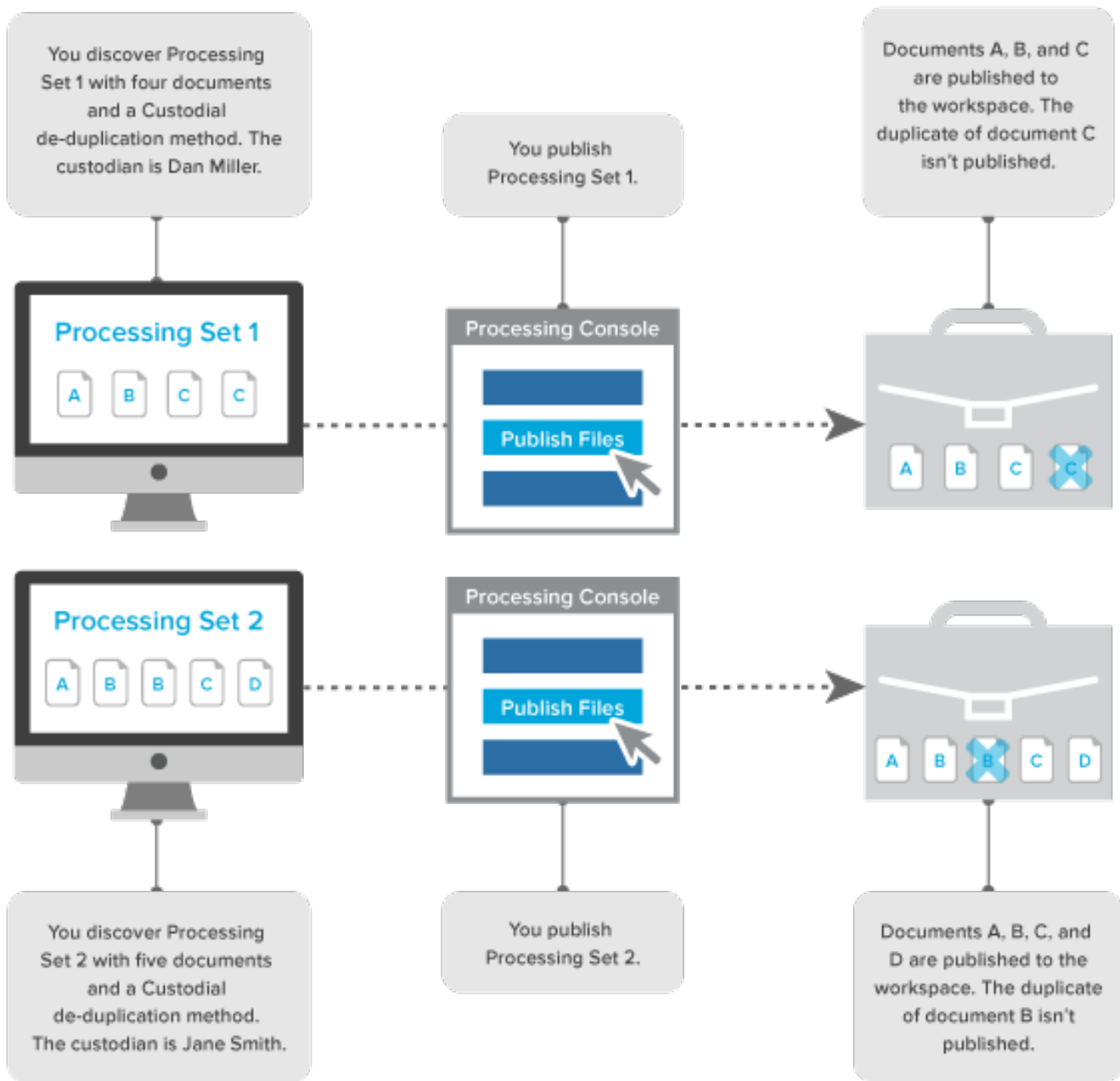
Global De-Duplication



16.2 Custodial deduplication

When you select Custodial as the deduplication method on the profile, documents that are duplicates of documents owned by the custodian specified on the data source aren't published to the workspace.

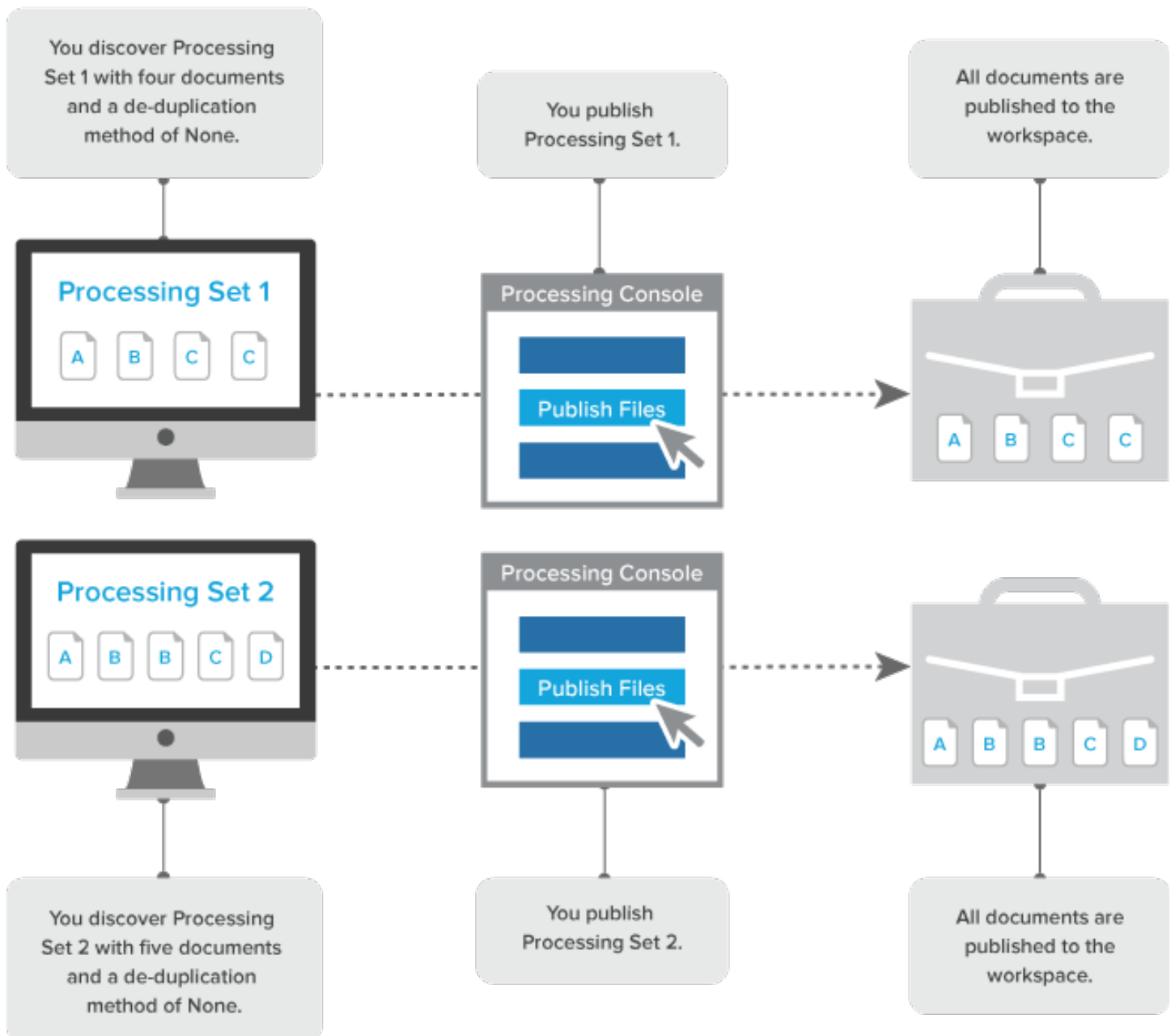
Custodial De-Duplication



16.3 No deduplication

When you select None as the deduplication method on the profile, all documents and their duplicates are published to the workspace.

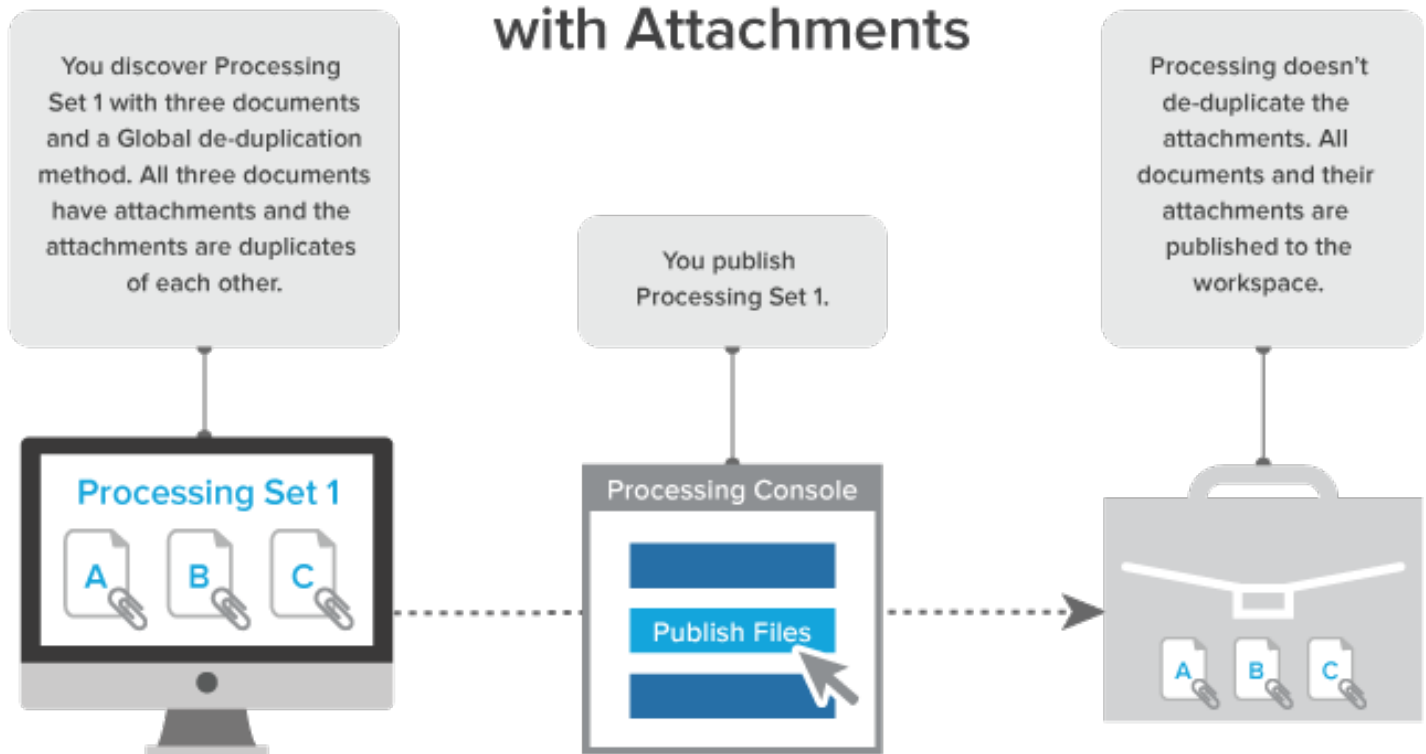
No De-Duplication



16.4 Global deduplication with attachments

When you select Global as the deduplication method on the profile and you publish a processing set that includes documents with attachments, and those attachments are duplicates of each other, all documents and their attachments are published to the workspace.

Global De-Duplication with Attachments



16.5 Global deduplication with document-level errors

When you select Global as the deduplication method on the profile, and you publish a processing set that contains a password-protected document inside a zip file, you receive an error. When you unlock that document and republish the processing set, the document is published to the workspace. If you follow the same steps with a subsequent processing set, the unlocked document is de-duplicated and not published to the workspace.



16.6 Technical notes for deduplication

The system uses the algorithms described below to calculate hashes when performing deduplication on both loose files (standalone files not attached to emails) and emails for processing jobs that include either a global or custodial deduplication.

The system calculates hashes in a standard way, specifically by calculating all the bits and bytes that make the content of the file, creating a hash, and comparing that hash to other files in order to identify duplicates.

The following hashes are involved in deduplication:

- **MD5/SHA1/SHA256 hashes**—provide a checksum of the physical native file.
- **Deduplication hashes**—the four email component hashes (body, header, recipient, and attachment) processing generates to de-duplicate emails.
- **Processing duplicate hash**—hash values generated during processing duplication are as follows: for loose files, a SHA256 hash is generated from the physical file's SHA256 hash. For emails, a hash is generated from the email's metadata properties. See [Calculating deduplication hashes for emails](#) for more details.

16.6.1 Fields used for comparing files in deduplication

The following fields are used when comparing emails for deduplication:

- With global deduplication, the Processing Duplicate Hash field is compared to the Processing Duplicate Hash fields for processed files, much like loose files.
- With custodial deduplication, the Processing Duplicate Hash field is compared to the Processing Duplicate Hash fields of specific custodians.
- The MD5, SHA256, SHA1, and Email Conversation Index are not considered in deduplication of email files.

Note: Office 2024 impact on EML Appointments in Processing

Starting in September 2025, Relativity handlers are using Office 2024 instead of Office 2016 for processing. This change may have an impact on the processing of EML Appointments. Hash values generated using Office 2016 may be different than hash values generated using Office 2024. As a result, some files previously identified as duplicates may no longer match if they are processed before and after the change. To ensure consistent hash values and discovery results, reprocess the files.

See the KB article, [Office 2024 impact on EML Appointments in Processing](#) for more detailed information.

The upgrade from Office 2016 to Office 2024 will be released via a phased roll-out, which means some customers will receive the update before others. We anticipate the roll-out phase being complete by the end of September. For more information on phased roll-outs, see the [Phased Rollout FAQ](#) on the Community.

16.6.2 Calculating MD5/SHA1/SHA256 hashes

To calculate a file hash for native files, the system:

1. Opens the file.
2. Reads 8k blocks from the file.
3. Passes each block into an MD5/SHA1/SHA256 collator, which uses the corresponding standard algorithm to accumulate the values until the final block of the file is read. Envelope metadata (such as filename, create date, last modified date) is excluded from the hash value.
4. Derives the final checksum and delivers it.

Note: Relativity cannot calculate the MD5 hash value if you have FIPS (Federal Information Processing Standards cryptography) enabled for the instance.

16.6.3 Calculating deduplication hashes for emails

Relativity calculates the MessageBody, Header, and Recipient hash values from the text metadata, which are case-sensitive. A hash generated from Bob.Smith@Email.com will be different from a hash generated from bob.smith@email.com. Attachment hash values are generated from the attached native and are not case sensitive.

16.6.3.1 MessageBodyHash

To calculate an email's MessageBodyHash, the system:

1. Captures the PR_BODY tag from the MSG (if it's present) and converts it into a Unicode string.
2. Gets the native body from the PR_RTF_COMPRESSED tag (if the PR_BODY tag isn't present) and either converts the HTML or the RTF to a Unicode string.
3. Removes all carriage returns, line feeds, spaces, and tabs from the body of the email to account for formatting variations. An example of this is when Outlook changes the formatting of an email and displays a message stating, "Extra Line breaks in this message were removed."

Note: The removal of all the components mentioned above is necessary because if the system didn't do so, one email containing a carriage return and a line feed and another email only containing a line feed would not be deduplicated against each other since the first would have two spaces and the second would have only one space.

4. Constructs a SHA256 hash from the Unicode string derived in step 2 or 3 above.

16.6.3.2 HeaderHash

To calculate an email's HeaderHash, the system:

1. Constructs a Unicode string containing Subject<crLf>SenderName<crLf>SenderEMail<crLf>ClientSubmitTime.
2. Derives the SHA256 hash from the header string. The ClientSubmitTime is formatted with the following: m/d/yyyy hh:mm:ss AM/PM. The following is an example of a constructed string:
 - RE: Your last email
 - Robert Simpson
 - robert@relativity.com
 - 10/4/2010 05:42:01 PM

16.6.3.3 RecipientHash

The system calculates an email's RecipientHash through the following steps:

1. Constructs a Unicode string by looping through each recipient in the email and inserting each recipient into the string. Note that BCC is included in the Recipients element of the hash.
2. Derives the SHA256 hash from the recipient string RecipientName<crLf>RecipientEMail<crLf>. The following is an example of a constructed recipient string of two recipients:
 - Russell Scarcella
 - rscarcella@relativity.com
 - Kristen Vercellino
 - kvercellino@relativity.com

16.6.3.4 AttachmentHash

To calculate an email's AttachmentHash, the system:

1. Derives a SHA256 hash for each attachment.
 - If the attachment is a not an email, the normal standard SHA256 file hash is computed for the attachment.
 - If the attachment is an e-mail, we use the e-mail hashing algorithm described to [Calculating deduplication hashes for emails](#) to generate all four de-dupe hashes. Then, these hashes are combined, as described in [Calculating the Relativity deduplication hash](#), to generate a single SHA256 attachment hash.
2. Encodes the hash in a Unicode string as a string of hexadecimal numbers without <crLf> separators.
3. Constructs a SHA256 hash from the bytes of the composed string in Unicode format. The following is an example of constructed string of two attachments:
 - 80D03318867DB05E40E20CE10B7C8F511B1D0B9F336
 - ↳ EF2C787CC3D51B9E26BC9974C9D2C0EEC0F515C770B8
 - ↳ 282C87C1E8F957FAF34654504520A7ADC2E0E23EA

ICS/VCF files are deduplicated not as emails but as loose files based on the SHA256 hash. Since the system now considers these loose files, Relativity is no longer capturing the email-specific metadata that it used to get as a result of ICS/VCF files going through the system's email handler.

16.6.4 Calculating the Relativity deduplication hash

To derive the Relativity deduplication hash, the system:

1. Constructs a string that includes the SHA256 hashes of all four email components described above, as seen in the following example. For more information, see [Calculating deduplication hashes for emails](#).
 - 6283cfb34e4831c97e363a9247f1f01beaaed01d
 - ↳ b3a65a47be310c27e3729a3ee05dce5acaec3696
 - ↳ c681cd7eb646a221a8fc376478b655c81214dca7
 - ↳ 419aabee6283cfb34e4831c97e363a9247f1f01b
 - ↳ eaaed01db3a65a47be310c27e3729a3ee3843222
 - ↳ f1805623930029bad6f32a7604e2a7acc10db9126e3
 - ↳ 4d7be289cf86e
2. We convert the above string to a UTF-8 byte array.
3. We then take that byte array and generate a SHA256 hash of it.

Notes:

- If two emails have an identical body, attachment, recipient, and header hash, they are duplicates.
 - For loose files, the Processing Duplicate Hash is a hash of the file's SHA256 hash.
-

16.6.5 Deduplication for short message formats

Relativity short message format (RSMF) files produced from Slack, Microsoft Teams, Cellebrite/UFDR, and Google Chat data are deduplicated in two separate layers. The layers run at different stages, use different comparison logic, and operate on different units of data:

- **Layer 1 - conversion-time deduplication** runs while RSMF files are being written. It operates on individual events (messages) inside a conversation, and its behavior is different for each source format.
- **Layer 2 - publish-time deduplication** runs during the Publish phase. It operates on whole RSMF files, and its behavior is the same regardless of source format.

Understanding which layer marked an item is the key to interpreting deduplication results. A message that survives Layer 1 can still end up inside a file that Layer 2 marks as a duplicate, and a file that Layer 2 marks as unique may still contain messages that Layer 1 collapsed.

16.6.5.1 Layer 1 — conversion-time deduplication

During conversion, each source format handles repeated or consecutive events differently before RSMF files are written. Only Cellebrite/UFDR performs true duplicate filtering. Slack merges consecutive events that share a timestamp. Teams and Google Chat apply no duplicate filtering at all, but both fold message edits and deletions into the original event.

Format	Duplicate filtering	Edit and deletion handling	Duplicate counts written
Slack	Consecutive events with the same time stamp are merged	Folded into the prior event	None
Teams	None	Consecutive events with the same message ID are folded into the prior event	None
Cellebrite/UFDR	Yes — matching events are skipped	Not applicable	Detected duplicates field and X-RSMF-DuplicatedEventCount header
Google Chat	None	Message-version entries are folded into the original event	None

Slack

The converter compares each event's timestamp against the previous written event in the same conversation. If a consecutive event carries the same timestamp, it is merged into the prior event — including any edits or deletions — rather than written as a new RSMF event.

The comparison looks only at the immediately preceding event, and the baseline resets at the start of each conversation. No duplicate count is retained, and no duplicate-count field or header is written.

Note: Slack's underlying event ID is a composite of user and timestamp. It is not the value used for this comparison.

Teams

No duplicate filtering is applied. No events are discarded, and unsupported or unrecognized event types are still emitted as placeholder events.

However, consecutive messages that share the same message ID are folded into the previous event as edit history, a deleted flag, or both, which reconstructs the edit and deletion history of the message. Because those versions are folded in, they do not appear as separate RSMF events. This is message-version handling rather than duplicate filtering.

Cellebrite/UFDR

The converter compares four values against the single previous written event:

- timestamp
- sender ID
- message body
- attachments, compared as an order-independent set

If all four match, the event is skipped instead of written.

Cellebrite/UFDR is the only format that reports duplicate counts, and it produces two distinct values:

- **Detected duplicates**—a per-event custom field recording the length of the current consecutive run, written on the primary event.
- **X-RSMF-DuplicatedEventCount**—a per-conversation total written to the RSMF header. This value is surfaced downstream as the field `Rsmf/DuplicatedMessageCount`.

Google Chat

No conversion-time deduplication is applied. Every original message is written to RSMF. The converter retains no duplicate count, emits no **Detected duplicates** field, and writes no `X-RSMF-DuplicatedEventCount` header.

Message edits and deletions arrive as separate message-version entries and are folded into the original event, keyed on version type and matching message ID. An edit adds edit history and updates the message body; a deletion sets a deleted flag. As with Teams, this is message-version handling rather than duplicate filtering.

16.6.5.2 Layer 2 — publish-time file deduplication

During the Publish phase, each file is assigned a deduplication identity built from a fixed, ordered set of SHA-256 hash slots: file, header, recipient, body, and attachment. Slots that are empty or disabled are zero-filled. The slots are concatenated in order — the identity is a conditional concatenation, not an arithmetic sum of the hashes.

For an RSMF file, which is treated as a chat or message item, the file-bytes slot is always empty. The effective identity is therefore content-only:

```
RSMF deduplication identity = header_hash + recipient_hash + body_hash + attachment_hash
```

The file-bytes slot still exists in the underlying structure, but it is populated only for physical files that are neither email nor chat items.

Each component hash is calculated as follows:

Component	Calculated from
Header	Subject, From name and address, and date normalized to UTC
Recipient	To names and email addresses
Body	Normalized body text, with whitespace stripped
Attachment	The byte hashes of the embedded attachments

Deduplication statuses

The RSMF file, not the individual messages inside it, is the unit of deduplication. One hash record is created per file.

Within the set of jobs in scope, files are assigned one of the following statuses:

Status	Meaning
Primary	One file per group of files sharing an identity
Duplicate	The remaining files in that group
Unique	A file whose identity matches no other file in scope
None	Container files

Where multiple files share an identity, a single primary is selected based on ranked path.

Deduplication works the same way no matter where the data came from. Messages from Slack, Teams, Celebrite/UFDR, and Google Chat are all treated identically. The only distinction it makes is between message items and non-message items.

16.6.5.3 Key limitations

Deduplication does not operate on individual messages at publish time

Because Layer 2 compares whole files, two RSMF files that overlap in content but are not byte-for-byte equivalent in the hashed components will both be retained. Duplicate messages spread across non-identical files are not collapsed at publish time.

Time-based slicing determines whether overlapping exports deduplicate

RSMF files are produced by a configurable time-window slicer. The default **Slice by** value is 24 hours, and Google Chat uses the same slicer. A conversation is also force-sliced when it reaches either 10,000 events or 2 GB of attachments.

Slices are anchored to a fixed calendar grid (midnight, hour multiples, and Monday for weekly slices) rather than to the start date of each export. Daylight saving time is not taken into account. See RSMF slicing configuration reference for details on slice intervals.

Because slices anchor to a grid, two exports that use the same **Slice by** setting and both fully cover a given interior window produce the same partition of events for that window. Identical content yields an identical hash, so those interior slices do deduplicate against each other.

Non-matches are confined to:

- boundary slices at the edges of each export's date range, where coverage of the window is partial
- date ranges covered by only one of the two exports
- cases where the two exports use different Slice by settings

In practice, cross-export deduplication degrades at the boundaries and when Slice by settings differ. It does not fail wholesale.

17 Processing sets

A processing set is an object to which you attach a processing profile and at least one data source and then use as the basis for a processing job. When you run a processing job, the processing engine refers to the settings specified on the data sources attached to the processing set when bringing data into Relativity.

Consider the following about processing sets:

- A single processing set can contain multiple data sources.
- Only one processing profile can be added to a processing set.
- You can't delete a workspace in which there is an in-progress inventory, discovery, or publish job in the Processing Queue.
- Don't add documents to a workspace and link those documents to an in-progress processing set. Doing this distorts the processing set's report data.
- When processing data, Relativity works within the bounds of the operating system and the programs installed on it. Therefore, it can't tell the difference between a file that's missing because it was quarantined by anti-virus protection and a file that was deleted after the user initiated discovery.
- Never stop Relativity services through Windows Services or use the IIS to stop a processing job.
- Processing jobs are not paused during scheduled maintenance periods. However, if Relativity encounters an issue during this time, it will automatically retry the job. If the job fails, you can manually retry the errors through the Files or Job Exceptions tabs.

17.1 Processing sets default view

Use the Processing Sets sub-tab to see a list of all the processing sets in your environment.

Note: You can manually search for any processing set in the workspace by entering its name in the text box at the top of the list and clicking **Enter**. Relativity treats the search terms you enter here as a literal contains search, meaning that it takes exactly what you enter and looks for any processing set that contains those terms.

This view provides the following information:

Processing Sets Processing Data Sources Processing Profile More

> 🔍

New Processing Set

All Processing Sets

Add Widget

1 - 21 of 21 100 per page

#	Name	Inventory Status	Inventoried files	Discover Status	Discovered files	Publish Status
	Filter	(All)	=	(All)	=	(All)
1	Processing set 003	Not started		Not started		Not started
2	Processing set 002	Not started		Completed with errors	26,353	Completed with errors
3	Processing set 001	Completed	3,774	Completed with errors	7,981	Completed

All 3 Delete

Total: 3

- **Discovered files**—the number of files across all data sources on the set that have been discovered.

Note: Discovery populates job and document level errors.

- **Publish Status**—the current status of the publish phase of the set. This field could display any of the following status values:
 - Not started
 - In progress
 - Completed
 - Completed with errors
 - Canceled
- **Published documents**—the number of files across all data sources on the set that have been published to the workspace.

Note: By adding the **Originating Processing Set** document field to any view, you can indicate which processing set a document came from.

From the Processing Sets sub-tab you can:

- Open and edit an existing processing set.
- Perform the following mass operations on selected processing sets:
 - Delete
 - Export to File
 - Tally/Sum/Average

Note: The Copy, Edit, and Replace mass operations are not available for use with processing sets.

17.2 Creating a processing set

When you create a processing set, you are specifying the settings that the processing engine uses to process data.

To create a processing set:

1. Navigate to the **Processing** tab and then click the **Processing Sets** sub-tab.
2. Click the **New Processing Set** button to display the Processing Set layout.
3. Complete the fields on the Processing Set layout. See [Fields](#).
4. Click **Save**.
5. Add as many Processing Data Sources to the set as you need. See [Adding a processing data source](#).

Note: The frequency with which the processing set console refreshes is determined by the ProcessingSetStatusUpdateInterval entry in the Instance setting table. The default value for this is 5 seconds. 5 seconds is also the minimum value. See the Instance setting guide for more information.

17.3 Processing set fields

To create a processing set, complete the following fields:

Processing Sets
Processing Data Sources
Processing Profile
More ▾

Save
Save and New
Save and Back
Cancel

Basic Settings

Name*

Processing Profile*

Choose... ▾

Add

Email notification recipients
Separate by semi-colon

- **Name**—the name of the set.
- **Processing profile**—select any of the profiles you created in the Processing Profiles tab. If you haven't created a profile, you can select the Default profile or click **Add** to create a new one. If there is only one profile in the workspace, that profile is automatically populated here. See [Processing profiles](#).
- **Email notification recipients**—the email addresses of those whom you want to receive notifications while the processing set is in progress. Relativity sends an email to notify the recipient of the following:
 - Inventory
 - Successful inventory completed
 - Inventory completed with errors
 - First discovery job-level error
 - Inventory error during job submission
 - Discovery
 - Successful discovery completed
 - Discovery completed with errors
 - First discovery job-level error
 - File discovery error during job submission
 - Retry—discovery
 - First discovery retry job-level error
 - Discovery retry error during job submission
 - Publish
 - Successful publish completed
 - Publish complete with errors
 - First publish job-level error
 - Publish error during job submission

- Retry—publish
 - First publish retry job-level error
 - Publish retry error during job submission

Note: Email notifications are sent per the completion of processing sets, not data sources. This ensures that a recipient doesn't receive excessive emails. The exception to this is job-level errors. If all data sources encounter a job-level error, then Relativity sends an email per data source.

After you save the processing set, the layout is updated to include the process set status display. The display remains blank until you start either inventory or file discovery from the console. The console remains disabled until you add at least one data source to the set.

The screenshot displays the 'Processing Sets' interface for 'Processing Set 004'. The top navigation bar includes tabs for 'Processing Sets', 'Processing Data Sources', 'Processing Profile', 'Password Bank | Processing', 'Inventory', 'Reports', 'Job Errors', and 'Files'. Below the navigation bar are buttons for 'Edit', 'Delete', 'Back', 'Edit Permissions', and 'View Audit', along with a 'Feedback' link and a pagination indicator '1 of 22'.

The main content area is divided into two sections: 'Processing Set 004 Basic Settings' and 'Processing Data Source Job Errors'.

Processing Set 004 Basic Settings

This section shows a workflow diagram with three steps: 'Inventory (Optional) Not Started', 'Discover Not Started', and 'Publish Not Started'. The steps are numbered 1, 2, and 3 respectively. To the right of the diagram is a table showing counts for each step:

Inventory count	Discover count	Publish count	Total exceptions
0	0	0	0

Below the workflow diagram are three expandable sections: 'Inventory (Optional)', 'Discovery', and 'Publish'.

Processing Data Source Job Errors

This section shows a table with columns for 'Status', 'Percent complete', 'Source path', 'Custodian', 'Document numberi...', 'Time zone', and 'OCI'. The table is currently empty, displaying 'No data.'.

Process Files

On the right side of the interface, there is a sidebar with the following sections:

- Process Files**: Includes buttons for 'Inventory Files', 'Filter Files', 'Discover Files', 'Retry File Exceptions', and 'Publish Files'.
- Reports**: Includes links for 'Inventory Report', 'Discovery Report', and 'All Reports'.
- Exceptions**: Includes links for 'Job Errors' and 'File Exceptions'.

The Processing Set Status section of the set layout provides data and visual cues that you can use to measure progress throughout the life of the processing set. This display and the information in the status section refresh automatically every five seconds to reflect changes in the job.

Note: To create a Quick-create set, see the Quick-create set(s) documentation for more information.

17.4 Adding a data source

A Processing Data Source is an object you associate with a processing set in order to specify the source path of the files you intend to inventory, discover, and publish, as well as the custodian who facilitates that data and other settings.

Notes:

- Be sure your data sources have unique names. If you have data sources with duplicate names, you may see processing errors.
 - Example of duplicate data source names:
 - \\[file.share]\Processing Source[SOURCE.NAME]
 - \\[file.share]\Processing Source[Source.Name]
 - Example of unique data source names:
 - \\[file.share]\Processing Source[SOURCE.NAME1]
 - \\[file.share]\Processing Source[Source.Name2]
- You have the option of using Integration Points to import a list of custodians from Active Directory into the Data Sources object. Doing this would give you an evergreen catalog of custodians to pick from when preparing to run a processing job.
- You can add multiple data sources to a single processing set, which means that you can process data for multiple custodians through a single set. There is no limit to the number of data sources you can add to a set; however, most sets contain ten or fewer.
- During publish, if you have multiple data sources attached to a single processing set, Relativity starts the second source as soon as the first source reaches the DeDuplication and Document ID generation stage. Previously, Relativity waited until the entire source was published before starting the next one.

To add a data source:

1. Create and save a new processing set, or navigate into an existing set. See [Creating a processing set](#).
2. In the **Processing Data Source** section, click **New**.

The screenshot displays the Relativity Processing Sets interface. At the top, there are tabs for Processing Sets, Processing Data Sources, Processing Profile, Password Bank | Processing, Inventory, Reports, Job Errors, and Files. Below the tabs, there are buttons for Edit, Delete, Back, Edit Permissions, and View Audit. The main content area shows a progress bar with three steps: Inventory (Optional) Not Started, Discover Not Started, and Publish Not Started. To the right of the progress bar, there is a table showing counts for Inventory count, Discover count, Publish count, and Total exceptions, all of which are 0. Below the progress bar, there are expandable sections for Inventory (Optional), Discovery, and Publish. At the bottom, there is a section for Processing Data Source and Job Errors. The Processing Data Source section has a 'New' button highlighted with a red box. Below this, there is a table with columns for Status, Percent complete, Source path, Custodian, Document number..., Time zone, and OCR language(s). The table currently shows 'No data.'

3. Complete the fields on the **Add Processing Data Source** layout. See [Fields](#).
4. Click **Save**. When you save the data source, it becomes associated with the processing set and the console on the right side is enabled for inventory and file discovery.

For details on what information is displayed in the data source view while the processing set is running, see [Processing Data Source](#).

Note: If you add, edit, or delete a data source associated with a processing set that has already been inventoried but not yet discovered, you must run inventory again on that processing set. You can't add or delete a data source to or from a processing set that has already been discovered or if there's already a job in the processing queue for the processing set.

17.5 Data source fields

To add a data source, complete the following fields:

The screenshot shows the 'Processing Data Source Layout' form. At the top, there are tabs for 'Processing Sets', 'Processing Data Sources', 'Processing Profile', 'Password Bank | Processing', 'Inventory', 'Reports', and 'More'. Below the tabs are buttons for 'Save and New', 'Save and Back', and 'Cancel'. The form is divided into three main sections:

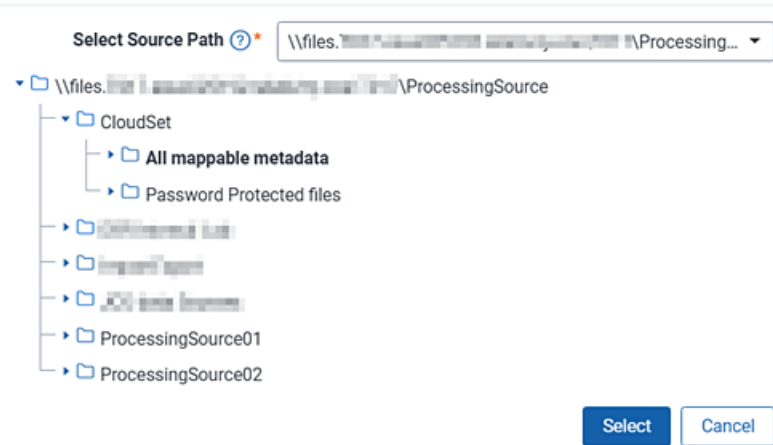
- Data**: Contains 'Source path' (with a browse button), 'Custodian' (with 'Select' and 'Add' buttons), and 'Destination folder' (with a text input and 'Select'/'Clear' buttons).
- Custodian Options**: Contains 'Time zone' (with a dropdown and 'Select'/'Clear' buttons), 'OCR language(s)' (with a dropdown and 'Select'/'Add'/'Clear' buttons), and 'Document numbering prefix' (with a text input and 'Start numbers' link).
- Advanced Options**: Contains 'Name' (with a text input) and 'Order' (with a text input).

17.5.1 Data

- **Source path**—the location of the data you want to process. Click **Browse** to select the path. The source path you select controls the folder tree below. The folder tree displays an icon for each file or folder within the source path. You can specify source paths in the resource pool under the Processing Source Location object. Click

Save after you select a folder or file in this field.

Browse for Files to Process



- The processing engine processes all the files located in the folder you select as your source as one job. This includes, for example, a case in which you place five different .PSTs from one custodian in a single folder.
- You can specify source paths in the resource pool under the Processing Source Location object. The Relativity Service Account must have read access to the processing source locations on the resource pool.
- Depending on the case sensitivity of your network file system, the source location that you add through the resource pool may be case sensitive and might have to match the actual source path exactly. For example, if the name of the file share folder is `\\files\\SampleShare\\Sample`, you must enter this exactly and not as `\\files\\SampleShare\\sample` or `\\files\\sampleshare\\Sample`, or any other variation of the actual name. Doing so will result in a document-level processing error stating, “The system cannot find the file specified.”
- If you process files from source locations contained in a drive that you have attached to your computer, you can detach those original source locations without issue after the processing set finishes. This is because Relativity copies the files from the source locations to the Relativity file repository. For a graphical representation of how this works, see [Copying natives during processing](#).

Note: Processing supports long file paths, but in the case of other Windows parsing issues outside of long path issues, Relativity won't be able to read that path. It is recommended that you pull documents out of subfolders that are nested in deep layers so that they are not hidden.

- **Custodian**—the owner of the processed data. When you select a custodian with a specified prefix, the default document numbering prefix field changes to reflect the custodian's prefix. Thus, the prefix from the custodian takes precedence over the prefix on the profile.
 - When you open the Add Entity window, the last accessed entity layout is selected by default in the layout drop-down list. For example, if you last created an entity with a Collections layout, that layout is selected here, even though you've accessed this window through the processing data source. To create a new custodian with processing-relevant fields, select the **Processing Entity** layout from the drop-down list.

Processing Entity ▾  Save Cancel

Custodian

Type* ☒ Person ☐ Other

First Name*

Last Name*

Classification ☒ Custodian – Processing
[Manage](#)

Document numbering prefix

Notes

- **Type**
 - **Person**—the individual acting as entity of the data you wish to process.
 - **Other**—the entity of the data you wish to process that isn't an individual but is, for example, just a company name. You can also select this if you wish to enter an individual's full name without having that name include a comma once you export the data associated with it. Selecting this changes the Entity layout to remove the required First Name and Last Name fields and instead presents a required Full Name field.
- **First Name**—the first name of the entity. This field is only available if you've set the Type above to Person.
- **Last Name**—the last name of the entity. This field is only available if you've set the Type above to Person.
- **Full Name**—the full name of the entity of the data you wish to process. This field is only available if you've set the Type above to Other. When you enter the full name of an entity, that name doesn't contain a comma when you export the data associated with it.
- **Classification**—differentiates among entity records created for Processing or Name Normalization.
 - Custodian – Processing—the indicator that this custodian was created for Processing.

Note: When new custodians are created using the Quick-Create Set(s) layout, the classification is set to Custodian – Processing.

- **Communicator**—the indicator that the record was created by Name Normalization. For more information see the Analytics Guide.
- **Document numbering prefix**—the prefix used to identify each file of a processing set once the set is published. The prefix entered on the entity appears as the default value for the required Document numbering prefix field on the processing data source that uses that entity. The identifier of the published file reads: <Prefix> #####.

- **Notes**—any additional descriptors of the entity.
 - If you add processing to an environment that already has custodian information in its database, Relativity doesn't sync the imported custodian data with the existing custodian data. Instead, it creates separate custodian entries.
 - If a single custodian has two identical copies of a document in different folders, only the primary document makes it into Relativity. Relativity stores a complete record internally of the duplicate, and, if mapped, the duplicate paths, all paths, duplicate custodian, all custodian fields in the primary record are published. Additionally, there may be other mapped fields available that can describe additional fields of the duplicates.

Note: One of the options you have for bringing custodians into Relativity is Integration Points (RIP). You can use RIP to import any number of custodians into your environment from Active Directory and then associate those custodians with the data sources that you add to your processing set.

- **Destination folder**—the folder in Relativity where the processed data is published. This default value of this field is pulled from the processing profile. If you edit this field to a different destination folder location, the processing engine reads this value and not the folder specified on the profile. You can select an existing folder or create a new one by right-clicking the base folder and selecting **Create**.
 - If the source path you selected is an individual file or a container, such as a zip, then the folder tree does not include the folder name that contains the individual file or container.
 - If the source path you selected is a folder, then the folder tree includes the name of the folder you selected.
 - After you create a destination folder and it is published, you cannot delete it.

17.5.2 Custodian Options

- **Time Zone**—determines what time zone is used to display date and time on a processed document. The default value is the time zone entered on the profile associated with this set. The default value for all new profiles is Coordinated Universal Time (UTC). If you wish to change this, click **Select** to choose from a picker list of available time zone values.
- **OCR language(s)**—determines what language is used to OCR files where text extraction isn't possible, such as for image files containing text.
 - Selecting multiple languages will increase the amount of time required to complete the OCR process, as the engine will need to go through each language selected.
 - The default value is the language entered on the profile associated with this set.
- **Document numbering prefix**—the prefix applied to the files once they are published. On published files, this appears as <Prefix>xxxxxxxx—the prefix followed by the number of digits specified. The numbering prefix from the custodian takes precedence over the prefix on the processing profile. This means that if you select a custodian with a different document numbering prefix than that found on the profile referenced by the processing set, this field changes to reflect the prefix of the custodian.
- **Start Number**—the starting number for the documents published from this data source.
 - This field is only visible if your processing set is using a profile with a Numbering Type field value of Define Start Number.
 - If the value you enter here differs from the value you entered for the Default Start Number field on the profile, then this value takes precedence over the value on the profile.

- The maximum value you can enter here is 2,147,483,647. If you enter a higher value, you'll receive an Invalid Integer warning next to field value and you won't be able to save the profile.
- If you leave this field blank or if there are conflicts, then Relativity will auto-number the documents in this data source. This means it will use the next available control number for the document numbering prefix entered. For example, if you've already published 100 documents to the workspace and you mistakenly enter 0000000099 as a start number, Relativity will automatically adjust this value to be 0000000101, as the value you entered was already included sequentially in the previously published documents.
- You can use the **Check for Conflicts** option next to this field. When you click this, you'll be notified that the start number you entered is acceptable or that it's already taken and that the documents in that data source will be auto-numbered with the next available control number. Note that this conflict check could take a long time to complete, depending on the number of documents already published to the workspace.

Note: When Level Numbering is selected, you can define the start number for each Processing Data Source.

- **Start Numbers**—allows you to define the first number to use on each level for this specific data source.

Start numbers

Level 2 (box number)	###
Level 3 (folder number)	1
Level 4 (document number)	0001

When you create a new profile or when there are no values on a field, the system will use # to indicate how many digits were configured for that level in the Processing Profile used on the Processing Set.

If a level was configured to take up to 3 digits, enter a start number with no padding, (e.g., 1) or with padding, (e.g., 0001).

17.5.3 Advanced Options

- **Name**—the name you want the data source to appear under when you include this field on a view or associate this data source with another object or if this data source encounters an error. Leaving this blank means that the data source is listed by custodian name and artifact ID. Populating this field is useful in helping you identify errors later in your processing workflow.

Note: The processing data source is saved with <Custodian Last Name>, <Custodian First Name> - <Artifact ID> populated for the Name field, if you leave this field blank when creating the data source. Previously, this field only displayed the artifact ID if it was left blank. This is useful when you need to identify errors per data source on an error dashboard, as those data sources otherwise wouldn't display a custodian name.

- **Order**—the priority of the data source when you load the processing set in the Inventory tab and submit the processing set to the queue. This also determines the order in which files in those sources are de-duplicated. This field is automatically populated. For more information, see [Order considerations](#).

Note: When you delete a document that has been published into Review, Processing will re-calculate deduplication to identify and publish the duplicate if there is one, and will not include the deleted document in subsequent deduplication logic.

17.5.4 Order considerations

- Processing jobs are carried out based on the availability of worker resources. The order number does not influence the allocation of worker resources or performance for Inventory or Discover jobs.
- Relativity does not recommend adding more than 100 data sources to a Processing Set. Doing so could cause performance issues during the submission of the set. If more than 100 data sources are required, it is recommended that you break up the job into multiple Processing Sets.
- The order number does play a part in data source priority for deduplication and assignment of control numbers. Again, the lower the order number, the higher the priority in which data sources are deduped and control numbers assigned.
- You can change the priority of data sources in the Processing and Imaging Queue. If you change the priority of a publish or republish job, you also update the priorities of all other jobs associated with the same data source.

17.5.5 Edit considerations for data sources

Note the following guidelines for modifying data sources:

- You can't add or delete a data source to or from a processing set if there's already a job in the queue for that set or if discovery of that set has already completed.
- If you add a data source to a processing set that has already been inventoried but not yet discovered, you must run inventory again on that processing set.
- If you edit a data source that is associated with a processing set that has already been inventoried but not yet discovered, you must run inventory again on that processing set.
- If you delete a data source from a processing set that has already been inventoried but not yet discovered, you must run inventory again on that processing set.
- If the processing set to which you've added a data source has already been inventoried, with or without errors, but not yet discovered, you're able to edit all fields on that data source; however, you must run inventory again on that processing set after you edit the source.
- If the processing set to which you've added a data source has already been discovered, with or without errors, you can only edit the Name and Document numbering prefix fields on that data source.
- If the processing set to which you've added a data source has already been published, with or without errors, you can only edit the Name field on that data source.

Note: When you make a change that merits a re-inventory job, Relativity applies a "Force reinventory" flag to the processing set's table in the workspace database.

17.5.6 Processing data source view

At the bottom of the processing set layout is the Processing Data Source view, which will display information related to the data sources you add.

Processing Set 004

Basic Settings

Inventory (Optional)

Not Started

1

Discover

Not Started

2

Publish

Not Started

3

Inventory count

0

Discover count

0

Publish count

0

Total exceptions

0

> Inventory (Optional)

> Discovery

> Publish

Processing Data Source

Job Errors

New

Delete

1

- 2 of 2

10

per page

	Status	Percent complete	Source path	Custodian	Document numeri...	Time zone	OCR language(s)
☐	New		\\files. [redacted]	Jones, Tom	Jones_	(UTC-06:00) Central Time (US & Canada)	English
☐	New		\\files. [redacted]	Johnson, Howard	Johnson_	(UTC-06:00) Central Time (US & Canada)	English

This view provides the following fields:

- **Status**—the current state of the data source as inventory, discovery, publish, or republish runs on the processing set. This and the Percent Complete value refresh automatically every five seconds. The status values are:
 - **New**—the data source is new and no action has been taken on the processing console.
 - **Waiting**—you've clicked Inventory, Discover, or Publish Files on the console and an agent is waiting to pick up the job.
 - **Initializing**—an agent has picked up the job and is preparing to work on it.
 - **Document ID Generation**—document ID numbers are being generated for every document. You'll see this status if the profile attached to the set has a deduplication method of None.
 - **DeDuplication and Document ID Generation**—the primary and duplicate documents are being identified, and the document ID number is being generated for every document. You'll see this status if the profile attached to the set has deduplication set to Global or Custodial. If you have multiple data sources attached to a single processing set, the second source is started as soon as the first set reaches the DeDuplication and Document ID generation stage. Previously, Relativity waited until the entire source was published before starting the next one.
 - **Deduped Metadata Overlay**—deduped metadata is being overlaid onto the primary documents in Relativity. This status was added in July 2017 as part of the distributed publish enhancement.
 - **Inventorying/Discovering/Publishing**—an agent is working on the job. Refer to the Percent Complete value to see how close the job is to being done.

- **Inventory/Discovery/Publish files complete**—the job is complete, and the Percent Complete value is at 100%.
- **Unavailable**—the data source is not accessible and no action can be taken on the processing console.
- **Percent Complete**—the percentage of documents in the data source that have been inventoried, discovered, or published. This and the Status value refresh automatically every five seconds.
- **Source path**—the path you selected for the source path field on the data source layout.
- **Custodian**—the custodian you selected for the data source.
- **Document numbering prefix**—the value you entered to correspond with the custodian on the data source layout. If you didn't specify a prefix for the data source, then this is the default prefix that appears on the processing profile.
- **Time zone**—the time zone you selected for the data source.
- **OCR language(s)**—the OCR language(s) you selected on the data source.

17.5.7 Job Errors view

At the bottom of the processing set layout is the Job Errors view, which displays information related to all job-level errors that occurred on all data sources associated with the set.

Processing Sets

Processing Data Sources

Processing Profile

Password Bank | Processing

Inventory

Reports

Job Errors

Files

>

🔍

Current Job Errors

- **Error Identifier**—the unique identifier of the error as it occurs in the database. When you click this message, you are taken to the error details layout, where you can view the stack trace and other information. Note that for Unresolvable errors, the console is disabled because you cannot take any actions on that error from inside Relativity. For more information, see [Processing error resolution](#).
- **Error Status**—the status of the error. This is most likely Unresolvable.
- **Message**—the cause and nature of the error. For example, "Error occurred while trying to overlay de-duplication details. Please resolve publish error or republish documents from data source below. DataSource Artifact Id: 1695700".
- **Custodian**—the custodian associated with the data source containing the file on which the error occurred.
- **Processing Set**—the name of the processing set in which the error occurred.
- **Data Source**—the data source containing the file on which the error occurred.
- **Error Created On**—the date and time at which the error occurred during the processing job.

- **Republish Required**—the error must be retried in order to be successfully published.
- **Notes**—any manually added notes associated with the error.

For more information on handling document errors, see [Processing error resolution](#).

17.6 Processing Data Sources tab

To see all data sources associated with all processing sets in the workspace, navigate to the Processing Data Sources sub-tab.

Processing Sets Processing Data Sources Processing Profile Password Bank Processing Inventory Reports Job Errors Files						
> 🔍 All Processing Data Sources						
#	Processing Data Source	ProcessingSet	Custodian	Preprocessed fil...	Preprocessed fil...	Nisted file count
	Filter	Filter	Filter	= ▾ Filter	= ▾ Filter	= ▾ Filter
1	Johnson, Howard - 1101979	Processing Set 004	Johnson, Howard			
2	Jones, Tom - 1101977	Processing Set 004	Jones, Tom			
3	Allen, Phillip - 1095903	Mega processing s...	Allen, Phillip	931,090,201.00	1,530	837

The default view on the Processing Data Sources tab includes the following fields:

- **Processing Data Source**—the name of the data source. If you originally left this blank, then this value will consist of the name of the custodian and artifact ID.
- **ProcessingSet**—the name of the processing set the data source is attached to.
- **Custodian**—the custodian attached to the data source.
- **Preprocessed file size**—the total size, in bytes, of all the files in the data source before you started the processing set.
- **Preprocessed file count**—the number of files in the data source before you started the processing set.
- **Nisted file count**—the number of files from the data source that were then removed, per the de-NIST setting.
- **Excluded file size**—the total size, in bytes, of all the documents from the data source that were excluded from discovery.
- **Excluded file count**—the number of files from the data source that were excluded from discovery.
- **Filtered file count**—the number of files from the data source that were filtered out before discovery.
- **Discover time submitted**—the date and time at which the files in the data source were last submitted for discovery.
- **Discovered document size**—the total size, in bytes, of all the documents from the data source that were successfully discovered.
- **Discovered document count**—the number of files from the data source that were successfully discovered.
- **Last publish time submitted**—the date and time at which the files in the data source were last submitted for publish.

- **Deduplication method**—the deduplication method set on the processing profile associated with the processing set.
- **Duplicate file count**—the number of files that were deduplicated based on the method set on the processing profile.
- **Published documents**—the number of documents from the data source that were successfully published.
- **Published document size**—the total size, in bytes, of all the documents from the data source that were successfully published.
- **Status**—the current status of the data source.

Additional fields to customize your Processing Data Sources view

- **Artifact ID**—the artifact ID of the workspace.
- **Auto-publish set**—arranges for the processing engine to automatically kick off publish after the completion of discovery, with or without errors. By default, this is set to No.
- **Container count**—the count of all native files classified as containers before extraction/decompression, as they exist in storage. This also includes nested containers that haven't been extracted yet.
- **Container size**—the sum of all native file sizes, in bytes, classified as containers before extraction/decompression, as they exist in storage. this value may be larger than the preprocessed file size because it also includes nested containers.
- **Custodian::Department**—the department of a Custodian-type entity
- **Custodian::Email**—the email address of a Custodian-type entity
- **Delimiter**—the delimiter you want to appear between the different fragments of the control number of your published child documents.
- **DeNIST Mode** -determines which files to include in the DeNISTing process. Options include:
 - **DeNIST all files**—parent/child groups are broken and any file on the NIST list is removed.
 - **Do not break parent/child groups**—parent/child groups are left in tact, regardless if the files are on the NIST list. Loose NIST files are removed.
- **Destination folder**—the folder in Relativity into which documents are placed once they're published to the workspace.
- **Discover time complete** -the date and time at which the files in the data source were discovered.
- **Discovered files**—the count of all the native files discovered that aren't classified as containers as they exist in storage.
- **Discovery group ID**—the unique identifier of the discovery group.
- **Document numbering prefix**—the prefix applied to each file in a processing set once it is published to a workspace. The default value for this field is REL.
- **Duplicate file size**—the sum of duplicate native file sizes, in bytes, associated to the user, processing set and workspace.
- **Excel Header/Footer Extraction**—header and footer information extracted from Excel files when you publish them.
- **Excel Text Extraction Method**—determines whether the processing engine uses Excel or dtSearch to extract text from Excel files during publish.

- **Extract children**—arranges for the removal of child items during discovery, including attachments, embedded objects and images and other non-parent files.
- **Filtered file size**—the total size, in bytes, of the files from the data source that were filtered out before discovery.
- **Inventoried files**—the number of files from the data source that were inventoried.
- **Is Start Number Visible**—true/false value for the starting number field toggle.
- **Last activity**—the date and time at which a job last communicated to the worker.
- **Last document error ID**—the unique identifier of an error attached to a document.
- **Last inventory group ID**—the unique identifier of a group of inventoried files.
- **Last inventory time submitted**—the date and time at which the files in the data source were last submitted for inventory.
- **Last run error**—the last job error that occurred in the running of the OCR set.
- **Nisted file size**—the total size, in bytes, of the files from the data sources that were then removed, per the de-NIST setting.
- **Number of Digits**—determines how many digits the document's control number contains. The range of available values is 1 and 10. By default, this field is set to 10 characters.
- **OCR**—enabled or disabled to run OCR during processing.
- **OCR Accuracy**—the desired accuracy of your OCR results and the speed with which you want the job completed.
- **OCR language(s)**—the language used to OCR files where text extraction isn't possible, such as for image files containing text.
- **OCR Text Separator**—a separator between extracted text at the top of a page and text derived from OCR at the bottom of the page in the Extracted Text view.
- **Order**—the priority of the data source when you load the processing set in the Inventory tab and submit the processing set to the queue. This also determines the order in which files in those sources are de-duplicated.
- **Parent/Child Numbering**—determines how parent and child documents are numbered relative to each other when published to the workspace.
- **Percent Complete**—the percentage of documents from the data source that have been discovered or published.
- **PowerPoint Text Extraction Method**—determines whether the processing engine uses PowerPoint or dtSearch to extract text from PowerPoint files during publish.
- **Preexpansion file count**—the number of files in the data source for all Level 1 non-container files at the first level after initial expansion.
- **Preexpansion file size**—the total size, in bytes, of all the files in the data source for all Level 1 non-container files at the first level after initial expansion.
- **ProcessingSet::Republish required**—errors attached to a processing set that need republishing.
- **Propagate deduplication data**—applies the deduped custodians, deduped paths, all custodians, and all paths field data to children documents, which allows you to meet production specifications and perform searches on those fields without having to include family or overlay those fields manually.
- **Publish group ID**—the unique identifier of a published group of documents.

- **Publish time complete**—time it took for a file to finish publishing.
- **Retry jobs remaining**—number of errors attached to a file that needs trying.
- **Rolled up file count**—the number of files rolled up during discovery. This setting references rolled up image text where child images have had their text rolled up into the parent document.
- **Security**—level of accessibility of files to users.
- **Source folder structure retained**—the folder structure of the source of the files you process when you bring these files into Relativity is maintained.
- **Source path**—the location of the data you want to process.
- **Start Number**—the starting number for documents that are published from the processing set(s) that use this profile.
- **Storage file size**—the sum of all file sizes, in bytes, as they exist in storage.
- **System Created By**—identifies the user who created the document.
- **System Created On**—the date and time when the document was created.
- **System Last Modified By**—identifies the user who last modified the document.
- **System Last Modified On**—the date and time at which the document was last modified.
- **Time zone**—determines what time zone is used to display date and time on a processed document.
- **Total file count**—the count of all native files (including duplicates and containers) as they exist after decompression and extraction.
- **Total file size**—the sum of all native file sizes (including duplicates and containers), in bytes, as they exist after decompression and extraction.
- **When extracting children, do not extract**—excludes MS Office embedded images, MS Office embedded objects, and/or Email inline images when extracting children.
- **Word Text Extraction Method**—determines whether the processing engine uses Word or dtSearch to extract text from Word files during publish.

17.7 Deleting a processing set

If your Relativity environment contains any outdated processing sets that haven't yet been published and are taking up valuable space, or sets that simply contain mistakes, you can delete them, depending on what phase they're currently in.

The following table breaks down when you're able to delete a processing set.

Point in processing	Can delete?
Pre-processing—before Inventory and Discovery have been started	Yes
While Inventory is in progress	No
After Inventory has been canceled	Yes
After Inventory has completed	Yes
While Discovery is in progress	No

Point in processing	Can delete?
After Discovery has been canceled	Yes
After Discovery has completed	Yes
While Publish is in progress	No
After Publish has been canceled	No
After Publish has completed	No

If you need to delete a processing set that is currently being inventoried or discovered, you must first cancel inventory or discovery and then delete the set.

Note: Deletion jobs will always take the lowest priority in the queue. If another job becomes active while the delete job is running, the delete job will be put into a paused state and will resume once all other jobs are complete.

17.7.1 Permissions

Note: This feature supports feature permission actions. Each action grants access to a specific task within the feature. You can grant groups access to individual actions, allowing them to perform only the tasks they need. To grant access to the entire feature, select all of its actions. Manage feature actions on the Features tab in a group's workspace permissions. For instructions, see [Setting workspace permissions](#). For more information about feature permissions, see [Feature permissions](#).

To set feature-level permissions:

1. Use the search bar to navigate to **Workspace Details**.
2. Click **Manage Workspace Permissions** (located in the right side column).

Note: Feature-level permissions are assigned to groups. Groups are managed through instance-level permissions. If you do not know which group to use for a feature-level permission, contact your Relativity administrator.

Tab	Name/Feature
Tab Visibility	- Processing - Processing Sets
Object Security	- Field - Edit - Add Field Choice By Link (on) - OCR Language (optional) - Delete - Processing Data Source - Delete - Processing Error - Delete - Processing Set - Delete

Tab	Name/Feature
Other Settings	- Admin Operations - Delete Object Dependencies. This is required to delete the processing set's child objects and linked associated objects.

To delete a processing set, perform the following steps:

1. In the processing set list, select the checkbox next to the set(s) you want to delete. If you're on the processing set's layout, click **Delete** at the top of the layout.

Note: If you use the Delete mass operation to delete a processing set, but then you cancel that deletion while it is in progress, Relativity puts the set into a canceled state to prevent you from accidentally continuing to use a partially deleted set. You can't process a set for which you canceled deletion or in which a deletion error occurred.

2. (Optional) Click **View Dependencies** on the confirmation window to view all of the processing set's child objects that will also be deleted and the associated objects that will unlink from the set when you proceed with the deletion.

Delete 1 Processing Set

Warning: Clicking delete will permanently delete this object and any children and unlink any associated objects. Click View Dependencies below for details.

Close Dependencies

Object Type	Action	Count	Connection
(All)	(All)	=	Filter
Processing Set	Delete	1	Parent
Processing Data Source	Delete	2	Child of: Processing Set
OCR Language	Unlink	1	Field: OCR Language

Delete
Cancel

3. Click **Delete** on the confirmation window. When you proceed, you permanently delete the processing set object, its children, and its processing errors, and you unlink all associated objects.

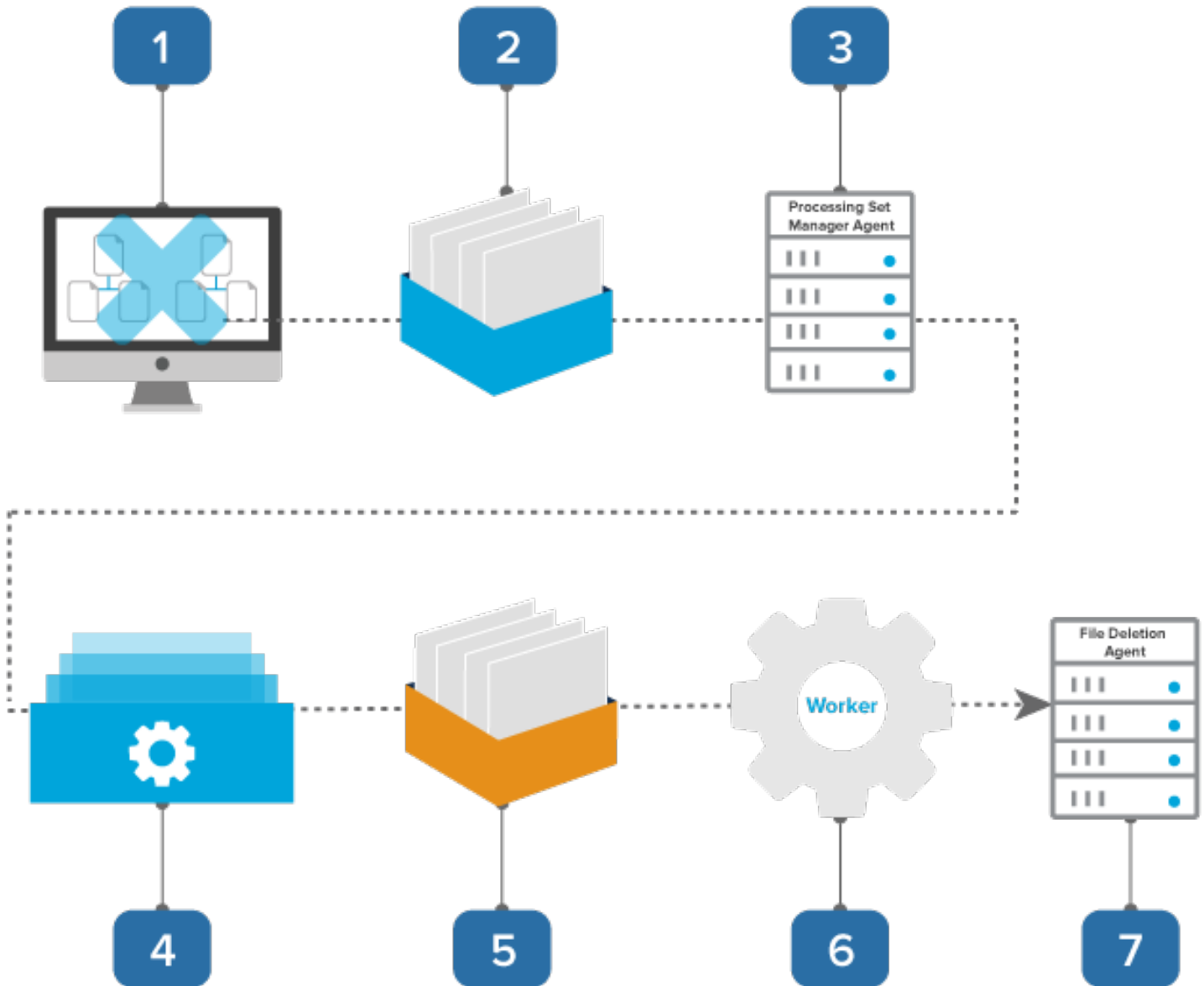
The following table breaks down what kinds of data is deleted from Relativity and Invariant when you delete a processing set in certain phases.

Phase deleted	From Relativity	From Invariant
Pre-processed (Inventory and Discovery not yet started)	Processing set object - data sources	N/A
Inventoried processing set	Processing set object - errors, data sources, inventory filters	Inventory filter data; inventoried metadata
Discovered processing set	Processing set object - errors, data sources	Discovered metadata

When you delete a processing set, the file deletion manager deletes all physical files and all empty sub-directories. Files that the system previously flagged for future deletion are also deleted.

The following graphic and accompanying steps depict what happens on the back end when you delete a processing set:

Processing Set Deletion Process



1. You click Delete on the processing set.
2. A pre-delete event handler inserts the delete job into the queue while Relativity deletes all objects associated with the processing set.
3. A processing set agent picks up the job from the queue and verifies that the set is deleted.
4. The processing set agent sends the delete job to Invariant.
5. The delete job goes into the Invariant queue, where it waits to be picked up by a worker.

6. A worker deletes the SQL data associated with the processing set and queues up any corresponding files to be deleted by the File Deletion agent.
7. The File Deletion starts up during off hours, accesses the queued files and deletes them from disk.

Note: If an error occurs during deletion, you can retry the error in the Discovered Files tab. see Retry Delete for more information.

17.8 Avoiding data loss across sets

Due to the way that processing was designed to deal with overwrites during error retry, there is the chance that you can inadvertently erase data while attempting to load documents into Relativity across different modes of import.

To avoid an inadvertent loss of data, do NOT perform the following workflow:

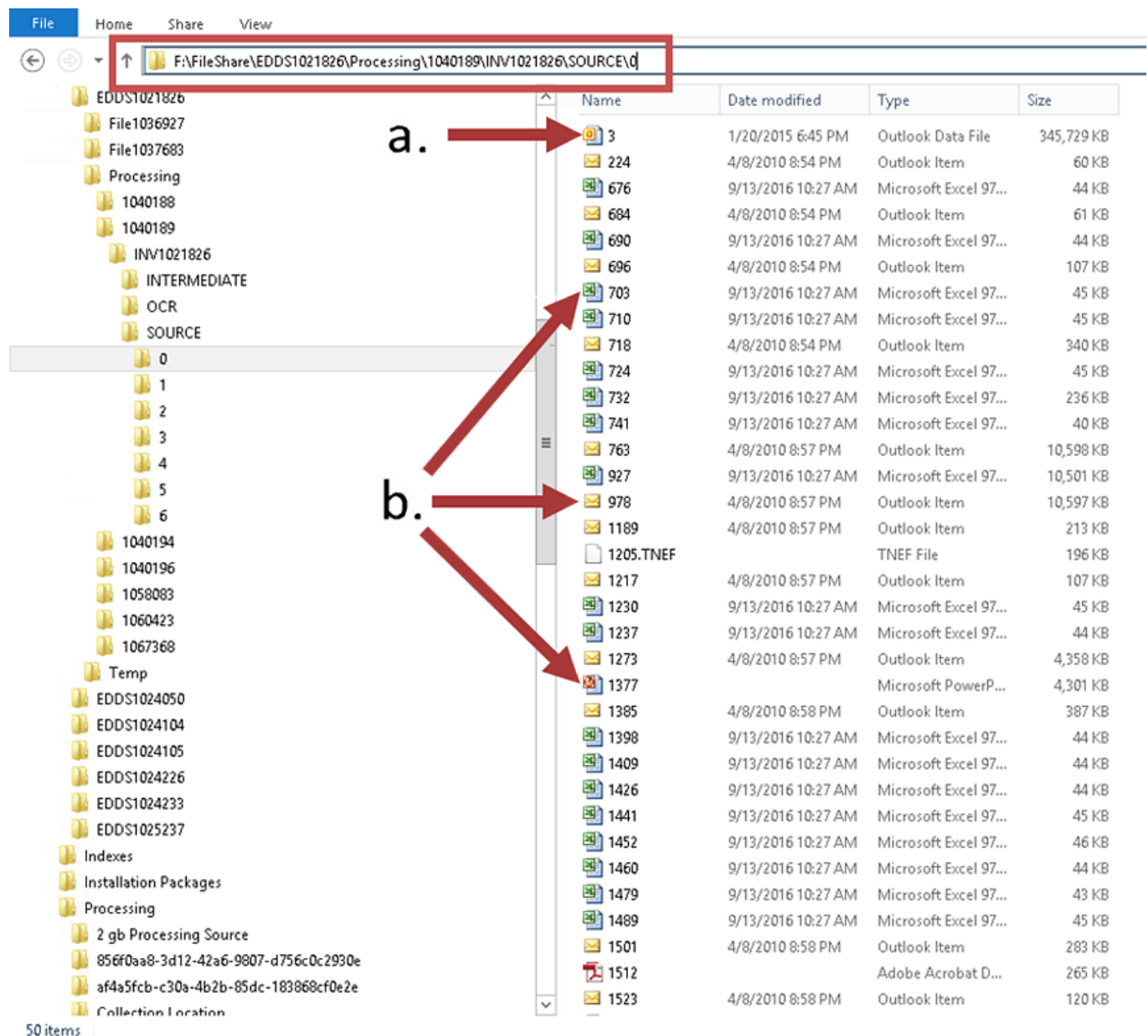
1. Run a processing set.
2. After the processing set is complete, import a small amount of data using Import/Export so that you can keep one steady stream of control numbers and pick up manually where the previous processing set left off.
3. After importing data through Import/Export, run another processing set, during which Relativity tries to start the numbering where the original processing job left off. During this processing set, some of the documents cause errors because some of the control numbers already exist and Relativity knows not to overwrite documents while running a processing set.
4. Go to the processing errors tab and retry the errors. In this case, Relativity overwrites the documents, as this is the expected behavior during error retry. During this overwrite, you lose some data.

17.9 Copying natives during processing

To gain a better understanding of the storage implications of copying natives during processing, note the behavior in the following example.

When you process a PST file containing 20,000 unique total documents while copying natives:

1. You copy the PST from the original source to your Processing Source Location, as this is the identified location where Relativity can see the PST. Note that you can make the original source a processing source by opening the original source to Relativity.



Note: If you run Inventory on this set, Relativity will identify all parents and attachments, but it will only extract metadata on the parent email.

1. The **EDDS12345\Processing\ProcessingSetArtifactID\INV12345\Source\0** folder displays as the original PST.
2. Relativity begins to harvest individual MSG files in batches and processes them. If an MSG has attachments, Relativity harvests files during discovery and places them in the queue to be discovered individually. Throughout this process, the family relationship is maintained.
2. Relativity discovers the files, during which the metadata and text are stored in Relativity Processing SQL.
3. Relativity publishes the metadata from the Relativity Processing SQL Datastore to the Review SQL Datastore and imports text into the text field stored in SQL or Relativity Data Grid. This metadata includes links to the files that were harvested and used for discovery. No additional copy is made for review.

4. Once processing is complete:

- You can delete the processing source PST.
- You can delete the PST file in the EDDS folder, assuming there are no errors.

Note: You can't automate the deletion of files no longer needed upon completion of processing. You need to delete this manually.

- You should retain files harvested during processing, as they are required for review.

The following graphic depicts what happens behind the scenes when the system copies native files to the repository during processing. Specifically, this shows you how the system handles the data source and EDDS repository across all phases of processing when that data source isn't considered permanent.

This graphic is designed for reference purposes only.

Source

EDDS Repository
\\filerpo\EDDS\GUID\ABC.msg

Inventory

Discovery

Publish



Inventory

Discovery

Publish



Inventory

Discovery

Publish



Inventory

Discovery

Publish



Inventory

Discovery

Publish

FILE TABLE	
Identifier	varchar(255)
Location	\\filerpo\EDDS\GUID\ABC.msg
InRepository	0
Size	bigint
Details	xml

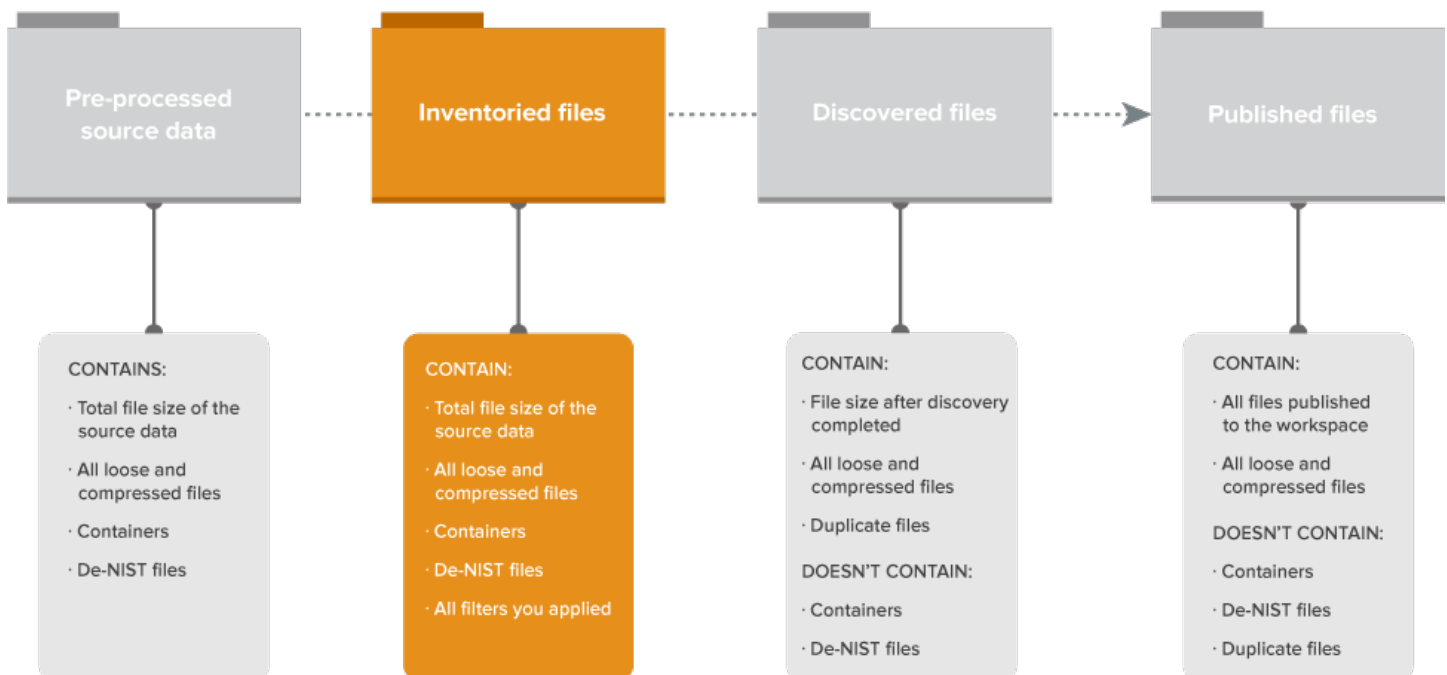
18 Inventory overview

Use Inventory to narrow down your files before discovering them by eliminating irrelevant raw data from the discovery process through a variety of preliminary filters. With inventory you can exclude certain file types, file locations, file sizes, NIST files, date ranges, and sender domains. Doing this gives you a less-cluttered data set when you begin to discover your files.

This topic provides an overview of the inventory process. For information on running an inventory job, applying filters, and managing inventory exceptions, see [Inventory processing](#).

The following graphic depicts how inventory fits into the basic workflow you would use to reduce the file size of a data set through processing. This workflow assumes that you're applying some method of deNIST and deduplication.

File Size Reduction through Processing



Inventory reads all levels of the data source, including any container files, to the lowest level. Inventory then only extracts data from first-level documents. For example, you have a .ZIP within a .ZIP that contains an email with an attached Word document, inventory only extracts data up to the email. Deeper level files are only extracted after you start Discovery. This includes the contents of a .ZIP file attached to an email and the complete set of document metadata.

You are not required to inventory files before you start file discovery. Note, however, that once you start file discovery, you can't run inventory on that processing set, nor can you modify the settings of an inventory job that has already run on that set.

The following is a typical workflow that incorporates inventory:

1. Create a processing set or select an existing set.
2. Add data sources to the processing set.
3. Inventory the files in that processing set to extract top-level metadata.
4. Apply filters to the inventoried data.

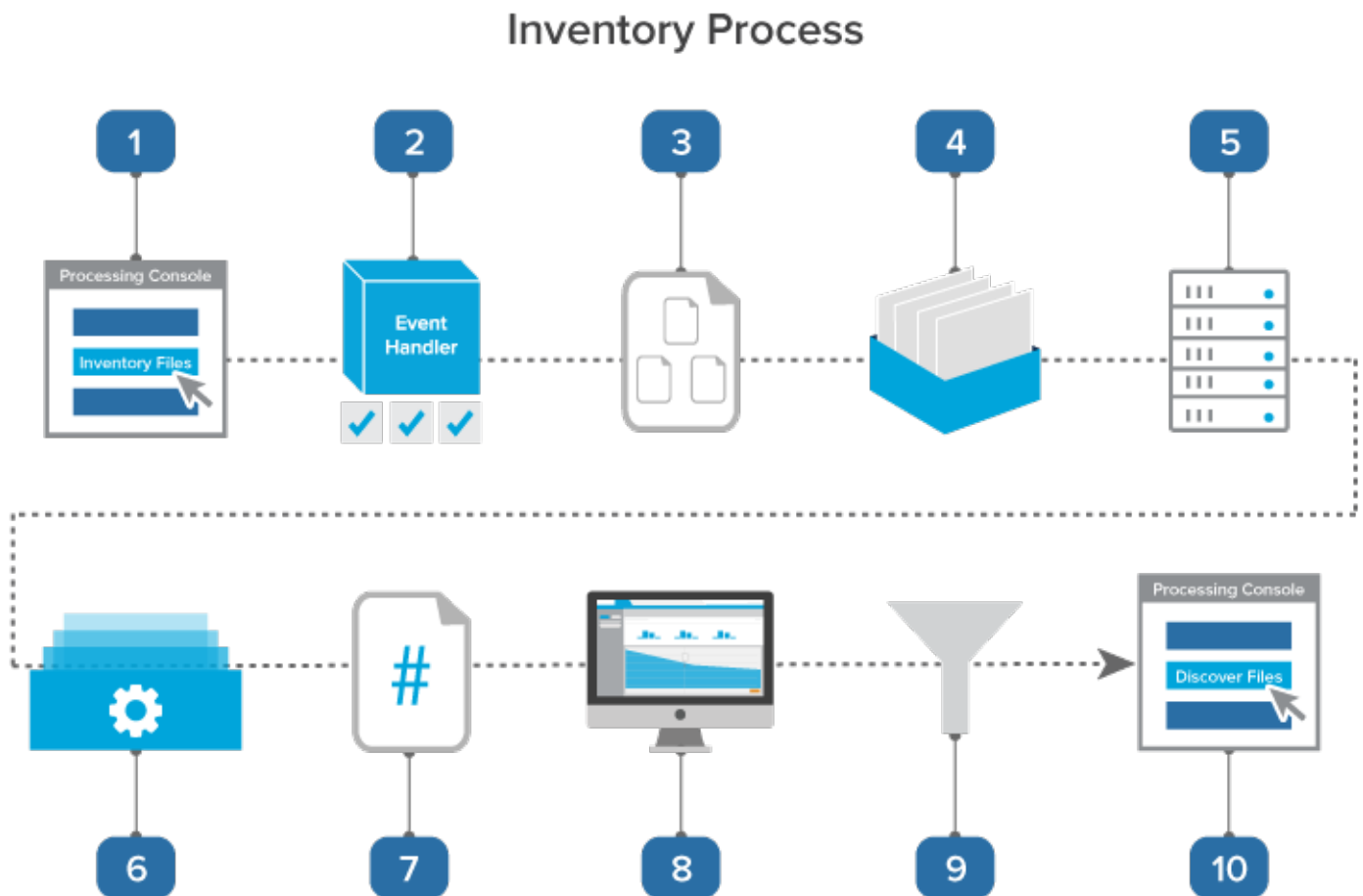
5. Run discovery on the refined data.
6. Publish the discovered files to the workspace.

To do this, you inventory your data sources, click Filter Files on the processing set console, load the inventoried set in the filtering files, and apply a Location filter to exclude the location of the 2012 Backup.PST container.

You can then move on to discover the remaining files in the set.

18.1 Inventory process

The following graphic and corresponding steps depict what happens behind the scenes when you start inventory. This information is meant for reference purposes only.



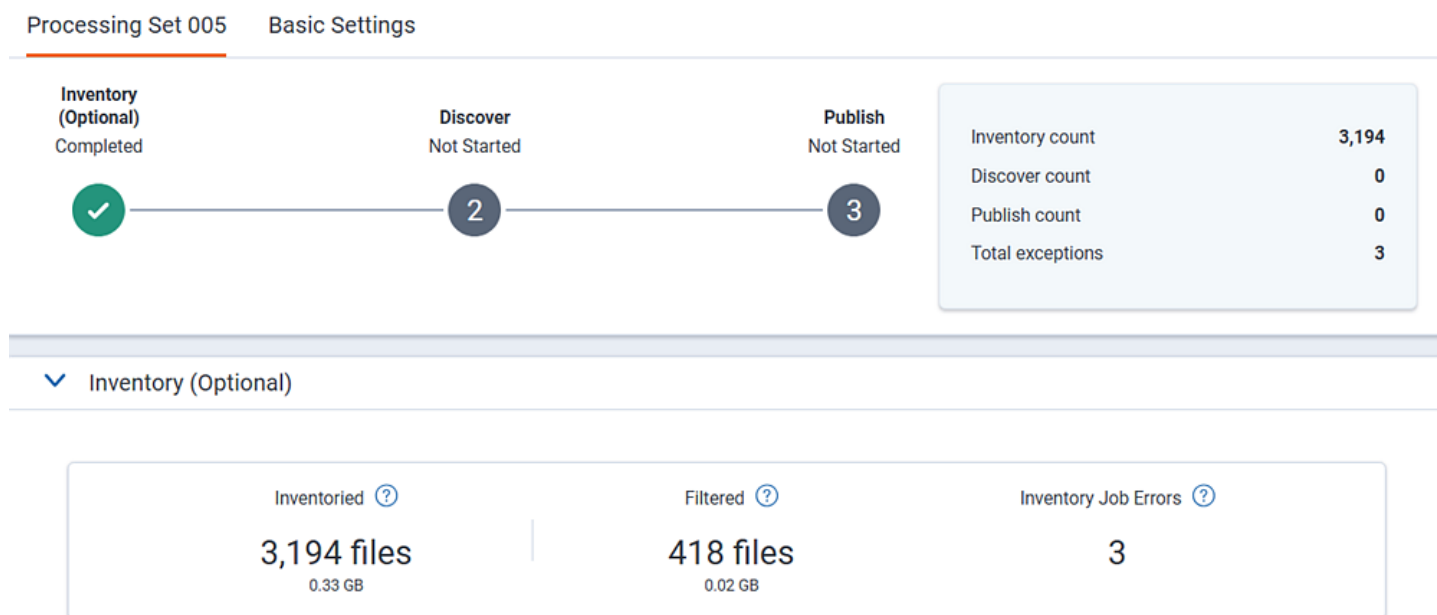
1. You click Inventory Files on the processing console.
2. A console event handler checks to make sure the processing set is valid and ready to proceed.
3. The event handler inserts the data sources to the processing queue.
4. The data sources wait in the queue to be picked up by an agent, during which time you can change their priority.
5. The processing set manager agent picks up each data source based on its order, all password bank entries in the workspace are synced, and the agent inserts each data source as an individual job into the processing engine. The agent provides updates on the status of each job to Relativity, which then displays this information on the processing set layout.

- The processing engine inventories each data source by identifying top-level files and their metadata and merges the results of all inventory jobs. Relativity updates the [reports](#) to include all applicable inventory data. You can generate these reports to see how much inventory has narrowed down your data set.
- The processing engine sends a count of all inventoried files to Relativity.
- You load the processing set containing the inventoried files in the Inventory tab, which includes a list of available filters that you can apply to the files.
- You [apply the desired filters](#) to your inventoried files to further narrow down the data set.
- Once you've applied all desired filters, you move on to [discovery](#).

18.2 Monitoring inventory status

You can monitor the job's progress on the Processing Set Details page. This page breaks down each of the processing phases (Inventory, Discover, Publish) into their own sections that include dashboard summaries, sub-job details, and counts.

18.2.1 Inventory job details



Dashboard numbers

- Inventoried**—the number of non-container, top level files identified.
- Filtered**—the number of file filtered from Inventory filters.
- Inventory Job Errors**—the number of errors that occurred during Inventory. You cannot retry these errors. You should re-run Inventory to resolve the issues.

Progress bars

Each processing phase—Inventory, Discover, and Publish—consists of distinct sub-jobs that execute sequentially until the phase is complete. Progress for each job is visually represented by progress bars that appear under the dashboard. Once all sub-jobs within a phase are finished, the progress bars are no longer displayed.

File expansion [?](#)



100%

Updating Processing Set fields [?](#)



100%

- **File expansion**—Relativity expands all source files to the first non-container level and extracts select data for pre-discovery filtering.
- **Updating Processing Set fields**—Relativity updates associated fields in the processing set with the job results.

See [Exception and error resolution](#) for details.

If you skip inventory, the status section displays a Skipped status throughout the life of the processing set.

Inventory
(Optional)

Skipped



Once inventory is complete, the status section displays a Complete status, indicating that you can move on to either filtering or discovering your files. For more information, see [Filtering files and Discovering files](#).

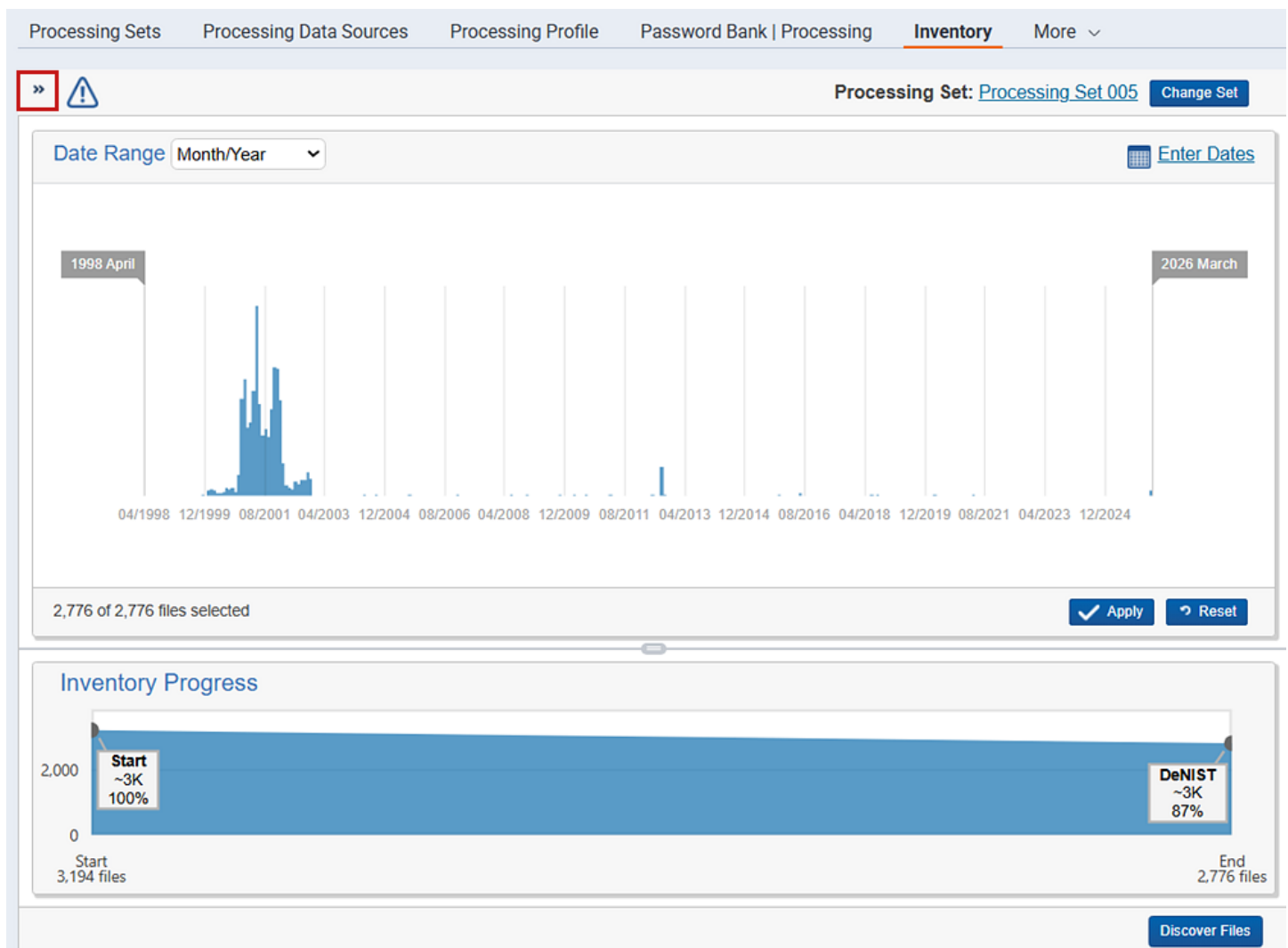
Inventory
(Optional)

Completed



18.3 Inventory filters

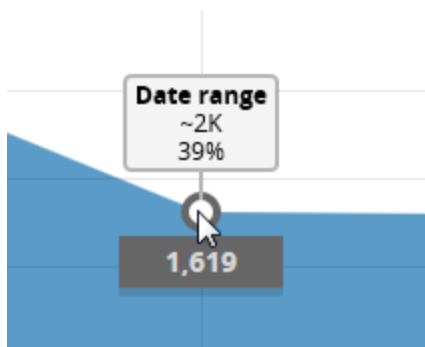
You can view the filters applied to the processing set in real-time as the job runs. Access this screen through the **Inventory** tab in the navigation bar, or by clicking the **Filter Files** button on the processing details screen. Hide or show the filters panel by clicking the arrow on the upper left side of the graph. See [Filtering Files](#) on the Inventory Processing page for details on how to apply and use filters.



This graph provides the following information to help you gauge the progress of your filtering:

- **Start # files**—lists the number of files in the data set before you applied any filters to it. This value sits in the bottom left corner of the pane.
- **End # files**—lists the current number of files in the data set now that you have excluded documents by applying filters. This value sits in the bottom right corner of the pane.
- **~#K**—lists the approximate number of files that remain in the data set under the filter type applied.
- **#%**—lists the percentage of files that remain from the data set under the filter type applied. If a filter excludes only a small number of files from the previous file count, this value may not change from the value of the preceding filter type.

You can view the exact number of files that remain in the data set by hovering over the gray dot above or below the file type box.



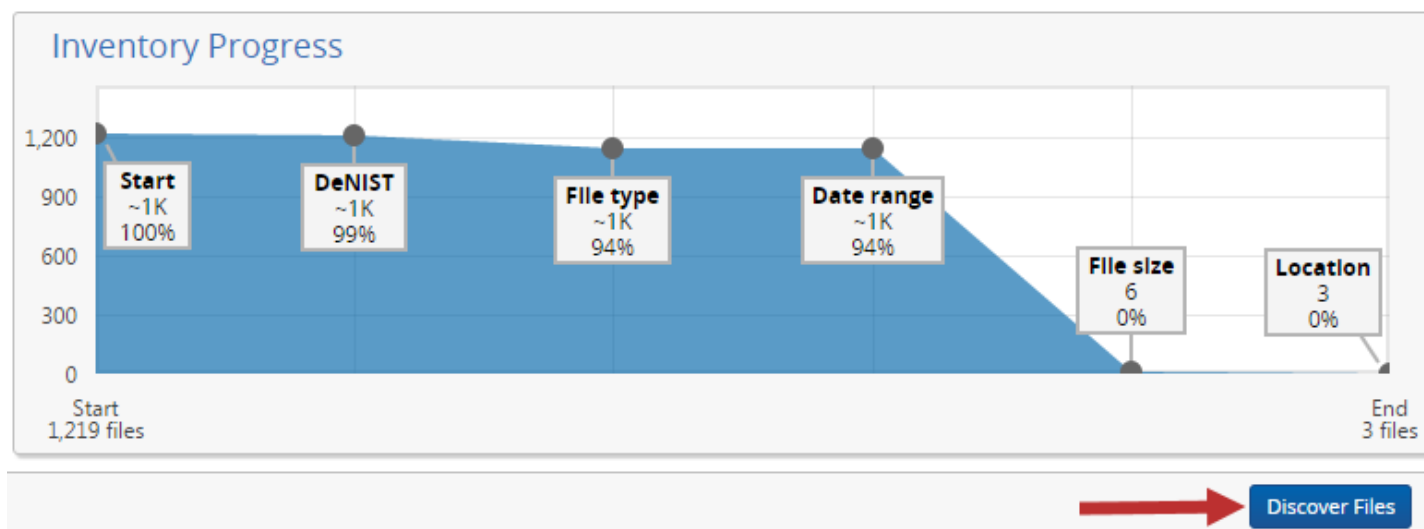
At any time before you discover the files reflected in the Inventory Progress pane, you can reset or delete any filters you already applied.

Once you determine that the filters you have applied have reduced the data set appropriately, you can discover the remaining files.

18.4 Discovering files from Inventory

You can discover files from the Inventory tab using the Discover Files button in the bottom right corner of the layout.

For more information on discovery, see [Discovery process](#).



Clicking Discover Files puts the discovery job in the queue and directs you back to the processing set layout, where you can monitor the job's progress.

The same validations that apply when you start discovery from the processing set layout apply when discovering from the Inventory tab.

19 Inventory processing

This topic provides details on inventory processing, applying filters, and managing inventory errors. For an overview of the inventory process, see [Inventory overview](#).

19.1 Running inventory

To inventory the files found in a processing set's data source(s), click **Inventory Files** on the processing set console. This option is only available if you have added at least one data source to the processing set.

Processing Sets Processing Data Sources Processing Profile Password Bank | Processing Inventory Reports Job Errors Files

Edit Delete Back Edit Permissions View Audit Feedback 1 of 24

Processing Set 005 Basic Settings

Inventory (Optional) Not Started Discover Not Started Publish Not Started

1 2 3

Inventory count 0
Discover count 0
Publish count 0
Total exceptions 0

Processing Data Source Job Errors

New Delete 1 - 4 of 4 10 per page

Status	Percent complete	Source path	Custodian	Document numbering prefix	Time zone
Inventoried with errors	100.00				(UTC-06:00) Central Time (US & Canada)
Inventory files complete	100.00				(UTC-06:00) Central Time (US & Canada)
Inventory files complete	100.00				(UTC-06:00) Central Time (US & Canada)
Inventory files complete	100.00				(UTC-06:00) Central Time (US & Canada)

Process Files

Inventory Files Filter Files Discover Files Retry File Exceptions Publish Files

Reports

Inventory Report Discovery Report All Reports Exceptions Job Errors File Exceptions

The Inventory Files button on the console is disabled in the following situations:

- There are no data sources associated with the processing set.
- The processing set is canceled.
- The processing set has already been discovered or published.
- A discovery, retry discovery, publish, republish, or retry publish job is in the queue for the processing set.

When you start inventory, the Inventory Files button changes to Cancel. You can use this to cancel the processing set. For more information, see [Canceling inventory](#).

Note: The processing set manager agent sends the password bank to the processing engine when you start inventory. If a custodian is associated with a Lotus Notes password bank, that custodian's information is sent with the inventory job.

You can re-inventory files any time after the previous inventory job is complete. For more information, see [Re-inventory](#).

19.1.1 Canceling inventory

If the need arises, you can cancel inventory before it is complete. To cancel the job, click **Cancel** in the Process Files console.

Process Files

Cancel

Filter Files

Discover Files

Retry File Exceptions

Publish Files

Reports

Inventory Report

Discovery Report

All Reports

Exceptions

Job Errors

File Exceptions

Consider the following regarding canceling inventory:

- If you click Cancel while the status is still Waiting, you can re-submit the inventory job.
- If you click Cancel after the job has already been sent to the processing engine, the entire processing set is canceled, meaning all options are disabled and it is unusable. Deduplication is not run against documents in canceled processing sets.
- Once the agent picks up the cancel inventory job, no more errors are created for the processing set.
- Errors that result from a job that is canceled are given a canceled status and cannot be retried.

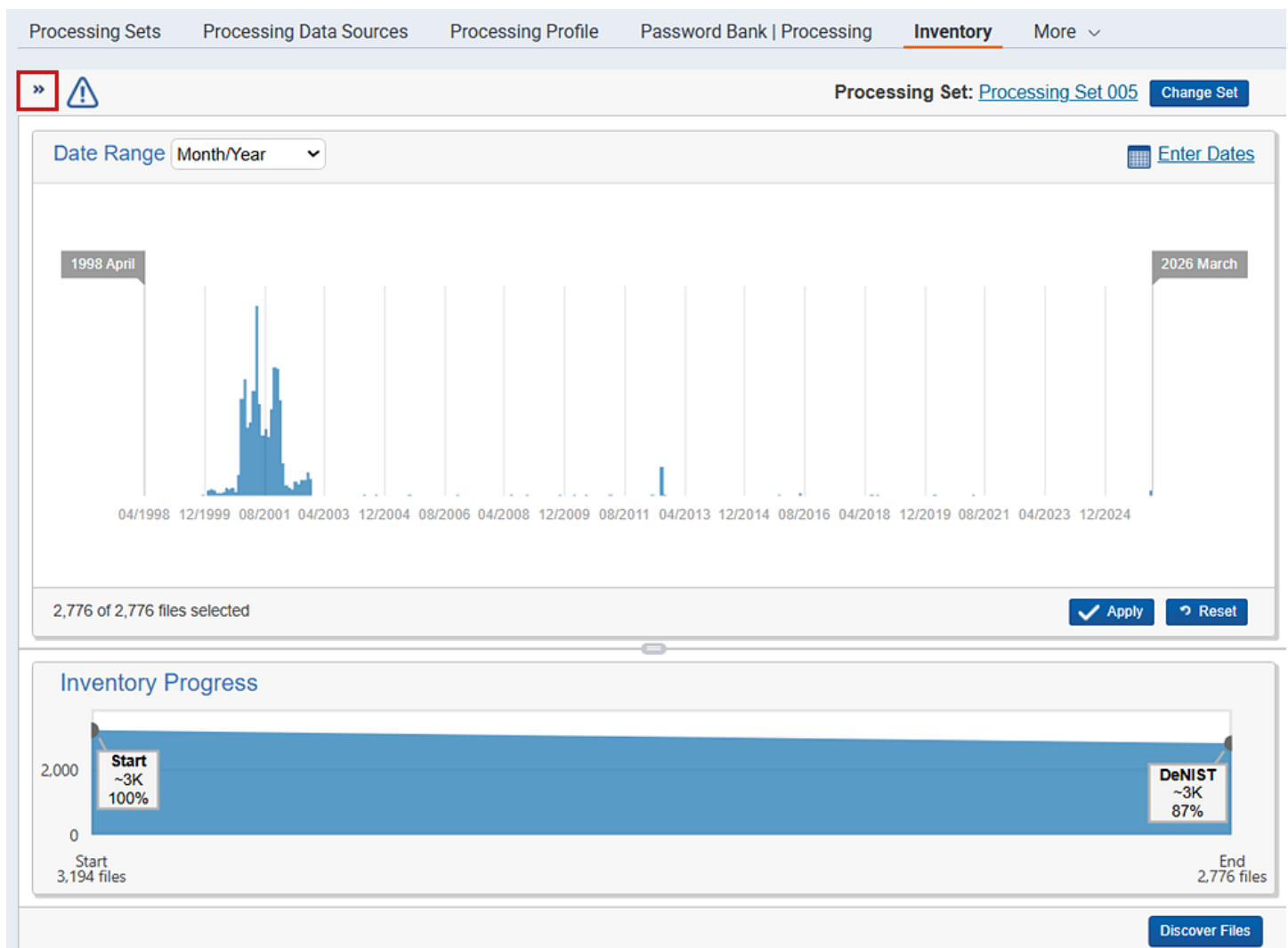
19.2 Filtering files

When inventory is complete you have the option of filtering your files in the Inventory tab before moving to discovery.

Note that Relativity only filters on the files that you have inventoried. Everything else that cascades down from the files that were discovered is not subject to the inventory filters that you set.

Note: Filtering is not supported for Relativity Short Message Format (RSMF) metadata during Inventory.

To do this, click **Filter Files** in the Process Files console. You are redirected to the Inventory tab. Use the arrow icon to the left of the charts to hide or show the Filters pane.



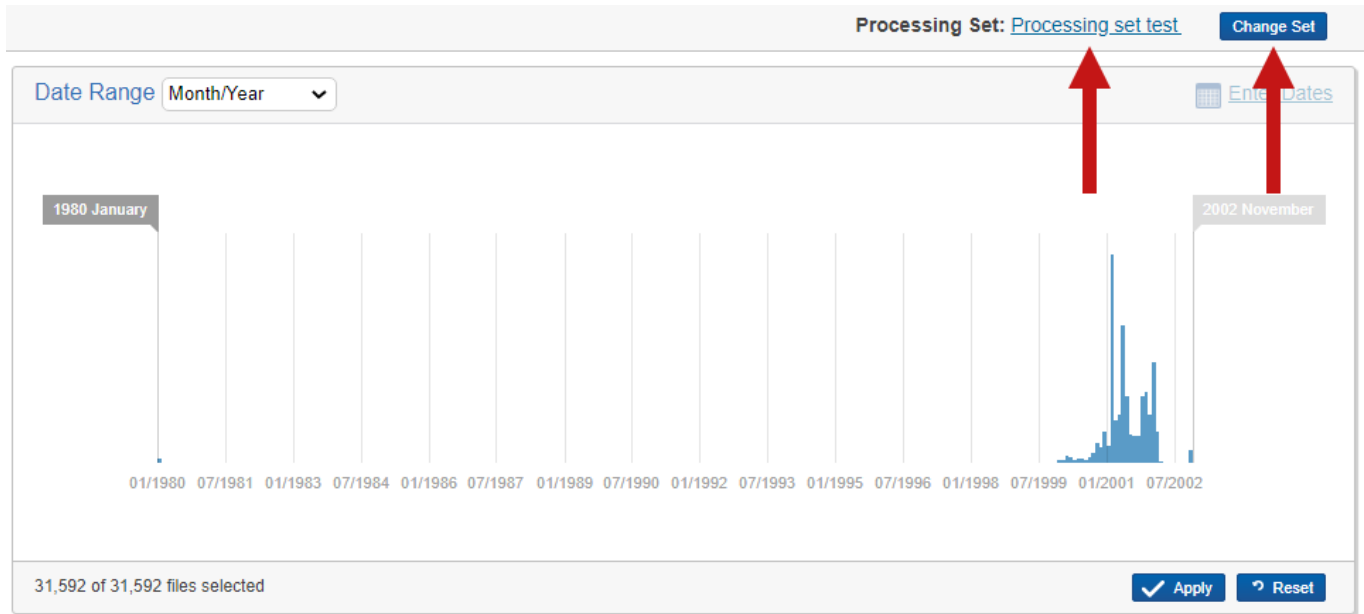
A processing set is not eligible for use in the Inventory tab if:

- You canceled the set.
- You already discovered or published the set.
- You have not yet run inventory on the set.
- A discovery, retry discovery, publish, republish, or retry publish job is in the queue for the set.

If no processing sets are eligible for use in the Inventory tab, you are directed to the Processing Sets tab to create a new set or check on the progress of an existing set.

The following considerations apply to all processing sets in Inventory:

- If you need to load a different processing set, click **Change Set** to display a list of other available sets.
- You can click the processing set name link in the top right to return to that set's layout.

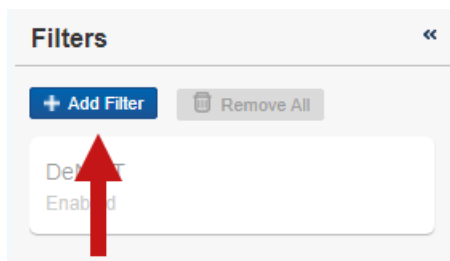


Note: If you leave the Inventory tab after having loaded a processing set, that set and any filters applied to it are preserved for you when you return to the Inventory tab.

You can add filters to the inventoried data and see how those filters affect the data in your processing set. You cannot add filters if inventory is not complete or if the processing set has already been discovered.

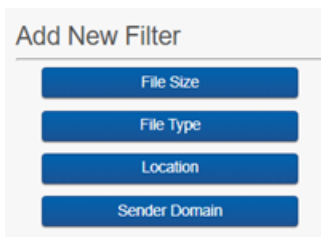
There are six different filters you can apply to a processing set. You can apply these filters in any order; however, you can only apply one filter of each type. This section describes how to apply file location, file size, file type, and sender domain filters. See [Applying a Date range filter](#) or [Applying a DeNIST filter](#) for instructions on filtering inventoried files by those properties.

To add a new filter, click **Add Filter**.



Note: Filters affect the data set only at the time at which you apply them. This means that if you apply a filter to exclude a certain file type from your data but someone from your organization adds more files to the set, including instances of the excluded type, then the recently added files are not actually removed when you start discovery. In order to exclude the added file types, you must first re-inventory the files in the updated data set. You can then run discovery and expect to see all instances of that file type excluded.

Clicking Add Filter displays a list of the following available filters:



- **File Size**—exclude files that are smaller and/or larger than the size range you specify. This filter uses a range graph, in which you click and drag range selectors to exclude files.
- **File Type**—include or exclude certain files based on their type. For example, you may want to remove all .exe or .dll files since they typically have no relevance in the review process. This filter uses a two list display, in which you choose which files to exclude by moving them from the Included list to the Excluded.

Note: Renaming a file extension has little effect on how Relativity identifies the file type. When processing a file type, Relativity looks at the actual file properties, such as digital signature, regardless of the named extension. Relativity only uses the named extension as a tie-breaker if the actual file properties indicate multiple extensions.

- **Location**—include or exclude files based on their folder location. This filter uses a two list display, in which you choose which files to exclude by moving them from the Included list to the Excluded.
- **Sender Domain**- include or exclude email files sent from certain domains. For example, you may want to rid your data set of all obvious spam or commercial email before those files get into your workspace. This filter uses a two list display, in which you choose which files to exclude by moving them from the Included list to the Excluded.

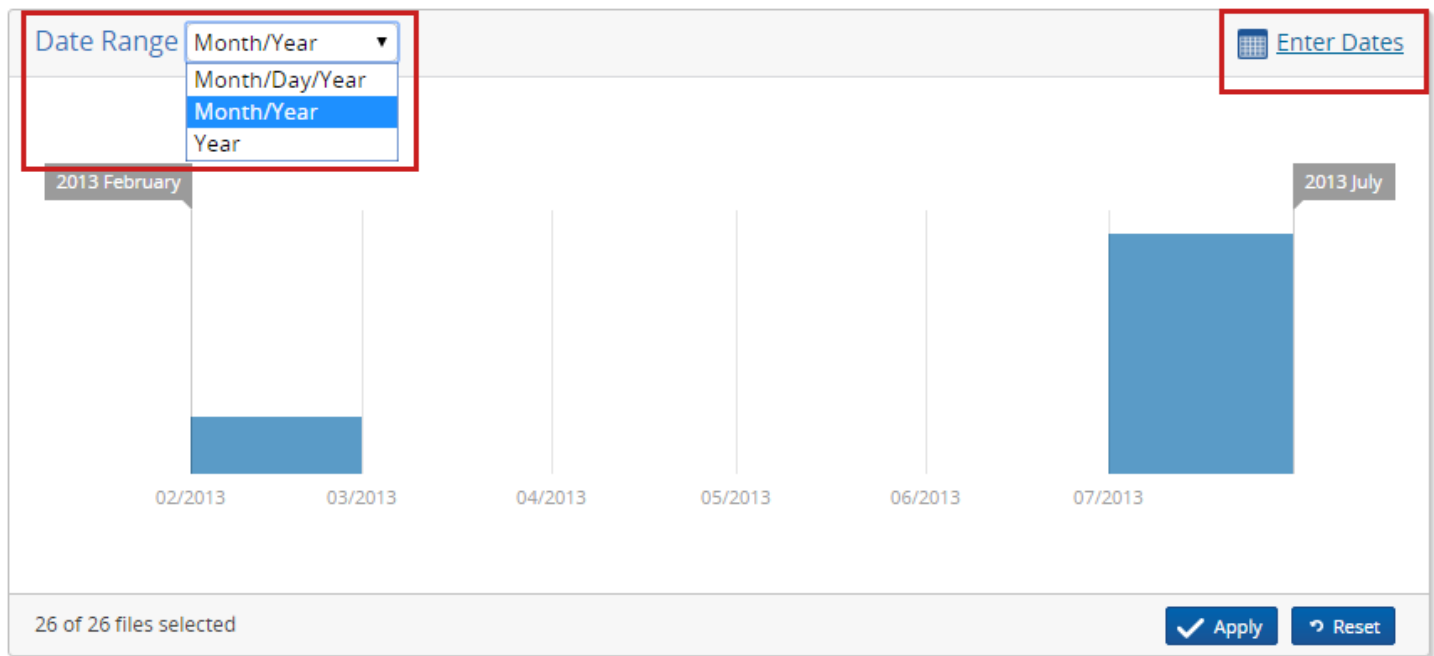
The following considerations apply to all filter types:

- If the applied filter conditions have excluded all files from the set, then there are no results for you to interact with and you cannot add or apply more filters.
- If a filter is already applied to the data, the corresponding button is disabled.
- The Inventory Progress graph displays the effect each filter has on your overall file count. The points on the graph indicate which filters you applied and the number of remaining files in your processing set.
- When you change the parameters of the filter, the number of documents automatically updates to reflect the change.
- Filters operate progressively, with each additional filter further narrowing down the total data set. For example, if you choose to include a certain file type and later filter out all file locations that contain those types of files, the discoverable data set does not ultimately include files of that type.
- To cancel a filter before you apply it, click **Cancel**. If you cancel, you lose all unsaved changes.
- You cannot save and reuse filters from one inventory set to another.

19.2.1 Applying a Date range filter

When the selected processing set loads, no filters are applied to the files by default; however, a graph displays the date range for all files in the processing set.

Note: The deNIST filter is applied by default if your processing profile has deNIST field set to Yes.



Note: When you filter for dates, you are filtering specifically on the Sort Date/Time field, which is taken from the file's Sent Date, Received Date, and Last Modified Date fields in that order of precedence. This happens on email messages repeated for the parent document and all child items to allow for date sorting.

You have the following options for applying a date range filter:

- Use the **Date Range** menu in the top left to select from **Month/Day/Year**, **Month/Year**, and **Year**. When you move the range selectors to a particular point on the graph, they will snap to the nearest whole number. Change the units of measurement to smaller denominations for more precise filter settings.

Note: When processing documents without an actual date, Relativity provides a null value for the following fields: Created Date, Created Date/Time, Created Time, Last Accessed Date, Last Accessed Date/Time, Last Accessed Time, Last Modified Date, Last Modified Date/Time, Last Modified Time, Primary Date/Time, and Sort Date/Time. The null value is excluded and not represented in the filtered list.

- The **Enter Dates** link in the top right, when clicked, displays a window in which you can select a **Start** and **End** date from the two calendars. Click **Apply** after specifying the new dates.

- Drag the right and left limits of the graph until you have a visualization of the desired range. When you do this, the areas that you have designated to exclude are light blue. Click **Apply** after dragging these limits to their appropriate spots.

Date Range Month/Year [Enter Dates](#)



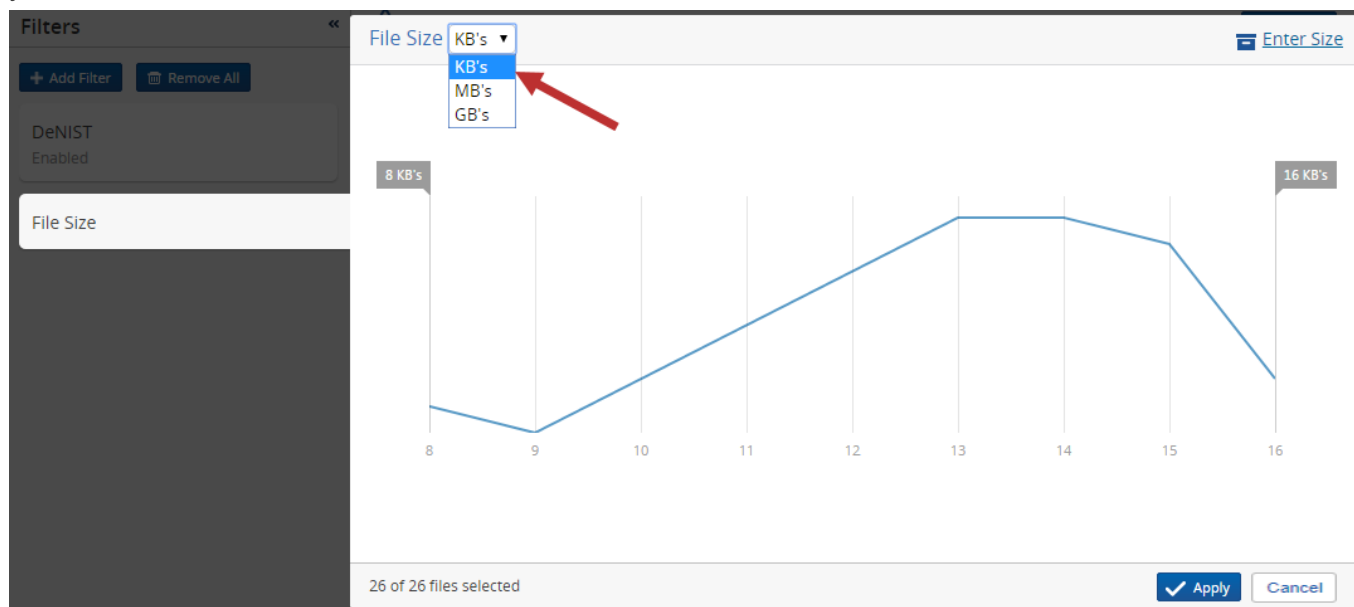
- To reset the parameters of a filter after you apply it, click **Reset**.

Note: If you run a re-inventory job on a processing set to which you have already added the date range filter, the date range display does not update automatically when you return to the Inventory tab from the processing set layout. You have to re-click the date range filter to update the range.

19.2.2 Applying a File Size filter

To filter your processing set files by size:

1. Click **Add Filter**.
2. Select **File Size** from the filter list.
3. Use the available options on the **File Size** range graph filter to specify how you want to apply the file size filter to your files.



- Use the **File Size** menu in the top left of the graph to select from **KB's**, **MB's**, and **GB's**. If all files in the current data set are from the same file size, for example 0 GB's, you cannot get a visualization for that size. When you move the range selectors to a particular point on the graph, they will snap to the nearest unit of measurement selected. Change the units of measurement to smaller denominations for more precise filter settings.
- Use the **Enter Size** link in top right of the graph to select **Start** and **End** values for the size range. By default, the lowest value in the data set appears in the Start field and the highest value appears in the End field.

Select File Size Range ×

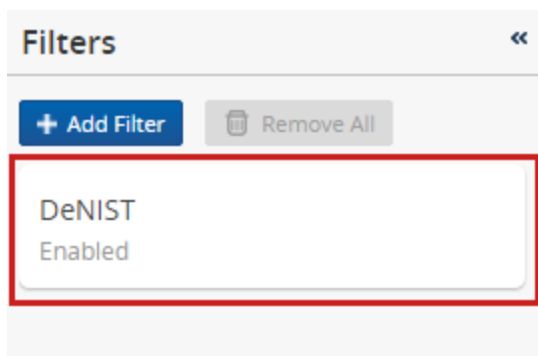
Start: × End:

4. Click **Apply** once you have designated all the file sizes you want to exclude. The Inventory Progress pane reflects the addition of the file size filter, as well as the percentage and number of files that remain from the original data set. For more information, see Inventory progress.

19.2.3 Applying a deNIST filter

You can toggle the deNIST Filter on or off to exclude commonly known computer system files that are typically useless in e-discovery. You will do this on the processing profile, and the selection you make there is reflected in the Inventory interface.

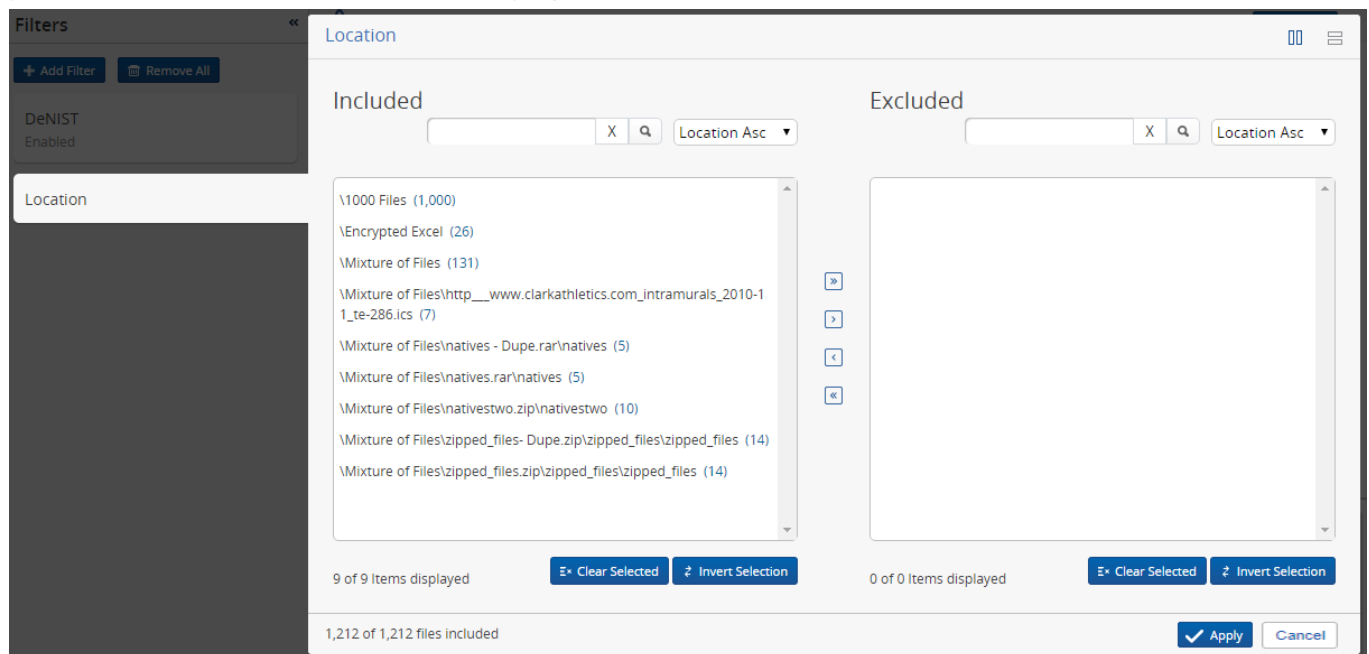
If the DeNIST field is set to No on the processing profile, the DeNIST filter does not appear by default in Inventory, and you do not have the option to add it. Likewise, if the DeNIST field is set to Yes on the profile, the corresponding filter is enabled in Inventory, and you cannot disable it for that processing set.



19.2.4 Applying a Location filter

To filter your processing set files by location:

1. Click **Add Filter**.
2. Select **Location** from the filter list.
3. Use the available options on the **Location** two-list filter to specify how you want to apply the location filter to your files. For more information, see Applying two-list filters.



4. Click **Apply** once you have designated all the file locations you want to exclude. The Inventory Progress pane reflects the addition of the location filter, as well as the percentage and number of files that remain from the original data set. For more information, see Inventory progress.

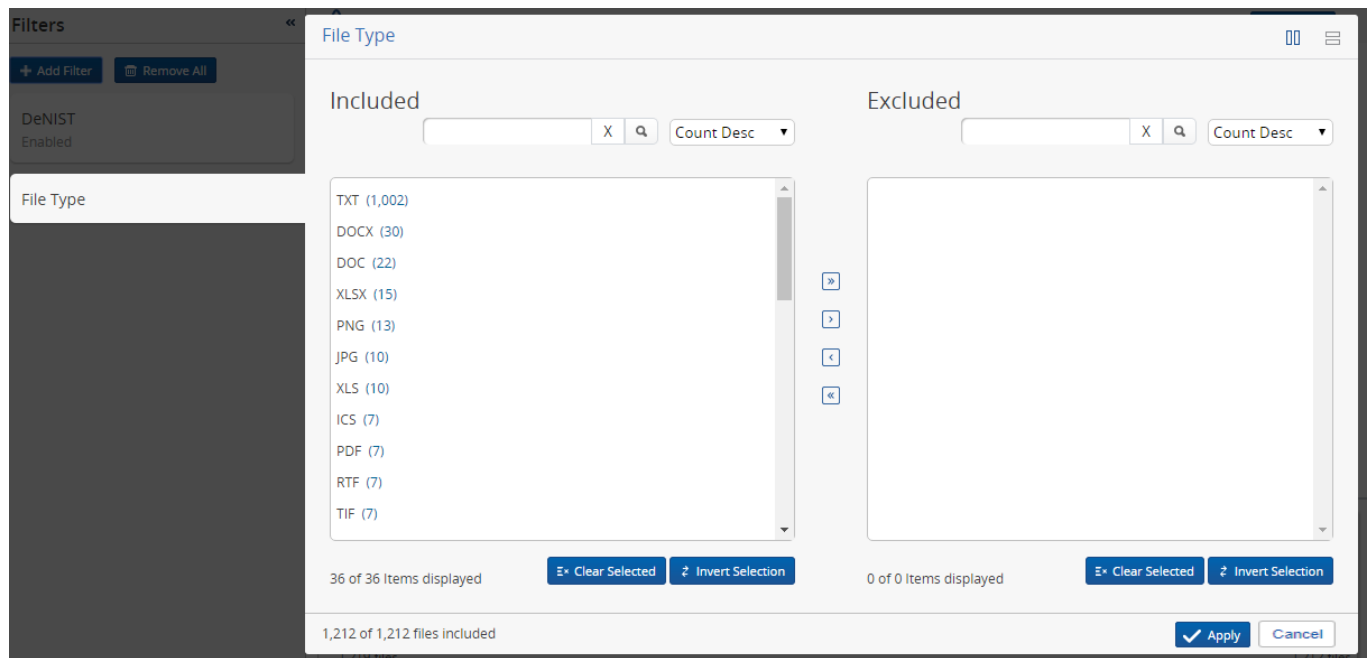
You can now apply an additional filter to further reduce the data set, or you can discover the files.

19.2.5 Applying a File Type filter

To filter your processing set files by type:

1. Click **Add Filter**.
2. Select **File Type** from the filter list.

3. Use the available options on the **File Type** two-list filter to specify how you want to apply the file type filter to your files. For more information, see Applying two-list filters.



4. Click **Apply** once you have designated all the file types you want to exclude. The Inventory Progress pane reflects the addition of the file type filter, as well as the percentage and number of files that remain from the original data set. For more information, see Inventory progress.

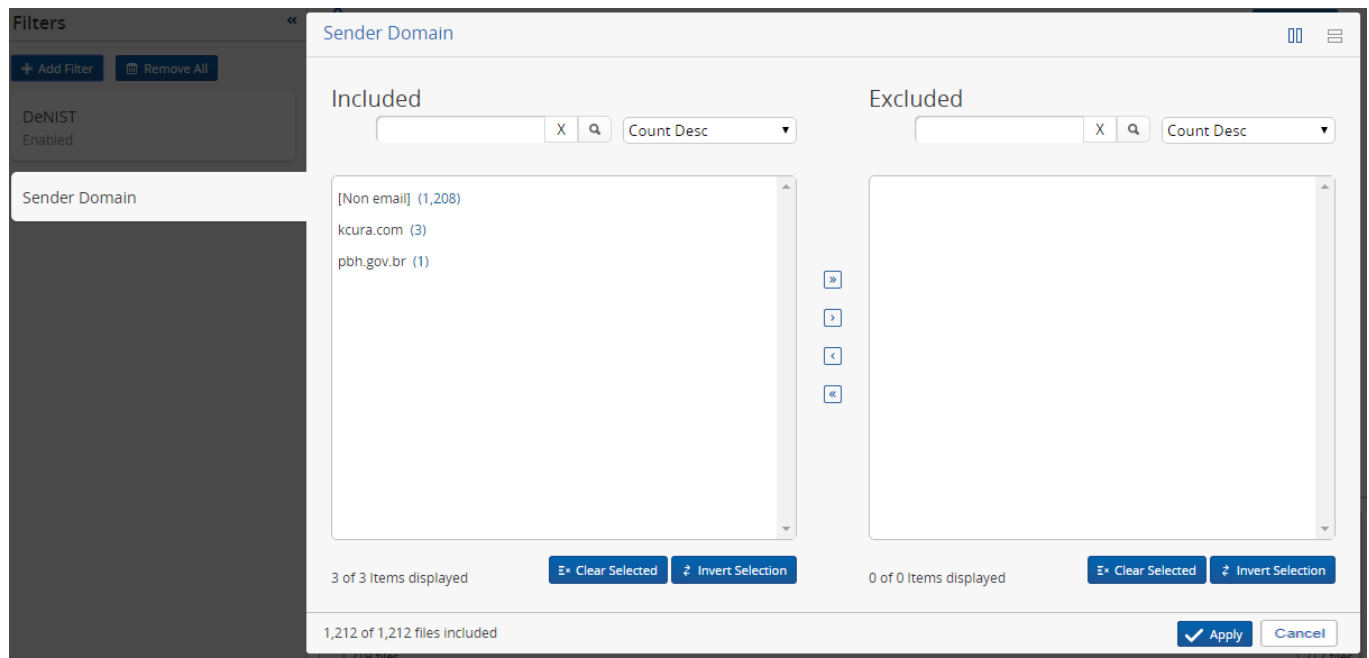
You can now apply an additional filter to further reduce the data set, or you can discover the files.

19.2.6 Applying a Sender Domain filter

To filter your processing set files by email sender domain:

1. Click **Add Filter**.
2. Select **Sender Domain** from the filter list.

3. Use the available options on the **Sender Domain** two-list filter to specify how you want to apply the sender domain filter to your files. For more information, see Applying two-list filters.



4. Click **Apply** once you have designated all the email domains you want to exclude. The Inventory Progress pane reflects the addition of the sender domain filter, as well as the percentage and number of files that remain from the original data set. For more information, see Inventory progress.

You can now apply an additional filter to further reduce the data set, or you can discover the files.

19.2.6.1 Unspecified domains

Some of the domain entries in your filter window might not be displayed in a traditional domain format. For example, if there are files from an unspecified domain in your processing set, these files appear as a number in parentheses without a domain name next to it. Note the other instances in which Relativity returns unspecified domains and how it handles those items:

- **[Non email]**—the item was not handled by the Outlook handler.
- **Blank**—the Outlook handler processed the file, but could not find a sender domain.
- **[Internal]**—Relativity parsed an LDAP-formatted email address because there was no other valid domain available. When the system cannot identify the domain, it attempts to extract the organization or company name from the address.

19.2.6.2 Applying a two-list filter

The two-list filter lets you filter processing set files by the following filter types:

- File Location
- File Type
- Sender Domain

When you add any of these filters, all instances of the variable being filtered for appear in the **Included** list to the left (or top). To exclude any instance, highlight it and click the single right arrow button to add it to the **Excluded** list on the right (or bottom).

Note: If you add items from the Included list to the Excluded or vice versa, and these additions affect the sort and search criteria of the modified list, you can refresh the list to re-apply the sort and search.

The screenshot shows a 'File Type' filter interface. At the top right, there are two icons: a vertical list icon and a horizontal list icon. Below these are two main sections: 'Included' and 'Excluded'. Each section has a search bar with an 'X' and a magnifying glass icon, and a sort dropdown menu. The 'Included' list contains items like 'XLSB (4)', 'XLSM (4)', 'XLSX (4)', 'XLTM (4)', 'XLTX (4)', 'XLS (3)', and 'XLT (3)'. The 'Excluded' list is currently empty. Between the two lists are four movement buttons: a double right arrow, a single right arrow, a single left arrow, and a double left arrow. At the bottom of each list, there are buttons for 'Clear Selected' and 'Invert Selection'. The status bar at the bottom shows '26 of 26 files included' and '0 of 0 items displayed' for the excluded list, along with 'Apply' and 'Cancel' buttons.

Note: Items removed from the data by edits to a previously applied filter are displayed in later filters with a value of (0) next to them. For example, if you apply the file type filter and then later narrow the date range to the extent that it filters out all files of the PDF type, then the next time you view the file type filter, PDFs are listed as having a count of (0).

You can use any of the following options in the two-list filter:

- Move over all items with double left and right arrows. Move over only the item(s) selected with the single left and right arrows.
- Toggle the two-list filter to display vertically or horizontally with the parallel line icons in the top right.
 - The vertical lines display all files in the left column, and those designated for exclusion in the right column.
 - The horizontal lines display all files in the top window, and those to be excluded in the bottom window.
- Double-click on any item to move it to the other list.
- Select multiple items in either list by clicking on the items, or select all items between two values in a list with the Shift key.
- Sort the Included and Excluded lists based on the following settings, depending on the filter type:
 - **Location Asc**—sorts a list of file locations in alphabetical order.
 - **Location Desc**—sorts a list of file locations in reverse alphabetical order.

- **Sender Domain Asc**—sorts a list of sender domains in alphabetical order.
- **Sender Domain Desc**—sorts a list of sender domains in reverse alphabetical order.
- **File Type Asc**—sorts a list of file types in alphabetical order.
- **File Type Desc**—sorts a list of file types in reverse alphabetical order.
- **Count Asc**—sorts a list of variables from the smallest count to the largest.
- **Count Desc**—sorts a list of variables from the largest count to the smallest.
- **Clear Selected**—marks the previously selected items in the Included or Excluded list as unselected.
- **Invert Selection**—marks the previously selected items in the Included or Excluded list as unselected while selecting the items that were not selected before.

19.3 Removing filters

Clicking **Remove All** under Filter Controls removes all the filters from the menu on the left side of the menu.

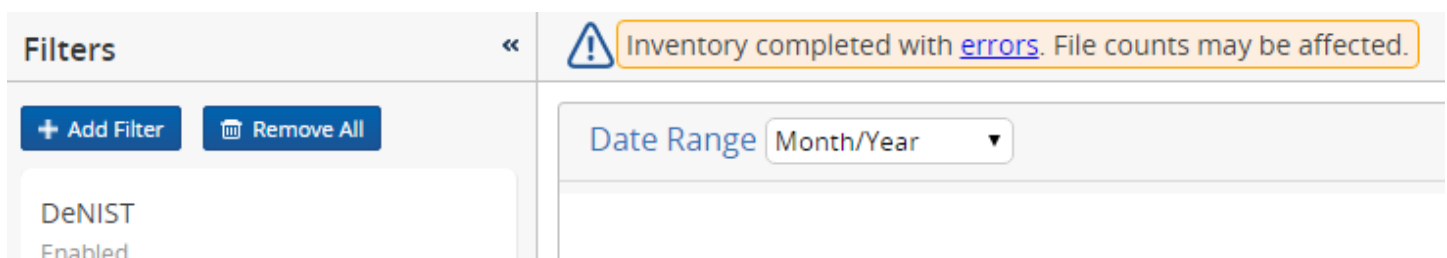
You can also remove filters individually by clicking the **X** on a single filter in the menu. You cannot delete a filter if you are currently working with it.

You are redirected to the processing set page if any of the following occur:

- Inventory or re-inventory is in process for the set
- The set has been canceled
- Discovery has been run for the set
- A job is in the queue for the set
- The set is secured or no longer exists

19.4 Inventory errors

If the processing set you select for inventory encountered any errors, the triangle icon appears in the upper left corner of the set. Hover over this icon to access a link to all applicable errors.



Clicking the link to view errors takes you to the Job Errors tab, which contains all errors for all processing sets in the workspace. By default, Relativity applies search conditions to this view to direct you to errors specific to your inventory data. Click any error message in the view to go to that error's details page, where you can view the stack trace and cause of the error.

All inventory errors are unresolvable. If you need to address an error that occurred during inventory, you must do so outside of Relativity and then re-run inventory on the processing set.

See [Processing error resolution](#) for details.

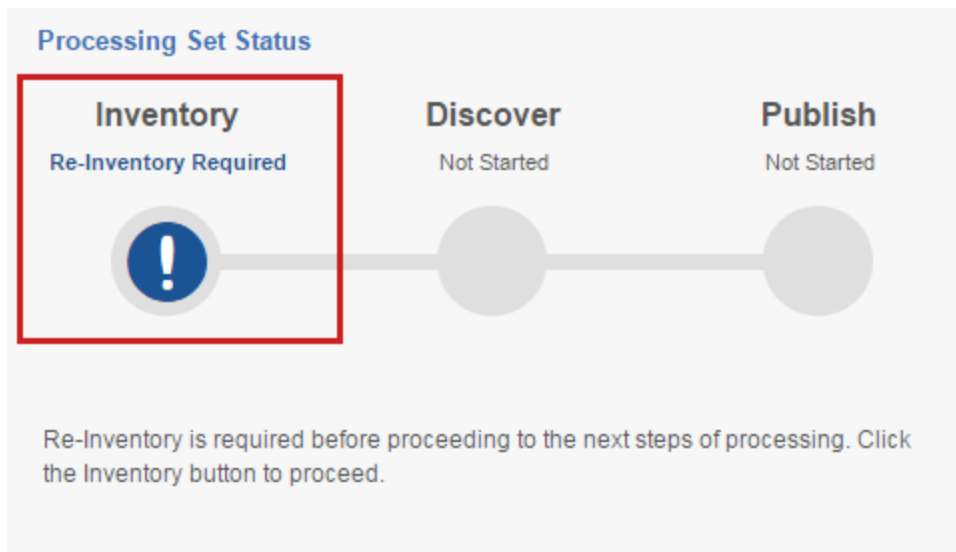
19.4.1 Inventory error scenarios

You receive an error when starting file inventory if any of the following scenarios occur:

- The processing license expires.
- You have an invalid processing license.
- The DeNIST table is empty and the DeNIST field on the profile is set to Yes.
- No processing webAPI path is specified in the Instance setting table.

19.5 Re-inventory

You may be prompted to run inventory again in the status display on the processing set layout.



You must run inventory again on a processing set if:

- You have added a data source to processing set that has already been inventoried but not yet discovered.
- You have edited a data source that is associated with a processing set that already been inventoried but not yet discovered.
- You have deleted a data source from a processing set that has already been inventoried but not yet discovered.

You can also voluntarily re-inventory a processing set any time after the Inventory Files option is enabled after the previous inventory job is complete.

To re-inventory at any time, click **Inventory Files**.

When you click Inventory again, you are presented with a confirmation message containing information about the inventory job you are about to submit. Click **Re-Inventory** to proceed with inventory or **Cancel** to return to the processing set layout.

Re-Inventory

This will clear all data from the inventory job previously run for this processing set.

Re-Inventory

Cancel

When you re-inventory files:

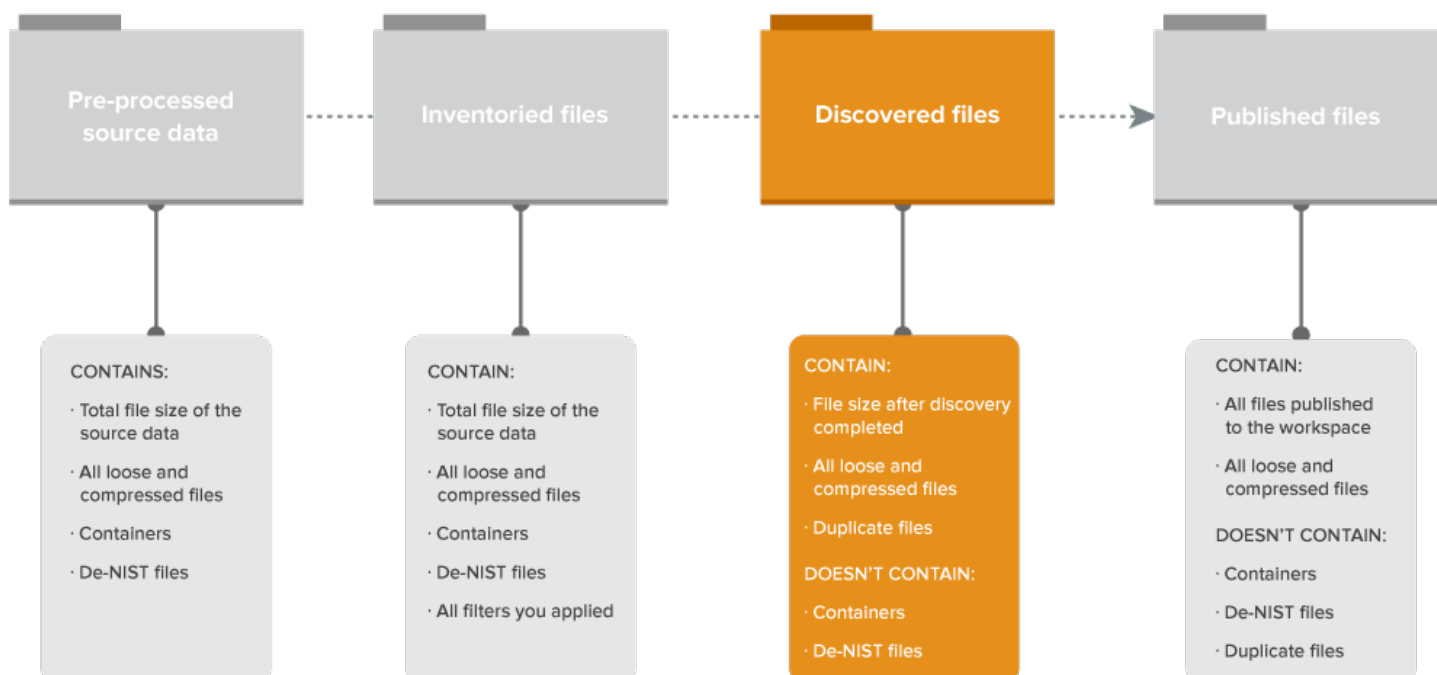
- Filters that you previously applied in the Inventory tab do not get cleared.
- Errors that you encountered in a previous Inventory job are cleared.

20 Discovering files

Discovery is the phase of processing in which the processing engine retrieves deeper levels of metadata not accessible during Inventory and prepares files for publishing to a workspace.

The following graphic depicts how discovery fits into the basic workflow used to reduce the file size of a data set through processing. This workflow assumes that you are applying some method of deNIST and deduplication.

File Size Reduction through Processing



The following is a typical workflow that incorporates discovery:

1. Create a processing set or select an existing set.
2. Add data sources to the processing set.
3. Inventory the files in that processing set to extract top-level metadata.

4. Apply filters to the inventoried data.
5. Run discovery on the refined data.
6. Publish the discovered files to the workspace.

20.1 Running file discovery

To start discovery, click **Discover Files** in the Process Files console. You can run the discover job whether or not you have inventoried or filtered your files.

Note: When processing documents without an actual date, Relativity provides a null value for the following fields: Created Date, Created Date/Time, Created Time, Last Accessed Date, Last Accessed Date/Time, Last Accessed Time, Last Modified Date, Last Modified Date/Time, Last Modified Time, Primary Date/Time, and Sort Date/Time. The null value is excluded and not represented in the filtered list.

The screenshot shows the 'Processing Sets' console with the 'Processing Set 005' selected. The 'Basic Settings' tab is active, displaying a progress bar with three steps: 'Inventory (Optional)' (Completed), 'Discover' (Not Started), and 'Publish' (Not Started). The 'Discover' step is highlighted with a '2' in a circle. To the right, a summary table shows the current counts:

Inventory count	3,194
Discover count	0
Publish count	0
Total exceptions	3

Below the progress bar, the 'Inventory (Optional)' section is expanded, showing three metrics:

Inventoried ?	Filtered ?	Inventory Job Errors ?
3,194 files 0.33 GB	418 files 0.02 GB	3

On the right side of the console, the 'Process Files' section contains buttons for 'Inventory Files', 'Filter Files', 'Discover Files' (highlighted with a red border), 'Retry File Exceptions', and 'Publish Files'. Below this, the 'Reports' section lists 'Inventory Report', 'Discovery Report', and 'All Reports'. The 'Exceptions' section lists 'Job Errors' and 'File Exceptions'.

A confirmation message appears with the discovery settings and filters applied. Click **Continue** to proceed with discovery or **Cancel** to return to the processing set layout. If you enabled auto-publish, the confirmation message provides the option to discover and publish the files.

Continue?

Once you start file discovery:

- You can't modify the settings of this processing set
- Inventory filters will be locked
- Publish will automatically begin once file discovery is complete
- Any unmapped fields will not have metadata populated
- You will be unable to cancel a processing set until any subsequent dependent processing sets are also canceled in this workspace.

The following filters will be applied:

- DeNIST

Auto-publish: Yes

DeNIST: Yes

Deduplication method: Global

Export Batch Size: 4000

Continue

Cancel

Consider the following when discovering files:

- Relativity does not re-extract text for a re-discovered file unless an extraction error occurred. This means that if you discover the same file twice and you change any settings on the profile, or select a different profile, between the two discovery jobs, Relativity will not re-extract the text from that file unless there was an extraction error. This is because processing always refers to the original/primary document and the original text stored in the database.
- If you have arranged for auto-publish on the processing set's profile, the publish process begins when discovery finishes, even if errors occur during discovery. This means that the Publish button is not enabled for the set until after the job is finished. You also see a status display for both discover and publish on the set layout.
- When discovering file types, Relativity refers to the file header information to detect the file type.
- You can't change the settings on any processing job at any point after file discovery begins. This means that once you click **Discover**, you can't go back and edit the settings of the processing set and re-click Discover Files. You would need to create a new processing set with the desired settings.
- You cannot start discovery while inventory is running for that processing set.
- When you start discovery or retry discovery for a processing job, the list of passwords specified in the password bank accompanies the processing job so that password-protected files are processed in that job. For more information, see [Password bank on page 1](#).

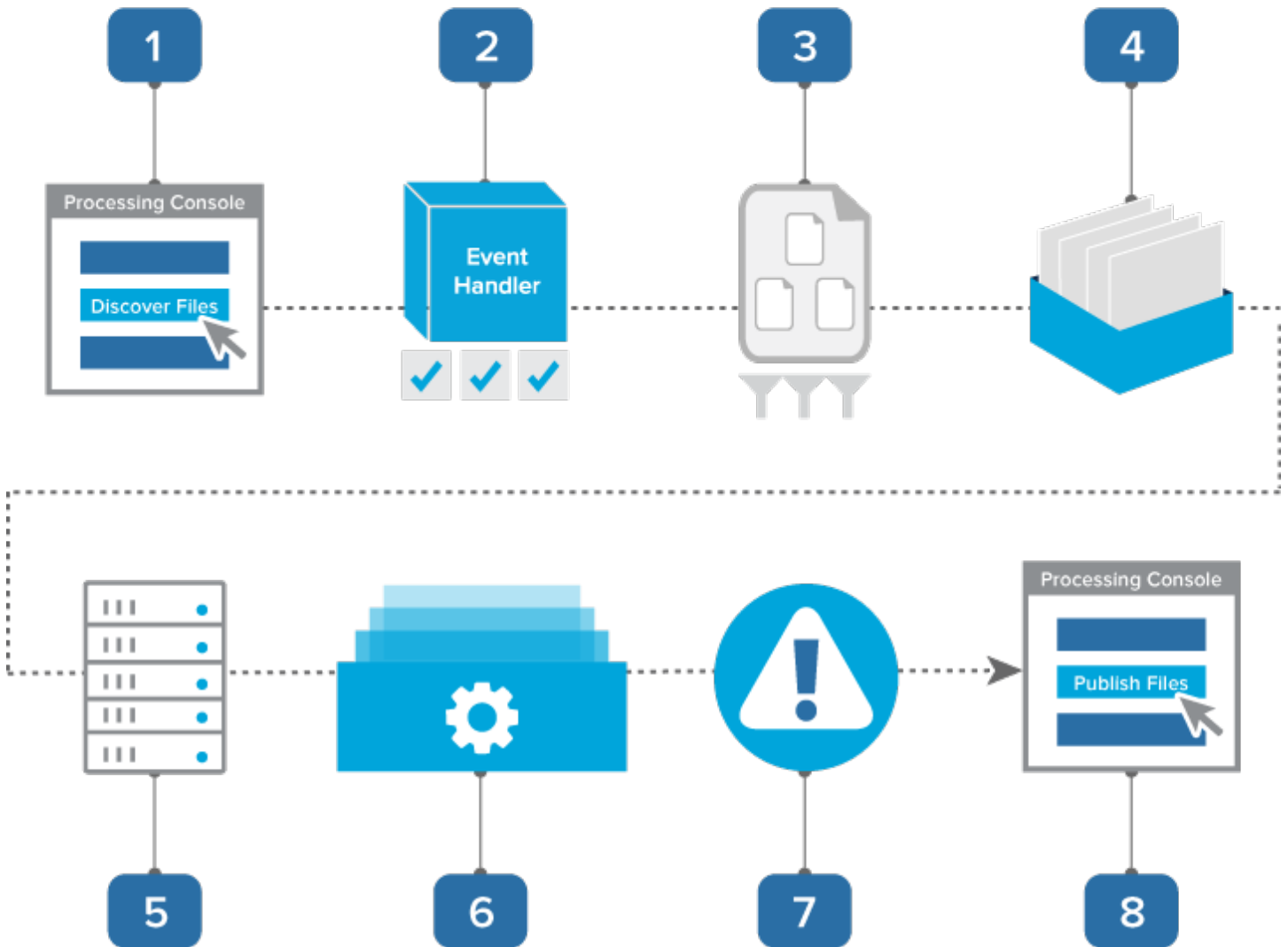
Note: Relativity prioritizes application metadata over operating system file properties where possible. For example, if a file type stores application metadata, such as Date Created and Date Modified, Relativity retains those values for the file. If the application metadata fields are empty or the file type does not store application metadata, Relativity uses the operating system's file properties instead. Application metadata is more reliable since it is stored in the file itself. Operating system file properties can often change. For example, moving a file from one folder to another may change property values. Examples of file types that store application metadata include Microsoft Office files such as Word or Excel.

When you start discovery, the Discover button changes to **Cancel**. Click this to stop discovery. See [Canceling discovery](#) for details.

20.1.1 Discovery process

The following graphic and corresponding steps depict what happens behind the scenes when you start discovery. This information is meant for reference purposes only.

Discovery Process



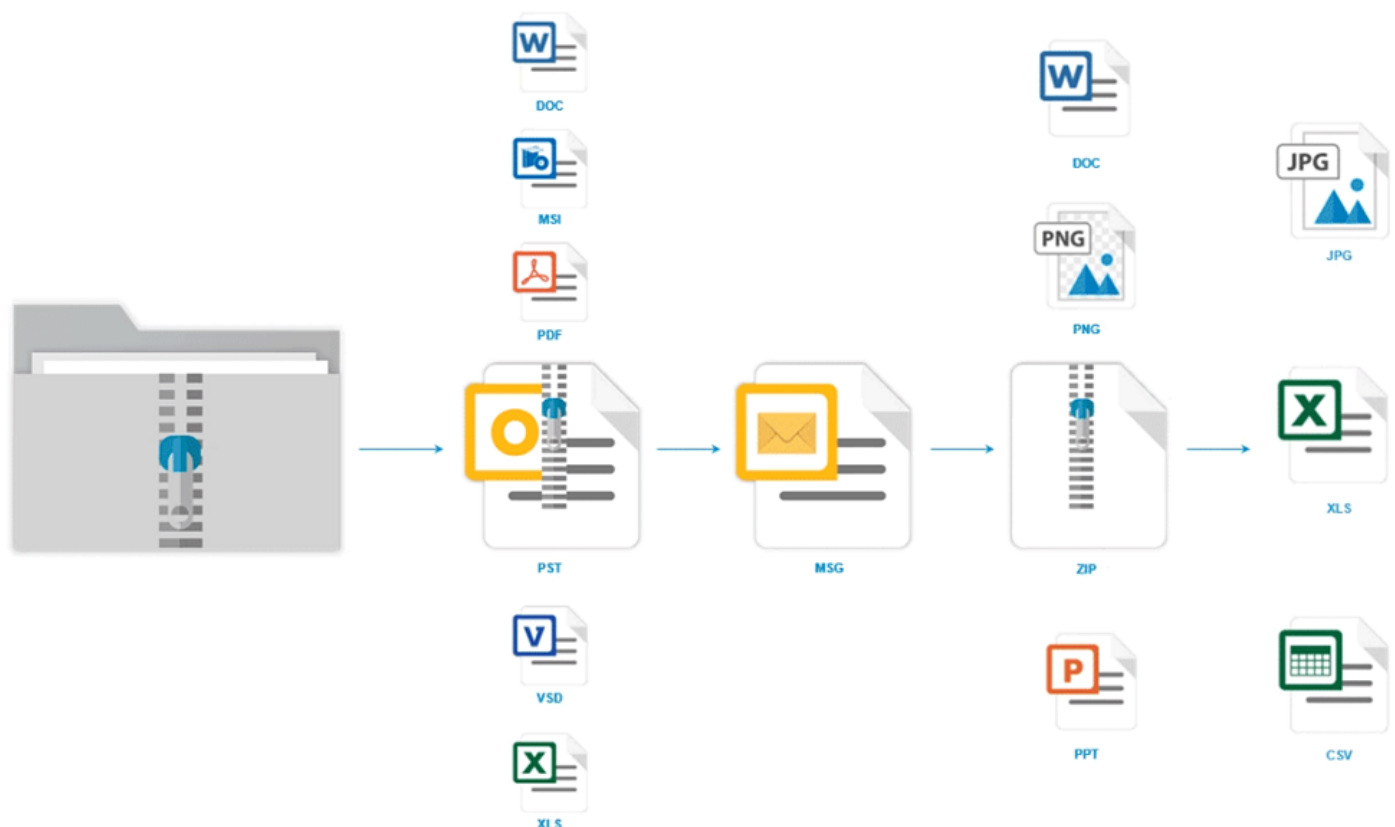
1. You click Discover Files on the processing set console.
2. A console event handler copies all settings from the processing profile to the data sources on the processing set and then checks to make sure that the set is valid and ready to proceed.
3. The event handler inserts all data sources into the processing set queue.
4. The data sources wait in the queue to be picked up by an agent, during which time you can change their priority.
5. The processing set manager agent picks up each data source based on its order, all password bank entries are synced, and the agent submits each data source as an individual discovery job to the processing engine. The agent then provides updates on the status of each job to Relativity, which then displays this information on the processing set layout.

6. The processing engine discovers the files and applies the filters you specified in the Inventory tab. It then sends the finalized discovery results back to Relativity, which then updates the reports to include all applicable discovery data.
7. Any errors that occurred during discovery are logged in the errors tabs. You can view these errors and attempt to retry them. See [Processing error resolution](#) for details.
8. You can now publish the discovered files to your workspace. If you've arranged for auto-publish after discovery, publish will begin automatically and you will not invoke it manually.

20.1.2 Container extraction

It may be useful to understand how the processing engine handles container files during discovery. Specifically, the following graphic depicts how the engine continues to open multiple levels of container files until there are no more containers left in the data source.

This graphic is meant for reference purposes only.



20.2 Special considerations - OCR and text extraction

Consider the following regarding OCR and text extraction during discovery:

- If enabled on the processing profile, Relativity performs OCR on the following qualifying files:
 - Pixel-based image formats, such as .tiff, .jpeg, .gif, .bmp, .png and so forth.
 - Vector-based drawing formats such as .cad, .vsd/x, .dwg, and so forth.
 - Note that for .svg files, Relativity only supports extraction of text-based elements. Images within these files are not subject to OCR.

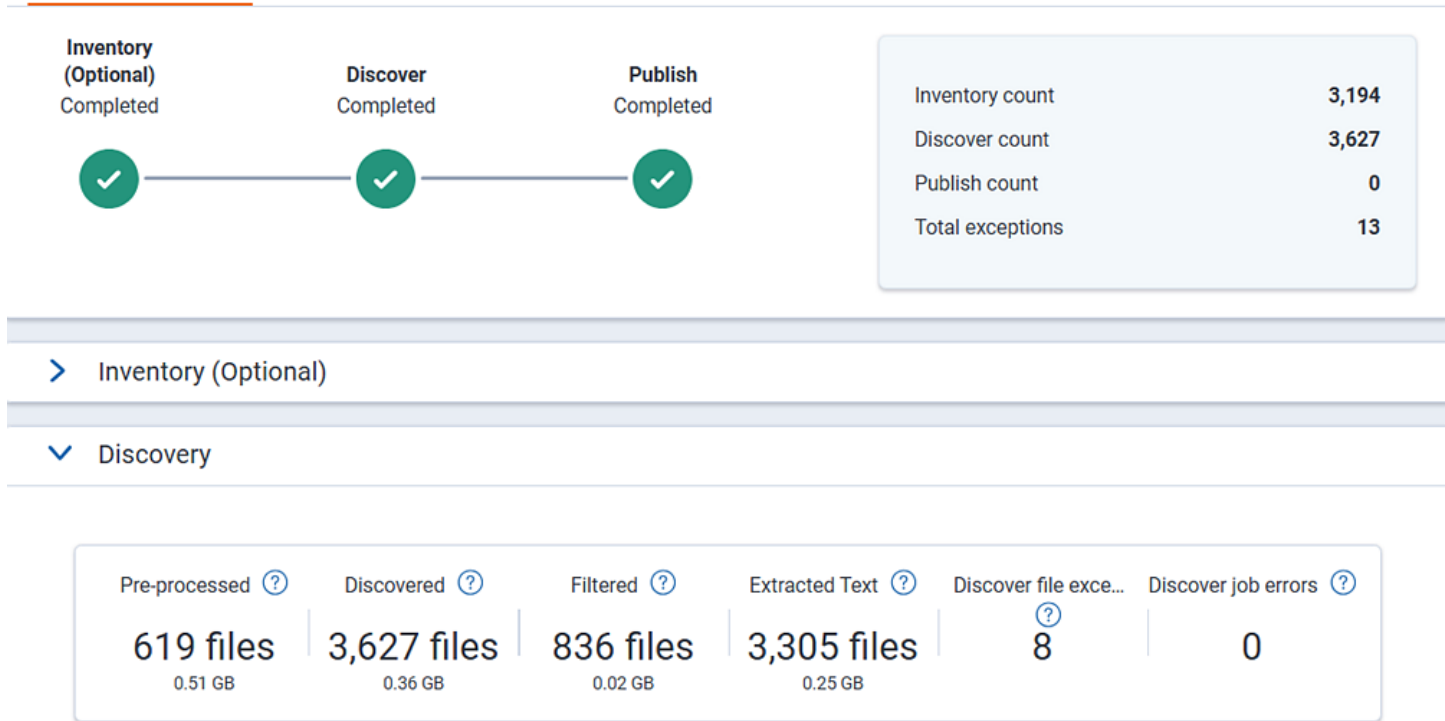
- PDF pages with image layers.
- Images embedded in documents, as well as email in-line images; also requires enabling the **Extract children** setting on the processing profile.
- Relativity populates the **Extracted Text** field with the results of OCR performed during processing discovery.
- Documents successfully OCR'd during processing can be identified in the workspace using the fields **Has OCR Text** and **Text Extraction Method**. For more information, see [Mapping processing fields](#).
- For multi-page records with a mix of native text and images, Relativity segments out OCR and extracted text at the page level, not the document level. Relativity combines all page-level text and OCR within the published **Extracted Text** field.
- In the case where a file contains both native text and OCR within the extracted text of the record, there is a header in the **Extracted Text** field indicating the text that was extracted through OCR. This header only exists if the **OCR Text Separator** option is enabled in the processing profile.
- Relativity extracts OCR to Unicode. Any and all language scripts represented in the OCR, including non-English characters, are Unicode-encoded in the published **Extracted Text** field.
- Relativity displays a missing extracted text error if it cannot OCR one or more pages in a document. The error reads: *An error occurred on X (number of) page(s) when attempting to OCR. Consider retrying. Page numbers missing OCR text : (page numbers)*
- The processing engine copies native files and OCR results to the document repository. Whether or not you publish these files, they remain in the repository, and they are not automatically deleted or removed.
- If OCR is not essential to your processing job, it is recommended to disable OCR and related settings on your processing profile. Forgoing OCR at this stage can significantly reduce processing time and prevent irrelevant documents from having OCR performed on them. You can then perform OCR on only relevant documents post-publish within the workspace. For more information regarding OCR settings and impact on processing, see [Processing profile - Extraction settings](#).

20.3 Monitoring discovery status

You can monitor the job's progress on the Processing Set Details page. This page breaks down each of the processing phases (Inventory, Discover, Publish) into their own sections that include dashboard summaries, sub-job details, and counts.

20.3.1 Discover job details

Processing Set 005 Basic Settings



Dashboard numbers

- **Pre-processed**—the number and size of all files in the data source location for the processing set.
- **Discovered**—the number of files discovered, excluding deleted files and containers.
- **Filtered**—the total number of non-container files filtered from the Discover job. This includes files filtered from Inventory, Inclusion/Exclusion, and DeNIST. This number includes both parent and child documents. To view specific filter counts, view the [Discovery Exclusion Results](#) report.
- **Extracted Text**—the number of files where Relativity obtained text either directly from the file or through OCR. Not all files discovered have text extracted, for example, image files that do not have text. Previously discovered files with extracted text are included in this count.
- **Discover file exceptions**—the number of files with Discover exceptions. Some Discover file exceptions are expected in cases of corruption or encryption.
- **Discover job errors**—the number of job errors that occurred during Discovery. Job errors are critical issues that must be remediated. If a job error occurs during the initial expanding files stage, it will be unresolvable by default. You should create a new processing set.

Note: In some rare cases, you may see an additional field called **Files removed after Retry**. If you retry discovery errors, this field appears when the total file count after retrying is less than the original discovery file count. For example, you see the **Files removed after Retry** field is 10. This number tells you the total number of files discovered after retrying errors is ten less than the original number of discovered files.

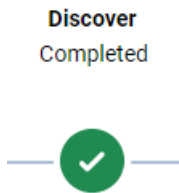
Progress bars

Each processing phase—Inventory, Discover, and Publish—consists of distinct sub-jobs that execute sequentially until the phase is complete. Progress for each job is visually represented by progress bars that appear under the dashboard. Once all sub-jobs within a phase are finished, the progress bars are no longer displayed.



- **Preparing files**—Relativity identifies and preps data from the data source(s) for expansion and extraction.
- **Expanding files**—Relativity expands all source records to locate files and extract metadata.
- **Text extraction/OCR**—Relativity extracts text from discovered files. If enabled, it performs OCR on applicable records to extract additional text. Be aware that OCR can affect performance and accounts for the majority of the overall job time.
- **Updating Processing Set fields**—Relativity updates associated fields in the processing set with the job results.

Once discovery is complete, the status section displays a check mark, indicating that you can move on to publishing your files or viewing your files in the Discovered files view. and [Discovered files view](#).



20.4 Canceling discovery

Once you start discovery, you can cancel it before the job reaches a status of Discovered with errors or Discover files complete.

To cancel discovery, click **Cancel** in the Process Files console.

Process Files

Inventory Files

Filter Files

Discover Files

Retry File Exceptions

Cancel

Reports

Inventory Report

Discovery Report

All Reports

Exceptions

Job Errors

File Exceptions

Consider the following regarding canceling discovery:

- If you click cancel while the status is still waiting, you can re-submit the discovery job.
- If you click cancel after the job has already been sent to the processing engine, the set is canceled, meaning all options are disabled and it is unusable. Deduplication isn't run against documents in canceled processing sets.
- If you have auto-publish enabled and you cancel discovery, file publishing does not start.
- Once the agent picks up the cancel discovery job, no more errors are created for the processing set.
- Errors resulting from a canceled job are given a canceled status and cannot be retried.
- Once you cancel discovery, you cannot resume discovery on those data sources. You must create new data sources to fully discover those files.

Once you cancel discovery, the status section is updated to display the canceled state.

Processing Set 005 Basic Settings

Inventory
(Optional)
Canceled

×

Discover
Canceled

×

Publish
Canceled

×

Inventory count	0
Discover count	0
Publish count	0
Total exceptions	0

21 Files tab

Upcoming changes to the Files tab

Relativity is updating the Files tab on a phased rollout schedule, which means some customers will see the changes before others. The most significant changes are to the Files tab user interface, the datasets returned, and the sorting functionality. Where the two tab views differ, this documentation describes both.

If you have any questions about these changes or your environment, contact Customer Support. The Files tab in Relativity Processing allows you to view and analyze a list of all discovered documents and their metadata before deduplication and publishing.

Processing SetsProcessing Data SourcesProcessing ProfilePassword Bank | ProcessingInventoryReportsJob ErrorsFilesMore

All Files

Add Widget

No Dashboard S...

Export

1 - 100 of 295

100 per page

#	☐	Details	File ID	File Name	File Type	File Extension - Text	Custodian	Data Source	Dedupe Status	Is Published?	File Size (KB)	Sender Domai..
			=	f	Filter	Filter	Filter	Filter	(All)	(All)	=	Filter
1	☐	Details	1	Sample Lotus...	Lotus Notes Database	NSF	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	21,504	
2	☐	Details	2	Excel Sheet - ...	Encrypted Office Document	XLSX	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	64	
3	☐	Details	3	Pre2007Basi...	Microsoft Word 97 - 2003 Document	DOC	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	22	
4	☐	Details	4	Word Sheet - ...	Encrypted Office Document	DOCX	Brule, Steve	Brule, Steve - 1548436	Duplicate	No	24	
5	☐	Details	5	PDF Sheet - T...	Adobe Portable Document Format	PDF	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	94	
6	☐	Details	6	Presentation ...	Encrypted Office Document	PPTX	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	1,122	
7	☐	Details	7	b4ubuild_sa...	Microsoft Project Document	MPP	Brule, Steve	Brule, Steve - 1548436	Duplicate	No	320	
8	☐	Details	8	flag.JTT	Ichitaro document	JTD	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	16	
9	☐	Details	9	outlook(this i...	XHTML Document	HTM	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	45	
10	☐	Details	10	Encrypted Me...	PKCS #7 Message	P7M	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	5	
11	☐	Details	11	Naming.wpd	WordPerfect 6 and higher document	WPD	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	6	
12	☐	Details	12	Test; Test.msg	Outlook Message File	MSG	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	26	[Internal] - KCUR CORPORATION
13	☐	Details	13	FW Congratul...	Outlook Message File	MSG	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	31	hotmail.com
14	☐	Details	14	ManyFields2...	Outlook Message File	MSG	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	15	

All 295

Export as CSV

Save filter

Total: 295

Files tab (MDS)

The Files tab (using MDS storage) differs slightly from the current interface. Details on these changes are described in the [All Files view \(MDS\)](#) section.

Processing Sets Processing Data Sources Processing Profile Password Bank | Processing Inventory Reports Processing Duplication Workflow Job Errors **Files**

> Q Replace All files Add Widget No Dashboard Wi... > Q

A viewing sample of 5,000 documents is displayed from the total returned results of 123,144.

.....

1 - 100 of 5,000 100 per page

#	Details	Logical ID	File ID	File Name	File Type	File Extension - Text	Custodian	Data Source	Dedupe Status	Level
1		7AD31AB5AEC902...	1	58c4f24e-ed3b-4ab7-ae45-1...	Rich Text File	RTF	Allen, Phillip	Phillip Allen 001	Primary	1
2		B72D278BCBAF4A...	2	790dc3cd-4d76-4fca-8b86-2...	Rich Text File	RTF	Allen, Phillip	Phillip Allen 001	Primary	1
3		92A82776DCFD7C5...	3	PALLEN_0013633.msg	Unknown	MSG	Allen, Phillip	Phillip Allen 001	Primary	1
4		36FF9F5D5B6F156...	4	58a31dd2-6c1d-49ab-850e-3...	Rich Text File	RTF	Allen, Phillip	Phillip Allen 001	Primary	1
5		6D8AB6614E59F17...	5	PALLEN_0008257.doc	Microsoft Word 97 - 2003 Document	DOC	Allen, Phillip	Phillip Allen 001	Duplicate	1
6		80B3D939895ADC3...	6	PALLEN_0010033.xls	Microsoft Excel 97- 2003 Worksheet	XLS	Allen, Phillip	Phillip Allen 001	Duplicate	1
7		D202AC7E15E5ECD...	7	Changes to Kiodex co-marke...	Outlook Message File	MSG	Allen, Phillip	Phillip Allen 001	Primary	1
8		9AB707561F954CD...	8	790faa89-3659-403a-9637-ff...	Microsoft Word 97 - 2003 Document	DOC	Allen, Phillip	Phillip Allen 001	Primary	1
9		9F70083BCAC56FC...	9	794e79b3-108c-4977-8511-a...	Adobe Portable Document Format	PDF	Allen, Phillip	Phillip Allen 001	Primary	1
10		2D7474F71CB11E4...	10	PALLEN_0008202.doc	Microsoft Word 97 - 2003 Document	DOC	Allen, Phillip	Phillip Allen 001	Primary	1

All 123,144 Select... Save Filter Viewing sample 5000 Total: 123,144

21.1 Views on the Files tab

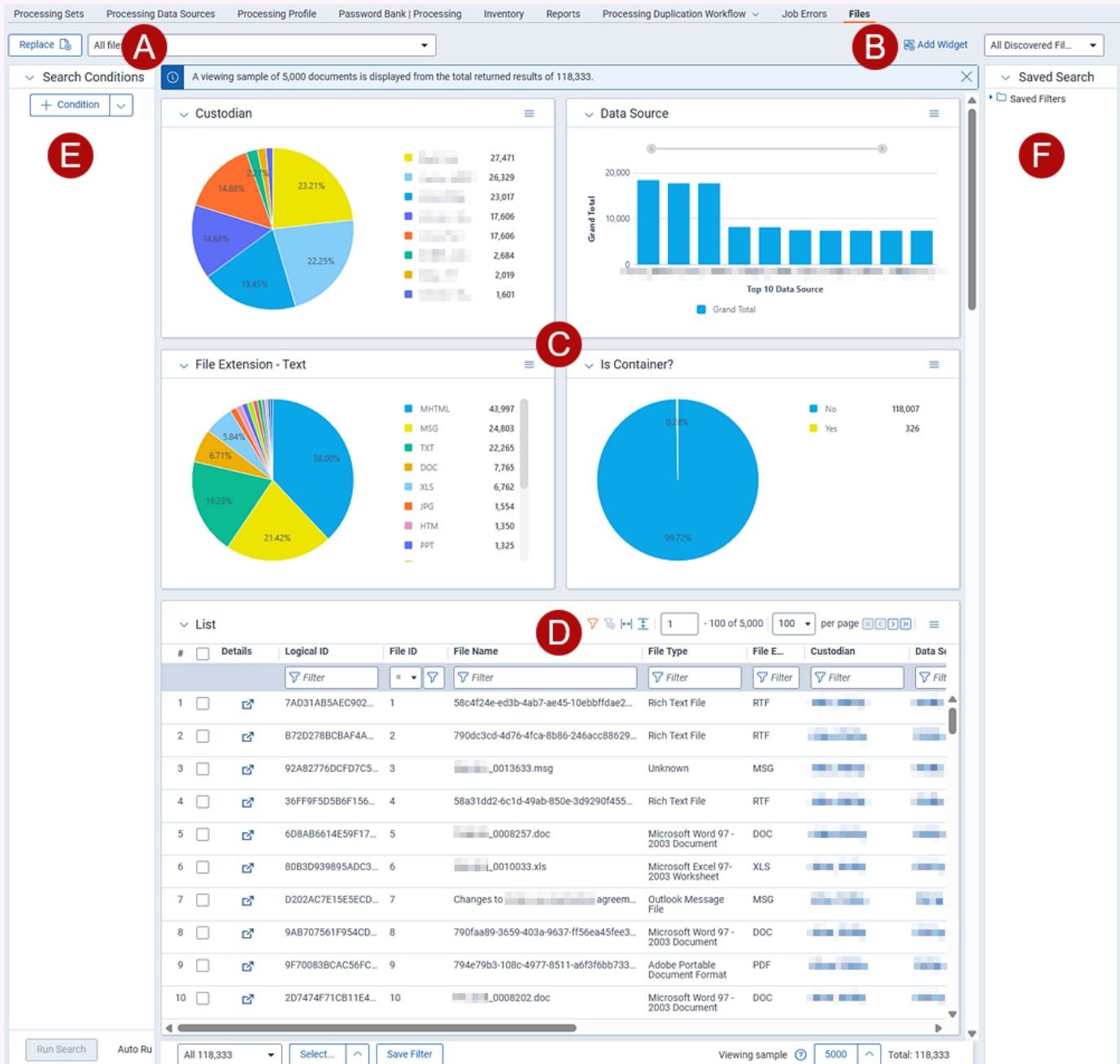
The Files tab contains the following views:

- **All files**—lists all the files in your workspace.
- **Current files with exceptions**—lists the documents that yielded exceptions in your workspace that currently have an Exception Status value of Not Resolved.
- **All files with exceptions**—lists the documents that yielded exceptions in your workspace, including those with a current Exception Status value of Resolved and Unresolved.
- **Deleted files**—lists the documents you deleted from your workspace.

21.1.1 All Files view

The All Files view contains all the discovered files in your workspace. This view does not contain documents that have been deleted and have a Yes value for the Processing Deletion? field. Those documents can only be found in the Deleted Documents view described in the next section.

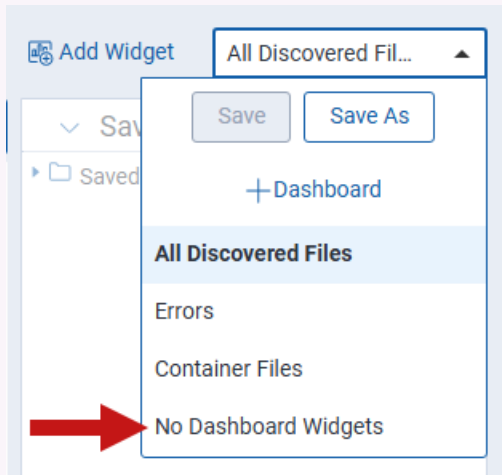
21.1.2 All Files view (MDS)



The Files tab has the following sections:

- Views**—select from predefined views or create a custom view.
- Dashboard widgets**—select from predefined dashboards or create a custom dashboard. Selecting **No Dashboard Widgets** hides the widgets from the current view so that only the list is displayed. Widgets display the results from the full data set, not just the sample set. They can give you an excellent overview of your processing status, for example, errors and exceptions.
Using dashboard widgets

You can change the widgets displayed in your view by selecting from a predefined dashboard or adding new widgets to a custom dashboard. If you prefer, you can hide all of the widgets as well. For information on how to add creating custom dashboards and widgets, see [Dashboards](#).



Examples of widgets:

- **All Discovered Files**—displays widgets showing custodians, data sources, file extensions, and a summary of the container files discovered.
- **Errors**—displays widgets showing exception categories, file extensions, and a summary of the container files discovered.
- **Container Files**—displays widgets showing a summary of the container files discovered.
- **No Dashboard Widgets**—hides all widgets from the view, so that you only see the document list.

- C. **Widget display area**—displays the dashboard's widgets.
- D. **Document list**—displays the files returned after discovery.
- E. **Search Conditions**—after creating a saved filter, any conditions applied appear in the Search Condition pane. To change a filter condition, first clear the conditions, then add one or more new conditions.
- F. **Saved Filters**—displays a list of saved filters. See [Saved Filters](#) for information on managing saved filters.

21.1.2.1 Files tab considerations

There are a few considerations when viewing documents on the Files tab. Some of the considerations differ from Relativity's default Documents list layout and functionality.

The following considerations apply to the Files tab list and navigation:

- The number of browsable documents (viewing sample) is limited to the **Viewing sample** setting. This number represents a sample of the data set. A banner just above the list lets you know the sample size and the total number of documents in the data set.

A viewing sample of 5,000 documents is displayed from the total returned results of 118,333.

1 - 100 of 5,000 100 per page

#	Details	Logical ID	File ID	File Name	File Type	File E...	Custodian
1		7AD31AB5AEC902...	1	58c4f24e-ed3b-4ab7-ae45-10ebbfdae2...	Rich Text File	RTF	
2		B72D278BCBAF4A...	2	790dc3cd-4d76-4fca-8b86-246acc88629...	Rich Text File	RTF	

To change the viewing sample size, use the viewing sample selector in the lower right corner of your screen.

5000
10000
25000
5000

Viewing sample 5000 Total: 118,333

- The total record count is displayed in the lower right corner.

All 118,333 Select... Save Filter Viewing sample 5000 Total: 118,333

- The mass actions count displays the number of documents selected for a mass action. Mass actions apply to the full data set and not just the sample set.

All 118,333 Select... Save Filter Viewing sample 5000 Total: 118,333

- Sorting is limited, depending on the number of documents in the sample. If the viewing sample exceeds the sorting threshold, sorting is disabled and a banner appears at the top of the list indicating the sample size exceeds the sorting threshold. You will also see a message when you attempt to apply sorting to a column.

Sorting limits

Sorting is unavailable due to the large result set.

The current filtered results exceed the sorting threshold of 10,000. To enable sorting, refine your filters further or export the dataset to a CSV for sorting outside of Relativity.

Acknowledge

Note: You can export the file and sort the contents outside of Relativity, then reimport. Another work-around is to add additional filters to achieve the results you need.

21.1.2.2 Documents list default fields

The documents list contains the following fields by default:

- **Details**—the details view of all fields, including compressed metadata, of the discovered file selected.
- **File ID**—the number value associated with the discovered file in the database.
- **File Name**—the original name of the discovered file.
- **File Type**—the file type of the discovered file.
- **File Extension - Text**—allows file extensions to be filtered by text.
- **Custodian**—the custodian associated with the discovered file.
- **Data Source**—the data source containing the discovered file.
- **Dedupe Status**—the deduplication status of a file. The Dedupe Status field is active after publishing completes and is useful for determining which files are duplicates and, therefore, not published. The Dedupe Status field is empty for container files, discovered files, and any other files previously processed (such as files published prior to the field being available.) If you republish files, the Dedupe Status field updates with the status of the republished files. Child documents inherit their dedupe status from their parent.
Dedupe Status field values include:
 - **Primary**—a file that has one or more duplicates after publication.
 - **Duplicate**—a duplicate of a primary file.
 - **Unique**—a primary file that does not have any duplicates.
- **Level**—a numeric value indicating how deeply nested the document is within the family. The higher the number, the deeper the document is nested.
- **File Size (KB)**—the size of the discovered file. To specify KB or MB, this field needs to be recreated as a fixed-length text field.
- **Is Published**—the yes/no value indicates if a discovered file is published.
- **Sender Domain**—the domain of the sender of an email.
- **Sort Date**—the date taken from the file's Sent Date, Received Date, and Last Modified Date fields in that order of precedence.
- **Virtual Path - Text**—the complete folder structure and path from the original folder or file chosen for processing to the discovered file.

21.1.3 Deleted Documents view

The Deleted Documents view contains the files you deleted from the Documents tab after the files were published.

Processing Duplication Workflow ▾ Processing Sets Processing Data Sources Processing Profile Password Bank Processing Inventory Reports More ▾								
> Q Replace Deleted files Add Widget All Discovered Fil... > Q								
1 - 3 of 3 100 per page ⏮ ⏪ ⏩ ⏭								
#	☐	File ID	File Name	Custodian	Data Source	Processing Deletion?	Is Published?	Error Message
		= ▾ f	Filter	Filter	Filter	(All) ▾	(All) ▾	Filter
1	☐	1083	332400.MSG		- 1550014	Yes	No	
2	☐	1098	333807.MSG		- 1550014	Yes	No	
3	☐	1121	333334.MSG		- 1550014	Yes	No	
All 3 Export as CSV Total: 3								

This view contains the following fields:

- **File ID**—the number value associated with the discovered file in the database.
- **File Name**—the original name of the discovered file.
- **Custodian**—the custodian associated with the discovered file.
- **Data Source**—the data source containing the discovered file.
- **Processing Deletion?**—the yes/no value indicating if a discovered file or partial family file is deleted.
- **Is Published?**—the yes/no value indicating if a discovered file is published.
- **Error Message**—the message that details the error, cause, and suggested resolution of the error prioritized by the following processing phases:
 - Delete
 - Publish
 - Discover
 - Text Extraction

Files tab (MDS)

21.1.4 Deleted files view

Processing Sets

Processing Data Sources

Processing Profile

Password Bank | Processing

Inventory

Reports

Processing Duplication Workflow

Job Errors

Files

> 🔍

Replace 

Deleted files

Add Widget

No Dashboard Wi...

> 🔍

< List



1

- 4 of 4

100

per page





#	☐	Logical ID	File ID	File Name	Custodian	Data Source	Processing Deletion?	Is Published?	Exception Message
		 Filter	= 	 Filter	 Filter	 Filter	(All)	(All)	 Filter
1	☐	EE31FCB89E33E6C...	209	687ecf5c-5c74-467...			Yes	No	
2	☐	C7D9BE6616EF83D...	211	691e007f-ba67-4fb...			Yes	No	
3	☐	1F9F0EE7107D502...	214	ffecd284-d648-4a6...			Yes	No	
4	☐	C562A6D3F01AA00...	217	03e67c2b-1b09-43...			Yes	No	

All 4

Select... ^

Save Filter

Viewing sample  5000 ^

Total: 4

This view contains the following fields:

- **Logical ID**—the unique identifier of a file in MDS storage.
- **File ID**—the number value associated with the discovered file in the database.
- **File Name**—the original name of the discovered file.

- **Custodian**—the custodian associated with the discovered file.
- **Data Source**—the data source containing the discovered file.
- **Processing Deletion?**—the yes/no value indicating if a discovered file or partial family file is deleted.
- **Is Published?**—the yes/no value indicating if a discovered file is published.
- **Error Message**—the message that details the error, cause, and suggested resolution of the error prioritized by the following processing phases:
 - Delete
 - Publish
 - Discover
 - Text Extraction

21.1.5 Current Files with Exceptions

The Current Files with Exceptions view contains all the documents that yielded exceptions in your workspace that currently have an Exception Status value of Not Resolved. By default, this view does not contain files with an Exception Status of Resolved, as those can be found in the All Files with Exceptions view.

Salt vs Pepper

Processing Files

Processing Duplication Workflow Processing Sets Processing Data Sources Processing Profile Password Bank | Processing Inventory Reports More

> 🔍 Replace Current files with exceptions Add Widget No Dashboard W... > 🗨

< List 1 - 25 of 44 25 per page

#	Details	File ID	Storage ID	File Name	Exception Level	Exception Message	Exception Phase	Exception Category	Exception Status
	= Filter	Filter	Filter	(All)	Filter	(All)	(All)	(All)	(All)
1		107	52	PasswordProtected_0.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
2		116	52	PasswordProtected_1.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
3		117	53	PasswordProtected_2.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
4		212	63	PasswordProtected_0.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
5		218	63	PasswordProtected_1.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
6		220	53	PasswordProtected_2.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
7		124	45	Presentation - Test1234!...	Error	There was an error during expansion of this OpenOffice XML file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved
8		202	45	Presentation - Test1234!	Error	There was an error during expansion of this	Discover	Password Protected File	Not Resolved

All 44 Export as CSV Save filter

- **Details**—the details view of all fields, including compressed metadata, of the discovered file selected.
- **File ID**—the unique identifier assigned to an instance of a file.

- **Storage ID**—the unique identifier assigned to a file.

Note: Differentiating between the File ID and the Storage ID. Every file is assigned a unique Storage ID. Every instance of a file is assigned a unique File ID. For example, you may have a Word document with a unique Storage ID. If that document is attached to two different emails, then each instance of the document attachments is assigned a unique File ID. Therefore, a single file with a Storage ID, may have multiple instances, each with their own unique File ID.

- **File Name**—the original name of the discovered file.
- **Exception Level**—this field displays the type of exception Relativity encountered. There are two levels:
 - Warning—the job continues; however, metadata may be incomplete.
 - Error—the job is blocked from completing. You must address the error and retry the job.
- **Exception Message**—the message that details the exception, cause, and suggested resolution of the exception.
- **Exception Phase**—the phase of processing in which the exception occurred. This field will display any of the following values, as dictated by the phases' precedence. For example, if a file has both Text Extraction and Publish exceptions associated with it, this field will display a value of Publish.
 - Delete
 - Publish
 - Discover
 - Text Extraction
- **Exception Category**—provides insight into the nature of the exceptions that have occurred on your processed files. For details, see [Exception categories and their probability of retry success](#).
- **Exception Status**—the current status of the exception. The Current Files with Exceptions view only displays files with an Exception Status of Not Resolved.
- **File Type**—the file type of the discovered file.
- **File Extension - Text**—allows file extensions to be filtered by text.
- **File Size (KB)**—the size of the discovered file. To specify KB or MB, this field needs to be recreated as a fixed-length text field.
- **Custodian**—the custodian associated with the discovered file.
- **Data Source**—the data source containing the discovered file.
- **Is Published**—the yes/no value indicating if a discovered file is published.

Files tab (MDS)

21.1.6 Current files with exceptions view

Processing Sets Processing Data Sources Processing Profile Password Bank Processing Inventory Reports Processing Duplication Workflow More									
> Replace Current files with exceptions Add Widget No Dashboard Wi... >									
< 1 - 100 of 149 100 per page									
#	Details	Logical ID	File ID	Storage ID	File Name	Exception Level	Exception Message	Exception Phase	
Filter = Filter Filter (All) Filter (All)									
1	☐	24453D8A23AA00E1...	85	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	
2	☐	0CD14CFA4D2BF7D...	10004	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	
3	☐	7C1CACCE49D9339...	10012	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	
4	☐	E85C680018D6ED4D...	24586	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	
5	☐	E132C60F54713FB...	24590	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	
6	☐	E2CCA29F0E517FF6...	37517	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	
All 149 Select... Save Filter Viewing sample 5000 Total: 149									

- **Details**—the details view of all fields, including compressed metadata, of the discovered file selected.
- **Logical ID**—the unique identifier of a file in MDS storage.
- **File ID**—the unique identifier assigned to an instance of a file.
- **Storage ID**—the unique identifier assigned to a file.

Note: Differentiating between the File ID and the Storage ID. Every file is assigned a unique Storage ID. Every instance of a file is assigned a unique File ID. For example, you may have a Word document with a unique Storage ID. If that document is attached to two different emails, then each instance of the document attachments is assigned a unique File ID. Therefore, a single file with a Storage ID, may have multiple instances, each with their own unique File ID.

- **File Name**—the original name of the discovered file.
- **Exception Level**—this field displays the type of exception Relativity encountered. There are two levels:
 - Warning—the job continues; however, metadata may be incomplete.
 - Error—the job is blocked from completing. You must address the error and retry the job.
- **Exception Message**—the message that details the exception, cause, and suggested resolution of the exception.
- **Exception Phase**—the phase of processing in which the exception occurred. This field will display any of the following values, as dictated by the phases' precedence. For example, if a file has both Text Extraction

- Delete
- Publish
- Discover
- Text Extraction

- ### 21.1.7 All Files with Exceptions

Salt vs Pepper

Processing Files

Processing Duplication Workflow

Processing Sets

Processing Data Sources

Processing Profile

Password Bank | Processing

Inventory

Reports

More

>

Q

Replace

Current files with exceptions

Add Widget

No Dashboard Wi...

+

+

+

List

1

- 25 of 44

25

per page

⏪

⏩

⏴

⏵

⏶

⏷

#	☐	Details	File ID	Storage ID	File Name	Exception Level	Exception Message	Exception Phase	Exception Category	Exception Status
			= Filter	Filter	Filter	(All)	Filter	(All)	(All)	(All)
1	☐		107	52	PasswordProtected_0.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
2	☐		116	52	PasswordProtected_1.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
3	☐		117	53	PasswordProtected_2.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
4	☐		212	63	PasswordProtected_0.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
5	☐		218	63	PasswordProtected_1.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
6	☐		220	53	PasswordProtected_2.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
7	☐		124	45	Presentation - Test1234!....	Error	There was an error during expansion of this OpenOffice XML file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved

All 44

Export as CSV

Save filter

Total: 44

This view contains the following fields:

- **Details**—the details view of all fields, including compressed metadata, of the discovered file selected.
- **File ID**—the unique identifier assigned to an instance of a file.
- **Storage ID**—the unique identifier assigned to a file.

Note: Differentiating between the File ID and the Storage ID. Every file is assigned a unique Storage ID. Every instance of a file is assigned a unique File ID. For example, you may have a Word document with a unique Storage ID. If that document is attached to two different emails, then each instance of the document attachments is assigned a unique File ID. Therefore, a single file with a Storage ID, may have multiple instances, each with their own unique File ID.

- **File Name**—the original name of the discovered file.
- **Exception Level**—this field displays the type of exception Relativity encountered. There are two levels:
 - Warning—the job continues; however, metadata may be incomplete.
 - Error—the job is blocked from completing. You must address the error and retry the job.
- **Exception Message**—the message that details the exception, cause, and suggested resolution of the exception.
- **Exception Phase**—the phase of processing in which the exception occurred. This field will display any of the following values, as dictated by the phases' precedence. For example, if a file has both Text Extraction and Publish exceptions associated with it, this field will display a value of Publish.
 - Delete
 - Publish
 - Discover
 - Text Extraction
- **Exception Category**—provides insight into the nature of the exceptions that have occurred on your processed files. For details, see [Exception categories and their probability of retry success](#).
- **Exception Status**—the current status of the exception. The Current Files with Exceptions view only displays files with an Exception Status of Not Resolved.
- **File Type**—the file type of the discovered file.
- **File Extension - Text**—allows file extensions to be filtered by text.
- **File Size (KB)**—the size of the discovered file. To specify KB or MB, this field needs to be recreated as a fixed-length text field.
- **Custodian**—the custodian associated with the discovered file.
- **Data Source**—the data source containing the discovered file.
- **Is Published**—the yes/no value indicating if a discovered file is published.

Files tab (MDS)

21.1.8 All files with exceptions view

Processing Sets Processing Data Sources Processing Profile Password Bank Processing Inventory Reports Processing Duplication Workflow More									
> All files with exceptions Add Widget No Dashboard Wi...									
▼ List 1 - 100 of 149 100 per page									
#	Details	Logical ID	File ID	Storage ID	File Name	Exception Level	Exception Message	Exception Phase	Exce
Filter = Filter Filter (All) Filter (All) (All)									
1		24453D8A23AA0...	85	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Pass Item
2		0CD14CFA4D2BF...	10004	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Pass Cont
3		7C1CACCE49D9...	10012	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Pass Cont
4		E85C680018D6E...	24586	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Pass Cont
5		E132C60F54713...	24590	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Pass Cont
6		E2CCA29F0E517...	37517	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Pass
All 149 Select... Save Filter Viewing sample 5000 Total: 149									

- **Details**—the details view of all fields, including compressed metadata, of the discovered file selected.
- **Logical ID**—the unique identifier of a file in MDS storage.
- **File ID**—the unique identifier assigned to an instance of a file.
- **Storage ID**—the unique identifier assigned to a file.

Note: Differentiating between the File ID and the Storage ID. Every file is assigned a unique Storage ID. Every instance of a file is assigned a unique File ID. For example, you may have a Word document with a unique Storage ID. If that document is attached to two different emails, then each instance of the document attachments is assigned a unique File ID. Therefore, a single file with a Storage ID, may have multiple instances, each with their own unique File ID.

- **File Name**—the original name of the discovered file.
- **Exception Level**—this field displays the type of exception Relativity encountered. There are two levels:
 - Warning—the job continues; however, metadata may be incomplete.
 - Error—the job is blocked from completing. You must address the error and retry the job.
- **Exception Message**—the message that details the exception, cause, and suggested resolution of the exception.
- **Exception Phase**—the phase of processing in which the exception occurred. This field will display any of the following values, as dictated by the phases' precedence. For example, if a file has both Text Extraction and Publish exceptions associated with it, this field will display a value of Publish.

- Delete
 - Publish
 - Discover
 - Text Extraction
- **Exception Category**—provides insight into the nature of the exceptions that have occurred on your processed files. For details, see [Exception categories and their probability of retry success](#).
 - **Exception Status**—the current status of the exception. The Current Files with Exceptions view only displays files with an Exception Status of Not Resolved.
 - **File Type**—the file type of the discovered file.
 - **File Extension - Text**—allows file extensions to be filtered by text.
 - **File Size (KB)**—the size of the discovered file. To specify KB or MB, this field needs to be recreated as a fixed-length text field.
 - **Custodian**—the custodian associated with the discovered file.
 - **Data Source**—the data source containing the discovered file.
 - **Is Published**—the yes/no value indicating if a discovered file is published.

21.2 Details modal

You can view file details by clicking the **Show details for this file** icon. Details include file data, file metadata, and content metadata.

File Data

File Name	Co-Marketing Agreement.zip
File ID	85
Logical ID	24453D8A23AA00E10765E3C5965A9C71BD073FDB5B2F5A52065E2A168AEEE8AB
Custodian	[REDACTED]
Data Source	[REDACTED]
Container Extension	
Container ID	0
Container Name	
Contains Embedded Items?	No
Details	
Exception Message	There was an error during expansion of this compressed container. It may be password prot
File Extension - Text	ZIP
File Size (KB)	2,402
File Type	ZIP compressed archive
Folder Path	[REDACTED] \Loose Files
Is Container	Error

File Metadata

Stored As	\\files.[REDACTED]\T017\Files\EDDS1021749\Processing\1061812\INV1
LastModified	Mon, 23 Nov 2015 18:14:28 UTC
CreatedOn	Mon, 23 Nov 2015 18:14:27 UTC
LastAccessed	
LiteralFileExtension	ZIP

Content Metadata

WasDecrypted	false
OracleFileID	1802
OracleFileType	.ZIP/JAR File
MediaType	application/zip

Click the **Processing Exceptions** tab to view a summary of exceptions, and the exception history.

Exception Summary

Current Exceptions	5
Discover Exception(s)	Password Protected Item (5)
Text Extraction Exception(s)	none
Publish Exception(s)	none
Delete Exception(s)	none

Exception History

Exception total: 5
Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:07:00 UTC
There was an error during expansion of this compressed container. It may be password protec
Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:07:00 UTC
There was an error during expansion of this compressed container. It may be password protec
Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:07:00 UTC
There was an error during expansion of this compressed container. It may be password protec
Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:07:00 UTC
There was an error during expansion of this compressed container. It may be password protec
Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:06:59 UTC
There was an error during expansion of this compressed container. It may be password protec

Close

- The Exception History section represents all exceptions> that have ever occurred on a file. This acts as a timeline of the record's exceptions, showing when they occurred, what they were about, and if any are still active. This includes exceptions resulting from retries of previous exceptions and contains category, phase, date/-time, and message information. All times are kept in UTC format.
- The Exception Summary section displays a count of all active exceptions along with their associated category and phase. This is especially important when investigating exceptions> relating to container files, as there can be many associated to the parent container during file extraction. This helps determine the level of impact the issue has as it may affect many files originating from it.

21.3 Retrying delete exceptions

Navigate to the Deleted Documents view to see a record of all deleted documents. The **Processing Deletion?** field is the yes/no indicator for deleted documents. You can filter by Exception Message to see the exceptions that occurred during deletion. Exceptions can be retried using the **>Retry Exceptions** mass operations option. Once deleted, documents are excluded from further processing operations (for example, deduplication, retry, and republish) and the next duplicate is published as the new primary document, if available. To see a summary of primary documents that

have been replaced, see the Master Document Replacement Summary report in Processing Reports. See Mass Delete for more information on deleting documents.

Processing Duplication Workflow							Processing Sets	Processing Data Sources	Processing Profile	Password Ban
> 🔍 Replace 🔄 Deleted files										
#	☐	File ID	File Name	Custodian	Data Source	Processing Deletion?				
		= ▼ f	Filter	Filter	Filter	(All) ▼				
1	☒	52482	PasswordPr...		- 1550012	No				
2	☒	12078	tahuja1-db ...		- 1550010	No				
3	☒	12036	tahuja1-db ...		- 1550010	No				
4	☐	52481	PWZipWith...		- 1550012	No				
5	☐	2354	MarkTest_...		- 1550010	No				
6	☐	2381			- 1550010	No				
7	☐	11833			- 1550010	No				
8	☐	2494			- 1550010	No				
9	☐	2306			- 1550010	No				
10	☐	6			- 1548436	No				
Checked (3) ▼							Export as CSV	⬆	🔍 Save filter	

21.4 Republishing files from the Files tab

Note: The following information is specific to republishing files at the file level via the Files tab and is distinct from republishing files via the processing set console. For details on republishing via the processing set console, see [Republishing a processing set](#) on the Publishing files topic.

The Republish mass operation provides the ability to publish specific documents on a more granular level compared to the processing set page. For example, you can select specific files and re-publish only that subset. In case only a few members of a family are selected, then this mass operation will automatically republish the whole family of documents.

Republish will overlay all metadata fields mapped at the time you started the operation. If fields are unmapped, Relativity will not remove the data from the field that was already published. The Extracted Text and Native file fields will be overwritten if they are different than the initially published document.

Navigate to the **All Files** view to see a list of discovered files. Use the **Is Published?** field filter to display published files. You can only republish files that have been previously published.

1

- 100 of 5,

Is Published?

Sender Domain -...

Yes

(1)

Filter

Select all - Clear

(Not Set)

Yes

No

Apply

Cancel

Use the **Republish** mass action to republish files at the file level.

Processing Duplication Workflow Processing Sets Processing Data S

> 🔍 [Replace](#) All files ▼

#	☐	Details	File ID	File Name	File Type	File Ext
			= ▼ f	Filter	Filter	Filter
1	☒		52482	PasswordPro...	ZIP compressed archive	ZIP
2	☒		12078	tahuja1-db wi...	Lotus Notes Database	NSF
3	☒		12036	tahuja1-db wi...	Lotus Notes Database	NSF
4	☐		52481	PWZipWithM...	ZIP compressed archive	ZIP
5	☐		2354	MarkTest_Mu...	ZIP compressed archive	ZIP
6	☐		2381	CorruptFile.exe	Win64 Executable Generic	EXE
7	☐			n...	ZIP compressed archive	ZIP
8	☐				RAR Archive	RAR
9	☐			to...	RAR Archive	RAR
10	☐			n ...	Encrypted Office Document	PPTX
11	☐			n ...	Encrypted	PPTX

Export as CSV

Single Download / Replace

Retry Exceptions

Ignore Exceptions

Undo Ignore Exceptions

Republish

Download

Checked (3) ▼ Export as CSV ^ 🔍 Save filter

When you click **Republish**, a confirmation modal appears containing the following information about the job you are about to submit:

- Selected republish count of eligible files
- Total number of files to be republished, including families
- Total number of mapped fields
- Number of documents per batch when importing documents during processing

Note: If you have selected files ineligible for republish, the confirmation message reflects this by stating that there are 0 files to be republished. Ineligible files include files from unpublished processing sets or data sources, containers (for example, PST, ZIP), duplicate files, and files where the Processing Deletion? field status is Yes.

Republish

Republish 2 files and families for a total of 2 files?
Export Batch Size: 4000

Republish

Close

If you have selected eligible files, click **Republish** to proceed or **Cancel** to return to the All Files view.

Note: Once you republish, you cannot cancel this job.

21.4.1 Republishing files where job errors are present

If you attempt to republish files with job errors, you will see a confirmation message indicating you must first resolve the errors before republishing.

To resolve job errors before republishing:

1. Click **Resolve job errors**.

Relativity attempts to fix the job errors and republish the files. For more information on unresolved job errors, see [Processing error resolution](#).

Republish

We found 38 outstanding job errors on 58 data sources.

One or more files attempting to be Republished is associated with an outstanding Job Error.
Republish will automatically resume upon successful resolution of job errors.

Resolve job errors

Cancel

21.4.2 Common use cases for using the Republish mass action

The following are common situations in which you would opt to use the Republish mass action:

- Additional metadata fields were mapped after the initial publish of the processing set/data source completed. For example, you did not map the File Name field during the initial publish, which resulted in no metadata being populated for the documents. Now, you can map the File Name fields, go to the Files tab, filter for that data source, select the returned files and republish them, which will result the File name field getting populated.
- Files that did not get published because of document or job level publish errors.
- Newly discovered files came from a retry discovery operation after the initial publish on the set. After the initial publish completes, you can still retry discovery errors, which could result in more files to be discovered. You can select unpublished files and republish only that subset.

21.5 Retrying exceptions

Navigate to the All Files view to see a record of all discovered files, including files with exceptions found during Discovery via the **Discover Error** field. The files that have ready to retry exceptions can be retried at file level using the **Retry Exceptions** mass operations option. When you click **Retry Exceptions**, you see a confirmation modal confirming the number of files you are about to submit for retry.

Retry

Retry 2 files?

Only selected files with an error status of **Ready to retry** will be retried.

Retry

Cancel

Click **Retry** on the confirmation modal.

Processing Duplication Workflow Processing Sets Processing Data S

> 🔍 Replace  All files ▾

#	☐	Details	File ID	File Name	File Type	File Ext
			= ▾ f	Filter	Filter	Filter
1	☒		52482	PasswordPro...	ZIP compressed archive	ZIP
2	☒		12078	tahuja1-db wi...	Lotus Notes Database	NSF
3	☒		12036	tahuja1-db wi...	Lotus Notes Database	NSF
4	☐		52481	PWZipWithM...	ZIP compressed archive	ZIP
5	☐		2354	MarkTest_Mu...	ZIP compressed archive	ZIP
6	☐		2381	CorruptFile.exe	Win64 Executable Generic	EXE
7	☐			n...	ZIP compressed archive	ZIP
8	☐				RAR Archive	RAR
9	☐			o...	RAR Archive	RAR
10	☐			n ...	Encrypted Office Document	PPTX
11	☐			n ...	Encrypted	PPTX

Export as CSV

Single Download / Replace

Retry Exceptions

Ignore Exceptions

Undo Ignore Exceptions

Republish

Download

Checked (3) ▾ Export as CSV ^ 🔍 Save filter

Navigate to the **All Files with Exceptions** view to see a record of all files with exceptions.

Salt vs Pepper Processing Files ☆ Search (CTRL + /) ☆

Processing Duplication Workflow Processing Sets Processing Data Sources Processing Profile Password Bank | Processing Inventory Reports More

> Q Replace **Current files with exceptions** Add Widget No Dashboard WI... > >

▼ List 1 - 25 of 44 25 per page

#	Details	File ID	Storage ID	File Name	Exception Level	Exception Message	Exception Phase	Exception Category	Exception Status
1		107	52	PasswordProtected_0.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
2		116	52	PasswordProtected_1.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
3		117	53	PasswordProtected_2.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
4		212	63	PasswordProtected_0.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
5		218	63	PasswordProtected_1.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
6		220	53	PasswordProtected_2.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
7		124	45	Presentation - Test1234!	Error	There was an error during expansion of this OpenOffice XML file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved
8		202	45	Presentation - Test1234!	Error	There was an error during expansion of this OpenOffice XML file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved

All 44 Export as CSV Save filter

21.6 Saved filters

You can save any filters you set on any of the views in the Files tab and reuse them in future workflows. To do this:

1. Filter on any of the fields in the view and click **Save Filters** at the bottom of the view.

All Files Add Widget No Dashboard S... Export

1 - 9 of 9 100 per page

#	Details	File ID	File Name	File Type	File Extension - Text	Custodian
1		10	Sample Lotus_ALT.nsf	Lotus Notes Database	NSF	Relativity ODA, LLC
2		16	Sample Lotus.nsf	Lotus Notes Database	NSF	Relativity ODA, LLC
3		17	EE13C20FBDBE6ED007...	Lotus Notes Message File	MSG	Relativity ODA, LLC
4		18	385E4A5955F4D66407...	Lotus Notes Message File	MSG	Relativity ODA, LLC
5		19	EE13C20FBDBE6ED007...	Lotus Notes Message File	MSG	Relativity ODA, LLC
6		20	385E4A5955F4D66407...	Lotus Notes Message File	MSG	Relativity ODA, LLC
7		32	Sample Lotus.nsf	Lotus Notes Database	NSF	Relativity ODA, LLC
8		39	EE13C20FBDBE6ED007...	Lotus Notes Message File	MSG	Relativity ODA, LLC
9		40	385E4A5955F4D66407...	Lotus Notes Message File	MSG	Relativity ODA, LLC

All 9 Export as CSV Save filter Total: 9

2. In the Saved Filter modal, complete the Name, Notes (optional), and Conditions fields, and click **Save**.

Saved Filter

Information *

Name *

Saved Filter 1

Notes

Conditions

+ Condition

▼

+ No Related Items ▼

1

1. File Type

is like Lotus Notes

Clear All Conditions

Save

Cancel

- **Name**—the name you want these saved filters to appear under in the saved filters browser.
- **Notes**—any notes you want to enter to clarify the purpose of these saved filters.
- **Conditions**—displays the filters applied to the list. To add additional condition, click **+ Condition**. The Add Condition - Select Field modal appears, where you can select addition fields and apply conditions.

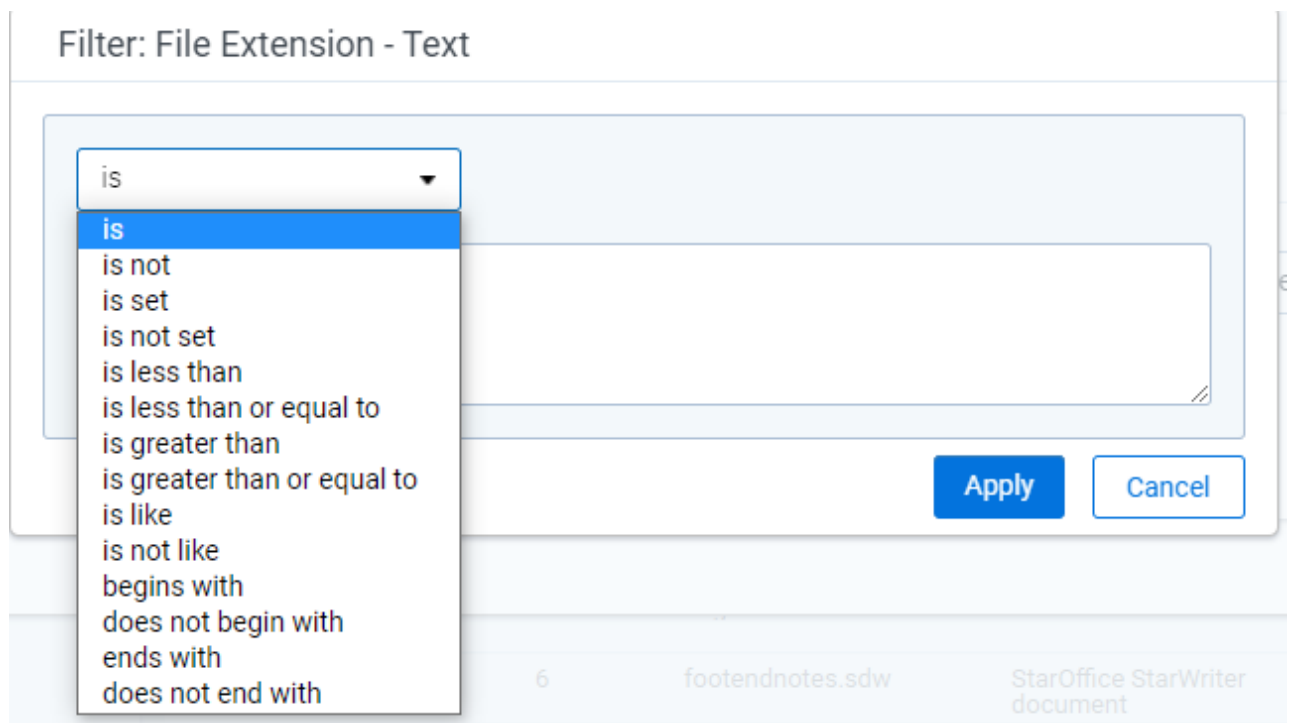
Add Condition - Select Field



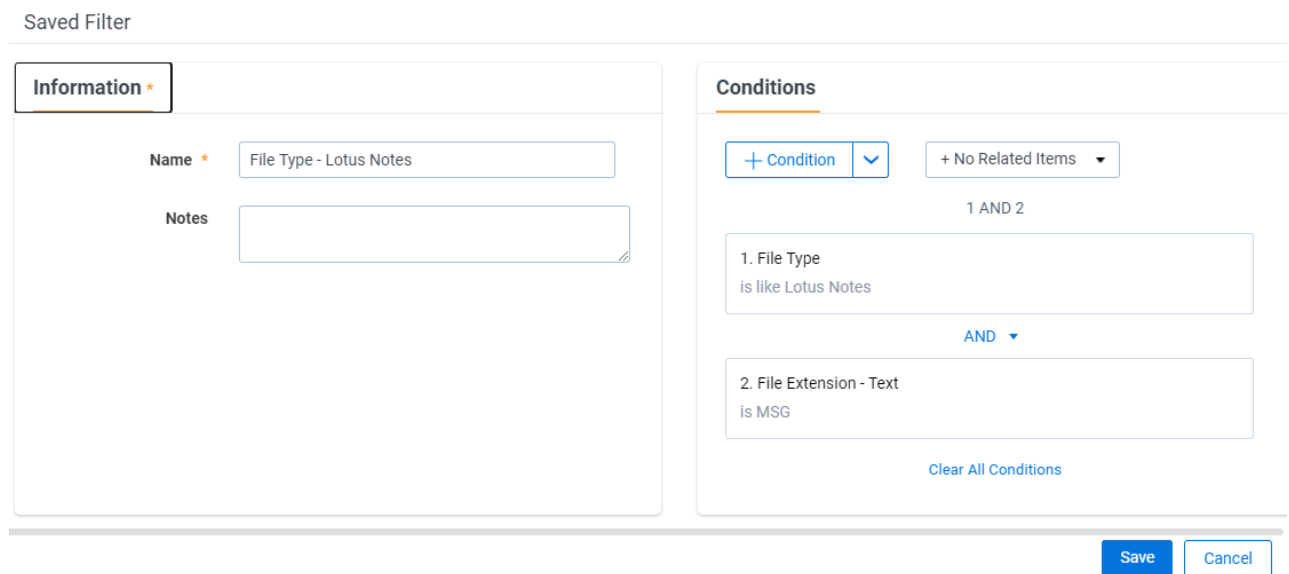
Container Extension	
Container ID	
Container Name	
Contains Embedded Items?	
Custodian	
Data Source	
Dedupe Status	
Details	
Discovery Group ID	
Error Category	

Cancel

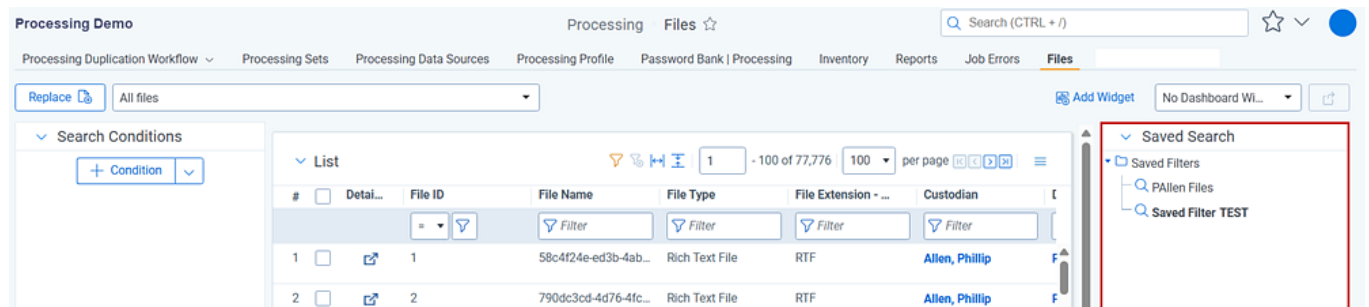
Once you select the additional fields you want to add to the saved filters set, specify the conditions you want to apply to those fields and click **Apply**.



The new field(s) and condition(s) are visible in the Saved Filter modal. Click **Save** to save the filter set.



3. The saved filter is added to the list of Saved Filters.



21.7 Saved filters (MDS)

You can save any filters you set on any of the views in the Files tab and reuse them in future workflows. To do this:

1. **Filter** the list using the filter boxes at the top of each column.

▼ List 1 - 100 of 5,000 100 per page

#	Details	Logical ID	File ID	File Name	File Type	File Extension - Text	Custodian	Data
		Filter	=	Filter	Filter	MSG		
1	☐	92A82776DCFD7C5...	3	PALLEN_0013633.msg	Unknown	MSG	Allen, Phillip	Philli
2	☐	D202AC7E15E5ECD...	7	Changes to KiodeX co-marketing agreem...	Outlook Message File	MSG	Allen, Phillip	Philli
3	☐	13C898880EBF174...	61	RE Updated policy.msg	Outlook Message File	MSG	Allen, Phillip	Philli
4	☐	40302B75DB6363E...	2256	00000000A03146A8A1BEE243851D8EF...	Outlook Message File	MSG	Beck, Sally	Beck 1069
5	☐	AD14B6A0A56F868...	2257	00000000A03146A8A1BEE243851D8EF...	Outlook Message File	MSG	Beck, Sally	Beck 1069
6	☐	3283ABE763D5D24...	2258	00000000A03146A8A1BEE243851D8EF...	Outlook Message File	MSG	Beck, Sally	Beck 1069
7	☐	C5F885815FBE37C...	2259	00000000A03146A8A1BEE243851D8EF...	Outlook Message File	MSG	Beck, Sally	Beck 1069

All 22,638

Viewing sample 5000 Total: 22,638

2. Click **Save Filter** at the bottom of the screen.
3. In the Saved Filter modal, complete the **Name** and (optional) **Notes** fields, and click **Save**.

Saved Filter

Information

Name*

Notes

Conditions

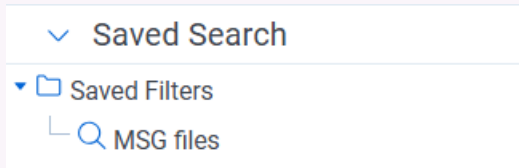
Conditions cannot be added to a Saved Filter from this modal. To add or change conditions, close the modal and use the conditions console. Once done, proceed with the saved filter.

File Extension - Text

is like "MSG"

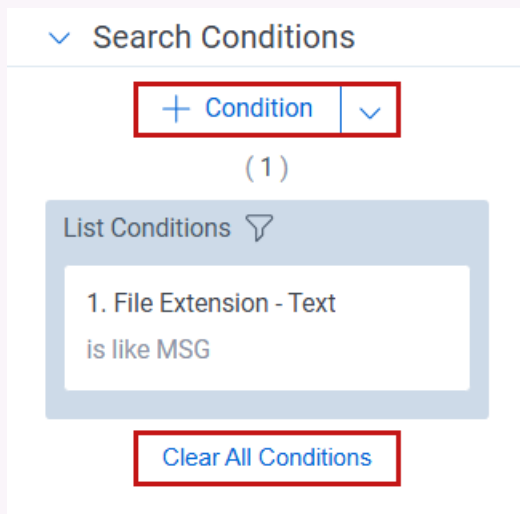
- **Name**—the filter name that is displayed in the Saved filters browser.
- **Notes**—any notes you want to enter to clarify the purpose of these saved filters.

4. The saved filter is added to the **Saved Search > Saved Filters** browser.



Notes:

- You cannot edit conditions in the Search Conditions browser. To edit a saved filter condition, clear the condition in the Search Conditions browser, then add a new condition, or create a new saved filter with the updated conditions.



- To create a new filter from an existing filter, right-click the filter name, then select Copy. The copied condition appears in the Saved Filters browser.
 - To edit a filter, right-click the filter name, then select Edit.
 - To delete a filter, right-click the filter name, then select Delete.
 - To stop filtering the list, click **Clear All Conditions** in the Search Conditions browser, then click **Run Search**. To reapply the filter, click the filter name in the Saved Filters browser. You can
-

enable the **Auto Run** option to run automatically commands whenever you make a change.

Search Conditions

+ Condition

▼

(1)

List Conditions

▼

1. File Extension - Text
is like MSG

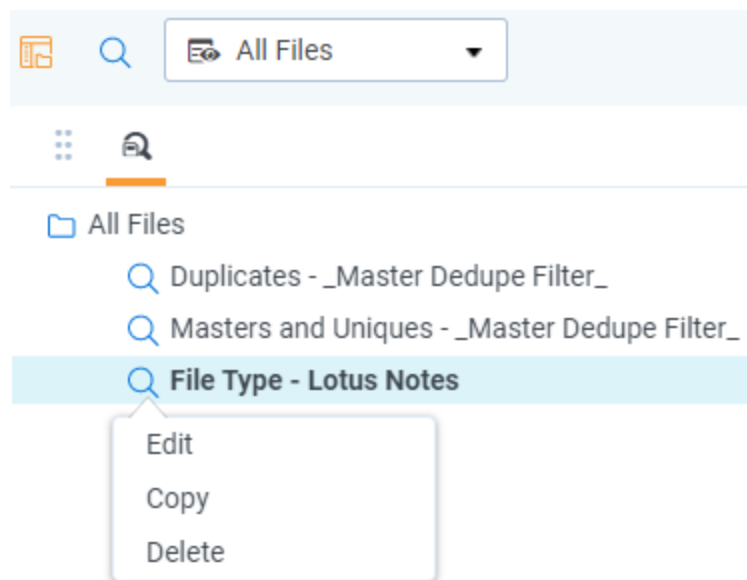
Clear All Conditions

Run Search

Auto Run ☒

21.7.1 Right-click options

You can right-click any saved filter and choose to Edit, Copy, or Delete it.



21.7.1.1 Edit

Clicking **Edit** opens the Saved Filter modal. You can add, remove, or modify fields and conditions, then click Save.

#	Details	File ID	File Name	File Type	File Extension ...	Custodian	Data Source
1		10	Sample Lotus_ALT.nsf	Lotus Notes Database	NSF	Relativity ODA, LLC	Relativity ODA, 1058393
2		16	Sample Lotus.nsf	Lotus Notes Database	NSF	Relativity ODA, LLC	Relativity ODA, 1058393
3		17	EE13C20FBDDBE6ED007...	Lotus Notes Message File	MSG	Relativity ODA, LLC	Relativity ODA, 1058393
4		18	385E4A5955F4D66407...	Lotus Notes Message File	MSG	Relativity ODA, LLC	Relativity ODA, 1058393
5		19	EE13C20FBDDBE6ED007...	Lotus Notes Message File	MSG	Relativity ODA, LLC	Relativity ODA, 1058393
6		20	385E4A5955F4D66407...	Lotus Notes Message File	MSG	Relativity ODA, LLC	Relativity ODA, 1058393
7		32	Sample Lotus.nsf	Lotus Notes Database	NSF	Relativity ODA, LLC	Relativity ODA, 1058439
8		39	EE13C20FBDDBE6ED007...	Lotus Notes Message File	MSG	Relativity ODA, LLC	Relativity ODA, 1058439

21.7.1.2 Copy

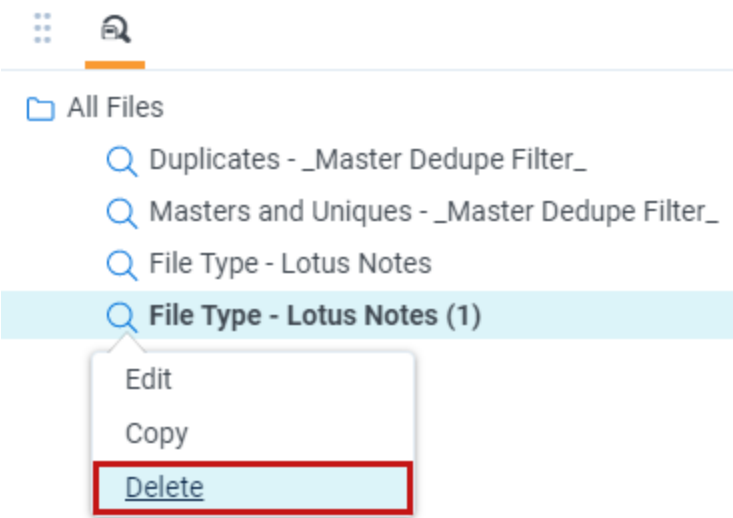
To make a copy of an existing filter set, right-click and select **Copy**.

#	Details	File ID	File Name	File Type	File Extension - Text	Cu
1		17	EE13C20FBDDBE6ED007...	Lotus Notes Message File	MSG	Re
2		18	385E4A5955F4D66407...	Lotus Notes Message File	MSG	Re
3		19	EE13C20FBDDBE6ED007...	Lotus Notes Message File	MSG	Re
4		20	385E4A5955F4D66407...	Lotus Notes Message File	MSG	Re
5		39	EE13C20FBDDBE6ED007...	Lotus Notes Message File	MSG	Re
6		40	385E4A5955F4D66407...	Lotus Notes Message File	MSG	Re

Use the Copy option when you need a new filter with similar fields and conditions. When you copy a filter set and save it, the new set name has a (1) appended to the file name. To rename the filter set, edit the filter, then change the name on the Saved Filter modal.

21.7.1.3 Delete

To remove a filter set from the saved filter browser entirely, right-click and select **Delete** .



21.8 Single Download / Replace

Single Download / Replace provides the ability to download a file to your local machine for investigation. It will also provide the ability to replace and retry an original file with a new version that has been fixed during error remediation. You can only replace and retry files with errors of a status of Not Resolved.

To perform the Single Download / Replace mass action, perform the following steps:

- 1. To take action on a specific file, select it, and then select the **Single Download / Replace** mass action.

Note: If you select multiple files, you will see an error as this action is meant for one file at a time. Similarly, you will see an error message if you do not select a file.

7	☐		7	Change
8	☐	Export as CSV Single Download / Replace Retry Exceptions Ignore Exceptions Undo Ignore Exceptions Republish Download		790faa
9	☐			794e79
10	☐			PALLEP
11	☐			58c246
12	☐			58b8d6
13	☐			PALLEP
14	☐			PALLEP
All 4,455		Export as CSV		Save filter

The Single Download / Replace dialog box opens allowing you to browse for or drop a replacement file.

2. Once you select or browse and drop the replacement file, click **Download**.

Single Download / Replace

Existing File

 Co-Marketing Agreement.zip 2,402 (KB)

Replacement File

Drop your files here or [browse for files](#).

Download

Cancel

3. Once the file downloads, resolve the error, and drag the resolved file back into the Single Download / Replace modal. A message displays that the file is ready for replace and retry. If the file extensions do not match and/or if the new file is larger than the original, you can still proceed with the replace and retry action.
4. Click **Replace & Retry**.

Single Download / Replace

Existing File

 1456_excel2007pwd.xlsx 12 (KB)

Replacement File

 Your file has been uploaded and is ready to be replaced and retried. 



- The file extensions do not match.
- The file you are attempting to upload is at least 10% larger than the existing file.



 For attorneys.pptx 276 (KB) 

Replace & Retry

Cancel

To see if the action was successful, you can check the Error History by going into the Details modal. If the retry was successful, the error displays a Resolved status.

21.9 Mass download and replacement of errored files

The Download mass action works in conjunction with the Upload Replacement(s) control to move errored files from Relativity to your local drive and then back again after error remediation. After selecting files with an exception status of Not Resolved, the Download mass action bundles the files in a .zip container and then saves the container to your local drive.

Relativity renames each file using the Storage ID metadata from the original file. The file extension remains intact. The Storage ID ensures that the downloaded file names are unique.

The Upload Replacement(s) control moves the remediated files from your local drive and replaces the original errored files. Relativity automatically retries the errors as part of the upload process.

During the upload process, Relativity matches the remediated file's name (now the Storage ID) with the Storage ID metadata from the original file. After replacement, the original file name and extension remain intact.

Note: Differentiating between the File ID and the Storage ID. Every file is assigned a unique Storage ID. Every instance of a file is assigned a unique File ID. For example, you may have a Word document with a unique Storage ID. If that document is attached to two different emails, then each instance of the document attachments is assigned a unique File ID. Therefore, a single file with a Storage ID, may have multiple instances, each with their own unique File ID.

21.9.1 Mass download of errored files

Use the steps below to download errored files to your local drive. You can only use the Download mass action on files that have an exception status of Not Resolved.

To mass download files:

1. Use the search bar to navigate to the **Files** tab.
2. Select **Current Files with Exceptions** from the Views drop-down menu.
3. Select individual files to include in the download package or use the **Select All** option to mark all files for downloading.

Note: Files must be in the Discover phase for mass download. You will see an error message if you attempt to download files in other phases, such as Publish or Text Extraction.

▼ List						
#	☒	Details	File ID	Storage ID	File Name	Error Message
1	☒		85	61	Co-Marketing Agreement....	There was a expansion of compressed may be pas protected. C passwords Bank.
2	☒		229	200	GM SG-220.pdf	Error in line The input va author sour length of 29

4. Select **Download** from the mass action drop-down menu.

▼ List 🔍 🔗

#	☐	Details	File ID	Storage ID	File Name	Error Message
1	☒	🔗	85	61	Co-Marketing Agreement...	There was an error expansion of this compressed content may be password protected. Consider passwords to Password Bank.
2	☐	🔗	229	200	GM SG-220.pdf	Error in line 23, column 1 The input value for author source field length of 296 characters This exceeds the the author destination field, which is currently 255 characters
3	☒	🔗	61	45	RE Updated policy.msg	There was an error expansion of this root cause is currently unknown. Consider retrying.
4	☐	🔗			08817 Pyrostop.doc	Error in line 2, column 1 The input value for author source field length of 335 characters This exceeds the the author destination field, which is currently 255 characters
5	☒	🔗			PALLEN_0010005.pdf	There was an error expansion of this It may be password protected. Consider passwords to Password Bank.

Export as CSV
Single Download / Replace
Retry Exceptions
Ignore Exceptions
Undo Ignore Exceptions
Republish
Download

Checked (4) ▼
Export as CSV ⬆ 🔍 Save filter

A confirmation message displays the number of files you are downloading.

Note: Relativity downloads unique files from each selection. The number of files selected may differ from the number of files downloaded if duplicates exist.

Download

You have selected 4 record(s) to be downloaded, 2 unique file(s) were identified. Only unique files will be downloaded.

Download

Cancel

5. Click **Download**.

6. A status message displays the overall progress.

Download

 Mass Download completed successfully.

You have selected 4 record(s) to be downloaded, 2 unique file(s) were identified. Only unique files will be downloaded.

DownloadClose

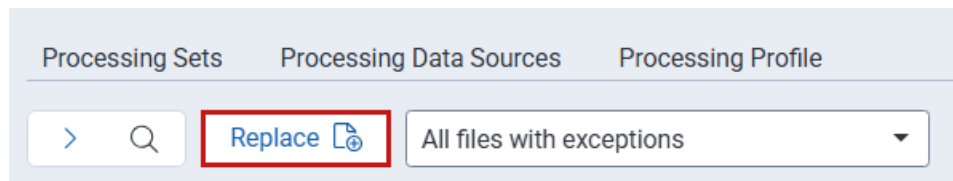
When complete, files are packaged in a .zip container and saved to your local drive.

Note: During the download process, Relativity changes the file name of downloaded files to the Storage ID. The file extension remains the same as the original. This results in each file having a unique file name, which is used later when replacing the original files on Relativity.

21.9.2 Mass replacement of errored files

Use the steps below to upload remediated files from your local drive back into Relativity. You must place the files in a .zip container before uploading. There is a 1 GB file size limit for each .zip container.

1. Use the search bar to navigate to the **Files** tab.
2. Click the **Replace** button on the **Files** tab.



Note: While the Replace action is a mass action, meaning multiple files are affected, it is not a part of the mass actions menu (at the bottom of the screen). Action items in the mass actions menu require selecting target files from the Files list. Since you are moving files from your local drive back into Relativity, there are no files from the list to preselect.

3. The Replace modal opens. Either drag and drop or browse for the .zip container holding your replacement files.

Replace

Please ensure that files are properly named after storage ID and the ZIP is up to 1 GB.



Drop your files here or [browse for files](#).



4. Click **Upload**.

When you upload the replacement file, Relativity performs the following checks:

- The replacement files are contained in a .zip file, which has a file size limit of 1 GB. If you exceed the file size limit, you will see an error. To resolve this error, redistribute the replacement files so the .zip

container is within the file size limit.

- The replacement file's file name matches the original file's Storage ID.
- The replacement file size is within 10% of the original (larger or smaller) file size.
- The replacement file extension matches the original file extension.
- The original file has an exception status of Not Resolved.

5. The Replace modal updates with the number of successful files, successful files with warnings, and unsuccessful files.

Replace

 Your files have been uploaded and are ready to be replaced and retried. 

Successful 

Successful with warning 

Unsuccessful 

2 files

0 files

0 files

Proceeding with Replace & Retry will include both Successful and Successful with Warning files. Warnings indicate inconsistencies with file extension and/or in file size, which is to be expected. Unsuccessful files are due to lack of a match or a match does not have an outstanding discovery exception.

For more information, use the Replace log button to download a full list of replacement files and matches.

Replace & Retry

Replace log

Close

- **Successful**—the number of files successfully replaced.
- **Successful with warning**—files where the candidate file name minus extension has a Storage ID match in the Files tab and has an Exception Status of Not Resolved. In addition, these files may have a different extension or exceed the 10% deviation in file size.
- **Unsuccessful**—files where the candidate has a Files tab match but no outstanding exception, the candidate does not have a Files tab match, or Relativity could not extract the file from the .zip container.

You can view a log of the files, along with their status, by clicking the **Replace log** link on the modal. The log displays the following information:

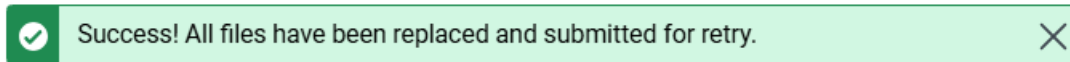
- Replacement file name
- Candidate file extension
- Candidate file size
- Files tab Storage ID
- File tab file extension
- File tab file size

- Exception (error) status
- Validation status
- Reason for warning or failure.

6. Click **Replace & Retry**.

7. A final confirmation message indicates the completion of the replacement process. Click **Close**.

Replace



To see the replaced files, navigate to the All Errors View or All Files View in the Files tab. To check progress of retry, refer to error status field or the associated processing sets.

Close

8. The last step is to **republish** your files so that Relativity updates metadata and extracted text fields.

22 Publishing files

Publishing files to a workspace is the step that loads processed data into the environment so reviewers can access the files. At any point after file discovery is complete, you can publish the discovered files to a workspace. During publish, Relativity:

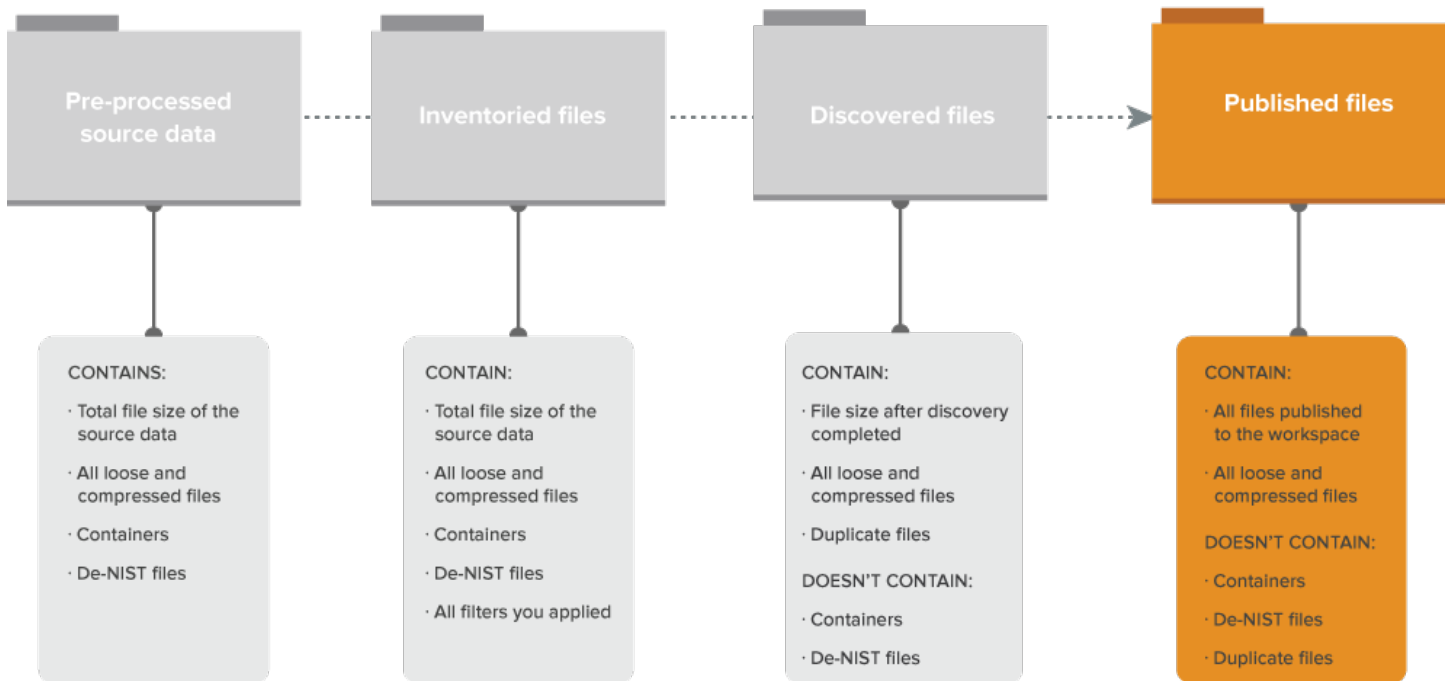
- Applies all the settings you specified on the profile to the documents you bring into the workspace.
- Determines which is the primary document and primary custodian and which are the duplicates.
- Populates the All Custodians, Other Sources, and other fields with data.

Note: For details on deleting files after publishing, see [Post-publish delete](#).

Publish includes the three distinct steps of deduplication document ID creation, primary document publish, and overlaying deduplication metadata. Because of this, it's possible for multiple processing sets to be publishing at the same time in the same workspace.

The following graphic depicts how publish fits into the basic workflow you use to reduce the file size of a data set through processing. This workflow assumes that you are applying some method of de-NIST and deduplication.

File Size Reduction through Processing



The following is a typical workflow that incorporates publish:

1. Create a processing set or select an existing set.
2. Add data sources to the processing set.
3. Inventory the files in that processing set to extract top-level metadata.
4. Apply filters to the inventoried data.
5. Run discovery on the refined data.
6. Publish the discovered files to the workspace.

22.1 Running file publish

To publish files, click **Publish Files**. You only need to manually start publish if you disabled the Auto-publish set field on the profile used by this processing set.

Note: When processing documents without an actual date, Relativity provides a null value for the following fields: Created Date, Created Date/Time, Created Time, Last Accessed Date, Last Accessed Date/Time, Last Accessed Time, Last Modified Date, Last Modified Date/Time, Last Modified Time, Primary Date/Time, and Sort Date/Time. The null value is excluded and not represented in the filtered list.

Processing Sets Processing Data Sources Processing Profile Password Bank | Pro

Edit Delete Back 1 of 24 ⏪ ⏩ ⏴ ⏵

Processing Set 005 Basic Settings

Inventory (Optional)
Completed

Discover
Completed

Publish
Not Started

- > Inventory (Optional)
- > Discovery
- > Publish

Processing Data Source Job Errors

Inventory Files

Filter Files

Discover Files

Retry File Exceptions

Publish Files

Reports

Inventory Report

Discovery Report

All Reports

Exceptions

Job Errors

File Exceptions

When you click Publish Files, you are presented with a confirmation message containing information about the job you are about to submit. If you have not mapped any fields in the workspace, the message reflects this. Click **Publish** to proceed or **Cancel** to return to the processing set layout.

Continue?

Publish 1,116 files?

Once you publish files:

- You won't be able to change the deduplication method
- You will be unable to cancel a processing set until any subsequent dependent processing sets are also canceled in this workspace.
- Any unmapped fields will not have metadata populated

Deduplication method: Global
Export Batch Size: 4000

Yes

Cancel

Consider the following when publishing files:

- During publish, Relativity assigns control numbers to documents. Duplicates do not receive unique control numbers.

Note: You cannot control the order in which Relativity assigns control numbers to files. The sorted order of files in the originating file system (generally, alphabetically) does not influence the order in which files are published, folders created, and control numbers assigned. Therefore, you may see sequential gaps in control numbers within folders.

- The publish process includes the three distinct steps of deduplication document ID creation, primary document publish, and overlaying deduplication metadata; as a result, it's possible for multiple processing sets to be publishing at the same time in the same workspace.
- After data is published, we recommend that you not change the Control Number (Document Identifier) value, as issues can arise in future publish jobs if a data overlay occurs on the modified files.
- If you have multiple data sources attached to a single processing set, Relativity starts the second source as soon as the first set reaches the DeDuplication and Document ID generation stage. Previously, Relativity waited until the entire source was published before starting the next one.
- The Publish option is available even after publish is complete. This means you can republish data sources that have been previously published with or without errors.
- If you have configured auto-publishing on the processing profile, then when you start discovery, you are also starting publish once discovery is complete, even if errors occur during discovery. This means that the Publish button is never enabled.
- Once you publish files, you are unable to delete or edit the data sources containing those files. You are also unable to change the deduplication method you originally applied to the set.
- When you delete a document, Relativity automatically recalculates deduplication and publishes a new document to replace the deleted one, if applicable.
- If you arrange to copy source files to the Relativity file share, Relativity no longer needs to access them once you publish them. In this case, you are not required to keep your source files in the location from which they were processed after you published them.
- Publish is a distributed process that is broken up into separate jobs, which leads to more stability by removing this single point of failure and allowing the distribution of work across multiple workers. These changes enable publish to operate more consistently like the other processing job types, where batches of data are processed for a specific amount of time before completing each transactional job and moving on. Note the upgrade-relevant details regarding distributed publish:
 - UpdateMastersWithDedupeInformation- the third phase of publish that finishes before metadata updates if no deduplication fields are mapped.
 - The deduplication fields are All Custodians, Deduped Custodians, All Paths/Locations, Deduped Count, and Deduped Paths.
 - If no deduplication fields are mapped for a publish job where the deduplication method is either Global or Custodial, then the UpdateMastersWithDedupeInformation job should finish before overlaying or updating any metadata.
 - The tracking log reads "Overlaying dedupe information will not be performed on the primary documents. The deduplication fields are not mapped."
 - The following instance settings have been added to facilitate the work of distributed publish. Due to the change in publish behavior caused by these new instance settings, we recommend contacting [Relativity support](#) for guidance on what values to specify for these settings before performing an upgrade.
 - ProcessingMaxPublishJobCountPerRelativitySQLServerProcessingMaxPublishJobCountPerRelativitySQLServer - the maximum number of publish jobs per

Relativity SQL server that may be worked on in parallel.

- The default value is 21. Leaving this setting at its default value results in increased throughput.
 - This updates on a 30-second interval.
 - If you change the default value, note that setting it too high could result in web server, SQL server, or BCP/file server issues. In addition, other jobs in Relativity that use worker threads may see a performance decrease, such discovery or imaging. If you set it too low, publish speeds may be lower than expected.
 - You cannot allocate more jobs per workspace than what is allowed per SQL server.
- **ProcessingMaxPublishSubJobCountPerWorkspacePro-**
cessingMaxPublishSubJobCountPerWorkspace - the maximum number of publish jobs per workspace that may be worked on in parallel.
 - The default value is 7. Leaving this setting at its default value results in increased throughput.
 - This updates on a 30-second interval.
 - If you change the default value, note that setting it too high could result in web server, SQL server, or BCP/file server issues. In addition, other jobs in Relativity that use worker threads may see a performance decrease, such discovery or imaging. If you set it too low, publish speeds may be lower than expected.
 - You cannot allocate more jobs per workspace than what is allowed per SQL server.

The following table provides the recommended values for each instance setting per environment setup:

Envir- onment setup	Pro- cess- ingMaxPub- lishSubJobCountPerWorkspace	Pro- cess- ingMaxPub- lishJobCountPerRelativitySQLServer
Tier 1Tier 1 (see the Sys- tem Require- ments Guide for details)	3	7
Tier 2Tier 2 (see the Sys- tem Require- ments Guide for details)	6	12
Relativity aiR baseline	3	7

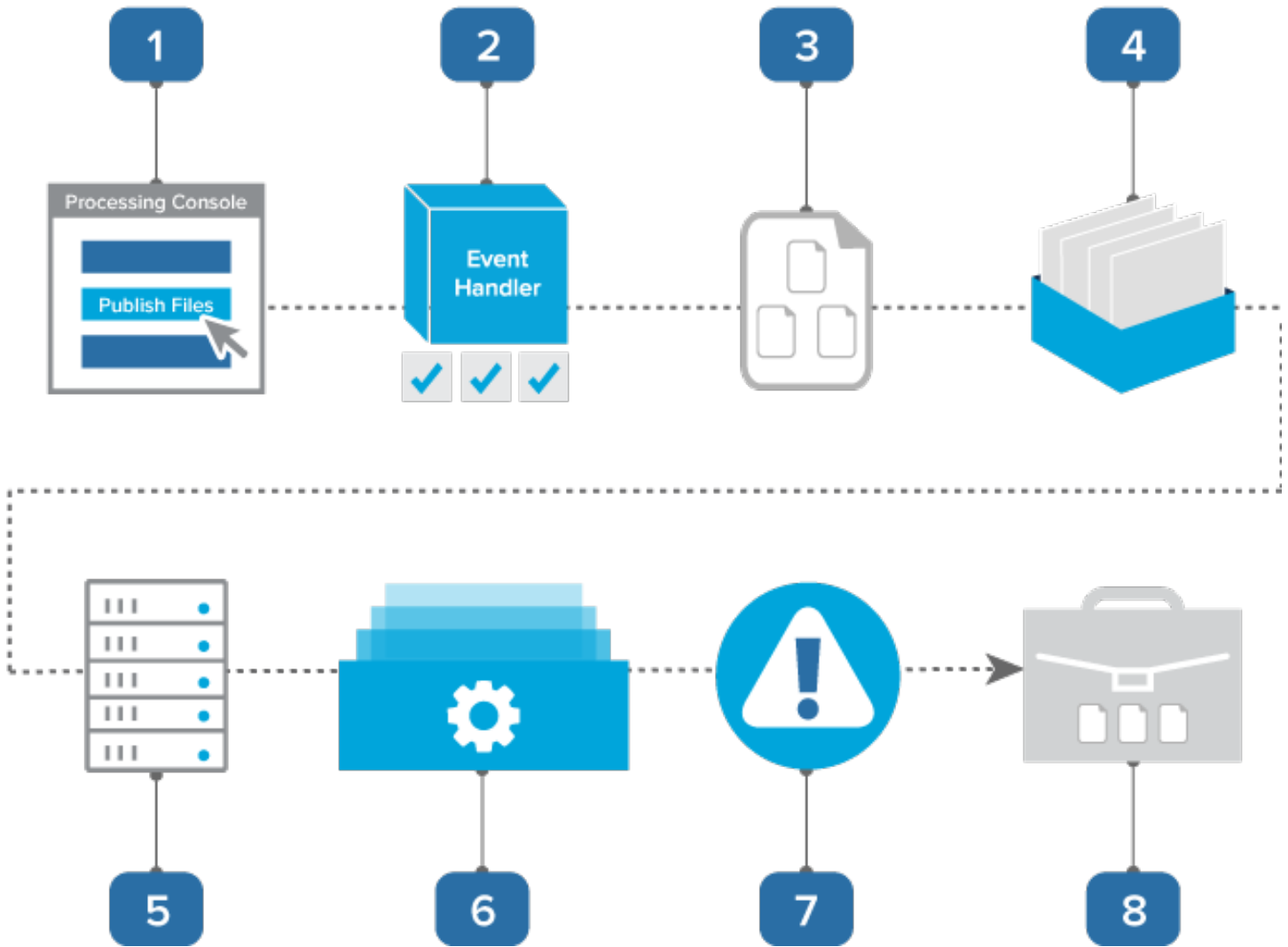
Note: Once you publish data into Relativity, you have the option of exporting it with Import/Export.

When you start publish, the Publish Files button changes to Cancel. You can use this to cancel the processing set. For more information, see [Canceling publish](#).

22.1.1 Publish process

The following graphic and corresponding steps depict what happens behind the scenes when you start publish. This information is meant for reference purposes only.

Publish Process



1. You click Publish Files on the processing set console. If you have set auto-publish after discovery, publish begins automatically and you are not required to start it manually.
2. A console event handler checks to make sure that the set is valid and ready to proceed.
3. The event handler inserts all data sources on the processing set into the processing set queue.
4. The data sources wait in the queue to be picked up by an agent, during which time you can change their priority.
5. The processing set manager picks up each data source based on its order, all password bank entries are synced, and the agent submits each data source as an individual publish job to the processing engine. The agent then provides updates on the status of each job to Relativity, which then displays this information on the processing set layout.

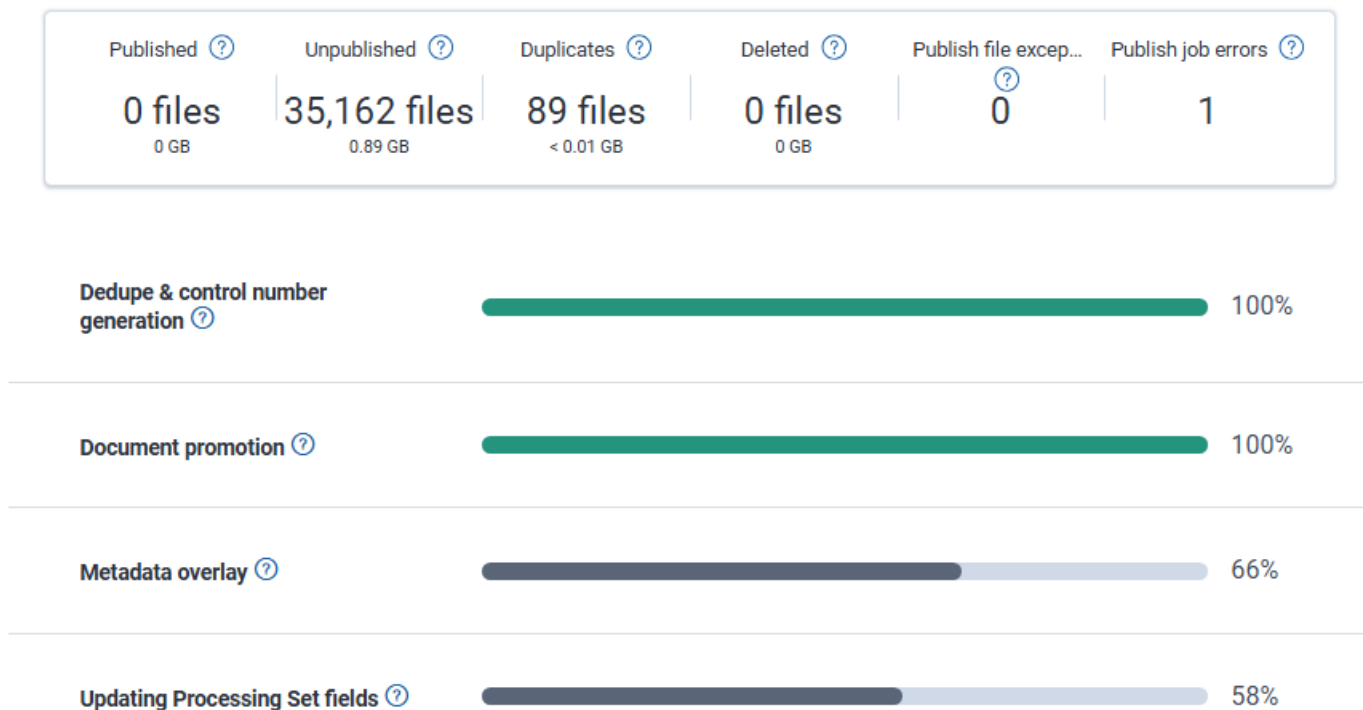
- The processing engine publishes the files to the workspace. Relativity updates the reports to include all applicable publish data. You can generate these reports to see how many and what kind of files you published to your workspace.
-
- Note:** Publish is a distributed process that is broken up into separate jobs, which leads to more stability by removing this single point of failure and improves performance by allowing the distribution of work across multiple workers. Thus, publish is consistent with the other types of processing jobs, in that it operates on batches of data for a specific amount of time before completing each transactional job and moving on.
-
- Any errors that occurred during publish are logged in the errors tabs. You can view these errors and attempt to retry them. See [Processing error resolution](#) for details.
 - You set up a review project on the documents you published to your workspace, during which you can search across them and eventually produce them.

22.2 Monitoring publish status

You can monitor the job's progress on the Processing Set Details page. This page breaks down each of the processing phases (Inventory, Discover, Publish) into their own sections that include dashboard summaries, sub-job details, and counts.

22.2.1 Publish job details

▼ Publish



Dashboard numbers

- **Published**—the number of files promoted to the Documents tab.
- **Unpublished**—the number of files that have either errored or not gone through the Publish phase yet. Some files may be unpublished if they had discover errors which were later resolved, with the files now waiting to be published.
- **Duplicates**—the number of files identified as duplicates. These include duplicate parent and child loose documents.
- **Deleted**—the number of files deleted from the Document tab after Publish completes.
- **Publish file exceptions**—the number of files with publish exceptions. These issues result in unpublished files and should be remediated.
- **Publish job errors**—the number of publish job errors. These issues can result in large groups of unpublished files and should be remediated.

Progress bars

- **Dedupe & control number generations**—Relativity identifies primary and duplicate records by comparing hash values against all published documents in the workspace. The resulting primary records receive a generated control number.
- **Document promotion**—Relativity imports primary documents and their families to the Documents tab.
- **Metadata overlay**—Relativity updates previously published records with information from duplicate documents.
- **Updating Processing Set fields**—Relativity updates associated fields in the processing set with results from the job.

See [Processing error overview](#) for details.

Once publish is complete, the status section displays a blue check mark and you have the option of republishing your files, if need be. For details, see [Republishing files](#).

22.3 Canceling publishing

If the need arises, you can cancel your publish job before it completes.

To cancel publish, click **Cancel**.

Processing Sets Processing Data Sources Processing Profile Password Bank | Processing

Edit Delete Back Edit Permissions 12 of 24

Processing Set 005 Basic Settings

Inventory (Optional)
Completed

Discover
Completed

Publish
1 Data Source

Inventory
Discover
Publish
Total e
Last a

> Inventory (Optional)

> Discovery

▼ Publish

Published ?	Unpublished ?	Duplicates ?	Deleted ?	Publish
0 files 0 GB	7,332 files 0.72 GB	0 files 0 GB	0 files 0 GB	

Process Files

Inventory Files

Filter Files

Discover Files

Retry File Exceptions

Cancel

Reports

Inventory Report

Discovery Report

All Reports

Exceptions

Job Errors

File Exceptions

Consider the following about canceling publish:

- You cannot cancel a republish job. The cancel option is disabled during republish.
- Once the agent picks up the cancel publish job, no more errors are created for the data sources.
- If you click Cancel Publishing while the status is still Waiting, you can re-submit the publish job.
- If you click Cancel Publishing after the job has already been sent to the processing engine, then the set is canceled, meaning all options are disabled and it is unusable. Deduplication isn't run against documents in canceled processing sets.
- Errors that result from a job that is canceled are given a canceled status and cannot be retried.
- Once the agent picks up the cancel publish job, you cannot delete or edit those data sources.
- When you cancel a processing set during publish, documents already published before canceling are not deleted. You must manually delete them to remove them.

Once you cancel publish, the status section is updated to display the canceled set.

- When you publish multiple sets with global deduplication, dependencies are put in place across the sets to ensure correct deduplication results. Because of this, cancel behavior for publish has been adjusted in the following ways.
- If you need to cancel three different processing sets that are all set to global or custodial deduplication, you must do so in the reverse order in which you started those publish jobs; in other words, if you started them 1-2-3 order, you must cancel them in 3-2-1 order.
- When Global deduplication is set, cancel is available on all processing sets in which the DeDuplication and Document ID generation phase has not yet completed. Once the DeDuplication and Document ID generation phase is complete for all data sources on the set and there are other processing sets in the workspace that are also set to be deduped, the cancel button is disabled on the processing set.

22.4 Republishing a processing set

Note: The following information is specific to republishing files via the processing set console and is distinct from republishing files at the file level via the Files tab. For details on republishing at the file level, see [Republishing files](#) on the Files tab topic.

You can republish a processing set any time after the Publish Files option is enabled after the previous publish job is complete. Republishing is required after retrying errors if you want to see the previously errored documents in your workspace.

To republish, click **Publish Files**. The same workflow for publishing files applies to republish with the exception that Relativity does not re-copy the settings from the profile to the data sources that you are publishing.

When you click Publish Files again, you are presented with a confirmation message containing information about the job you are about to submit. If you haven't mapped any fields in the workspace, the message reflects this. Click **Publish** to proceed or **Cancel** to return to the processing set layout.

Continue?

Publish 3,627 files?

All ready-to-retry errors resulting from previous publish jobs will be retried.

Deduplication method: Global
Export Batch Size: 4000

Continue

Cancel

The status section is updated to display the in-progress republish job.

The Publish dashboard and status bars reflect the new job's progress.

Published [?]	Unpublished [?]	Duplicates [?]	Deleted [?]	Publish file except... [?]	Publish job errors [?]
0 files 0 GB	3,627 files 0.36 GB	3,627 files 0.36 GB	0 files 0 GB	0	0

Dedupe & control number generation [?]



Document promotion [?]



Metadata overlay [?]



Updating Processing Set fields [?]



Consider the following when republishing files:

- All ready-to-retry errors resulting from this publish job are retried when you republish.
- Deduplication is respected on republish.
- When you resolve errors and republish the documents that contained those errors, Relativity performs an overlay, meaning that there is only one file for the republished document in the Documents tab.
- Once published, a processing set may not be republished if the numbering type (default or level) on the set's profile has been changed.
- Once published, the start number(s) on a processing set may not be changed. Attempting to do so is disallowed.
- Changes made to numbering type in a processing profile are not respected after initial publishing. Data Source information cannot be changed after initial publishing.

22.5 Retrying exceptions after publish

During the Discover phase, some files may be identified as having exceptions, such as password-protected files that cannot be decrypted. You can still publish these files, but the published data may have incomplete metadata or missing extracted text. If you resolve exceptions for files that have already been published, you must republish those files so the updated version replaces the original in the workspace. To republish, see Republishing files from the Files tab or Common Use cases for using the Republish mass action .

23 Post-publish delete

There are a number of reasons you may want to delete documents after they are published. For example, you may find your profile settings are not what you intended or they are incorrect. Maybe it's an incorrect custodian assignment, or you ran processing in the wrong workspace. It may just be a cost consideration where you need to reduce the amount of data for review.

23.1 Overview

Use the Delete mass action on the Documents view to delete multiple files (after they have been discovered and published). To view a list of previously deleted files, navigate to the Files tab and open the Deleted Documents view.

You can configure which documents to delete, including the option to remove all duplicates and associated files from the same data source. Relativity automatically includes any duplicate records of a primary document deleted from the same processing data source.

Since Relativity removes all duplicates within the same data source, you do not need to work with any automatically promoted files.

For information on reporting what was deleted, see [Master Document Replacement Summary](#).

23.2 Post-publish delete and duplicate documents

Post-publish delete (PPD) evaluates published documents at the document level, not at the processing set or data source level. PPD reviews document deduplication relationships during this process and removes a document only when doing so will not break an existing deduplication relationship. In other words, PPD operates on published documents and their deduplication relationships, not on everything that originated from a particular data source.

Example

If a document in data source A is a duplicate of a document in data source B, deduplication links them into the same family. When you try to delete data source A using PPD, PPD attempts to remove the published documents from data source A but cannot if those documents belong to deduplication families that include documents from data source B. PPD cannot remove the published documents because of the shared deduplication families across data sources.

Result

You can delete some documents from data source A. However, you cannot delete the entire processing set or data source if any documents are in deduplication families shared with another data source. This behavior is expected and reflects the workspace-wide deduplication model.

23.3 Running post-publish delete

To mass delete documents after discovery and publish, perform the following:

1. Navigate to the Documents tab and select the documents you want to delete.

1 - 100 of 15,613 | 100 per page

#		Control Number	Processing File Id	Group Identifier	Custodian	Unified Title
1	☒	[redacted]_00000006	187	[redacted]_00000006	[redacted]	697acf0d-4aae-4fbd-a050-2bd9047cc288.rtf
2	☒	[redacted]_00000007	190	[redacted]_00000007	[redacted]	maul_und_klauenseu che.mpeg.mpeg
3	☒	[redacted]_00000008	193	[redacted]_00000008	[redacted]	03ead885-36df-47e8-9361-5e9e9f4c7593.MP3
4	☐	[redacted]_00000009	197	[redacted]_00000009	[redacted]	677e1c02-1900-4ffd-9d9d-e0ee4231c347.rtf
5	☐	[redacted]_00000010	198	[redacted]_00000010	[redacted]	669cc7a8-8782-48db-9aef-a1e3d32b3517.pdf
6	☐	[redacted]_00000011	202	[redacted]_00000011	[redacted]	03f76293-f72d-4917-a315-14e77a8259ed.pdf
7	☐	[redacted]_00000012	205	[redacted]_00000012	[redacted]	690c2cbe-3f3d-409b-b29d-ard??1f6hac0 doc

Checked (3) | Edit | Save Search | Total: 15,613Total: 81,421

2. Select the **Delete** mass operation.

Edit

Move

Delete

Replace

Tally/Sum/Average

Export to File

Cluster

Save as List

Convert

Re-produce

Image

PDF

Detect Sentiment

Edit | Save

3. You see a confirmation modal with the following settings options:

- **Delete settings**—selects the scope of the delete action for deleting files in the same data source. The following options are available:

- **Documents, duplicates, and all associated files**—deletes all documents, duplicates, images, and natives from within the same data source. Associated objects are unlinked. This option is the default.
- **Only image(s)**—deletes only images and leaves parent documents in place.
- **Only native(s)**—deletes only natives and leaves parent documents in place.
- **Image(s) and native(s)**—deletes only images and natives and leave parent documents in place.
- **Force delete**—files are deleted even if there are redactions, highlights, or transcript tags. You can only disable this option when deleting images or natives.
- **Create audit snapshot**—when enabled, this option creates an additional audit record of every field value associated with the document before it is deleted. Use caution when enabling this option as it can increase on the size of the audit table and will slow down the mass delete process.
- **Promote duplicates**—when enabled, duplicates from other data sources are retained and published, replacing deleted files. This option only applies to data sources outside of the original data source.

Delete 3 Documents



- Warning: Clicking delete will permanently delete this object and any children and unlink any associated objects. Click View Dependencies below for details.
- For any document originating from Relativity Processing: The file, its duplicates in the same data source along with its children will be marked as deleted on the Files page.

Delete settings Select the scope of the delete action. By default, documents will be deleted along with any duplicates, images, natives, and all associated objects will be unlinked. Selecting only images, only natives, or images and natives will leave document records in place.	Document, duplicates, and all associated files ▾
Force delete Yes When enabled, files will be deleted even if Redactions, Highlights, Links, or Transcript Tags are present. Can only be disabled for image/native deletions.	
Create audit snapshot ☒ When enabled, an audit record of every field associated with the document will be created before each file is deleted. This can significantly increase the delete job duration.	
Promote duplicates ☒ When enabled, duplicates from other data sources will not be deleted, and will be published to replace the deleted file.	

[> View Dependencies](#)

Delete

Cancel

23.4 Interacting with deleted documents

To view deleted documents, navigate to the **Files** tab and select the **Deleted Files** view.

To view files with exceptions, navigate to the **Files** tab and select either the **Current Files with Exceptions** or **All Files with Exceptions** view. See [Exception and error overview](#) for more details on processing exceptions.

23.5 Retrying delete errors

Navigate to the Deleted Files view to see a record of all deleted documents. The **Processing Deletion?** field is the yes/no indicator for deleted documents. You can filter by Exception Message to see the exceptions that occurred during deletion. Some exceptions can be retried using the **Retry Exceptions** mass actions option. Once deleted, documents are excluded from further processing operations (such as deduplication, retry, and republish.)

EntitiesProcessing Sets**Files**ReportsAdvanced

> 🔍

Replace

Deleted files

Add Widget

No Dashboard Wi...

▼ List

1-5 of 5100 per page

#	☐	Logical ID	File ID	File Name	Custodian	Data Source	Processing Deletion?
		Filter	= Filter	Filter	Filter	Filter	(All)
1	☒	48614BCF5E8AAA...	18431	A1CAEBFC8D10A7...			Yes
2	☒	8106FF175C20713...	18435	0E5AACDA0DB7FF...			Yes
3	☐	3C068E92B07409D...	22630	xls and word.zip			Yes
4	☐	9480C27EDE74...		xls-valid.xls			Yes
5	☐	E0B7276C40D1...		valid-word.doc			Yes

Export as CSV

Single Download / Replace

Retry Exceptions

Ignore Exceptions

Undo Ignore Exceptions

Download

Republish

Checked (2)

Select...

Save Filter

Viewing sample

5000

Total: 5

24 Exception and error overview

You may encounter exceptions and errors from time to time while running processing jobs. Exceptions and errors can occur during any processing phase, affecting single files or the entire job. This topic provides a high-level overview of the types of exceptions and errors you may encounter and where to find exception and error .

For information on resolving processing exceptions, see [Processing exception resolution](#).

This topic details the following:

- [Types of processing exceptions and errors](#)
- [Processing phases where exceptions and errors may occur](#)
- [Processing exception tabs](#)
 - [Files](#)
 - [File exception fields](#)
 - [Viewing file exception details](#)

- [Job errors](#)
 - [Job error fields](#)
 - [Viewing job error details](#)
- [Viewing exceptions from a processing set](#)
- [Retrying exceptions after making changes to the processing profile](#)
- [Required security permissions](#)

24.1 Types of processing exceptions

There are two types of processing issues that may affect the outcome of your processing job: file exceptions and job exceptions.

- **File exceptions**—affect a single file. Overall, the processing job continues, even if file exceptions exist. File exceptions are resolvable and have a status of **Not Resolved** in the Files tab. An example of a file exception might is a password-protected file where Relativity cannot find the password. By supplying the password you can retry the exception.
- **Job exceptions**—affect the entire processing job and block publishing. You must resolve job exceptions before you can continue processing.

24.2 Exception levels

Discover and publish exceptions are further defined by the exception level, which indicates severity of the issue. Each phase has two exception levels:

- Discover
 - Warning—the job continues; however, metadata may be missing.
 - Error—the job is blocked from completing.
- Publish
 - Warning—the job continues; however, metadata may be missing.
 - Error—the job is blocked from completing.

24.3 Processing phases where exceptions may occur

Exceptions can occur during any processing phase, such as Inventory, Discover, and Publish.

- **Inventory**—all Inventory exceptions are unresolvable, and you cannot retry them. To address inventory exceptions, rerun Inventory from the same processing set.
- **Discover**—the Discover phase covers file discovery and text extraction.
 - **Discover**—job exceptions that occur during discovery are unresolvable, and you cannot retry them. You should always address unresolvable discover job exceptions. Not addressing unresolvable exceptions may result in missing documents or incomplete metadata in published documents. Address these exceptions by creating a new processing set (or sets) and rediscovering the data.
 - **Text-extraction**—job exceptions that are resolvable have an exception status of **Ready to retry**. Retry the exception from the same processing set.

- **Publish**—job exceptions during the publishing phase that are resolvable have an exception status of **Ready to retry**. Retry the exception from the same processing set.

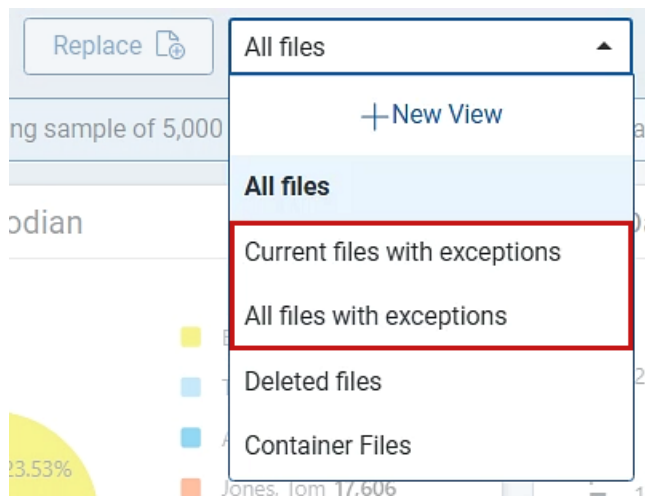
24.4 Processing exception tabs

Use the [Files](#) tab and [Job Exceptions](#) tab to view exceptions and errors that occurred during processing. Relativity notifies you of exceptions and errors on the processing set page when the job completes.

24.4.1 Files tab

The **Files** tab displays a list of all documents within your workspace. The Files tab has two views related specifically to file exceptions and errors:

- **Current Files with Exceptions**—displays all outstanding exceptions from processing jobs. This is the primary location for workflows like exception retry, ignore, and file replacement.
- **All Files with Exceptions**—displays a historical report of exceptions from all processing sets, regardless of exception resolution.



For more information on the Files tab, see the [Files](#) topic.

24.4.1.1 File exception fields

The **Current Files with Exceptions** and **All Files with Exceptions** views display the following fields:

Processing Sets

Processing Data Sources

Processing Profile

Password Bank | Processing

Inventory

Reports

Processing Duplication Workflow

Job Errors

Files

>

Q

Replace

All files with exceptions

Add Widget

No Dashboard Wi...

>

🔍

>

List

#	Details	Logical ID	File ID	Storage ID	File Name	Exception Level	Exception Message	Exception Phase	Exception Category	Exception Status
		Filter	=	Filter	Filter	(All)	Filter	(All)	(All)	(All)
8		F5767A43C93A9...	79274	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
9		0F423E930C658...	96595	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
10		761888373DE55...	52576	3016	SuperSecretStore.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
11		E999A1E4375A7...	36891	3016	SuperSecretStore.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
12		B6D2A1FC806D8...	22769	3016	SuperSecretStore.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved
13		917472EFDDA08...	22768	3016	SuperSecretStore.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Container	Not Resolved

1

- 100 of 147

100

per page

<

>

🔍

All 147

Select...

Save Filter

Viewing sample

5000

Total: 147

- **Details**—click this icon to open a dialog containing information on the file properties and exceptions. See [Viewing file exception details](#) for more detailed information.
- **Logical ID**—the unique identifier of the file in MDS storage.
- **File ID**—the unique identifier assigned to an instance of a file.
- **Storage ID**—the unique identifier assigned to a file.

Note: Differentiating between the File ID and the Storage ID. Every file is assigned a unique Storage ID. Every instance of a file is assigned a unique File ID. For example, you may have a Word document with a unique Storage ID. If that document is attached to two different emails, then each instance of the document attachments is assigned a unique File ID. Therefore, a single file with a Storage ID, may have multiple instances, each with their own unique File ID.

- **File Name**—the original name of the file.
- **Exception Level**—this field displays the type of exception Relativity encountered. There are two levels:
 - Warning—the job continues; however, metadata may be incomplete.
 - Error—the job is blocked from completing. You must address the error and retry the job.
- **Exception Message**—displays a message about the exception, along with possible resolution.
- **Exception Phase**—displays the phase in which the exception occurred. Exception phase options include Discover, Text Extraction, Publish, and Delete.
- **Exception Category**—displays the exception category such as a password protected file where Relativity cannot find the password. See the table below the field list for a list of Exception Category values.
- **Exception Status**—provides information on where the file is in exception remediation. The exception status value helps in determining if you need to take action on a file, or to see if the file had previous exceptions. When a file has all its exceptions resolved, the **Exception Message**, **Exception Category**, and **Exception Phase** fields no longer display any content. However, the **Exception Status** field displays Resolved to indicate that the file had issues that are now resolved. Exception status values include:

- **Not Resolved**—the exception is still outstanding.
- **Resolving**—the exception is in the process of being retried.
- **Resolved**—exception resolved.
- **Ignored**—exception ignored.

Note: In some cases, the **Exception Status** field on the **Files** tab may not be in sync with the **Exception Status** field on the **Documents** tab. The mismatch occurs when you resolve an exception but do not republish the file. For example, after discovery, you notice a file exception due to a missing password. The Exception Status fields on the Files and Documents tabs show the same Exception Status. At some point, you fix the password and retry the exception. The exception status on the Files tab now shows resolved, while the Documents tab still shows an exception. Republishing syncs the fields once again.

The following table provides a listing of values for the **Exception Category** field, along with a description of the issues associated with the exception.

Category name	Description
(Not Set)	Relativity did not set a category for this issue.
Corrupt Container	These exceptions are exclusive to container files that have encountered corruption when attempting to open and locate files within the container itself. <u>When containers have these exceptions associated to them, you will not see any extracted loose files.</u> The exceptions are either ignored or downloaded offline for you to investigate whether the corruption is fixable, then subsequently replaced and retried.
Corrupt File / Corrupt Item	These exceptions are exclusive to non-container files that have found elements of corruption during Processing. The exceptions are either ignored or downloaded offline for you to investigate whether the corruption is fixable, then subsequently replaced and retried.
Empty File	Empty files contain 0 bytes and therefore have no content to retrieve.
Environmental	Environmental exceptions are a result of issues in the Relativity Processing environment. Retry and resolve these exceptions when encountered.
File Read-Write Error	These exceptions are a subset of Environmental issues specifically caused by file system issues. Retry and resolve these exceptions when encountered.
Missing Attachment	Relativity could not extract an attachment from a document or email from its file.
Missing Extracted Text	These exceptions represent issues that occurred during Text Extraction jobs that have caused a file to be missing some or all associated text. Relativity could not identify a root cause. Retry and resolve the exceptions where possible.
Missing File Metadata	A file is missing a piece of metadata.
Password Protected Container	These exceptions are exclusive to container files that have encountered some form of password protection or encryption security measures. You must add the password or encryption key to the Password Bank to resolve the exception. For more information, see Password Bank. Note: When investigating publish exceptions, if you see five password protected exceptions associated with an .msg file, but the email and all its contents opens and displays correctly in the viewer, the email had a password-protected container attached.

Category name	Description
Password Protected File	These exceptions are exclusive to non-container files that have encountered some form of password protection or encryption security measures. You must add the password or encryption key to the Password Bank to resolve the exception. For more information, see Password Bank .
Unidentified File Type	Relativity was not able to determine the file type.
Unknown	Relativity Processing could not identify the file during discovery. The file may be corrupt, but an exact cause was not known at the time of discovery.
Unsupported	Relativity Processing cannot obtain metadata or text from unsupported files. You can publish the files to your workspace, but they may not be accessible from the viewer.

24.4.1.2 Viewing file exception details

You can view file details by clicking the **Show details for this file** icon. Details include file data, file metadata, and content metadata.

File Data

File Name	Co-Marketing Agreement.zip
File ID	85
Logical ID	24453D8A23AA00E10765E3C5965A9C71BD073FDB5B2F5A52065E2A168AEEE8AB
Custodian	[REDACTED]
Data Source	[REDACTED]
Container Extension	
Container ID	0
Container Name	
Contains Embedded Items?	No
Details	
Exception Message	There was an error during expansion of this compressed container. It may be password prot
File Extension - Text	ZIP
File Size (KB)	2,402
File Type	ZIP compressed archive
Folder Path	[REDACTED] \Loose Files
Is Container	Error

File Metadata

Stored As	\\files.[REDACTED]\T017\Files\EDDS1021749\Processing\1061812\INV1
LastModified	Mon, 23 Nov 2015 18:14:28 UTC
CreatedOn	Mon, 23 Nov 2015 18:14:27 UTC
LastAccessed	
LiteralFileExtension	ZIP

Content Metadata

WasDecrypted	false
OracleFileID	1802
OracleFileType	.ZIP/JAR File
MediaType	application/zip

Click the **Processing Exceptions** tab to view a summary of exceptions, and the exception history.

Exception Summary

Current Exceptions	5
Discover Exception(s)	Password Protected Item (5)
Text Extraction Exception(s)	none
Publish Exception(s)	none
Delete Exception(s)	none

Exception History

Exception total: 5

Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:07:00 UTC
There was an error during expansion of this compressed container. It may be password protec

Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:07:00 UTC
There was an error during expansion of this compressed container. It may be password protec

Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:07:00 UTC
There was an error during expansion of this compressed container. It may be password protec

Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:07:00 UTC
There was an error during expansion of this compressed container. It may be password protec

Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:06:59 UTC
There was an error during expansion of this compressed container. It may be password protec

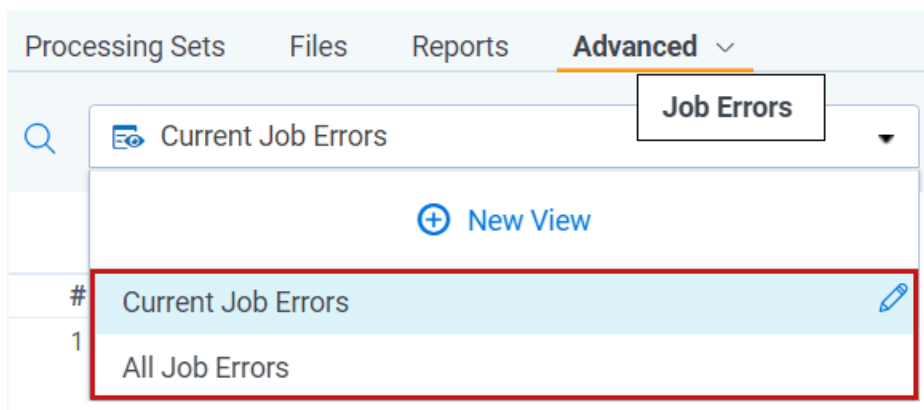
Close

- The Exception History section represents all exceptions> that have ever occurred on a file. This acts as a timeline of the record's exceptions, showing when they occurred, what they were about, and if any are still active. This includes exceptions resulting from retries of previous exceptions and contains category, phase, date/-time, and message information. All times are kept in UTC format.
- The Exception Summary section displays a count of all active exceptions along with their associated category and phase. This is especially important when investigating exceptions> relating to container files, as there can be many associated to the parent container during file extraction. This helps determine the level of impact the issue has as it may affect many files originating from it.

24.4.2 Job Errors tab

The Job Errors tab displays a list of job errors that occurred on processing sets run in your workspace. These errors are usually not associated with any specific files within a processing set, but rather the entire set itself. The Job Errors tab has two views:

- **Current Job Errors**—displays all unresolved job errors.
- **All Job Errors**—displays job errors that occurred throughout the lifecycle of the matter.



24.4.2.1 Job error fields

The Current Job Errors and All Job Errors views display the following fields:

- **Error Identifier**—the unique identifier of the error as it occurs in the database. When you click this message, you are taken to the error details layout, where you can view the stack trace for detailed error information. Note that for Unresolvable , the console is disabled because you cannot take any actions on that error from inside Relativity.
- **Error Status**—provides the status of the error.

Processing Sets Files Reports Advanced ▾						
Q 🔍 All Job Errors ▾ 🔍 All Job Errors ▾ 🔍 All Job Errors ▾ 🔍 All Job Errors ▾						
#	☐	Error Identifier	Error Status	Message	Custodian	Processing Set
1	☐	000000000000000000003068	Retried	The token supplied in the request is invalid.	TPT, 1	TPT Sets - TPT, 1
2	☐	000000000000000000003062	Retried	No WebApi nor Kepler service found to perform the request.	Brule, Steven	Rules
3	☐	000000000000000000003061	Retried	No WebApi nor Kepler service found to perform the request.	Brule, Steven	Rules
4	☐	000000000000000000003060	Retried	Error occurred when attempting to import the document.	Brule, Steven	Rules
5	☐	00000000000000000000998	Unresolvable	Error occurred when attempting to open this document. The document may be password protected.	Hetfield, James	BI Inventory
6	☐	00000000000000000000997	Unresolvable	Error occurred when attempting to open this document. The document may be password protected.	Hetfield, James	BI Inventory
7	☐	00000000000000000000996	Unresolvable	There was an error during expansion of this PDF file. It may be corrupted. Consider repairing or replacing the file and retrying.	Hetfield, James	BI Inventory
8	☐	00000000000000000000995	Unresolvable	Error occurred when attempting to open this document. The document	Hetfield, James	BI Inventory

- The most common error status values include:
 - **Unresolvable**—errors that you cannot retry.
 - **Ready to Retry**—resolvable errors that you can retry.
 - **Retried**—errors that have been retried and are unresolvable.
 - **In Progress**—the error is currently being retried.

- **Message**—displays an overview of the error.

Processing Sets Files Reports Advanced ▾						
All Job Errors						
🔍 🔊 🔗 🔧						
#	☐	Error Identifier	Error Status	Message	Custodian	Processing Set
1	☐	000000000000000003068	Retried	The token supplied in the request is invalid.	TPT, 1	TPT Sets - TPT, 1
2	☐	000000000000000003062	Retried	No WebApi nor Kepler service found to perform the request.	Brule, Steven	Rules
3	☐	000000000000000003061	Retried	No WebApi nor Kepler service found to perform the request.	Brule, Steven	Rules
4	☐	000000000000000003060	Retried	Error occurred when attempting to import the document.	Brule, Steven	Rules
5	☐	000000000000000000998	Unresolvable	Error occurred when attempting to open this document. The document may be password protected.	Hetfield, James	BI Inventory
6	☐	000000000000000000997	Unresolvable	Error occurred when attempting to open this document. The document may be password protected.	Hetfield, James	BI Inventory
7	☐	000000000000000000996	Unresolvable	There was an error during expansion of this PDF file. It may be corrupted. Consider repairing or replacing the file and retrying.	Hetfield, James	BI Inventory
8	☐	000000000000000000995	Unresolvable	Error occurred when attempting to open this document. The document	Hetfield, James	BI Inventory

- **Custodian**—the custodian associated with the data source containing the file on which the error occurred.
- **Processing Set**—the name of the processing set in which the error occurred.
- **Data Source**—the data source containing the file on which the error occurred.
- **Error Created On**—the date and time at which the error occurred during the processing job.
- **Republish Required**—the error must be retried in order to be successfully published.
- **Notes**—any manually added notes associated with the error.

24.4.2.2 Viewing job error details

Click the **Error Identifier** to open the **Job Error Details** dialog.

Note that the **Error Actions** console is not available for unresolvable job , since you cannot retry or ignore job as you can with document .

Job Error Details

Advanced

Error Identifier

00000000000113600061

Custodian

Thief, Processing

Processing Set

Namit Test Processing Set

Data Source

Thief, Processing - 1058973

Error Status

Unresolvable

Message

Error occurred when attempting to read from the RAR file. The container may be password protected.

Notes

Error Created On

11/27/2019 7:03 PM

Republish required

ERROR ACTIONS

Retry

To see the job error stack trace, click the **Advanced** tab. Here, you can parse through even more detailed information about the error, including timestamps, messages, and error phase.

Job Error Details

Advanced

Stack Trace

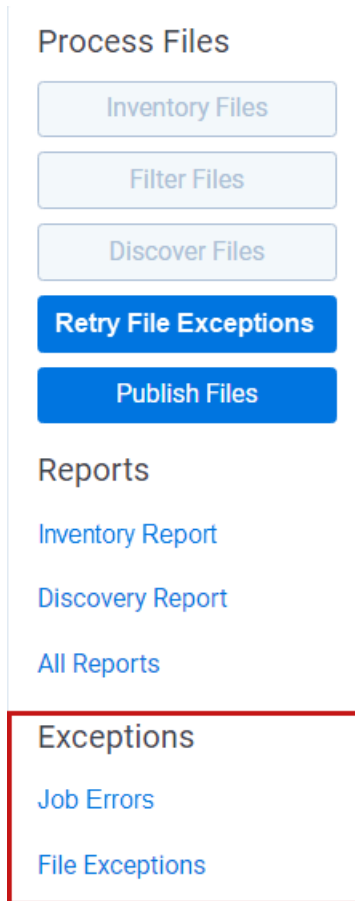
```

<?xml version="1.0" encoding="utf-16"?>
<Exceptions>
  <Exception>
    <Timestamp>2019-11-27T19:03:36.977</Timestamp>
    <Type>System.Exception</Type>
    <Message>Error occurred when attempting to read from the RAR file. The container may be password protected.</Message>
    <StackTrace>
      at Invariant.RarExpander.CheckRarError(Int32 n) in d:\sourcecode\Invariant\Invariant.handler.compression\Rarexpander.cpp:line 400
      at Invariant.RarExpander.Inventory(IHandlerParameters inventoryParameters, IInventoryCallbacks inventoryCallbacks) in d:\sourcecode\Invariant\Invariant.handler.compression\Rarexpander.cpp:line 543
      at Invariant.ExpanderParametersAdapter.Inventory(IHandlerBase handler, Object[] parameters) in D:\SourceCode\Invariant\Core\HandlerAdapter\ParametersAdapters\ExpanderParametersAdapter.cs:line 55
      at Invariant.HandlerAdapter.Invoke(String type, String method, Object[] parameters) in D:\SourceCode\Invariant\Core\HandlerAdapter\HandlerAdapter.cs:line 495
    --- End of stack trace from previous location where exception was thrown ---
      at Invariant.HandlerAdapter.Invoke(String type, String method, Object[] parameters) in D:\SourceCode\Invariant\Core\HandlerAdapter\HandlerAdapter.cs:line 500
      at Invariant.Worker.<DisplayClass176,0.>.<MonitoredInvoke>_0() in D:\SourceCode\Invariant\Core\Worker.cs:line 2312
      at Invariant.Worker.MonitoredInvoke(Action action) in D:\SourceCode\Invariant\Core\Worker.cs:line 2332
      at Invariant.Worker.MonitoredInvoke(HandlerAdapter adapter, String type, String invokeMethod, Object[] parameters) in D:\SourceCode\Invariant\Core\Worker.cs:line 2312
      at Invariant.Worker.DoInnerJob(Job jobToRun, String invokeMethod, Type dispatchType, Object[] parms, Int32 attempts) in D:\SourceCode\Invariant\Core\Worker.cs:line 2192
    --- End of stack trace from previous location where exception was thrown ---
      at Invariant.ErrorExtension.ThrowAsSerializableException(Exception exception) in D:\SourceCode\Invariant\Core\Errors\ErrorExtension.cs:line 29
      at Invariant.Worker.DoInnerJob(Job jobToRun, String invokeMethod, Type dispatchType, Object[] parms, Int32 attempts) in D:\SourceCode\Invariant\Core\Worker.cs:line 2277
      at Invariant.Inventory.InventoryWrapper(Job j, IdentificationResult result, InventoryMatter inventoryMatter) in D:\SourceCode\Invariant\Core\Inventory\Inventory.cs:line 357
      at Invariant.Inventory.InventoryFile(Job job, String filePath, String filename, String virtualPath, Boolean isFromContainer, String uniqueHash) in D:\SourceCode\Invariant\Core\Inventory\Inventory.cs:line 318</StackTrace>
    <Worker>ss025wrkr00001J\9</Worker>
    <Jobid>1839808</Jobid>
    <DispatchType>Invariant.InventoryHandler</DispatchType>
    <DispatchMethod>InventorySubJob</DispatchMethod>
    <Flags>32780</Flags>
    <Unresolvable>True</Unresolvable>
  </Exception>
</Exceptions>

```

24.5 Viewing exceptions from a processing set

From within your processing set, use the links in the **Process Files** console to view job errors and file exceptions. Click the **Job Errors** link to open the **Job Errors** tab > **Current Job Errors** view. Click the **File Exceptions** link to open the **Files** tab > **Current Files with Exceptions** view.



24.6 Retrying exceptions after making changes to the processing profile

When you retry exceptions, Relativity attempts to reprocess exceptioned files using the processing profile settings associated with the processing set. If you change a field setting in the processing profile and then retry exceptions, some changes will be respected (or recognized,) and some will not. In other words, after changing the processing profile, Relativity uses the new settings for some fields during retry and ignores new settings for others. In the table below, fields marked Yes are fields where changes override the original setting. Fields marked No are fields where the original setting precedes any changes.

Front-End Setting	Respected on retry?
Inventory Discover Settings	
DeNIST and DeNIST Mode	Yes
Default OCR languages	No
Default time zone	No
Include/Exclude and Inclusion/Exclusion	Yes
Mode	Yes
File Extensions	Yes
Extraction Settings	

Front-End Setting	Respected on retry?
Extract children	No
When extracting children, do not extract	No
Roll up image text	No
Email Output	Yes
Excel Text Extraction Method	Yes
Excel Header/Footer Extraction	Yes
PowerPoint Text Extraction Method	Yes
Word Text Extraction Method	Yes
OCR	Yes
OCR Accuracy	Yes
OCR Text Separator	Yes
Short Message Conversion Settings	
Slice by	Yes
Slack	Yes
Download attachments	Yes
Teams	Yes

24.7 Required security permissions

The following security permissions are required to perform actions on file errors:

Object Security	Tab Visibility
- Discovered File—view, edit - Download and Replace files with processing errors - Processing Error—enable Download and Upload files with processing errors	- Processing - Files

Note: This feature supports feature permission actions. Each action grants access to a specific task within the feature. You can grant groups access to individual actions, allowing them to perform only the tasks they need. To grant access to the entire feature, select all of its actions. Manage feature actions on the Features tab in a group's workspace permissions. For instructions, see [Setting workspace permissions](#). For more information about feature permissions, see [Feature permissions](#).

25 Exception and error resolution

Two distinct types of exceptions occur during processing: exceptions and errors. Exceptions are resolvable errors that occur when Relativity cannot process a job or file. For example, a password-protected file where Relativity cannot find

the password. In this case, you can resolve the issue by providing the password and rerunning the process. Errors occur when the processing engine cannot complete the job. For example, an error may occur when there is a break in network connectivity where Relativity cannot access file storage. In this case, the resolution is to rerun the processing job after you restore network connectivity. Other errors may be more complex and require support intervention.

For information on where to find processing exception and error information, see [Processing exception overview](#).

This topic details the following:

- [Types of processing exceptions](#)
- [Processing phases where exceptions may occur](#)
- [File exceptions](#)
 - [File exception actions](#)
 - [Resolving file exceptions](#)
 - [Retrying an exception from the processing set](#)
 - [Exception categories and their probability of retry success](#)
 - [Retry file exception mass action](#)
 - [Replacing corrupted files](#)
 - [Resolving a password-protected file exception](#)
 - [Publish documents with field length exceptions](#)
 - [Email file processing time out](#)
 - [Troubleshooting corrupted .msg files in a .pst container](#)
 - [Corrupt PDF files not flagged during discovery](#)
- [Job errors](#)
 - [Job error workflow](#)
 - [Resolving job errors](#)

25.1 Types of processing exceptions

There are two types of processing issues that may affect the outcome of your processing job: file exceptions and job exceptions.

- **File exceptions**—affect a single file. Overall, the processing job continues, even if file exceptions exist. File exceptions are resolvable and have a status of **Not Resolved** in the Files tab. An example of a file exception might is a password-protected file where Relativity cannot find the password. By supplying the password you can retry the exception.
- **Job exceptions**—affect the entire processing job and block publishing. You must resolve job exceptions before you can continue processing.

25.2 Exception levels

Discover and publish exceptions are further defined by the exception level, which indicates severity of the issue. Each phase has two exception levels:

- Discover
 - Warning—the job continues; however, metadata may be missing.
 - Error—the job is blocked from completing.
- Publish
 - Warning—the job continues; however, metadata may be missing.
 - Error—the job is blocked from completing.

25.3 Processing phases where exceptions may occur

Exceptions can occur during any processing phase, such as Inventory, Discover, and Publish.

- **Inventory**—all Inventory exceptions are unresolvable, and you cannot retry them. To address inventory exceptions, rerun Inventory from the same processing set.
- **Discover**—the Discover phase covers file discovery and text extraction.
 - **Discover**—job exceptions that occur during discovery are unresolvable, and you cannot retry them. You should always address unresolvable discover job exceptions. Not addressing unresolvable exceptions may result in missing documents or incomplete metadata in published documents. Address these exceptions by creating a new processing set (or sets) and rediscovering the data.
 - **Text-extraction**—job exceptions that are resolvable have an exception status of **Ready to retry**. Retry the exception from the same processing set.
- **Publish**—job exceptions during the publishing phase that are resolvable have an exception status of **Ready to retry**. Retry the exception from the same processing set.

25.4 File exceptions

After completing the discovery or publish phases, you notice the processing job encountered file exceptions. You want to investigate and resolve the exceptions to maximize the data published to your workspace.

25.4.1 File exception actions

You can take action on file exceptions from the mass actions drop-down menu located at the bottom of the **Files** tab page.

Processing Sets
Processing Data Sources
Processing Profile
Password Bank | Processing
Inventory
Reports
Job Errors
Files

>
Q
Replace
All files with exceptions
Add Widget
No Dashboard Wi...
>
Q

List
1
- 100 of 147
100
per page

#	Details	Logical ID	File ID	Storage ID	File Name	Exception Level	Exception Message	E
		Filter	=	Filter	Filter	(All)	Filter	
1		24453D8A23AA0...	85	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	D
2		0CD14CFA4D2BF...	10004	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	D
3		7C1CACCE49D9...	10012	61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	D
4		E8...		61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	D
5		E1...		61	Co-Marketing Agreement.zip	Error	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	D

All 147
Select...
Save Filter
Viewing sample
5000
Total: 147

Export as CSV
Single Download / Replace
Retry Exceptions
Ignore Exceptions
Undo Ignore Exceptions
Download
Republish

The following mass actions are available:

Mass operation	Description
Export as CSV	This exports the list of processing exception as a CSV file.
Single Download/Replace	Use the Single Download/Replace mass operation to download a file to your local machine for investigation. After fixing the exception, you can use the same mass operation to replace the original file with a newer version. Note the following regarding downloading and replacing files: - You can only download or replace a single file at a time. - You can only perform these actions on files with an exception status of Not Resolved. - There is no file size limitation when downloading files. - There is a one-gigabyte file size limitation when uploading replacement files. - Performing a file replacement retries associated exceptions once completed. - After uploading a new document, and you select Replace & Retry, the native file updates before you republish. - The retry action for job exceptions only retries exceptions in a Ready to Retry state.

Mass operation	Description
Retry exceptions	This action provides the ability to resolve issues occurring during discover and publish. You can see these issues on the Current Files with Exceptions and All Files with Exceptions views within the Files tab. For details on how to retry exceptions from the Deleted Documents view in the Files tab, see Retrying delete exceptions. - You must have edit permissions on the Discovered Files object to be able to retry file exceptions. - Note the following regarding retrying exceptions: - Auto-publish is not enabled when you retry exceptions. If you resolve any discover or text extraction exceptions, you must manually publish them into your workspace by navigating back to the processing set and clicking the Retry button. - The discovery retry of exceptions process has a longer timeout period than the initial discovery process. It is not uncommon for the retry process to run longer than the initial discovery process. - You should always resolve all publish exceptions as these exceptions represent data not in review. - If an exception occurs on a loose file during discovery, Relativity still attempts to publish it. For example, if a Password Protected exception occurs on a PDF file during discovery, you can still publish the file in its current state. The resulting record may have metadata or extracted text missing depending on the issue, but still referenced during review. - Relativity automatically retries all publish exceptions for a set when you retry any exception within that set. - Do not work on multiple retry attempts simultaneously. If you submit a secondary retry while the initial retry is still in progress, the second retry waits in a queue until the first retry completes. - You can only submit exceptions with an exception Status of Not Resolved for retry.
Ignore exceptions	This provides the ability to set a file's exception Status to Ignored, which removes it from the Current Files with Exceptions view. The record is visible in the All Files and All Files with Exceptions views.
Undo Ignore exceptions	Sets a file's exception Status field back to its original value after you previously ignored it.
Download	Use the Download mass action to download multiple exceptioned files to your local drive. Note the following regarding the downloading mass operation: - You can only use the Download mass operation on files with an exception status of Not Resolved. - When complete, files are packaged in a zip container and saved to your

Mass operation	Description
	local drive. - Use the Replace button to mass replace files. Note: The Replace button is located in the upper left side of the screen. It is not a part of the mass actions menu because mass actions require you to select files from the document list to apply the action. There are no files to select when replacing, as you are uploading them from your local drive. - There is a 2.14 GB limitation on the zip container file size. For more information on replacing multiple files at a time, see Replacing corrupted files.
Republish	Gives you the option of republishing exceptioned files after you resolve the exceptions. If the republish is successful, the files are available in the Documents list and removed from the Current Files with Exceptions view. For details on how to republish files from the Files tab, see Republishing files.

25.5 Resolving file exceptions

The following sections describe actions you may take to resolve file exceptions.

25.5.1 Retrying an exception from the processing set

You can retry file exceptions from the processing set by clicking the **Retry File exceptions** button located in the **Process Files** console.

Process Files

Inventory Files

Filter Files

Discover Files

Retry File Exceptions

Publish Files

Reports

Inventory Report

Discovery Report

All Reports

Exceptions

Job Errors

File Exceptions

A confirmation message appears with information about the number of exceptions you are about to retry, along with the associated exception categories. Click **Yes** to proceed or **Cancel** to return to the processing set page.

Continue?

Processing will be retried for 3657 files, to address errors in the following categories

- Unknown (3573)
- Missing Extracted Text (74)
- Configuration Error (10)

Errors that require additional intervention will not be retried here.

Yes

Cancel

25.5.1.1 Container file exceptions

File exceptions on container files, such as a .zip file, may result in missing data from one or more files within the container. In some cases, you may have many files in exception, even though you see a single file exception on the container. For container file exceptions, extract the files and then create a new processing set with the extracted files.

25.5.1.2 Exception categories and their probability of retry success

Relativity retries file exceptions that have a high chance of success. The exception category associated with the file determines the probability of success. exception categories such as Corruption or Password Protection are not retried as they are not likely to be successful without manual intervention. For example, adding passwords to the Password Bank or replacing corrupt files. The table below lists the types of exceptions that Relativity retries.

Exception Category	Included in Retry Button
Corrupt Container	No
Corrupt File	No
Environmental	Yes
File Read-Write Error	Yes
Missing Attachment	Yes
Missing Extracted Text	Yes
Missing File Metadata	Yes
Password Protected Container	No
Password Protected File	No
Unknown	Yes
Unsupported	No

25.5.2 Retry file exceptions mass action

Starting from your processing set, perform the following steps:

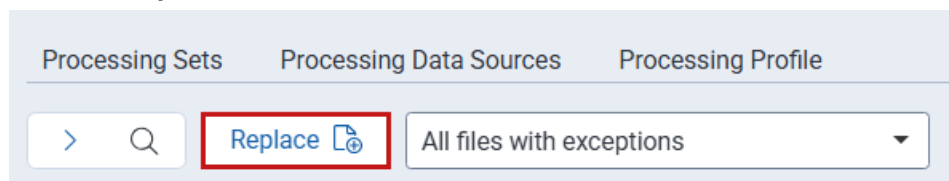
1. From the **Process Files** console, click **File exceptions** to open the **Current Files with Exceptions** view. The list is automatically filters for your processing set.
2. You can filter the exceptioned files that are the most important to resolve. Some common filters are:
 - **Exception Category**—group issues of a similar type.
 - **Exception Phase**—groups issues occurring during a particular part of processing.
 - **Custodian**—when you have a priority Custodian you want investigated first.
 - **Sort Date**—to retry files within the matter's relevant date range.
3. Once you identify a group of records to resolve, select the **Retry Exceptions** mass action to begin the process. Alternatively, you can retry all exceptioned files without filtering.

You can now track the progress of the exception retry through the processing set's progress bar or by navigating to the [Processing and Imaging Queue](#) and viewing active jobs.

25.5.3 Replacing corrupted files

Use the steps below to upload remediated files from your local drive back into Relativity. You must place the files in a .zip container before uploading. There is a 1 GB file size limit for each .zip container.

1. Use the search bar to navigate to the **Files** tab.
2. Click the **Replace** button on the **Files** tab.



Note: While the Replace action is a mass action, meaning multiple files are affected, it is not a part of the mass actions menu (at the bottom of the screen). Action items in the mass actions menu require selecting target files from the Files list. Since you are moving files from your local drive back into Relativity, there are no files from the list to preselect.

3. The Replace modal opens. Either drag and drop or browse for the .zip container holding your replacement files.

Replace

Please ensure that files are properly named after storage ID and the ZIP is up to 1 GB.



Drop your files here or [browse for files](#).

Upload

Close

4. Click **Upload**.

When you upload the replacement file, Relativity performs the following checks:

- The replacement files are contained in a .zip file, which has a file size limit of 1 GB. If you exceed the file size limit, you will see an error. To resolve this error, redistribute the replacement files so the .zip container is within the file size limit.
- The replacement file's file name matches the original file's Storage ID.
- The replacement file size is within 10% of the original (larger or smaller) file size.
- The replacement file extension matches the original file extension.
- The original file has an exception status of Not Resolved.

5. The Replace modal updates with the number of successful files, successful files with warnings, and unsuccessful files.

Replace

 Your files have been uploaded and are ready to be replaced and retried. 

Successful 	Successful with warning 	Unsuccessful 
2 files	0 files	0 files

Proceeding with Replace & Retry will include both Successful and Successful with Warning files. Warnings indicate inconsistencies with file extension and/or in file size, which is to be expected. Unsuccessful files are due to lack of a match or a match does not have an outstanding discovery exception.

For more information, use the Replace log button to download a full list of replacement files and matches.

[Replace & Retry](#)[Replace log](#)[Close](#)

- **Successful**—the number of files successfully replaced.
- **Successful with warning**—files where the candidate file name minus extension has a Storage ID match in the Files tab and has an Exception Status of Not Resolved. In addition, these files may have a different extension or exceed the 10% deviation in file size.
- **Unsuccessful**—files where the candidate has a Files tab match but no outstanding exception, the candidate does not have a Files tab match, or Relativity could not extract the file from the .zip container.

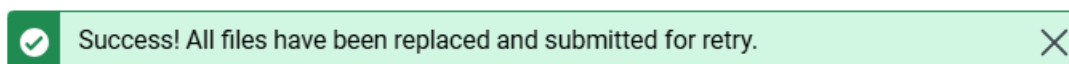
You can view a log of the files, along with their status, by clicking the **Replace log** link on the modal. The log displays the following information:

- Replacement file name
- Candidate file extension
- Candidate file size
- Files tab Storage ID
- File tab file extension
- File tab file size
- Exception (error) status
- Validation status
- Reason for warning or failure.

6. Click **Replace & Retry**.

7. A final confirmation message indicates the completion of the replacement process. Click **Close**.

Replace



To see the replaced files, navigate to the All Errors View or All Files View in the Files tab. To check progress of retry, refer to error status field or the associated processing sets.

Close

8. The last step is to **republish** your files so that Relativity updates metadata and extracted text fields.

25.5.4 Resolving a password-protected file exception

Some files are password-protected. If Relativity cannot find the password during the Discover phase, you will see an exception with the category, Password Protected Container. The resolution for this exception is to add the password to the Password Bank.

To add a password to the bank, perform the following steps:

1. Navigate to the **Password Bank | Processing** tab.
2. Click **New**.
3. Select **Passwords** for the **Type** value.
4. Add a meaningful description to help you differentiate between other password entries.
5. Add one or more passwords. You must enter each password on a separate line.
6. Click **Save**.
7. **Retry** the file exception from either the processing set Process Files console, or from the Files tab mass actions drop-down menu.

For completed details on storing passwords, see [Password Bank](#).

25.5.5 Publish documents with field length exceptions

When metadata exceeds the field length limit for optional fields, Relativity publishes the file, excluding the offending metadata, not as an error but as an exception warning. The published content is available for review or other downstream functions such as indexing and search terms reports. Files with metadata exceeding the field length limit for required fields have an exception errorstatus where the entire file is held back and not published.

25.5.5.1 Resolving field length exception warnings

You cannot successfully retry field length exception warnings without some intervention. For example, suppose you have an optional email field with a field length of 250 characters and try to publish a file with an email containing 300 characters. You see the field length exception warning. You can change the email field length to 350 characters and retry the warning, resulting in the email metadata being published with its parent file.

You cannot change required (system-mapped) fields with a fixed-length text field type. In this case, Relativity displays an exception error, and you cannot publish the file in its current state. For example, Container Name is a system-mapped field with a maximum limit of 255 characters. If you exceed that limit, you will see an exception error. Since you cannot change the field properties, you must change the container name length to meet the field requirements. After that, you can retry the error.

Steps

1. Use the search bar to navigate to the **Fields** tab.
2. Use the **Name** filter to locate the field you want to change. You can also scroll through the field names if you are not sure of the field name. In the image below, we want to change the **Email From** field.

The screenshot shows the 'Fields' tab interface. At the top, there is a search bar with a magnifying glass icon and a 'New Field' button. Below the search bar is a dropdown menu labeled 'All Non-System Document Fields'. The main area displays a table of fields. The first column is '#', the second is 'Name', and the third is 'Field Type'. The 'Email From' field is highlighted with a red box. Below the table, there are two rows of field details. The first row is for 'Domains (Email From)' with a 'Multiple Object' field type. The second row is for 'Email From' with a 'Fixed-Length Text' field type.

3. Click the field **Name** (in this example, Email From) to open the fields details.
4. Click **Edit**.
5. Change the Maximum Length field so that it exceeds the field characters that threw the exception.

The screenshot shows the 'Field Settings' tab. The 'Maximum Length' field is highlighted with a red box and set to 350. Below it, there are three toggle switches: 'Required' (off), 'Include in Text Index' (off), and 'Identifier' (No).

6. Click **Save**.

25.5.6 Email file processing timeout

Processing uses an email-specific timeout when handling email files (.eml, .msg). This timeout operates independently from instance timeout settings and defaults to 20 minutes.

The time allowance helps to early identify problematic email files that have a very low likelihood of processing successfully and ensures availability of the resources to process other files within processing set.

If the timeout is exceeded, the following file-level exception appears:

*Relativity terminated processing this file after exceeding the email time limit. The cause of the extended processing time is unknown. Retry the error. If the issue persists, see the **Exception and error resolution** page in the documentation for suggested actions.*

25.5.6.1 Troubleshooting a persistent email timeout error

If the same error occurs after retrying, the cause may be related to unidentified file-specific issues or corrupt data. To resolve this issue, use a new copy of the file and retry processing. If issue persists, contact support for assistance.

25.5.7 Troubleshooting corrupted .msg files in a .pst container

You can use the following steps to troubleshoot errors. about corrupted .msg files in .pst containers encountered during processing.

The error. message will be similar to:

For cases where Relativity extracted and discovered a message, but it might be corrupt

Relativity encountered an issue while extracting <entry_id.msg> from a container. This file was discovered in Relativity, but it might be corrupt. Try repairing or replacing it, then retry.

For cases where Relativity failed to extract and discover a message

Relativity failed to extract <entry_id.msg> and its attachments from the container, resulting in the file not being discovered. Possible causes of this error include corrupted data in the file or a Relativity system issue. First, retry the error. If the issue continues, consider including the file in a smaller container or processing it as a loose file.

The error. returns the File Name. The File Name is the equivalent of the Entry ID of the corrupted file. Use the Entry ID to locate and troubleshoot the issue using the tools described in the steps.

Example File Name: 00000000301D96839118B94DA10CED6A093B254944002000.MSG

Example Entry ID: 00000000301D96839118B94DA10CED6A093B254944002000

Note: This documentation contains references to third-party software, or technologies. While efforts are made to keep third-party references updated, the images, documentation, or guidance in this topic may not accurately represent the current behavior or user interfaces of the third-party software. For more considerations regarding third-party software, such as copyright and ownership, see [Terms of Use](#).

The following topics are covered in this section:

- [Generating a list of impacted files](#)
- [Troubleshooting tools](#)
 - [Option 1: Outlook Spy](#)
 - [Option 2: MFCMapi](#)
 - [Option 3: PST Walker](#)
- [Troubleshooting using O](#)
- [Troubleshooting using MFCMapi](#)
- [Troubleshooting using PST Walker](#)

25.5.7.1 Generating a list of impacted files

There are two options for generating a list of impacted files. You can access the processing history for a specific file or you can generate a report that displays a list of errored files and their Entry IDs.

Accessing the processing history: (This example looks for a single message file in a .pst container.

Note: Use this option to locate a single message file.

1. Navigate to the **Files** tab.
2. Change the view to **All files with exceptions** to filter the document list.
3. Scroll through the document list until you locate the file in error.
4. Click the **Show details** icon **Details** column for the .pst file.

5. Select the **Processing Exceptions** tab. In the Exception History section, you see a list of files with their file names. The file name is the equivalent of the Entry ID which you can use to further explore exceptions.

Details

Processing Exceptions

Text Extraction Exception(s) none

Publish Exception(s) none

Delete Exception(s) none

Exception History

Exception total: 785

Showing 100 of 785

Corrupt Item (Discover) - Not Resolved

Fri, 11 Apr 2025 11:59:23 UTC

Relativity encountered an issue while extracting 0000000030891BF1BCF4934E8434744ED161F51C843E2000.MSG from a container. This email file may be corrupt. Try repairing or replacing it, then retry. Need help? Look for "Troubleshooting corrupted .MSG files in a .PST container" in documentation.

Corrupt Item (Discover) - Not Resolved

Fri, 11 Apr 2025 11:59:22 UTC

Relativity encountered an issue while extracting 0000000030891BF1BCF4934E8434744ED161F51C643E2000.MSG from a container. This email file may be corrupt. Try repairing or replacing it, then retry. Need help? Look for "Troubleshooting corrupted .MSG files in a .PST container" in documentation.

Corrupt Item (Discover) - Not Resolved

Fri, 11 Apr 2025 11:59:22 UTC

Relativity encountered an issue while extracting 0000000030891BF1BCF4934E8434744ED161F51C443E2000.MSG from a container. This email file may be corrupt. Try repairing or replacing it, then retry. Need help? Look for "Troubleshooting corrupted .MSG files in a .PST container" in documentation.

Corrupt Item (Discover) - Not Resolved

Fri, 11 Apr 2025 11:59:22 UTC

Relativity encountered an issue while extracting 0000000030891BF1BCF4934E8434744ED161F51C243E2000.MSG from a container. This email file may be corrupt. Try repairing or replacing it, then retry. Need help? Look for "Troubleshooting corrupted .MSG files in a .PST container" in documentation.

Corrupt Item (Discover) - Not Resolved

Fri, 11 Apr 2025 11:59:22 UTC

Close

6. Return to the **Files** tab document list and filter the list by **File Name**. This is the corrupted document.

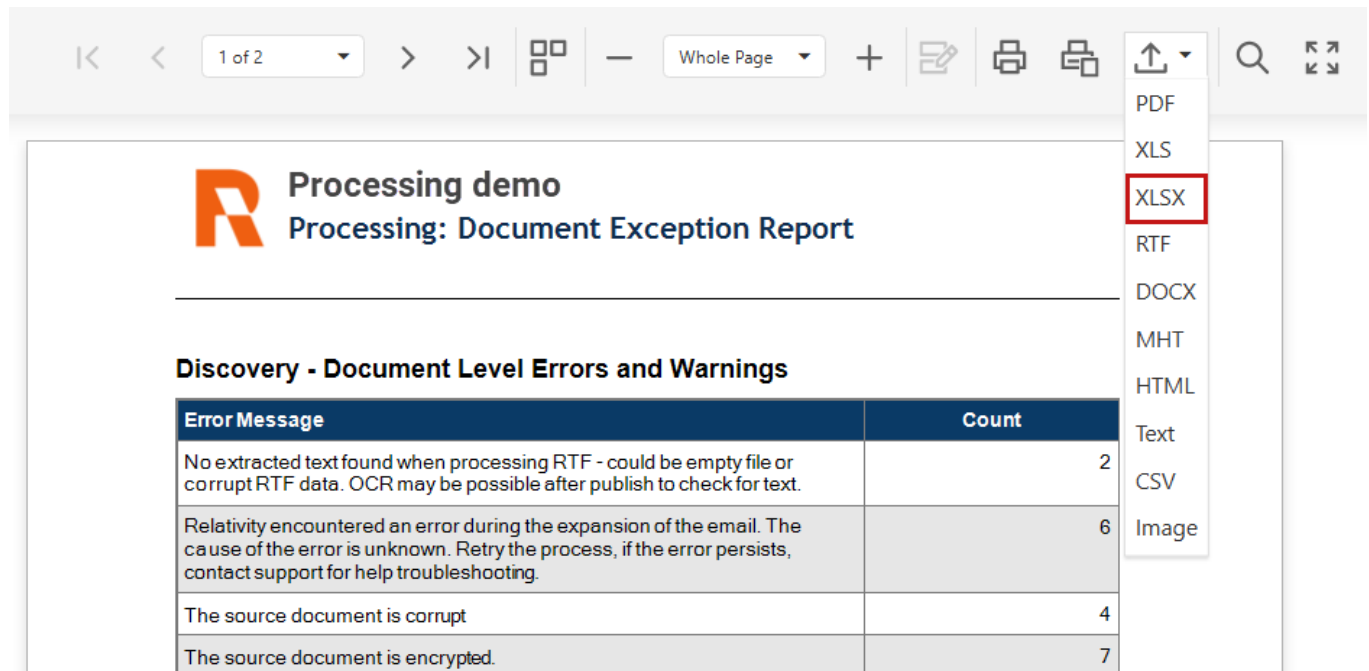
#	Details	File ID	File Name	File Type	File Extension - Text
		= Filter	0000000030891BF...	Filter	Filter
1		24152	0000000030891BF1BCF...	Outlook Message File	MSG

Generating a report that displays a list of errored files:

Note: Use this option to locate multiple files.

1. Navigate to the **Reports** tab.
2. Select the **Document Exception** report.
3. Select the appropriate **processing set**.

4. Click **Generate Report**.
5. **Export** the report to your local drive. Use either the CSV or XLSX options.

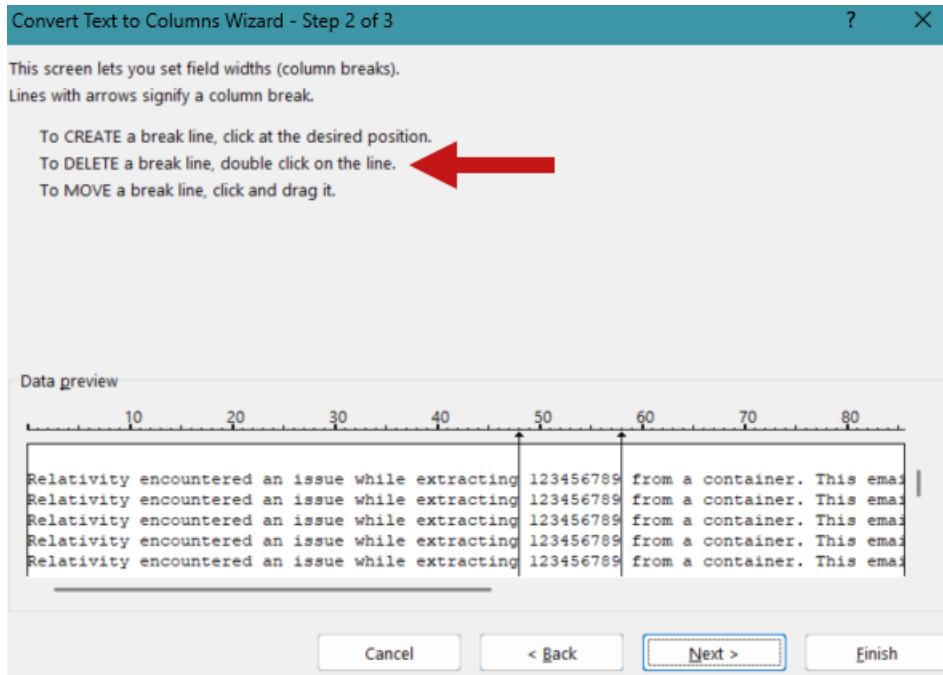


The screenshot shows a web application interface for a "Processing demo". The main heading is "Processing: Document Exception Report". Below this, there is a section titled "Discovery - Document Level Errors and Warnings" which contains a table of error messages and their counts. To the right of the table, an export menu is open, listing various file formats: PDF, XLS, XLSX (highlighted with a red box), RTF, DOCX, MHT, HTML, Text, CSV, and Image.

Error Message	Count
No extracted text found when processing RTF - could be empty file or corrupt RTF data. OCR may be possible after publish to check for text.	2
Relativity encountered an error during the expansion of the email. The cause of the error is unknown. Retry the process, if the error persists, contact support for help troubleshooting.	6
The source document is corrupt	4
The source document is encrypted.	7

6. Open the report, then use Excel filters and functions to isolate the Entry ID from the rest of the error message. To isolate a list of Entry IDs:
 - a. **Filter** all of the error messages so that only errors with the phrase *Troubleshooting corrupted .msg* are displayed.
Use the **Text to Columns** function to isolate file names from their respective error messages.
 - b. Click the **Text to Columns** option in the **Data** tab.
The Convert Text to Columns Wizard opens.
 - i. Select the **Fixed width** option and click **Next**.

- ii. In the Data preview pane, you see break lines for each column. Locate the break lines before and after the file name. **Double-click each break line** (except the file name lines) to delete them.



- iii. **Finish** the wizard.

Original	Entry ID	Count
Error Message	0000000030891BF1BCF4934E8434744ED161F51C04012000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C04022000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C04032000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C04042000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C04052000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C04062000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C04072000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C04082000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C04092000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C040A2000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C040B2000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C040C2000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C040D2000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C040E2000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C040F2000.MSG	1
Relativity encountered an issue while extracting	0000000030891BF1BCF4934E8434744ED161F51C04102000.MSG	1

The file names are now isolated from the error message.

Sample file name: 00000000301D96839118B94DA10CED6A093B254944002000.MSG.

The Entry ID is the file name without the file extension.

Sample Entry ID: 00000000301D96839118B94DA10CED6A093B254944002000.

Note: See [Split text into different columns with the Convert Text to Columns Wizard](#) on the Microsoft website for more information on the *Text to columns* function.

25.5.7.2 Troubleshooting tools

To work with .pst files, you must have Microsoft Outlook installed on your system, along with one of the following tools:

Option 1: OutlookSpy

OutlookSpy provides access to Outlook objects, including properties, call functions, and object hierarchy. You can download the software from the [OutlookSpy](#) website. You must have a license to use it.

Option 2: MFCMapi

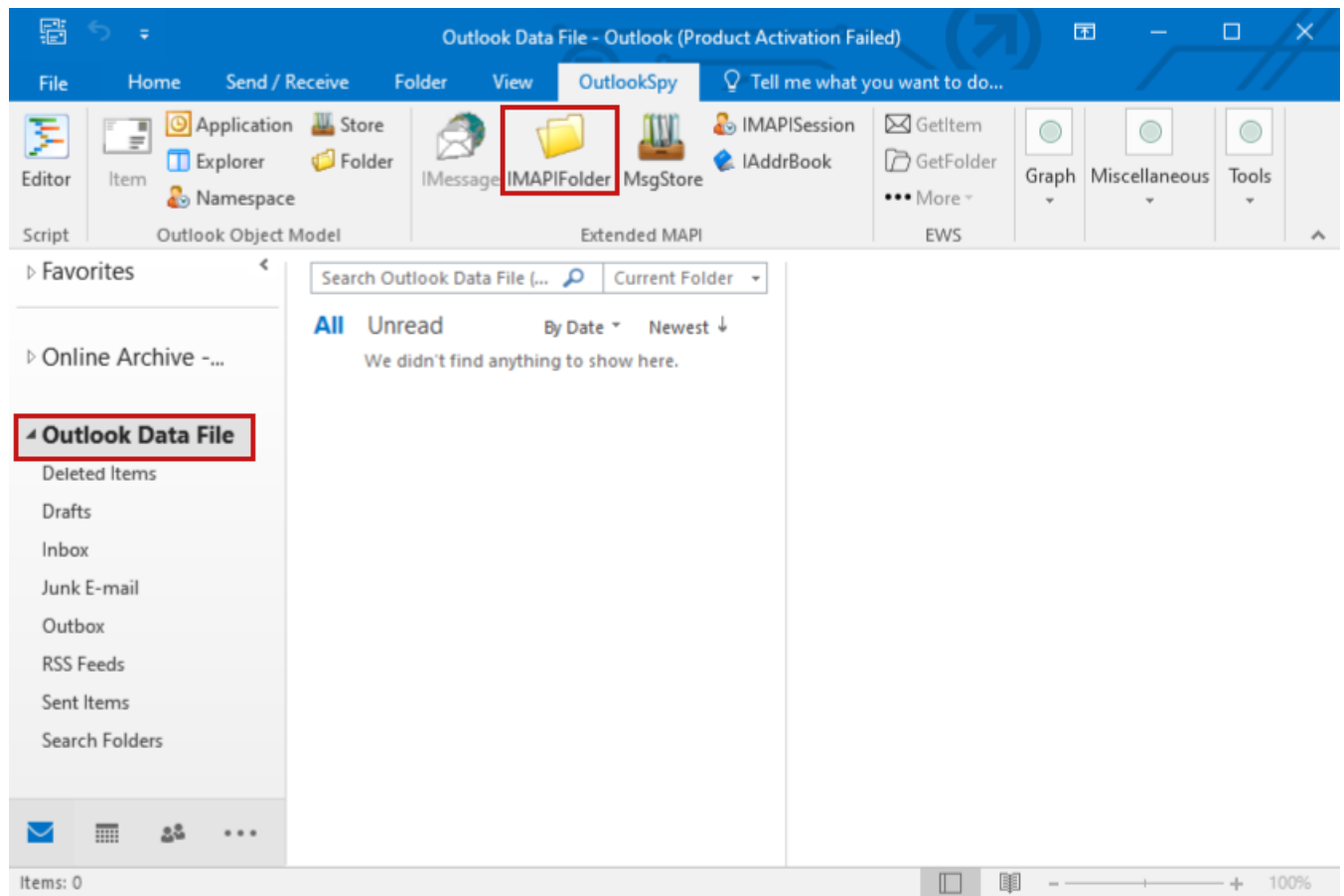
This tool provides access to Messaging Application Programming Interface (MAPI) stores. You can download the software from [GitHub](#). It is free.

Option 3: PST Walker

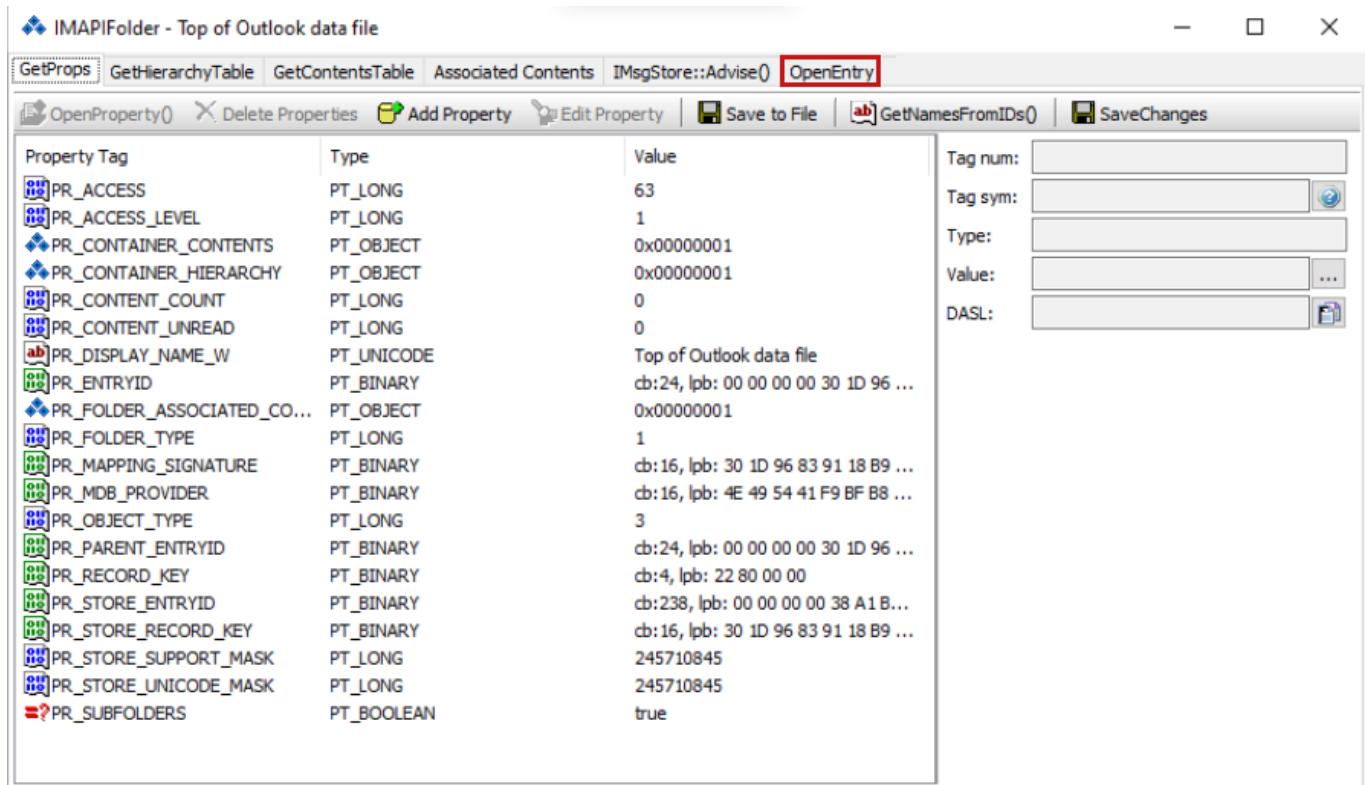
The PST Walker tool enables batch analysis of Entry IDs. Visit the [Apps Store](#) on the PST Walker website to download the software. You must have a license to use it.

25.5.7.3 Troubleshooting using Outlook with OutlookSpy

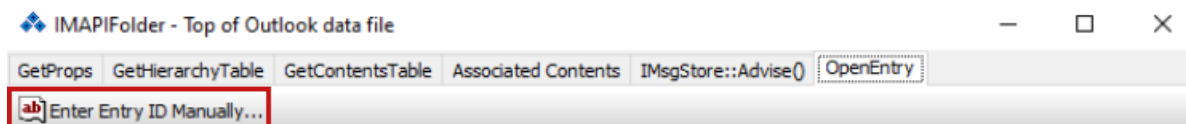
1. Open your local **Outlook**.
2. Click **File > Open & Export > Open Outlook Data File** and choose a .pst file to open.
On the left side of your screen, you will see a menu with the file's name.
3. Click the file name, and from the **OutlookSpy** tab, select **IMAPIFolder**.



4. Click the **OpenEntry** tab.

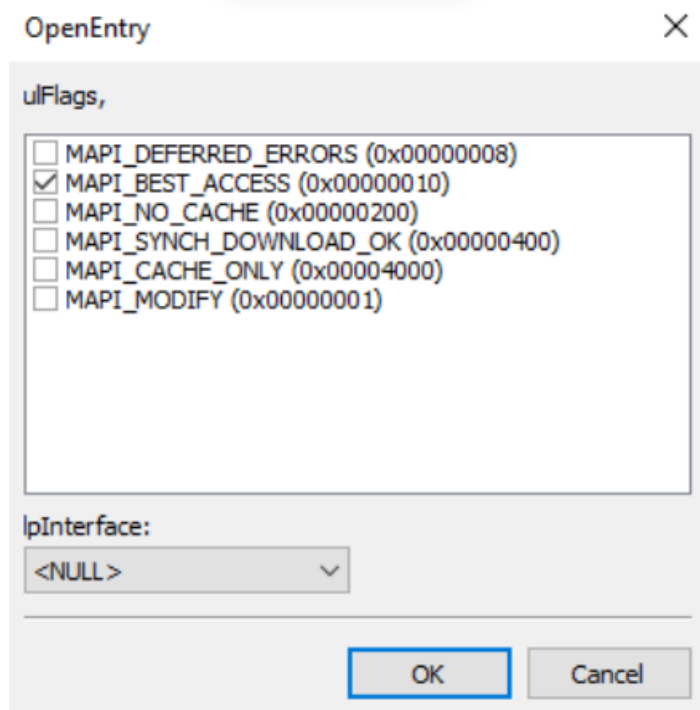


5. Click **Enter Entry ID Manually**.



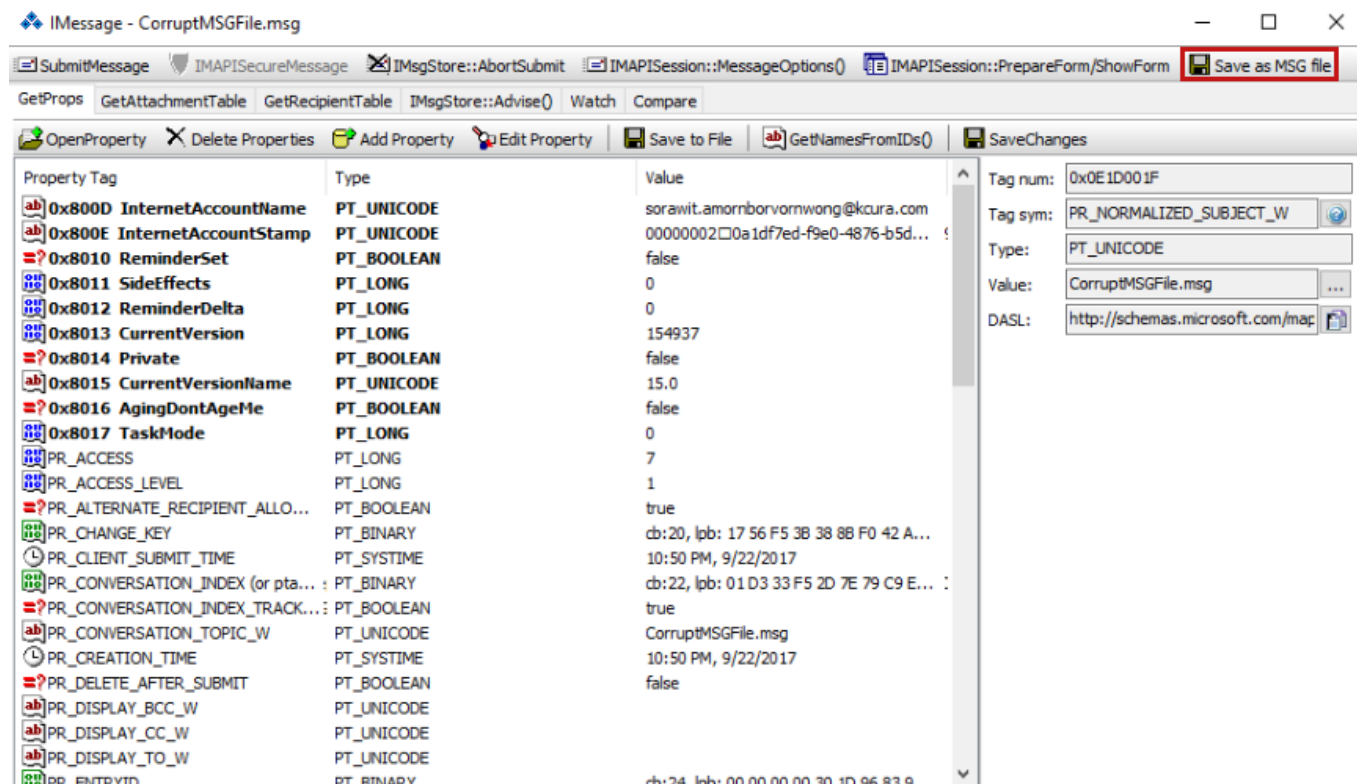
6. Enter the **Entry ID** in the text box.
7. Click **OK**.

- The OpenEntry flags screen opens. Keep the default **MAPI_BEST_ACCESS** option.



A screen displaying the .msg file's properties opens.

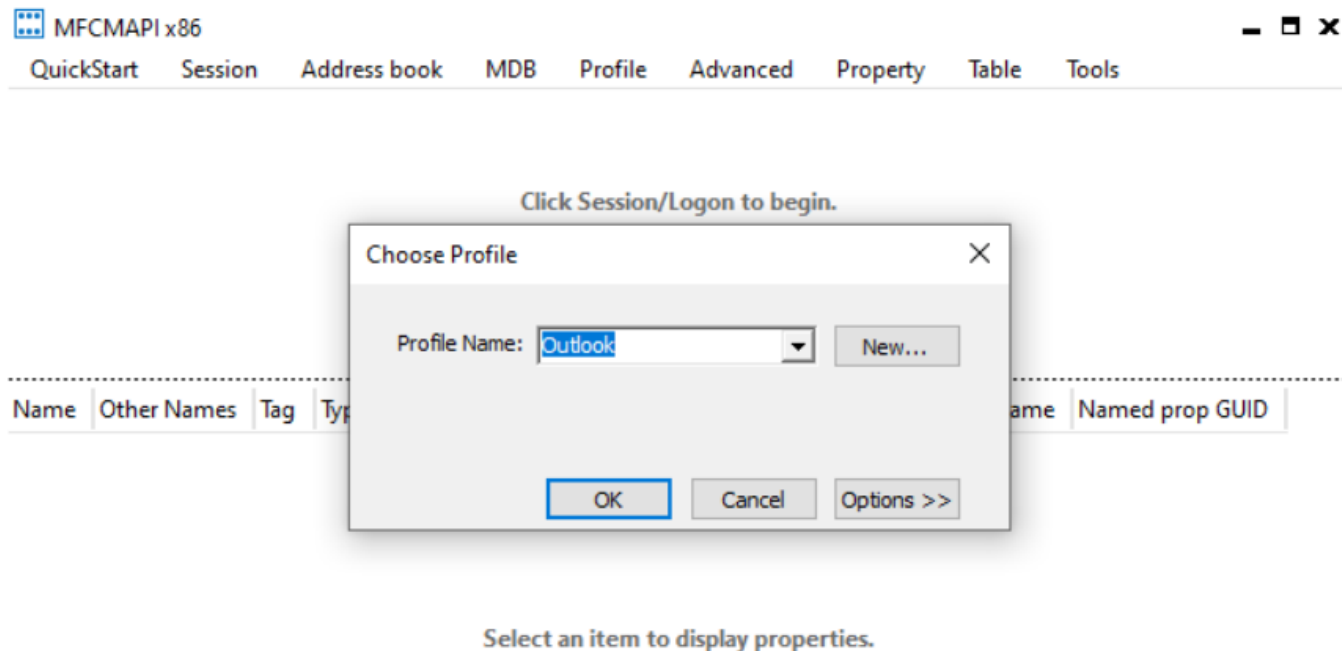
- Click **Save as MSG file** to export the properties.



25.5.7.4 Troubleshooting using MFCMapi

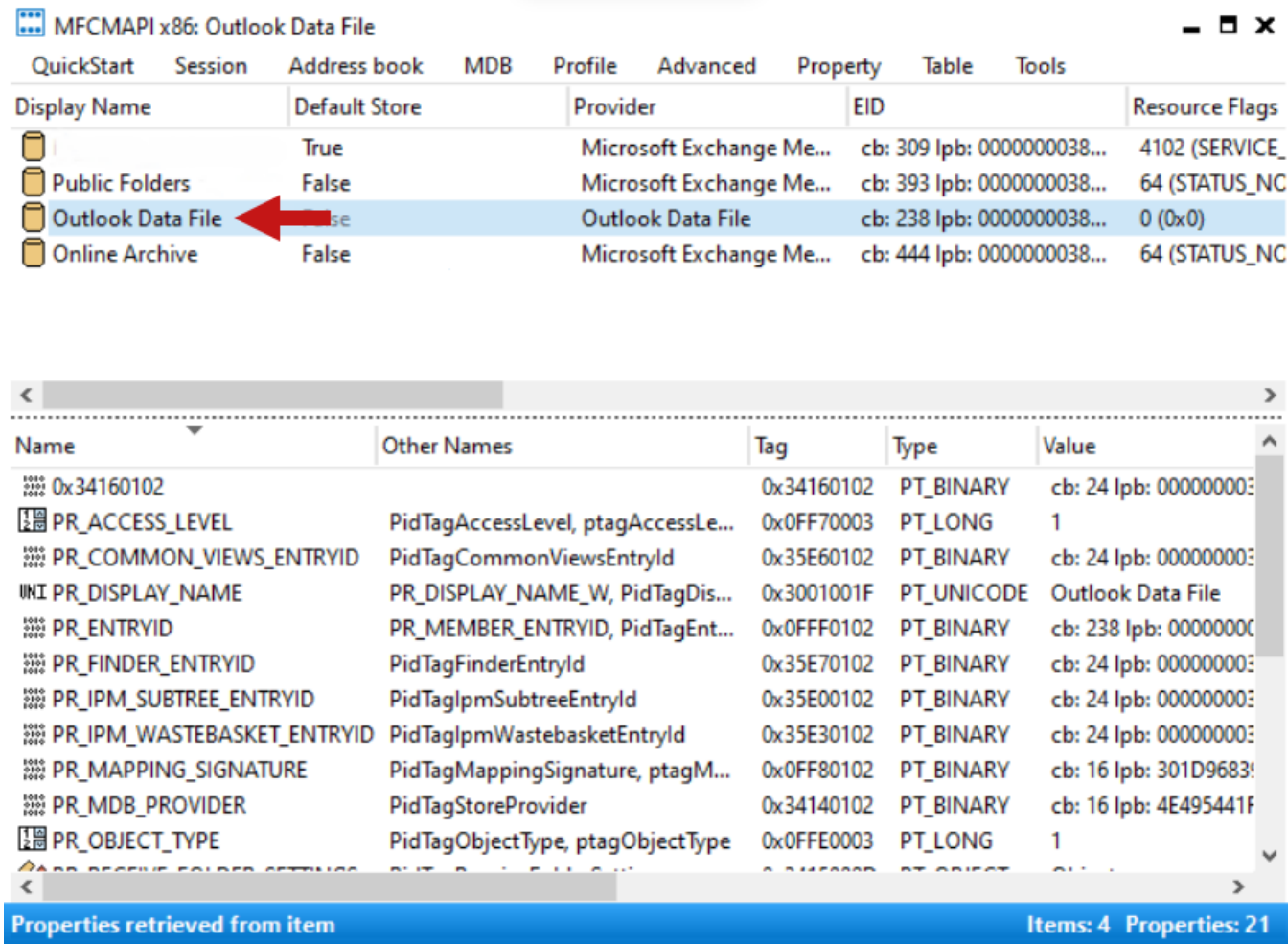
The first few steps mirror the OutlookSpy process.

1. Open your local **Outlook**.
2. Click **File > Open & Export > Open Outlook Data File** and choose a .pst file to open.
On the left side of your screen, you will see a menu with the file's name.
3. Open **MFCMapi** (downloaded from GitHub).
4. Click **Session > Logon...** .
5. Select your **Outlook profile**.



6. Click **OK**.
A screen displaying MAPI stores opens.

- Double-click the .pst store containing the files you want to work with.



The screenshot shows the 'MFCMAPI x86: Outlook Data File' window. The left pane lists the following items:

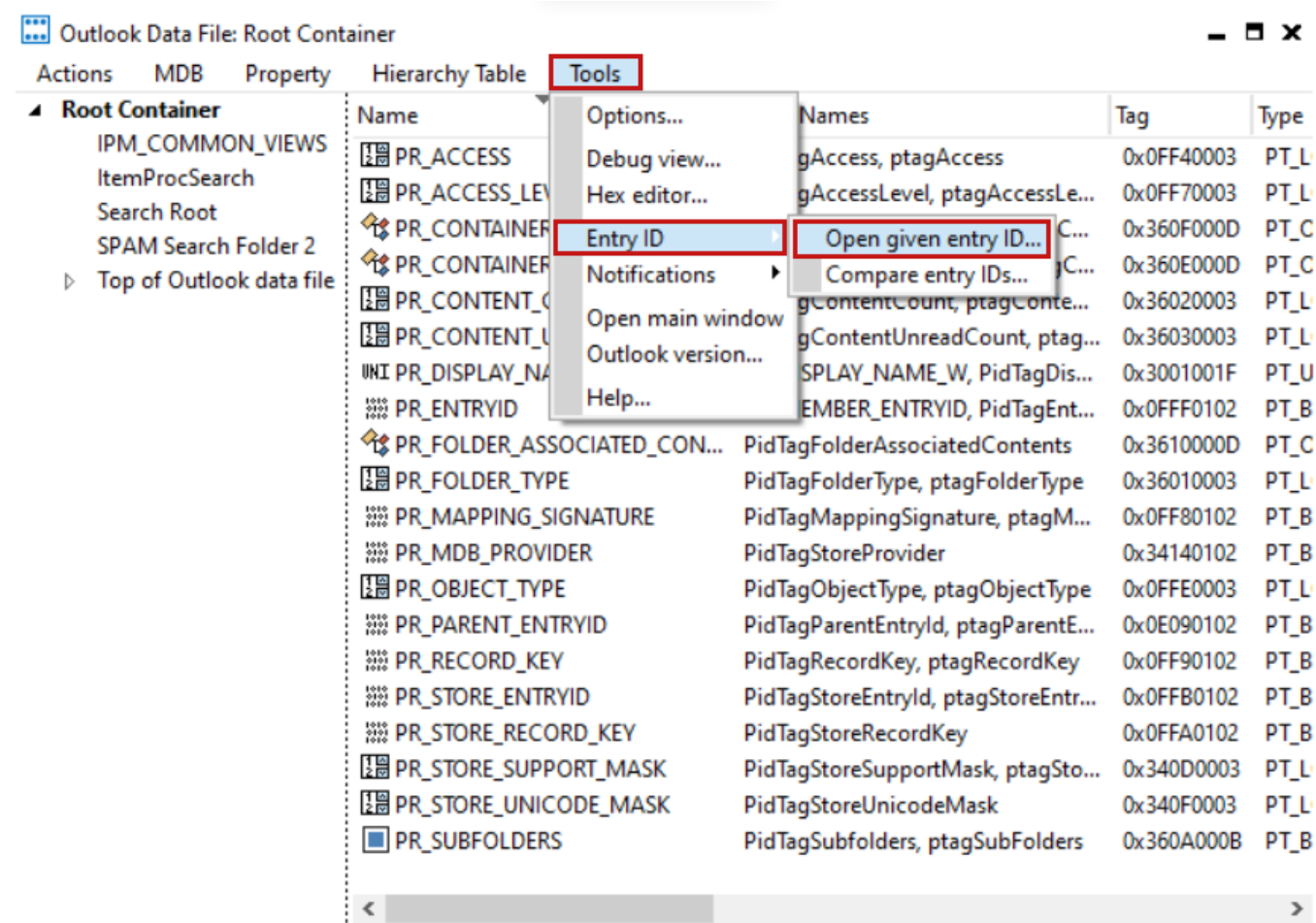
- Public Folders
- Outlook Data File (selected, indicated by a red arrow)
- Online Archive

The right pane displays the properties for the selected 'Outlook Data File' item:

Name	Other Names	Tag	Type	Value
0x34160102		0x34160102	PT_BINARY	cb: 24 lpb: 000000003
PR_ACCESS_LEVEL	PidTagAccessLevel, ptagAccessLe...	0x0FF70003	PT_LONG	1
PR_COMMON_VIEWS_ENTRYID	PidTagCommonViewsEntryId	0x35E60102	PT_BINARY	cb: 24 lpb: 000000003
PR_DISPLAY_NAME	PR_DISPLAY_NAME_W, PidTagDis...	0x3001001F	PT_UNICODE	Outlook Data File
PR_ENTRYID	PR_MEMBER_ENTRYID, PidTagEnt...	0x0FFF0102	PT_BINARY	cb: 238 lpb: 00000000
PR_FINDER_ENTRYID	PidTagFinderEntryId	0x35E70102	PT_BINARY	cb: 24 lpb: 000000003
PR_IPM_SUBTREE_ENTRYID	PidTagIpmSubtreeEntryId	0x35E00102	PT_BINARY	cb: 24 lpb: 000000003
PR_IPM_WASTEBASKET_ENTRYID	PidTagIpmWastebasketEntryId	0x35E30102	PT_BINARY	cb: 24 lpb: 000000003
PR_MAPPING_SIGNATURE	PidTagMappingSignature, ptagM...	0x0FF80102	PT_BINARY	cb: 16 lpb: 301D9683!
PR_MDB_PROVIDER	PidTagStoreProvider	0x34140102	PT_BINARY	cb: 16 lpb: 4E495441F
PR_OBJECT_TYPE	PidTagObjectType, ptagObjectType	0x0FFE0003	PT_LONG	1

At the bottom of the window, a status bar indicates: 'Properties retrieved from item' and 'Items: 4 Properties: 21'.

8. Click **Tools > Entry ID >** and then **Open given entry ID**.

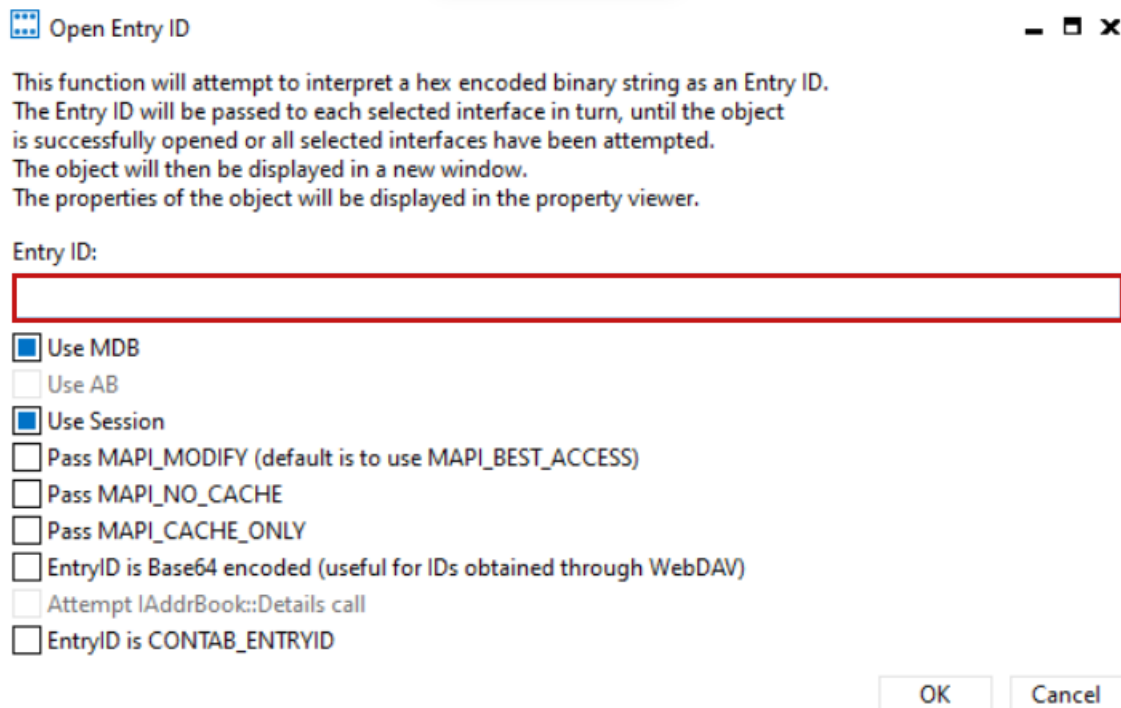


The screenshot shows the 'Outlook Data File: Root Container' window. The 'Tools' menu is open, and the 'Entry ID' option is selected, which has opened a sub-menu where 'Open given entry ID...' is highlighted. The main window displays a list of properties with columns for Name, Names, Tag, and Type.

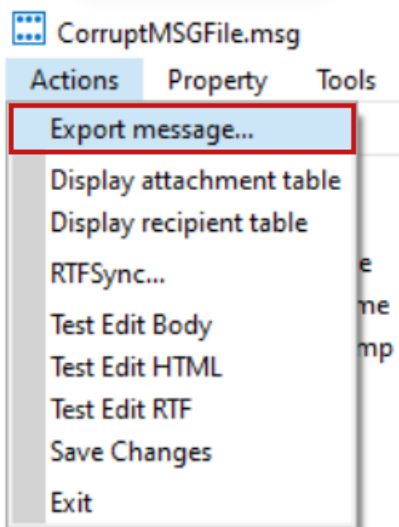
Name	Names	Tag	Type
PR_ACCESS	gAccess, ptagAccess	0x0FF40003	PT_L
PR_ACCESS_LEVEL	gAccessLevel, ptagAccessLe...	0x0FF70003	PT_L
PR_CONTAINER	C...	0x360F000D	PT_C
PR_CONTAINER	C...	0x360E000D	PT_C
PR_CONTENT_COUNT	gContentCount, ptagConte...	0x36020003	PT_L
PR_CONTENT_UNREAD_COUNT	gContentUnreadCount, ptag...	0x36030003	PT_L
PR_DISPLAY_NAME	SPLAY_NAME_W, PidTagDis...	0x3001001F	PT_U
PR_ENTRYID	EMBER_ENTRYID, PidTagEnt...	0x0FFF0102	PT_B
PR_FOLDER_ASSOCIATED_CONTENTS	PidTagFolderAssociatedContents	0x3610000D	PT_C
PR_FOLDER_TYPE	PidTagFolderType, ptagFolderType	0x36010003	PT_L
PR_MAPPING_SIGNATURE	PidTagMappingSignature, ptagM...	0x0FF80102	PT_B
PR_MDB_PROVIDER	PidTagStoreProvider	0x34140102	PT_B
PR_OBJECT_TYPE	PidTagObjectType, ptagObjectType	0x0FFE0003	PT_L
PR_PARENT_ENTRYID	PidTagParentEntryId, ptagParentE...	0x0E090102	PT_B
PR_RECORD_KEY	PidTagRecordKey, ptagRecordKey	0x0FF90102	PT_B
PR_STORE_ENTRYID	PidTagStoreEntryId, ptagStoreEntr...	0x0FFB0102	PT_B
PR_STORE_RECORD_KEY	PidTagStoreRecordKey	0x0FFA0102	PT_B
PR_STORE_SUPPORT_MASK	PidTagStoreSupportMask, ptagSto...	0x340D0003	PT_L
PR_STORE_UNICODE_MASK	PidTagStoreUnicodeMask	0x340F0003	PT_L
PR_SUBFOLDERS	PidTagSubfolders, ptagSubFolders	0x360A000B	PT_B

At the bottom of the window, a status bar displays: 'Display dialog to enter Entry ID to call OpenEntry with' and 'Contents: 0 Assoc. Contents: ? Properties: 20'.

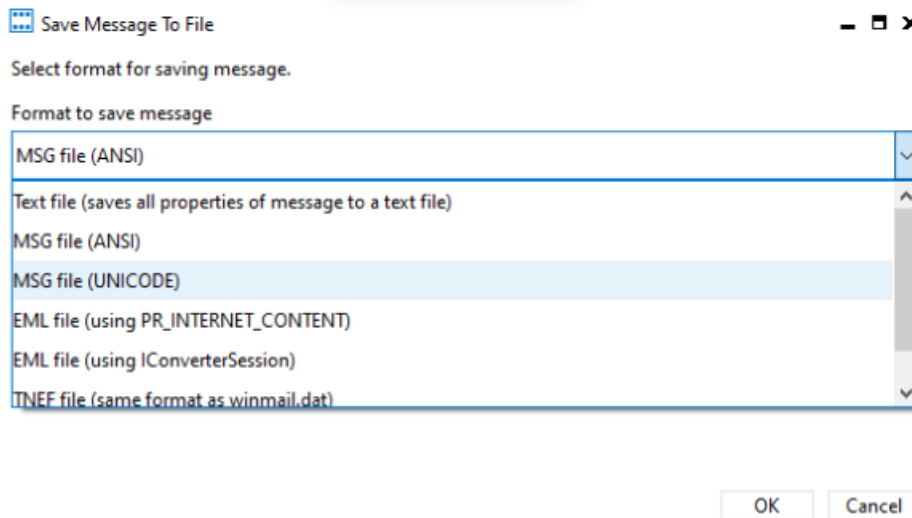
9. Enter the **Entry ID** in the Entry ID text box.



10. Click **OK**.
You see a new window displaying the file's data.
11. To export the data, click **Actions > Export message**.



12. Select a format.



13. Click **OK**.

25.5.7.5 Corrupt PDF files not flagged during discovery

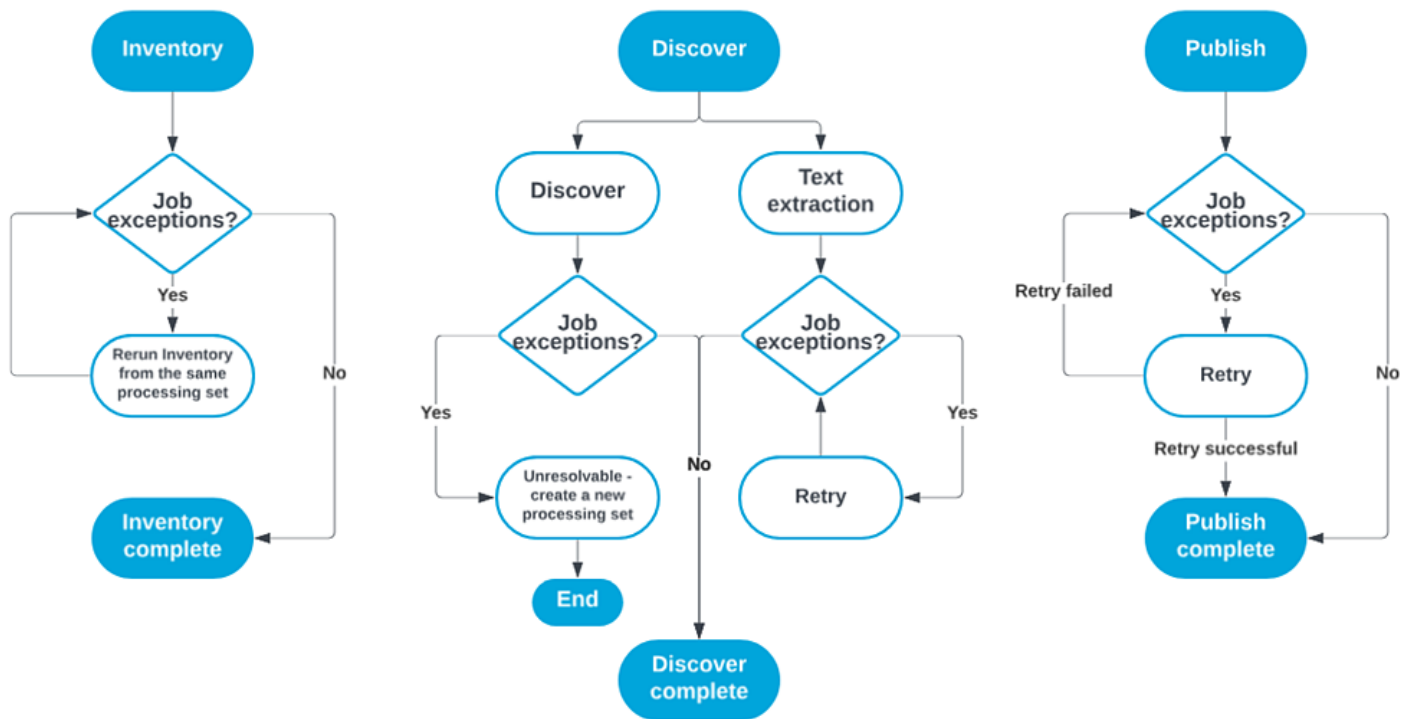
In some cases, corrupt PDF files are not flagged during discovery and therefore do not appear in the File Exceptions tab. During discovery, Relativity attempts to identify the file type, extract basic metadata, and extract text when possible, but it does not validate PDFs against external viewer standards. If the file structure is sufficiently readable and the corruption does not interfere with discovery or text extraction, the file may be successfully discovered and processed through publishing, even if it is otherwise damaged and cannot be opened in an external PDF viewer.

25.6 Job errors

Navigate to the **Job Errors** tab to view a list of job errors. The **Current Files with Exceptions** view displays all outstanding job errors. The **All Files with Exceptions** view displays all job errors across time, regardless of their resolution. Review the **Exception Status** and **Message** fields for an overview of the error and the potential for remediation.

25.6.1 Job error workflow

The following diagram shows the job error workflows for Inventory, Discover, and Publish jobs.



25.6.2 Resolving job errors

The steps for resolving job errors depend on the phase running when the error occurred. For some phases, errors are unresolvable; others require creating and running a new processing set. You can resolve some errors using mass actions or by retrying the errors from the processing set. The first step in resolving job errors is determining the phase in which the error occurred.

To identify the phase where the error occurred, perform the following steps:

1. Navigate to the **Job Errors** tab.
2. Click the **Error Identifier** for the error you are researching. The Job Error Details pane opens.
3. Click the **Advanced** tab.

4. Locate the **Data** section, then locate the **Relativity.Processing.Error.Phase** entry key. The key value is the phase where the error occurred. In the image below, the phase is Inventory.

Job Error Details Advanced

```
Stack Trace <?xml version="1.0" encoding="utf-16"?>
<Exceptions>
<Exception>
<Version>
<AssemblyVersion>7.3.1741.1</AssemblyVersion>
<ProductVersion>7.3.22570-bfc6305ed8406fec2abce8d4529b6396e342fb6f</ProductVersion>
</Version>
<Timestamp>2026-03-12T18:43:25.279</Timestamp>
<Type>System.Exception</Type>
<Message>There was an error during expansion of this PDF file. It may be corrupted. Consider repairing or replacing the file and retrying.</Message>
<Worker>p001wrk009PNM/18</Worker>
<JobId>49130</JobId>
<DispatchType>Invariant.InventoryHandler</DispatchType>
<DispatchMethod>InventorySubJob</DispatchMethod>
<Flags>32780</Flags>
<Unresolvable>True</Unresolvable>
<FileExtension>PDF</FileExtension>
<FileSize>49026</FileSize>
<InnerException>
<Timestamp>2026-03-12T18:43:25.279</Timestamp>
<Type>System.Exception</Type>
<Message>The PDF file is encrypted and a password was incorrect or not supplied.</Message>
<Source>Invariant.Rendering</Source>
<Worker>p001wrk009PNM/18</Worker>
<Data>
<Entry Key="HRESULT" Value="-2147467259" />
<Entry Key="ErrorCode" Value="4" />
<Entry Key="Relativity.Processing.Error.Category" Value="4" />
</Data>
</InnerException>
<FileName>[REDACTED]_0010005.pdf</FileName>
<StoredAs>\\[REDACTED]
[REDACTED]files.[REDACTED]relativity.one\T017\Files\EDDS1021749\Processing\1101988\INV1021749\PROCESSING_WOR
8fb00e93-8703-4009-9e3b-65f942286dd2.PDF</StoredAs>
<VirtualPath>[REDACTED]\Loose Files</VirtualPath>
<LogicalPath>
</LogicalPath>
<ImportSource>
</ImportSource>
<Data>
<Entry Key="Filename" Value="[REDACTED]_0010005.pdf" />
<Entry Key="StoredAs" Value="[REDACTED]
[REDACTED]files.t017.esus025192.relativity.one\T017\Files\EDDS1021749\Processing\1101988\INV1021749\PROCESSING_WOR
8fb00e93-8703-4009-9e3b-65f942286dd2.PDF" />
<Entry Key="VirtualPath" Value="[REDACTED]\Loose Files" />
<Entry Key="Relativity.Processing.Error.Category" Value="5" />
<Entry Key="Relativity.Processing.Error.Phase" Value="Inventory" />
</Data>
</Exception>
</Exceptions>
```

25.6.2.1 Inventory

All inventory errors are unresolvable. You must address the error outside of Relativity, then rerun Inventory from the processing set.

To re-inventory files, perform the following steps:

1. From the processing set, click the **Inventory Files** button in the **Process Files** console.
2. A confirmation message appears. Click **Yes**.

When you rerun the Inventory phase, Relativity clears any errors from the previous job.

For detailed information on the Inventory phase and errors, see [Inventory processing](#).

Note: When working with questionable data sets, run Inventory to get a pre-assessment of corrupt files. This way, you can review the error messages returned for possible remediation before continuing to other processing phases.

25.6.2.2 Discover

The Discover phase includes file discovery and text extraction. You should treat errors that occur during the Discover phase with high priority, as they are generally unexpected. Resolving these errors depends on where they occurred during the discovery process. Not resolving Discover errors may result in incomplete metadata when the files are published.

To address errors that occur during file discovery, perform the following steps:

1. If possible, retry the errors from within the processing set.
2. If the errors are unresolvable, create a new processing set using the same data source as the errored set.
3. Rerun Discover.

To address job errors that occur during text extraction, perform the following steps:

1. From the processing set, click the **Job Errors** link in the **Process Files** console.
2. You can retry errors with a status of **Ready to retry**. From the mass actions drop-down menu, select **Retry Errors**.
3. On the Retry error(s) confirmation dialog, click **Retry**.

The status changes to Resolving. You can refresh the page until the status updates with either Resolved or Not Resolved. If the error is not resolved, you must rediscover the files from a new processing set.

For detailed information on discovering files, see [Discovering files](#).

25.6.2.3 Publish

Job exceptions during the publishing phase that are resolvable have an exception status of **Ready to retry**. Retry the exception from the same processing set.

For detailed information on the Publish phase and exceptions, see [Publishing files](#).

26 Processing duplication workflow

The Processing Duplication Workflow is a way to identify and mark duplicate documents after discovery. This way you can review only one copy of each document, while leaving all copies of the document in place across the workspace. You can control how documents are published through the processing profile selected when creating a processing set. You can also create your own processing profile, then choose the deduplication method (global, custodial, or none) to use.

- **Global deduplication**—documents that are duplicates of documents that were already published to the workspace in a previous processing set aren't published again. For more information, see [Deduplication considerations > Global deduplication](#).
- **Custodial deduplication**—documents that are duplicates of documents owned by the custodian specified on the data source aren't published to the workspace. For more information, see [Deduplication considerations > Custodian deduplication](#).

- **No deduplication**—all documents and their duplicates are published to the workspace. For more information, see [Deduplication considerations > No deduplication](#).

Note: The scripts described in this topic help identify duplicate files and their source locations. Use the scripts if you have deduplication set to *No Deduplication* or are working with files uploaded outside of Relativity. You should not use these scripts if you are running deduplication within Relativity processing.

26.1 Downloading and installing the solution to your instance

Before you begin, confirm you have the Processing Duplication Workflow application in your environment's application library. If you do, you can skip to section, [Adding the solution to a workspace](#). If not, download the solution from the Community site, then deploy it to your environment. After it's deployed, you can add it to any workspaces within your environment.

26.1.1 Supported versions

Click the following link to access the solution files on the Community site.

Note: Some versions of this application may not be eligible for support by Relativity Customer Support. For more information, see the Version support policy.

Solution version	Supported Relativity version
2024.2.0.1	All supported versions of Relativity.

Note: You must have valid Relativity Community credentials to download files from the Community site. After accessing the **ProcessingDuplicationWorkflow** page, click the **Download** tab. If the file successfully downloads to your local drive, you will not see any other dialog. If you see an error stating "URL No Longer Exists," it may be due to a single sign-on error related to the SAML Assertion Validator, and you should contact your IT department.

26.1.2 Components

This custom solution consists of the following components:

- The Processing Duplication Workflow application.
- Item-level and family-level scripts that run at the workspace level. The scripts are described in greater detail in the [Before you run the scripts](#) section. They include:
 - All Custodians (item-level and family-level)
 - All Source Locations (item-level and family-level)
 - Update Duplicate Status (item-level and family-level)

26.1.3 Considerations

- This script should only be run by a system administrator. If you are not a system administrator, you should not run this script.
- The solution may require you to create fields in your environment. Be sure to read the section [Before you run the scripts](#) to confirm you have the necessary fields and saved searches in place.
- You have the option of tagging a document as Responsive and then propagating that value to the document's family members. For more information, see [Applying propagation to documents](#).

26.2 Deploying and configuring the solution

After downloading the solution from the Community site, you must deploy it to your environment's Application Library. To deploy the solution to your environment:

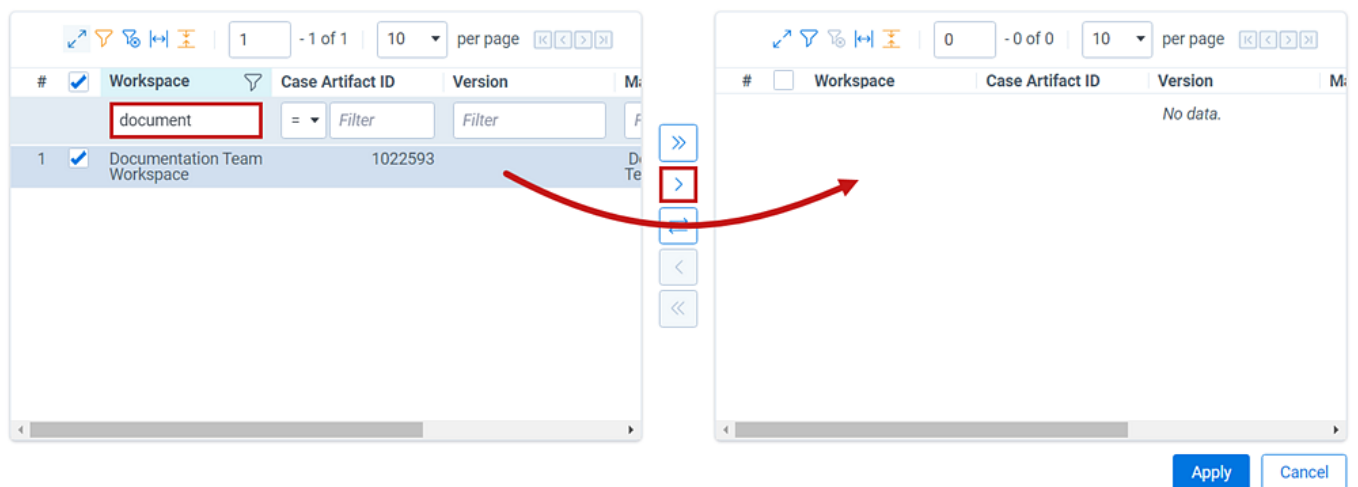
1. Use the search bar to navigate to your environment's **Application Library**.
2. Click the **New Library Application** button.
3. Next to **Application File**, click **Select File**.
4. Navigate to and select the **ProcessingDuplicationWorkflow.rap** file you downloaded from the Community site.
5. Click the **Save** button.

26.3 Adding the solution to a workspace

The Processing Duplication Workflow solution is already deployed to all instances. To install the solution on a workspace, perform the following:

1. Use the search bar to navigate to the **Application Library** in the Admin environment.
2. Use the **Name** filter to locate the **Processing Duplication Workflow** application.
3. Select the application to open the application information page
4. Locate the **Workspaces Installed** section and click **Select**.
5. Use the Workspace filter to locate your target workspace(s), then use the **Move Selected Left to Right** arrow to lock in your selection(s).

Select Items - Select Workspaces



6. Click **Apply**.
You are redirected back to the Workspaces Installed screen. When the installation complete, you should see

your workspace with a status of Installed.

Application Information

Application Artifacts

Record History

Name

Processing Duplication Workflow

User-Friendly URL

Application File

Processing_Duplication_Workflow.rap

Developer

Version

2024.1.2.1

Is Global Application

No

Schema Version

2024.1.0.0

Status

Installed

Release Notes

Workspaces Installed

Select

Upgrade Installed Workspaces

Cancel Installs

1 - 3 of 3 10 per page

Workspace	Schema Version	Status	Message	Case Artifact ID	Installed By	Installed On
Demo Workspace	2024.1.0.0	Installed		1038734	Service Account, Relativity	4/23/2024, 7:31 PM
RelativityOne Quick Start Templ...	2024.1.0.0	Installed		1037412	Service Account, Relativity	4/23/2024, 7:31 PM
Documentation Team Workspace	2024.1.0.0	Installed		1022593	Relativity User	9/19/2024, 9:54 PM

26.4 Before you run the scripts

To avoid delays later, take a few minutes to confirm you have the prerequisites required by the scripts. These include required fields, populated entities, and saved searches. If you have one or more missing items, create them now before proceeding.

26.4.1 Fields

The following is a list of all the fields used in the Processing Duplication Workflow scripts, along with their equivalent Relativity fields. We recommend you confirm that the Relativity field equivalents exist before proceeding. If they do not, use the field type and associative objects to create them.

Field name in script	Relativity field equivalent	Used in script	Field type	Associative Object	Field purpose
Saved Search	[Your Saved Search Name]	All scripts	Single Object	Search	This field serves two purposes. First, it defines the data set you want to run the script against. Second, Relativity updates the fields with the results of the script after you run it. You can create one saved search for both item-level and family-level scripts since the results output have the same fields for both.

Field name in script	Relativity field equivalent	Used in script	Field type	Associative Object	Field purpose
Custodian Object	Custodian	All Custodians - Item All Custodians - Family	Single Object	Entity	The custodian associated with a file. More than one custodian can be associated with a single file.
Custodian Object Field	Full Name	All Custodians - Item All Custodians - Family	Fixed-Length Text (255)	Entity	This processing field displays the first and last name of the custodian.
Destination Field (output)	All Custodians (Long Text)	All Custodians - Item All Custodians - Family All Source Locations - Item All Source Locations - Family	Long Text	Document	When used with the All Custodians scripts, this field displays a semi-colon delimited list of all custodians associated with a file.
Destination Field (output)	All Paths/Locations	All Custodians - Item All Custodians - Family All Source Locations - Item All Source Locations - Family	Long Text	Document	When used with the All Source Locations scripts, this field displays a semi-colon delimited list of the source locations for a file.
Duplicate Hash Field	Processing Duplicate Hash or EDRM MIH . See the EDRM MIH topic for more information on using the EDRM hash.	All scripts	Fixed-length Text (64)	Document	The identifier of the physical native file. Alternatively, you can select another hash field in its place, such as MD5 or SHA1.
Duplicate Sort Order Field	Custodian Sort Order	Update Duplicate Status - Item Update Duplicate Status - Family	Whole Number	Entity	This is an option field that when selected, sorts the script results according to the custodian sort order rank. To use this field, you must manually assign a

Field name in script	Relativity field equivalent	Used in script	Field type	Associative Object	Field purpose
					sort order value to the custodians in your saved search. See Other considerations for details on how to assign a sort order to custodians.
Duplicate Status Field	Duplicate Status or EDRM Duplicate Status See the EDRM MIH topic for more information on using the EDRM hash.	Update Duplicate Status - Item Update Duplicate Status - Family	Single Choice	Document	Displays one of three values: Unique, Master, or Duplicate based on the relational field (hash.) Unique files have a relational ID not shared with any other files. Master files have a relational ID shared by more than one file where the ID is the lowest of the files. Duplicate files have a relational ID shared by more than one file where the IDs are not the lowest in the files.
Family Identifier Field	Family Group	All Custodians - Family All Source Locations - Family Update Duplicate Status - Family	Fixed-Length Text (64)	Document	Identifies the family group a file belongs to.
Level Field	Level	All Custodians - Family Update Duplicate Status - Family	Whole Number	Document	Numeric value that represents how deeply nested a file is within a family. The higher the number, the deeper the file is nested.
Source Field	Source Path	All Source Locations -	Long Text	Document	The location of the file in the data set.

Field name in script	Relativity field equivalent	Used in script	Field type	Associative Object	Field purpose
		Item All Source Locations - Family			

26.4.2 Saved searches

Saved searches have two functions when using the duplication scripts:

- They identify the data set you are running the script against.
- After the script runs, it updates the saved search so that you can identify duplicate documents and their locations.

You can create three saved searches, one for each script type, and use them for both item and family levels.

26.4.2.1 All custodians

Create a saved search with the following fields:

- **Control Number**—the document identifier.
- **Custodian**—the custodian associated with the document.
- **Family Group**—the relational field that defines groups of related documents.
- **All Custodians (Long Text)**—the script uses this field to output a semicolon delimited list of custodians associated with the document or relational group (for family-level results).

26.4.2.2 All source locations

Create a saved search with the following fields:

- **Control Number**—the document identifier.
- **Source Path**—the location of the document in the data set.
- **Family Group**—the relational field that defines groups of related documents.
- **Level**—the nested level of the document. This field is useful when running the family-level script and tells you how deeply nested the document is.
- **All Paths/Locations**—the script uses this field to output a semicolon delimited list of source locations associated with the document.

26.4.2.3 Update duplicate status

Create a saved search with the following fields:

- **Artifact ID**—the unique identifier of the database object in Relativity.
- **Control Number**—the document identifier.
- **Processing Duplicate Hash**—the identifier of the physical native file, such as MD5 or SHA1.
- **Custodian**—the custodian associated with the document.
- **Level**—the numeric value that represents how deeply nested a file is within a family. The higher the number,

the deeper nested.

- **Duplicate Status**—the script updates this field with one of the following options: Unique, Master, Duplicate.

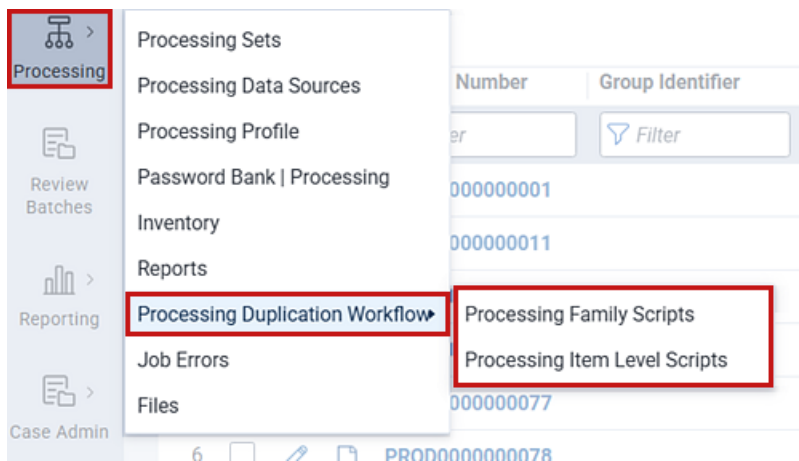
26.4.3 Other considerations

Finally, there are a few other housekeeping items to consider.

- **Script access and visibility**—after installing the Processing Duplication Workflow application, Relativity adds links to the scripts in the workspace's general menu. You can access the menu by clicking the hamburger link at the bottom of the left column. However, if you intend to use the scripts often and want a quicker way to access them, read the section on [Enabling the workflow tabs](#).
- **Populated entities**—if you uploaded your data from another solution outside of Relativity, you should confirm you have populated entities. If you have imported data from a solution outside of Relativity, read the section on [Populating entities for data uploaded outside of Relativity](#).
- **Custodian sort order**—(optional) some scripts give you the option of ranking the results based on the custodian's sort order. To use this setting, custodians must have a sort order value, which you manually assign to the custodian's entity object. For example, Custodian A has a sort order of 10, while Custodian B's sort order is 20. If everything else is equal between the two custodians, Custodian A will rank higher in the results than Custodian B, based on the sort order value. If you want to use the sort order to rank custodians, read the section on [Setting a custodian's sort order](#).

Enabling the workflow tabs

During the installation, Relativity creates three new tabs: **Processing Duplication Workflow**, **Processing Item Level Scripts**, and **Processing Family Scripts**. The tabs are not visible in the left-side menu by default.



To enable workflow tab visibility:

1. Open the workspace where you installed the Processing Duplication Workflow application.
2. Use the search bar to navigate to the **Tabs** page.

- Use the filter on the **Name** field to locate the processing tabs.

Workspace Details Search Indexes Object Type Fields Choices Field Categories Layouts Views Tabs				
> 🔍 New Tab All Tabs				
#		Name	Tab Order	Parent
		proc	=	(All)
1	☐	Processing	200	
2	☐	Processing Sets	10	Processing
3	☐	Processing Data Sources	15	Processing
4	☐	Processing Profile	20	Processing
5	☐	Password Bank Processing	40	Processing
6	☐	Processing Duplication Workflow		Processing
7	☐	Processing Item Level Scripts		Processing Duplication Workflow
8	☐	Processing Family Scripts		Processing Duplication Workflow

- Edit the **Processing Duplication Workflow** tab.
- In the **Tab Location** section, select **Processing** as the **Parent** value.
(Optional) Change the **Order** value to **100**. Keep the remaining default values.

Tab Information

Name* Processing Duplication Work

Tab Type Parent

Visible ☒

Relativity Applications

Select

Tab Location

Show in Sidebar* ☐

Parent Processing

Order* 100

- Click **Save and Back** to set your changes.
- Edit the **Processing Item Level Scripts** tab.

8. Select **Processing Duplication Workflow** as the **Parent** value.
(Optional) Change the **Order** to **100**.

Tab Location

Show in Sidebar* ☐

Parent - Processing Duplication Workflow ▼

Order*

9. Click **Save and Back**.

10. Use the preceding steps to enable the **Processing Family Scripts** tab.

You are now ready to run the solution against a set of documents.

Populating entities for data uploaded outside of Relativity

Entities are the individuals communicating within a data set. For example, custodians. When you upload data outside of Relativity, you may not have entities created. If you uploaded data with Relativity, the entities are created automatically. If you do not have any entities, use one of the following methods to create or import them into Relativity:

- Manually create an entity from the Entities tab. For more information, see [Creating and editing an entity](#).
- Load and sync entities with Integration Points. For more information, see [Integration Points](#).
- Load entities with Import/Export. For more information, see [Importing entities through Import/Export](#).
- Run Analytics name normalization to identify entities from email headers. For more information, see [Name normalization](#).
- Manually create a key entity in the Review Interface.
 - A key entity is one that has the multi-choice field Classification set to either Key Person or Key Organization on the Entity object.
 - These values are set automatically when you create an entity with either one of the respective views (Key People and Key Organization.) For more information, see [Case Dynamics](#).

Setting a custodian's sort order

Use the Custodian Sort Order field to add a sort order value to custodians in your data set. You can use this field in scripts and saved searches to rank results by the field's value. The lower the value, the higher the rank.

To summarize the process of adding the sort value:

1. Create a new field based on the Entity object.
2. Create a new layout with the new field.
3. Edit a custodian and change the layout so you can see the sort order field.
4. Add the sort order value and save.

Step-by-step instructions:

1. Use the search bar to navigate to the **Fields** tab.
2. Create and save a new field with the following settings:
 - **Name**—Custodian Sort Order
 - **Object Type**—Entity
 - **Field Type**—Whole Number

3. Use the search bar to navigate to the **Layouts** tab.
4. Create a new layout with the following settings:
 - **Object Type**—Entity
 - **Name**—Custodian Sort Order Layout
 - **Order**—1000

Layout Information

Object Type* Entity ▼

Name* Custodian Sort Order Layout

Order* 1000

Overwrite Protection Disabled

Relativity Applications Select

5. Click **Save**.
After you click Save, the screen reloads and the **Build Layout** button is visible in the **Build Layout** console on the right side of the screen.
6. Click the **Build Layout** button.
7. From the list of fields in the **Layout Options** console, add the **Custodian Sort Order** field to the layout canvas.

Edit Layout Information: Custodian Sort Order Layout 1121

Save Save and Close Close

Default Category

Full Name:

Custodian Sort Order:

Drop Field

Drop Field

Drop Field

Drop Field

LAYOUT OPTIONS

Properties ▼

Field: Custodian Sort Order 1121

Show Field Name: ☒ Yes ☐ No

Custom Label: ...

Add Category ▼

Fields

Custom Text

All Custodians

Custodian Sort Order

DeDuplicated Custodians

8. Click **Save and Close**.
Next, you will assign a sort order value to the appropriate custodians.
9. Use the search bar to navigate to the **Entities** tab.
 - Use the filters to find the appropriate custodian record to edit.
 - You do not have to assign a sort order value to all custodians. For example, you will create a saved search that defines the data set for the script you are running.
 - You might assign sort order values to only the custodians in the saved search to save time.

10. **Edit** a custodian record.
11. Change the layout to the **Custodian Sort Order Layout**.
If you see a message about leaving the page, click **OK**.

Processing Entity ▾ 

Processing Entity

- Custodian Sort Order Layout**
- Collect Custodian
- Custodian Deduplication Order

First Name*

Last Name*

Classification ☒ Custodian – Processing [Manage](#)

Document numbering prefix

Notes

Save **Save and Previous** **Save and Next**

12. Add a numeric value for the **Custodian Sort Order** field.
Lower numbers equate to higher custodian rankings.

Default Category

Full Name*

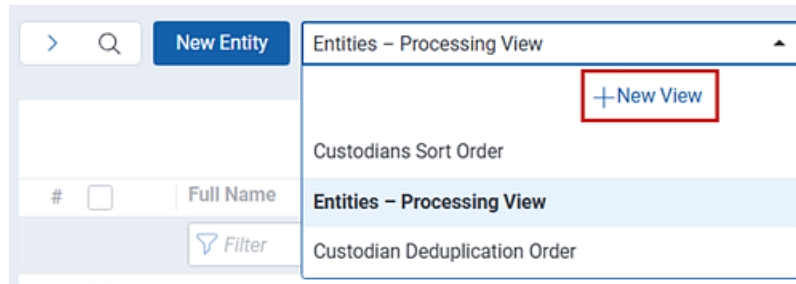
Custodian Sort Order

13. Click **Save and Back**.
Repeat these steps for all appropriate custodians.

Next, you will create a new view to display the custodians and their assigned sort order values. You can review the list and confirm the sort order values are correct for your needs.

To create a new view:

1. Still on the **Entities** tab, select **New View** from the **Views** drop-down menu.



2. Name the view **Custodian Sort Order View**.
3. Add the following fields: **Full Name**, **Custodian Sort Order**.

View

Information * Other

Object Type* Entity

Order* 9999 View Order

Name* Custodian Sort Order View

Dashboard Choose...

Owner* Public Me

Fields * Conditions Sort

Unselected

- custod
- All Custodians::Relativity Image Count
- All Custodians::Unified Title
- Custodian Sort Order
- DeDuplicated Custodians
- DeDuplicated Custodians::Level

Selected

- Filter
- Full Name
- Custodian Sort Order

Save Cancel

4. Click **Save** to add the view.
5. Test the view by selecting the **Custodian Sort Order View** from the **View** drop-down menu. You see a list of all custodians and their sort order assignments. Some custodians may not have a value if you did not assign one.

#	☐	Full Name	Custodian Sort Order
		Filter	= Filter
1	☐	Howard, Johnson	10
2	☐	Rosen, Pam	
3	☐	Porter, David V	20
4	☐	Newlin, Ross	
5	☐	Stinebaugh, John	
6	☐	Hoff, Jonathan	

6. To view a list of custodians having a sort order assigned, use the filter on the **Custodian Sort Order** field to eliminate custodians not having an assigned value.

#	☐	Full Name	Custodian Sort Order
		Filter	<input <input="" type="text" value="10"/>
1	☐	Howard, Johnson	10
2	☐	Porter, David V	20
3	☐	Smith, Bob	30

Note: The sort order field is optional in the scripts where it's available. If you choose to map the script field to the Custodian Sort Order field, the script results will display a list based on the custodian sort order rank. The Custodian Sort Order field itself does not appear in the results.

26.5 Running the scripts

You can run the processing duplication scripts at the item level or family level. When deciding which level you should choose, consider the following:

- **Item level** scripts identify all instances of duplicate documents, regardless of email families.
 - For example, an email sent to five custodians has one master and four duplicates.
 - An Excel file attached separately to two different emails is identified as a duplicate.
- **Family level** scripts only identify entire duplicate email families, such as an email and its attachments.
 - For example, an email sent to five custodians has one master and four duplicates.
 - An Excel file attached separately to two different emails is not identified as a duplicate.

For each level, there are three scripts:

- **All Custodians**—populates a field with a list of all custodians owning a document. Use this script to identify all duplicates in your data set, regardless of family identity. For more information, see [All custodians script](#).
- **All Source Locations**—identifies source locations for duplicates. Use this script to identify source locations of duplicates. For more information, see [All Source Locations script](#).
- **Update Duplicate Status**—populates a single-choice field with one of three values: Master, Duplicate, or Unique. Rerunning the script overwrites any existing values. Use this script to update the duplicate status of the document. Running this script at the family level also identifies the level, such as parent or child, of the document. For more information, see [Update Duplicate Status script](#).

26.5.1 All Custodians script

The All Custodians script populates a field with the names of all custodians who own a document (item level) or own a document within a given family (family level). At the family level, the script populates a field with the level of document. For example, a parent document (1) or child (2 or greater). You can use this script to identify duplicates regardless of family. The steps for running the script are similar for the item-level and family-level searches. The instructions below use the item-level UI as a base, with any family script differences highlighted.

To run the All Custodians script:

1. Navigate to the **Processing Item Level Scripts** sub-tab.
 - Family-level: Navigate to the **Processing Family Scripts** sub-tab.
2. From the landing page, select the **All Custodians (Item Level)** option.
 - Family-level: Select the **All Custodians (Family)** option.

All custodians item-level script

Processing Item Level

01. All Custodians (Item Level)

02. All Source Locations (Item Level)

03. Update Duplicate Status (Item Level)

When run at an item level, this script will executed against all files in the saved search population regardless of family relationship. This script will create a semicolon delimited list of all Custodians for an identified list of related documents in a saved search.

Script Inputs

Saved Search:*

DUP All Custodian Item/Fam... ▾

Duplicate Hash Field:*

Processing Duplicate Hash

Custodian Object:*

Custodian ▾

Custodian Object Field:*

Full Name ▾

Destination Field (output):*

All Custodians (Long Text) ▾

Batch Size: (default-50000)

Set required inputs in the sidebar to run the script.

Run

Preview

Back

All custodians family-level script

Processing Family

01. All Custodians (Family)

02. All Source Locations (Family)

03. Update Duplicate Status (Family)

When run at a family level, this script requires you to populate a Level field which indicates what the type of file it is (ex: 1 = parent 2 or greater = child), as well as a Family Identifier field (a relational field indicating what family a file is a part of). This script will create a semicolon delimited list of all Custodians for an identified list of related documents in a saved search.

DUP All Custodian Item/F...

Duplicate Hash Field: *

Processing Duplicate Hash

Custodian Object: *

Custodian

Family Identifier Field: *

Family Group

Level Field: *

Level

Custodian Object Field: *

Full Name

Destination Field(output): *

All Custodians (Long Text)

Batch Size: (default-50000)

Run

Preview

Back

Set required inputs in the sidebar to run the script.

3. Complete the following fields:

- **Saved Search**—select the saved search that has the group of documents to run the script against. This should include all files you wish to run the script against regardless of family relationships. This is because the script will run on each document independently of one another.
- **Duplicate Hash Field**—select the relational field which defines groups of duplicate documents. For example, MD5, SHA1, or Processing Duplication Hash.
- **Family Identifier Field** (Family-level only)—select the relational field which defines groups of family documents.
- **Level Field** (Family-level only)—select the whole number field which defines a numeric value indicating how deeply nested the document is within the family. This is commonly called the Level field if you discovered and published the data through Relativity Processing.
- **Custodian Object**—select the Custodian object that has your custodian information.
- **Custodian Object Field**—select the field that has the custodian's full name.
- **Destination Field (output)**—select the long text field to store the semi-colon delimited list of custodians.
- **Batch Size**—(Optional—not all product versions have this field.) This field is optional and is the number of files in a saved search batch. Leave this field blank to use the default value of 50,000 or enter a value below 50,000 to increase the script execution speed.

4. Click **Run**.

26.5.1.1 Viewing the results

Return to the saved search used in the script and refresh the list if necessary.

All custodians item-level script results

#	☐	Control Number	Custodian	Family Group	All Custodians (Long Text)
		Filter	Filter	Filter	Filter
1	☐	REL000000002_0001	Howard, Johnson	REL000000002	Johnson, Howard; Smith, Bob;
2	☐	REL000000002_0002	Howard, Johnson	REL000000002	Johnson, Howard; Smith, Bob;
3	☐	REL000000002_0003	Howard, Johnson	REL000000002	Johnson, Howard; Smith, Bob;
4	☐	REL000000002_0004	Howard, Johnson	REL000000002	Johnson, Howard; Smith, Bob;
5	☐	REL000000003	Howard, Johnson	REL000000003	Johnson, Howard; Smith, Bob;
6	☐	REL000000003_0001	Howard, Johnson	REL000000003	Johnson, Howard; Smith, Bob;
7	☐	REL000000004	Howard, Johnson	REL000000004	Johnson, Howard; Smith, Bob;
8	☐	REL000000004_0001	Howard, Johnson	REL000000004	Johnson, Howard; Smith, Bob;
9	☐	REL000000005	Howard, Johnson	REL000000005	Johnson, Howard; Smith, Bob;
10	☐	REL000000005_0001	Howard, Johnson	REL000000005	Johnson, Howard; Smith, Bob;
11	☐	REL000000006	Howard, Johnson	REL000000006	Johnson, Howard; Smith, Bob;
All 73,590 Edit ^					Total: 73,590

All custodians family-level script results

#	☐	Control Number	Custodian	Family Group	All Custodians (Long Text)
		Filter	Filter	Filter	Filter
1	☐	REL000000002_0001	Howard, Johnson	REL000000002	
2	☐	REL000000002_0002	Howard, Johnson	REL000000002	
3	☐	REL000000002_0003	Howard, Johnson	REL000000002	
4	☐	REL000000002_0004	Howard, Johnson	REL000000002	
5	☐	REL000000003	Howard, Johnson	REL000000003	Johnson, Howard; Smith, Bob;
6	☐	REL000000003_0001	Howard, Johnson	REL000000003	Johnson, Howard; Smith, Bob;
7	☐	REL000000004	Howard, Johnson	REL000000004	Johnson, Howard; Smith, Bob;
8	☐	REL000000004_0001	Howard, Johnson	REL000000004	Johnson, Howard; Smith, Bob;
9	☐	REL000000005	Howard, Johnson	REL000000005	Johnson, Howard; Smith, Bob;
10	☐	REL000000005_0001	Howard, Johnson	REL000000005	Johnson, Howard; Smith, Bob;
11	☐	REL000000006	Howard, Johnson	REL000000006	Johnson, Howard; Smith, Bob;
All 73,590 Edit ^					Total: 73,590

The following table describes the two report results:

Column	Description
Control Number	This is the file's ID. Child documents have an underscore and secondary number that tells you where the file is located within a family. For example, _0003 indicates the file is the third duplicate in the family group.
Custodian	The primary custodian associated with the document. The primary custodian is the entity having the lowest Artifact ID where there are more than one custodians associated with a file.
Family Group	The relational field that defines groups of related documents. Use this field to identify and filter families of documents.
All Custodians (Long Text)	The semi-colon delimited list of all custodians associated with the document or relational group.

26.5.2 All Source Locations script

The All Source Locations script identifies all source locations for a duplicate document. The steps for running the script are similar for the item-level and family-level searches. The instructions use the item-level UI as a base, highlighting family script differences.

To run the All Source Locations script:

1. Navigate to the **Processing Item Level Scripts** sub-tab.
 - (Family-level) Navigate to the **Processing Family Scripts** sub-tab.
2. From the landing page, select the **All Source Locations (Item Level)** option.
 - (Family-level) Select the **All Source Locations (Family)** option.*All source locations item-level script*

Processing Item Level

● 01. All Custodians (Item Level)

○ 02. All Source Locations (Item Level)

● 03. Update Duplicate Status (Item Level)

When run at an item level, this script will be executed against all files in the saved search population regardless of family relationship. This script will update a destination long text field with a list of semicolon delimited values from a source long text field for an identified list of related documents in a saved search located on the Document object.

Script Inputs

Saved Search:*

DUP All Source Locations Ite... ▼

Duplicate Hash Field:*

Processing Duplicate Hash

Source Field:*

Source Path ▼

Destination Field (output):*

All Paths/Locations ▼

Batch Size: (default-50000)

Set required inputs in the sidebar to run the script.

Run

Preview

Back

All source locations family-level script

Processing Family

● 01. All Custodians (Family)

○ 02. All Source Locations (Family)

● 03. Update Duplicate Status (Family)

This script will update a destination long text field with a list of semicolon delimited values from a source long text field for an identified list of related documents in a saved search located on the Document object.

Script Inputs

Saved Search:*

DUP All Source Locations It... ▼

Duplicate Hash Field:*

Processing Duplicate Hash

Family Identifier Field:*

Family Group ▼

Source Field:*

Source Path ▼

Destination Field (output):*

All Paths/Locations ▼

Level Field:*

Level ▼

Set required inputs in the sidebar to run the script.

Run

Preview

Back

3. Complete the following fields:

- **Saved Search**—select the saved search that has the group of documents to run the script against. This should include all files you wish to run the script against regardless of family relationships. This is because the script will run on each document independently of one another.
- **Duplicate Hash Field**—select the relational field which defines groups of duplicate documents. For example, MD5, SHA1, or Processing Duplication Hash.
- **Family Identifier Field** (Family-level only)—select the relational field which defines groups of family documents.
- **Source Field**—select the long text field that has the source location for documents.
- **Level Field** (Family-level only)—select the whole number field which defines a numeric value indicating how deeply nested the document is within the family. This is commonly called the Level field if you discovered and published the data through Relativity Processing.
- **Destination Field (output)**—select the long text field to store the semi-colon delimited list of source paths.
- **Batch Size**—(Optional—not all product versions have this field.) This field is optional and is the number of files in a saved search batch. Leave this field blank to use the default value of 50,000 or enter a value below 50,000 to increase the script execution speed.

4. Click **Run**.

Return to the saved search used in the script and refresh the list if necessary.

#	Control Number	Source Path	Family Group	All Paths/Locations
	Filter	Filter	Filter	Filter
1	☐ REL000000002_0001	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent\FWD: Athletes that DID NOT Qualify for the Olympics	REL0000000002	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\All documents; \\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Discussion threads; \\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\sent mail; \\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent;
2	☐ REL000000002_0002	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent\FWD: Athletes that DID NOT Qualify for the Olympics	REL0000000002	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\All documents; \\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Discussion threads; \\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\sent mail; \\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent;
3	☐ REL000000002_0003	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent\FWD: Athletes that DID NOT Qualify for the Olympics	REL0000000002	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\All documents; \\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Discussion threads; \\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\sent mail; \\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent;
4	☐ REL000000002_0004	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent\FWD: Athletes that DID NOT Qualify for the Olympics	REL0000000002	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\All documents; \\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Discussion threads; \\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\sent mail; \\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent;
5	☐ REL000000003	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent	REL0000000003	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent
6	☐ REL000000003_0001	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent\Fw: cat clip.....	REL0000000003	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent
7	☐ REL000000004	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent	REL0000000004	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent
8	☐ REL000000004_0001	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent	REL0000000004	\\Enron_CleaspedPII\Andrea_Ring\andrea_ring_000_1_1.pst\ving-a\Andrea_Ring_Jun2001\Notes Folders\Sent

All 73,590

Edit

Total: 73,590

[illegible]

Column	Description
Control Number	This is the file's ID. Child documents have an underscore and secondary number that tells you where the file is located within a family. For example, _0003 indicates the file

Column	Description
	is the third duplicate in the family group.
Source Path	The file location associated with the document.
Family Group	The relational field that defines groups of related documents. Use this field to identify and filter families of documents.
Level	Use this field value when running the family-level script. This number tells you the nested level of the document. The higher the number, the deeper the document is nested.
All Paths/Locations	The semi-colon delimited list of all source locations associated with the document or relational group.

26.5.3 Update Duplicate Status script

The Update Duplicate Status script assigns a duplicate status value to each document: Unique, Master, or Duplicate. Rerunning the script overwrites any existing values with new values. The steps for running the script are similar for the item-level and family-level searches. The instructions use the item-level UI as a base, highlighting family script differences.

To run the Update Duplicate Status script:

1. Navigate to the **Processing Item Level Scripts** sub-tab.
 - (Family-level) Navigate to the **Processing Family Scripts** sub-tab.
2. From the landing page, select the **Update Duplicate Status (Item Level)** option.
 - (Family-level) Select the **Update Duplicate Status (Family)** option.

Update duplicate status item-level script

Processing Item Level

● 01. All Custodians (Item Level)

● 02. All Source Locations (Item Level)

○ 03. Update Duplicate Status (Item Level)

When run at an item level, this script will be executed against all files in the saved search population regardless of family relationship. This Relativity case script will repopulate the existing single-choice field with one of three values - Master, Duplicate, or Unique. This script also contains an optional Duplicate Sort Order input field, which can determine the Master document via the custodian's order number (the lowest number for the relational group is Master). If there are duplicate or no custodians, the document with the lowest Artifact ID will be the Master.

Script Inputs

Saved Search:*

DUP Update Duplicate Status... ▼

Duplicate Status Field:*

Duplicate Status ▼

Duplicate Hash Field:*

Processing Duplicate Hash

Duplicate Sort Order Field:

Custodian Sort Order ▼

Batch Size: (default-50000)

Set required inputs in the sidebar to run the script.

Run

Preview

Back

Update duplicate status family-level script

Processing Family

- 01. All Custodians (Family)
- 02. All Source Locations (Family)
- 03. Update Duplicate Status (Family)

This Relativity case script will repopulate the existing single-choice field with one of three values - Master, Duplicate, or Unique. This script also contains an optional Duplicate Sort Order input field, which can determine the Master document via the custodian's order number (the lowest number for the relational group is Master). If there are duplicate or no custodians, the document with the lowest Artifact ID will be the Master.

Script Inputs

Saved Search:*

DUP Update Duplicate Status... ▼

Duplicate Status Field:*

Duplicate Status ▼

Duplicate Hash Field:*

Processing Duplicate Hash

Family Identifier Field:*

Family Group ▼

Level Field:*

Level ▼

Duplicate Sort Order Field:

Custodian Sort Order ▼

Batch Size: (default-50000)

Set required inputs in the sidebar to run the script.

Run

Preview

Back

Note: If using EDRM MIH, your script settings will be:

Update duplicate status item-level script

- 01. All Custodians (Item Level) ● 02. All Source Locations (Item Level) ○ 03. Update Duplicate Status (Item Level)

When run at an item level, this script will be executed against all files in the saved search population regardless of family relationship. This Relativity case script will repopulate the existing single-choice field with one of three values - Master, Duplicate, or Unique. This script also contains an optional Duplicate Sort Order input field, which can determine the Master document via the custodian's order number (the lowest number for the relational group is Master). If there are duplicate or no custodians, the document with the lowest Artifact ID will be the Master.

Script Inputs

Saved Search:*

EDRM Saved Search ▼

Duplicate Status Field:*

EDRM Duplicate Status ▼

Duplicate Hash Field:*

EDRM MIH ▼

Duplicate Sort Order Field:

Choose... ▼

Batch Size: (default-50000)

Run

Preview

Back

Set required inputs in the sidebar to run the script.

3. Complete the following fields:

- **Saved Search**—select the saved search that has the group of documents to run the script against. This should include all files you wish to run the script against regardless of family relationships. This is because the script will run on each document independently of one another.

Note: Map this field to the **EDRM Saved Search** if using the EDRM MIH to identify potential duplicates.

- **Duplicate Status Field**—select the field where Relativity outputs the duplicate status.

Note: Map this field to the **EDRM Duplicate Status** field if using the EDRM MIH to identify potential duplicates.

- **Duplicate Hash Field**—select the relational field which defines groups of duplicate documents. For example, MD5, SHA1, or Processing Duplication Hash.

Note: Map this field to **EDRM MIH** if you are using the EDRM MIH to identify potential duplicates

- **Family Identifier Field** (Family-level only)—select the relational field which defines groups of family documents.

- **Level Field** (Family-level only)—select the whole number field which defines a numeric value indicating how deeply nested the document is within the family. This is commonly called the Level field if you discovered and published the data through Relativity Processing.
- **Duplicate Sort Order Field**—this is a field located on the Entity object. Leave this blank unless you have set up a field on the Entity object to store a priority sort order value for the custodian. When this field is blank, the system sorts on the document Artifact ID field from the Document object. The first document loaded in the workspace becomes the Primary document when duplicates are identified. See [Viewing the Results](#) for more information on how this field operates.
- **Batch Size**—(Optional—not all product versions have this field.) This field is optional and is the number of files in a saved search batch. Leave this field blank to use the default value of 50,000 or enter a value below 50,000 to increase the script execution speed.

4. Click **Run**.
5. You see a message indicating the results are permanent. Select **Accept**.
6. Return to your saved search to view the results.

26.5.3.1 Viewing the results

When the script runs, it clears the Duplicate Status field for all documents in the workspace. After clearing, the field updates with the one of the following values for the documents in the saved search:

Note: (If you used the EDRM MIH for your script, see Viewing EDRM results for sorting and filtering the EDRM Saved Search.)

- **Unique**—the document in the saved search has a relational identifier that is unique compared to all the other documents.
- **Master**—documents in the saved search where more than one document has the same relational identifier.
 - If you specify the Duplicate Sort Order field, the document with the lowest order of the associated custodian is the master. If multiple documents in the same group share the same custodian, the document with the lowest Artifact ID becomes the master.
 - If you do not specify the Duplicate Sort Order field, the document having the lowest document Artifact ID in the relational group is the master.
- **Duplicate**—documents in the saved search where more than one document has the same relational identifier.
 - If you specify the Duplicate Sort Order field, documents not having the lowest ordered custodian in the relational group are duplicates.
 - If you do not specify the Duplicate Sort Order field, documents not having the lowest Artifact ID in the relational group are duplicates.
- **Not Set**—the relational identifier for the document in the saved search is not set.

Any documents not included in the selected saved search are excluded from the logic to calculate duplicate status and the Duplicate Status field is not populated.

You will see an **Update Complete** message when the script completes. Return to the saved search selected in the script to view the results.

Update duplicate status item-level script results

#	☐	Artifact ID	Control Number	Processing duplicate hash	Custodian	Level	Duplicate Status	Duplicate Sort Order
		= ▼ Filter	Filter	Filter	Filter	= ▼ i	(All) ▼	= ▼ Filter
1	☐	1523661	REL0000000002_0001	D251A31AB45BE5778028913B32E89129D771F9F93C58680D4AB4F6965D6085D4	Howard, Johnson	2	Master	
2	☐	1523662	REL0000000002_0002	2F25D1933DB269D43878E3569EDCE26F3B5B10F1A6D007F7F98201BF81C391C4	Howard, Johnson	2	Master	
3	☐	1523663	REL0000000002_0003	A35EF246CEF40AAA9D9FE59471D5706EFF5588357905BE0F33C8C6CC618A4BA2	Howard, Johnson	2	Master	
4	☐	1523664	REL0000000002_0004	DCF1AE3329B07A8A4CBB68B0F43DDCC557D9CAB30190D8A69483092CD9E3CA78	Howard, Johnson	2	Master	
5	☐	1523665	REL0000000003	E23BF21963620ACF4A42AA78CECA8B5F2583BCD8C4A097579B63B882796D92E8	Howard, Johnson	1	Master	
6	☐	1523666	REL0000000003_0001	D9474A90773EBBE22D3D9578D561265707F127EEE9FCA17602B881819AB0CB20	Howard, Johnson	2	Master	
7	☐	1523667	REL0000000004	57507F452AA8B1F313926E13E8F8AAE2943B9E6547F527018AAC57E163BD8564	Howard, Johnson	1	Master	
8	☐	1523668	REL0000000004_0001	D9474A90773EBBE22D3D9578D561265707F127EEE9FCA17602B881819AB0CB20	Howard, Johnson	2	Duplicate	
All 73,590 Edit ^								
								Total: 73,590

Update duplicate status family-level script results

#	☐	Artifact ID	Control Number	Processing duplicate hash	Custodian	Level	Duplicate Status...	Duplicate Sort Order KL
		= ▼ Filter	Filter	Filter	Filter	= ▼ i	(All) ▼	= ▼ Filter
18	☐	1523678	REL0000000009_0001	D251A31AB45BE5778028913B32E89129D771F9F93C58680D4AB4F6965D6085D4	Howard, Johnson	2	Master	
19	☐	1523679	REL0000000009_0002	2F25D1933DB269D43878E3569EDCE26F3B5B10F1A6D007F7F98201BF81C391C4	Howard, Johnson	2	Master	
20	☐	1523680	REL0000000009_0003	A35EF246CEF40AAA9D9FE59471D5706EFF5588357905BE0F33C8C6CC618A4BA2	Howard, Johnson	2	Master	
21	☐	1523681	REL0000000009_0004	DCF1AE3329B07A8A4CBB68B0F43DDCC557D9CAB30190D8A69483092CD9E3CA78	Howard, Johnson	2	Master	
22	☐	1523682	REL0000000010	57507F452AA8B1F313926E13E8F8AAE2943B9E6547F527018AAC57E163BD8564	Howard, Johnson	1	Duplicate	
23	☐	1523683	REL0000000010_0001	D9474A90773EBBE22D3D9578D561265707F127EEE9FCA17602B881819AB0CB20	Howard, Johnson	2	Duplicate	
24	☐	1523684	REL0000000011	E23BF21963620ACF4A42AA78CECA8B5F2583BCD8C4A097579B63B882796D92E8	Howard, Johnson	1	Duplicate	
25	☐	1523685	REL0000000011_0001	D9474A90773EBBE22D3D9578D561265707F127FFF9FCA17602B88	Howard, Johnson	2	Duplicate	
All 73,590 Edit ^								
								Total: 73,590

Column	Description
Control Number	This is the file's ID. Child documents have an underscore and secondary number that tells you where the file is located within a family. For example, _0003 indicates the file is the third duplicate in the family group.
Processing Folder Path	The file location associated with the document.
Family Group	The relational field that defines groups of related documents. Use this field to identify and filter families of documents.

Column	Description
All Paths/Locations	The script uses this field to output a semicolon delimited list of source locations associated with the document.

Note: If you mapped the Duplicate Sort Order Field in the script to the Custodian Sort Order field, the script results will display a list based on the custodian sort order rank. The Custodian Sort Order field itself does not appear in the results.

27 Reports

In the Reports tab, you can generate reports in Relativity to understand the progress and results of processing jobs. You can't run reports on processing sets that have been canceled. When you generate a processing report, this information is recorded in the History tab.

Note: This topic contains several references to progressive filtration. For context, consider the following explanation: A count based on absolute filtration counts the total number of files each filter eliminates as if it were the only filter applied to the entire data set. A count based on progressive filtration counts the total number of files each filter actually eliminates by accounting for all previously applied filters. For example, a file type filter eliminates 3000 PDF files from a data set, but a previously applied date range filter also eliminated 5000 PDF files. A count based on absolute filtration would report the file type filter as eliminating 8000 PDF files because the count is based on the total number of files each filter eliminates as if it were the only filter applied to the entire data set. However, a count based on progressive filtration counts the total number of files each filter actually eliminates by accounting for all previously applied filters. In the previous example, a progressive filtration count only reports 3000 PDF files eliminated by the file type filter, because the other 5000 PDF documents were progressively eliminated by the date range filter.

27.1 Generating a processing report

1. Navigate to the **Processing** tab.
2. Click the **Reports** sub-tab. You can also access the Reports tab by clicking the View All Reports link on your processing set console.
3. From the **Select Report** section, select the report type you want to generate. When you select a report type, the processing set list to the right is updated to reflect only those sets that are applicable to that report type. For example, if you haven't yet discovered the files in your set, that set won't show up when you select either of the Discovered Files reports. Canceled processing sets aren't available when you're running reports.
4. From the **Select Processing Set** section, select the set on which you want to report.

5. Click **Generate Report**.

The screenshot shows the 'Reports' section of a web application. At the top, there is a navigation bar with links: 'Processing Sets', 'Processing Data Sources', 'Processing Profile', 'Password Bank | Processing', 'Inventory', 'Reports' (highlighted with a red box), and 'More'. Below the navigation bar, a dropdown menu is open, showing 'Select Report & Processing Set(s)'. The main content area is divided into three columns: 'Select Report', 'Select Processing Set(s)', and 'Generate Report'. The 'Select Report' column contains a list of report types, with 'Job Exception' selected. The 'Select Processing Set(s)' column contains a table with two rows: 'Processing Set 002' (selected) and 'Processing Set 001'. The 'Generate Report' column contains a 'Generate Report' button. A progress indicator at the top shows three steps: 'Select Report' (completed), 'Select Processing Set(s)' (completed), and 'Generate Report' (current step, indicated by a blue circle with the number 3).

Processing Sets Processing Data Sources Processing Profile Password Bank | Processing Inventory **Reports** More ▾

▼ Select Report & Processing Set(s)

Select Report **Select Processing Set(s)** **Generate Report**

1 2 3

Select Report ?

- ☐ Data Migration
- ☐ File Size Summary
- ☐ Discovered Files By Custodian
- ☐ Discovered Files By File Type
- ☐ Document Exception
- ☐ Inventory Details
- ☐ Inventory Exclusion Results
- ☐ Inventory Summary
- ☒ Job Exception
- ☐ Text Extraction
- ☐ Inventory Exclusion Results By Custodian
- ☐ Inventory Details By Custodian
- ☐ Discovery Exclusion Results
- ☐ Master Document Replacement Summary

Select Processing Set(s)

☐	Name	Status
☒	Processing Set 002	Publish - Completed with errors
☐	Processing Set 001	Publish - Completed

Review

Generate Report

6. At the top of the report display, there are several icons that control your interaction with the report. For example, you can toggle on and off single or multiple page display, print the page, print the report, export the report, and view the report in full screen mode.

Note: If you export a report that contains Unicode characters as a PDF, and the web server you're logged in to does not have the font *Arial Unicode MS Regular* installed (regardless of whether the server the workspace resides on has this font installed), you see blocks in the generated PDF file. To resolve this issue, you can purchase and install the font separately, or you can install Microsoft Office to the web server, which installs the font automatically.

You can generate a new report at any time by clicking **New Report** at top right of the collapsed report generation window.

Note: If you choose to print a processing report through your browser, that report won't be displayed correctly, in that it will be incomplete. Therefore, it's recommended that you always use Relativity's print button to print reports and not print through the browser.

27.2 Data Migration

This report provides information on how data was migrated into Relativity, including details about excluded files and a summary of the number of starting files, published documents, and documents included in the workspace for each custodian associated with the selected processing sets. You can run this report on published processing sets.

Summary Statistics: Data Migration

This table provides a summary of the files excluded during data migration by custodian and contains the following information:

- **Custodian**—the name of each custodian associated with the migrated files.
- **Starting Files**—each custodian's initial number of discovered files in the processing set. This includes containers and denisted files.
- **Excluded Files**—each custodian's total number of excluded files. The list of exclusion types is included in a separate table.
- **Published Documents**—each custodian's total number of published documents.
- **Documents in Workspace**—each custodian's total number of documents in the workspace.

Note: Differences between Published Documents and Documents in Workspace counts could indicate that documents were deleted after publishing.

Excluded Files

This table lists all files excluded during data migration by custodian and provides the following information:

- **Custodian**—the name of the custodian associated with excluded files.
- **DeNIST**—the number of NIST files excluded.
- **Containers**—the number of container files excluded.
- **Duplicates**—the number of duplicate files excluded.
- **Embedded Images**—the number of embedded images excluded.
- **Publishing Errors**—the number of files excluded due to errors during publication.
- **Total Excluded Files**—each custodian's total number of excluded files.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

27.3 Master Document Replacement Summary

This report provides a summary of documents deleted and resulting replacements during deduplication recalculation during the Post-Publish Delete process. You can run this report on processing sets.

Deleted Master Documents

This table lists all files deleted during the discovery process by control number and provides the following information:

- **Control Number**—the identifier of the deleted document.
- **File ID**—the number value associated with the deleted file in the database.
- **Custodian**—the name of each custodian associated with the deleted file.
- **Published Control Number**—the identifier of the document published as a result of deduplication recalculation.
- **Published Custodian**—the custodian associated with the replacement document published as a result of deduplication recalculation.

Replacements Master Documents

This table lists all files replaced by control number and provides the following information:

- **Control Number**—the identifier of the replacement document published as a result of deduplication recalculation.
- **File ID**—the number value associated with the replacement file in the database.
- **Custodian**—the name of each custodian associated with the replacement document published as a result of deduplication recalculation.
- **Deleted Control Number**—the identifier of the deleted document.
- **Deleted Custodian**—the custodian associated with the deleted document.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

27.4 Discovery Exclusion Results

This report provides filtering summaries for exclusion or inclusion filter types applied during Discovery including file extensions, file types, file size, excluded file count, and processing sets filtered. You can run this report on discovered processing sets. See Processing profiles for more information on Inclusion/Exclusion Discovery filters.

Discover Filter Settings

This table provides a summary of the filter settings specified in Inventory | Discover settings within the Processing Profile and contains the following information:

- **Filter Type**—the filter type applied.
- **File Extensions**—all file extensions entered into the Inclusion/Exclusion File List.

File Type | File Size | Excluded File Count

This table lists the file types filtered out of the document list and the number and size (GB) of files per type that were excluded.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

27.5 Discovered Files by Custodian

This report provides information on the file types discovered during processing for the custodians associated with the selected processing sets. This report identifies the total processable and unprocessable file types discovered and categorizes the totals by custodian. You can run this report on discovered or published processing sets.

Discovered Files by Custodian

This table provides a summary of all files discovered per custodian and contains the following information:

- **Custodian**—the name of the custodian whose files were discovered.
- **Discovered Files**—the number of each custodian's discovered files.

File Types Discovered - Processable

This table provides a summary of the processable discovered files per file extension and contains the following information:

- **File Extension**—all file extensions discovered.
- **Discovered Files**—the number of files discovered with that file extension.

File Types Discovered - Processable (By Custodian)

This table provides a summary of the processable discovered file counts per file extension by custodian and contains the following information:

- **Custodian**—the name of the custodian whose processable files were discovered.
- **File Extension**—all file extensions of each custodian's processable discovered files.
- **Discovered Files**—the number of each custodian's processable discovered files by file extension.

File Types Discovered - Unprocessable

This table provides a summary of the discovered file counts per file extension and contains the following information:

- **File Extension**—all unprocessable discovered file extensions.
- **Discovered Files**—the number of unprocessable files discovered with that file extension.

File Types Discovered - Unprocessable (by Custodian)

This table provides a summary of the processable discovered file counts per file extension by custodian and contains the following information:

- **Custodian**—the name of the custodian whose unprocessable files were discovered.
- **File Extension**—all file extensions of each custodian's unprocessable files.
- **Discovered Files**—the number of each custodian's unprocessable discovered files by file extension.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

27.6 Discovered Files by File Type

This report provides information on the file types discovered during processing for the custodians associated with the selected processing sets. This report identifies the total processable and unprocessable file types discovered and categorizes the totals by file type. You can run this report on discovered or published processing sets. See [Supported file types](#) for a list of file types and extensions supported by Relativity for processing.

Discovered Files by Custodian

This table provides a summary of all files discovered per custodian and contains the following information:

- **Custodian**—the name of the custodian whose files were discovered.
- **Discovered Files**—the number of each custodian's discovered files.

Note: The file count for this report does not include container files (such as ZIP files).

File Types Discovered - Processable

This table provides a summary of the files discovered per file extension and contains the following information:

- **File Extension**—all file extensions discovered.
- **Discovered Files**—each file extension's number of files discovered.

File Types Discovered - Processable (By File Type)

This table provides a summary of the discovered file counts per file type and contains the following information:

- **File Extension**—the file extension of all discovered files.
- **Custodian**—the custodians of each file extension's discovered files.
- **Discovered Files**—the number of each file extension's discovered files by custodian.

File Types Discovered - Unprocessable

This table provides a summary of the files discovered per file extension and contains the following information:

- **File Extension**—all file extensions discovered.
- **Discovered Files**—each file extension's number of files discovered.

File Types Discovered - Unprocessable (By File Type)

This table provides a summary of unprocessable discovered file counts per file type and contains the following information:

- **File Extension**—the file extension of all unprocessable discovered files.
- **Custodian**—the custodians of each file extension's unprocessable discovered files.
- **Discovered Files**—the number of each file extension's unprocessable discovered files by custodian.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

27.7 Document Exception

This report provides details on the document level errors encountered during processing, broken down by those that occurred during the discovery process and those that occurred during the publishing process. You can run this report on discovered or published processing sets.

Document Level Errors - Discovery

This table lists all document level errors that occurred during discovery and contains the following information:

- **Error Message**—all error messages encountered during discovery.
 - **Total**—the total number of errors encountered during discovery.
 - **Total Distinct Documents with Discovery Errors**—the total number of documents that encountered errors during discovery. Because any single document can have multiple errors, this count might be lower than the total number of errors.
- **Count**—the number of instances the corresponding error occurred during discovery.

Document Level Errors - Publishing

This table lists all document level errors that occurred during publish and contains the following information:

- **Error Message**—all error messages encountered during publishing.
- **Total**—the total number of errors encountered during publishing.
- **Total Distinct Documents with Publishing Errors**—the total number of documents that encountered errors during publishing. Because any single document can have multiple errors, this count might be lower than the total number of errors.
- **Count**—the number of instances the corresponding error occurred during publishing.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

27.8 File Size Summary

This report provides information on file sizes for pre-processed, processed, and published data sets. Run this report after publishing a processing set.

Pre-Processed File Size

This table lists the pre-processed file size for all loose and compressed file sizes in the source location.

Processed File Size

This table lists the processed file size once Discovery is complete.

It includes:

- all loose and uncompressed files
- duplicate files

It excludes:

- container files
- DeNISTed files

Published File Size

This table lists the published file size for review.

It includes:

- all loose and uncompressed files

It excludes:

- container files
- DeNISTed files
- duplicate files

27.9 Inventory Details

This report provides detailed information on date range included, file size limitations, and deNIST settings. The report lists the number of files excluded by each filter applied. All excluded file counts reflect progressive filtration. See [Progressive and absolute filtration counts](#) for more information. You can run this report on inventoried processing sets.

Inventory Filter Settings

This table provides a summary of the filter settings specified in the Inventory tab and contains the following information:

- **DeNIST Files Excluded**—whether or not NIST files were excluded from the processing set after inventory.
- **Date Range Excluded**—the span of time set by the date range filter after inventory.
- **File Size Range Excluded**—the file size limitations set by the file size filter.
- **Inventory Errors**—the number of errors encountered during the inventory process.
- **Files With Unknown Dates**—the number of files with invalid dates.

Excluded by File Type Filter | Excluded File Count

This table lists all file types filtered out of the document list and the number of files per type that were excluded.

Excluded by Location Filter | Excluded File Count

This table lists all file locations filtered out of the document list and the number of files per location that were excluded.

Excluded by Sender Domain Filter | Excluded File Count

This table lists all sender domains filtered out of the document list and the number of files per domain that were excluded.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

27.10 Inventory Details by Custodian

This report provides detailed information on date range included, file size selection, and deNIST settings. The report lists the files and counts for each filter applied and also breaks down these counts by custodian. All excluded file counts reflect progressive filtration. You can run this report on inventoried processing sets.

Inventory Filter Settings

This table provides a summary of the filter settings specified in the Inventory tab and contains the following information:

- **DeNIST Files Excluded**—whether or not NIST files were excluded from the processing set after inventory.
- **Date Range Selected**—the span of time set by the date range filter after inventory.
- **File Size Range Selected**—the file size limitations set by the file size filter.
- **Inventory Errors**—the number of errors encountered during the inventory process.
- **Files With Unknown Dates**—the number of files with invalid dates.

This report contains the same tables as the Inventory Details Report, but it also includes the following:

Custodian | Excluded by File Type Filter | Excluded File Count

This table lists the file types filtered out of the document list per custodian and the number of files per type that were excluded.

Custodian | Excluded by File Location Filter | Excluded File Count

This table lists all file locations filtered out of the document list per custodian and the number of files per location that were excluded.

Custodian | Excluded by Sender Domain | Excluded File Count

This table lists all sender domains filtered out of the document list per custodian and the number of files per domain that were excluded.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

27.11 Inventory Exclusion Results

This report provides detailed information on date range excluded, file size limitations, and deNIST inventory settings. This report also provides counts of files excluded by applied filters and categorizes the results by file type, sender domain, and file location. All excluded file counts are absolute. See [Progressive and absolute filtration counts](#) for more information. You can run this report on inventoried processing sets.

Inventory Filter Settings

This table provides a summary of the filter settings specified in the Inventory tab and contains the following information:

- **DeNIST Files Excluded**—whether or not NIST files were excluded from the processing set after inventory.
- **Date Range(s) Selected**—the span of time set by the date range filter after inventory.
- **File Size Range(s) Selected**—the file size limitations set by the file size filter.
- **Total Files Excluded**—the number of files excluded by all applied filters.

File Type | Excluded File Count

This table lists all file types that were filtered out and the number of files per type that were excluded.

Location | Excluded File Count

This table lists all file locations that were filtered out and the number of files per location that were excluded.

Sender Domain | Excluded File Count

This table lists all sender domains that were filtered out and the number of files per domain that were excluded.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

27.12 Inventory Exclusion Results by Custodian

Provides detailed information on date range excluded, file size limitations, and deNIST inventory settings. This report also provides counts of files excluded by applied filters and categorizes the results by file type, sender domain, file location, and custodian. All excluded file counts are absolute. You can run this report on inventoried processing sets.

This report contains the same tables as the Inventory Exclusion Results report, but it also includes the following:

Custodian | Excluded by File Type Filter | Excluded File Count

This table lists the file types filtered out of the document list per custodian and the number of files per type that were excluded.

Custodian | Excluded by File Location Filter | Excluded File Count

This table lists all file locations filtered out of the document list per custodian and the number of files per location that were excluded.

Custodian | Excluded by Sender Domain | Excluded File Count

This table lists all sender domains filtered out of the document list per custodian and the number of files per domain that were excluded.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

27.13 Inventory Summary

This report provides filtering summaries for each filter type including applied order, file count excluded, percentage of files excluded, total documents remaining, and total percentage of files remaining. All excluded file counts reflect progressive filtration. See [Progressive and absolute filtration counts](#) for more information.

Final inventory results include file count after filtering, file size after filtering, total number of excluded files, and total percentage of files excluded. You can run this report on inventoried processing sets. Note that, because inventory affects only parent files, this report accounts for parent files only and therefore not necessarily all files in a processing set.

Initial Inventory Results

This table provides a general summary of the inventoried processing set before filtration and contains the following information:

- **Processing Set**—the name of the inventoried processing set.
- **Status**—whether or not errors occurred during inventory.
- **File Count**—the number of files in the unfiltered processing set.
- **File Size** (unit of measurement)—the size of the unfiltered processing set.

Filtering Summary

This table provides a general summary of all filters applied to the inventoried processing set and contains the following information:

- **Applied Order**—the order that the filters were applied.
- **Filter Type**—the filter type applied.
- **File Count Excluded by Filter**—the number of files excluded by the filter.
- **% of Files Excluded by Filter**—the percentage of the initial processing set excluded after filter is applied.
- **Total Remaining File Count**—the number of files remaining after filter is applied.
- **Total % of Files Remaining**—the percentage of the initial processing set remaining after filter is applied.

Final Inventory Results

This table provides summary totals on inventory filtration and contains the following information:

- **File Count After Filtering**—the number of files left after all filters are applied to the processing set.
- **File Size After Filtering** (unit of measurement)—reports the size of the filtered processing set.
- **Total Excluded Files**—the number of files excluded after all filters are applied.
- **Total % of Files Excluded**—the percentage of the initial inventoried processing set excluded after all filters are applied.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

27.14 Job Exception

This report provides details on the job level errors encountered during processing. You can run this report on discovered or published processing sets.

Job Level Errors

This table provides a summary of all errors encountered during processing and contains the following information:

- **Error Message**—the error message.
- **Phase of Processing**—the phase of processing in which the error occurred (inventory, discovery, or publish).
- **Count**—the number of instances each error occurred.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

27.15 Text Extraction

This report provides information, broken down by custodian and file type, on the number and percentage of published files that contain and don't contain extracted text and the total number of files published into Relativity. This also provides details on error messages encountered during processing. You can run this report on published processing sets. This report includes both OCR and extracted text.

Note: The Text Extraction report reads information from Data Grid if the Extracted Text field is enabled for Data Grid.

Text Extraction by Custodian

This table provides a summary of text extraction by custodian and contains the following information:

- **Custodian**—the name of the custodian.
- **With Text**—the number of files for that custodian with extracted text.
- **Without Text**—the number of files for that custodian without extracted text.
- **Percentage without Text**—the percentage of documents for that custodian with no extracted text.
- **Total Published Files**—the number of published files of that custodian.

Text Extraction by File Type

This table provides a summary of text extraction by file type and contains the following information:

- **File Extension**—the file type extension.
- **With Text**—the number of files of that file type with extracted text.
- **Without Text**—the number of files of that file type with no extracted text.
- **Percentage without Text**—the percentage of files of that file type without extracted text.
- **Total Published Files**—the number of published files of that file type.

Breakdown by Error Message

This table provides a summary of the number of files that received each error and contains the following information:

- **Error Message**—the error message.
- **File Count**—the number of files that encountered that error.

Note: The File Count value will never decrease, even if you resolve errors and retry documents. This is because Relativity gets this value directly from the Errors table in the database, which doesn't communicate error status, only that errors that are present. In other words, even an error that is resolved is still present in the Errors table in the database and therefore will display as being present in the Text Extraction report.

- **Percentage**—the percentage of documents that encountered that error.

Note: If you publish processing sets without mapping the File Extension processing field, the Text Extraction report won't accurately report document counts by file type.

Processing Sets

This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **Custodian**—the custodians attached to the data sources used by the processing set.
- **Source path**—the location specified in the Source path field on the data sources used by the processing set.

28 Processing administration

Processing administration consists of monitoring and managing active processing jobs through the Processing Queue tab. This page explains how to review job status and progress, perform available mass operations, understand potential performance impacts, and configure the permissions required to view or modify processing jobs.

28.1 Monitoring active jobs

By default, all active jobs are listed for all workspaces in the instance. If no jobs are visible, no jobs are currently running. Use the **Show Workspace Tree** toggle (A) to view a list of all workspaces currently running jobs. You can click a workspace to filter the job list. If you do not have permission to a workspace, you will see *Item Secured* instead of the workspace name and details.

The following columns appear:

Production Queue Processing and Imaging Queue Processing Queue Transcription Queue Branding Queue OCR Queue More ▾									
A 📄 🔍 🔄 🔄 🔄 🔄 1 > >> 1 - 4 of 4 100 ▾									
#	Workspace	Processing Set	Data Source	Job Type	Status	Last Activity	Documents Remaining	Priority	
	☐ Filter	☐ Filter	☐ Filter	☐ Filter	☐ Filter	☐ Filter	= ▾ ☐ Filter	= ▾ ☐ Filter	
1	☐ Documentation	Processing Set 1	Custodian, Redact - 1569230	Processing	Waiting		-1	90	
2	☐ Documentation	Processing Set 1	Custodian, Redact - 1569233	Processing	Waiting		-1	90	
3	☐ Documentation	Processing Set 2	Custodian, Redact - 1569230	Processing	Discovering	9/17/2026, 8:57 AM MDT	0	100	
4	☐ Documentation	Processing Set 3	Custodian, Redact - 1569233	Processing	Discovering	9/17/2026, 8:57 AM MDT	0	100	

All 4 ▾
Export to CSV ▾
Total: 4

- **Workspace**—the workspace in which the job was created. Click the name of a workspace to navigate to the main tab in that workspace.
- **Processing Set**—the name of the processing set. Click a set name to navigate to the Processing Set Layout on the Processing Sets tab. From here you can cancel publishing or edit the processing set.
- **Data Source**—the data source containing the files you're processing. This appears as either the name you gave the source when you created it or an artifact ID if you didn't provide a name.
- **Job Type**—the type of job running, for example, Processing.

Note: Filtering jobs aren't represented in the queue.

- **Status**—the status of the set. If you're unable to view the status of any processing jobs in your environment, check to make sure the Server Manager agent is running. This field could display any of the following status values:
 - Waiting
 - Canceling
 - Finalizing
 - Unavailable
 - Inventorying
 - Discover
 - Publish
 - Initializing

- Retrieving/Retrying Errors
- Submitting Job
- **Last Activity**—the date and time at which a job last communicated to the worker.
- **Documents Remaining**—the number of documents that have yet to be inventoried, discovered, or published. The value in this field goes down incrementally as data extraction progresses on the processing set.

Note: This column displays a value of **-1** if you have clicked **Inventory Files**, **Discover Files**, or **Publish Files** and the job has not yet started.

- **Priority**—the order in which jobs in the queue are processed. Lower priority numbers result in higher priority. This is determined by the value of the Order field on the data source. You can change the priority of a data source with the **Change Priority** button at the bottom of the view. Changing the priority only changes the priority for that immediate job.

- Resources are split equally between processing sets of the same priority.

Note: Resource distribution is also considered at the Workspace level to make sure that all jobs are making progress.

- Discovery and publishing jobs are multi-threaded and can run in parallel, depending on the number of agents available.
- Job types have the following priorities set by default:
 - Publishing jobs have a priority of 90.
 - Inventory and discovery jobs all have a priority of 100 in the queue. These jobs have resources shared equally as long as they are the same priority.

If you have reason for globally setting certain types of jobs to always run at a lower priority, please contact [Relativity support](#).
- **Paused**—the Yes/No value indicates whether or not the job was paused. A paused job typically occurs if there is an issue with the processing agent. You cannot manually pause a processing job.
- **Time Paused**—the time at which the job was paused, based on local time.
- **Failed Attempts**—the number of times an automatic retry was attempted and failed.
- **Submitted Date**—the date and time the job was submitted, based on local time.
- **Submitted By**—the name of the user who submitted the job.

At the bottom of the screen, the active jobs mass operations buttons appear.

28.2 Active jobs mass operations

Use the Mass Actions menu at the bottom of the page to run actions on one or more jobs.

Processing Queue ☆

☆
🔊

🔍
🔄
⏮
⏭
1
1 - 3 of 3
25

#		Workspace	Processing Set	Data Source	Job Type	Status	Last Activity	Documents Remaining	Priority
	☐	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	☐		Documentation Team	Processing Set 1		Processing	Waiting	-1	52
2	☐		Documentation Team	Processing Set 1		Processing	Waiting	-1	52
3	☐		Documentation Team	Processing Set 2		Processing	Waiting	-1	100

Cancel Imaging Job

Resume Processing Job

Change Priority

Export to CSV

Select...

Checked (0)

Total: 3

- **Export to CSV**—Export the Active Jobs list in CSV format.
- **Resume Job**—resumes any paused processing jobs that have exceeded the failed retry attempt count. You can resume multiple jobs at the same time. When you select this option, non-processing jobs are skipped, as are jobs that aren't currently paused.
- **Change Priority**—change the priority of processing jobs in the queue.
 - When you click **Change Priority**, you must enter a new priority value in the **Priority** field. Then click **Change Priority** to proceed with change.

Change priority for job

Priority:

- Changing an inventory job's priority updates the priorities of other inventory jobs from the same processing set. This ensures that filtering files is available for the processing set.
- Changing a publish or republish job's priority updates the priorities of other publish or republish jobs from the same processing set. This ensures the set's de-duplication order is respected.

Accept Cancel

- If you change the priority of a publish or republish job, you update the priorities of other publish and republish jobs from the same processing set. This ensures that deduplication is performed in the order designated on the set.
- When you change the priority of an inventory job, you update the priorities of other inventory jobs from the same processing set. This ensures that filtering files is available as expected for the processing set.
- While there is no option to pause discovery, changing the priority of a discovery job is a viable alternative.

28.3 Post-publish delete job performance

Post-publish and processing set deletion jobs can impact system performance. If delete processes are running, you will see a warning banner indicating the possibility of performance impact and the workspaces affected.



Warning: One or more Post-Publish Delete or Processing Set deletion jobs are running. This may impact performance of other Processing jobs running in the same workspace. These jobs are running in the following workspaces:

- Salt and Pepper



28.4 Permissions

Note: This feature supports feature permission actions. Each action grants access to a specific task within the feature. You can grant groups access to individual actions, allowing them to perform only the tasks they need. To grant access to the entire feature, select all of its actions. Manage feature actions on the Features tab in a group's workspace permissions. For instructions, see [Setting workspace permissions](#). For more information about feature permissions, see [Feature permissions](#).

Consider the following items related to security and client domains (formerly multi-tenancy):

- If you are the system admin for a client domain environment, your tenants can only see jobs in their client domain. This eliminates the possibility of information leaks for workers that do not actually operate within your client domain.
- In client domain environments, users from one client domain cannot see workers from other client domains.
- In client domain environments, users from one client domain can only see work from their workspace. All other threads show an Item secured value for the Workspace field, and the rest of the columns are blank.

Note: To change the priority of a job where **Customer lockbox** is enabled, you must be in a permission group having access to the workspace where the job originated, along with the System Administrator's group.

Note: Only System Administrators can modify processing jobs on the Processing Queue tab. Other users can see the Processing Queue tab with instance level permissions, but see an error when attempting to modify processing jobs.

Groups do not have access to the Processing Administration tab or sub-tabs by default.

To grant group access to the Processing Administration tab:

1. Use the search bar to navigate to the **Instance Details** tab.
2. In the Security section, click **Manage Permissions**.

- From the Admin Security pane, select the **Features** tab. For detailed information on setting permissions by feature, see Feature permissions.
- From the **Selected Group** drop-down menu , select the group to grant access.
- Scroll to the **Job Queues** option, then select **See details**.

The screenshot shows the 'Permissions' pane in the Relativity Admin interface. The 'Admin Security' section is active, and the 'Features' tab is selected. The 'Job Queues' feature is chosen, showing 0/17 of actions. The 'Description' for Job Queues is: 'Gives a bird's-eye view of all the background jobs running in the instance (like imaging jobs, productions, index builds, etc.) and the ability to reprioritize or manage those jobs for optimal throughput.' The 'Actions' list includes: 'Select all', 'View OCR Queue', 'View PDF Queue', 'View Processing and Imaging Queue' (highlighted), 'View DtSearch Queue Admin', 'View Import/Export Queue', 'View Branding Queue', 'View Production Queue', 'View Processing History' (highlighted), 'View Translate Jobs', 'View Workspace Upgrade Queue', 'View Transcription Queue', 'Manage Text Migration Jobs', 'Manage OCR Queue Priority', 'Manage Processing and Imaging Queue Priority', 'Manage Branding Queue', 'Manage Production Queue', and 'Manage Workspace Upgrade Queue Priority'. At the bottom, there are buttons for 'Review and Save', 'Revert All', and 'Close'.

- (Optional) Select a preset. A preset gives you a starting point for a feature permissions. When you select a preset, Relativity selects the actions defined for that preset. You can then select or clear individual actions before you save your permissions.
 - Administrator**—preselects all of the actions.
 - Viewer**—preselects only the viewing actions.

7. At a minimum, confirm you have selected **View Processing and Imaging Queue** and **View Processing History**.

Job Queues

Preset

Administrator

Description

Full access to view and manage priority for all job queues.

Actions

☒ Select all

☒ View OCR Queue

☒ View PDF Queue

☒ View Processing and Imaging Queue

☒ View DtSearch Queue Admin

☒ View Import/Export Queue

☒ View Branding Queue

☒ View Production Queue

☒ View Processing History

☒ View Translate Jobs

☒ View Workspace Upgrade Queue

☒ View Text Migration Jobs

☒ View Transcription Queue

☒ Manage OCR Queue Priority

☒ Manage Processing and Imaging Queue Priority

☒ Manage Branding Queue

☒ Manage Production Queue

☒ Manage Workspace Upgrade Queue Priority

>> Granular permissions mapping

Review and Save

Revert All

Close

8. Click **Review and Save**.
9. Click **Save**.
10. Click the **Admin Operations** tab.

11. Select **View Admin Repository**.

Permissions

Admin Security

Groups

Features NEW

Object Security

Tab Visibility

Admin Operations

Admin Operations

☐ Delete Object Dependencies ?

☐ Force Logout on User Status ?

☐ Manage Cold Storage ?

☐ Manage Object Types ?

☐ Send Email Notification ?

☐ Send Message ?

☐ Use Quick Nav

☒ View Admin Repository

☐ View Audits ?

☐ View Client Domain Staging Reports ?

☐ View Client Domain Usage Reports ?

☐ View Cost Explorer ?

☐ View Instance-wide Staging Reports ?

☐ View Instance-wide Usage Reports ?

Review and Save

Revert All

Close

12. Click **Review and Save**.

13. Click **Save**.

29 Processing to Data Grid

By processing directly into Data Grid, you have the opportunity to improve your publishing speeds.

29.1 Enabling processing to Data Grid

After you install Data Grid, the only requirement for setting up your workspace to process to Data Grid is enabling both the workspace and the extracted text field in your environment.

To enable your workspace for Data Grid, perform the following steps:

Note: We recommend you only enable Data Grid for fields storing extracted text, OCR text, or translated text.

1. Navigate to the **Workspace Details** tab, and then click **Edit**.
2. Enable the **Is Data Grid Enabled** field.

Workspace Information *
Advanced Settings *

Name * Salt Vs Pepper

Matter * RelativityOne Matter

Select Clear

Client RelativityOne Client

Resource Information

Resource Pool * ResourcePool_T029B

Database Location ctus01412829W01.sql-
y001.r1.kcura.com\ctus01412829W01

Default File Repository * \\files\T029B\Files\

Is Data Grid Enabled * ☒

Data Grid File Repository

Default Cache Location * \\files\T029B\Files\cache\

Download Handler URL * Relativity.Distributed

3. (Optional) Next to **Data Grid File Repository**, select the path for the physical location of the text files used by Data Grid. If no file repository is specified for this field, and Data Grid is enabled, Data Grid stores text in the default file repository.

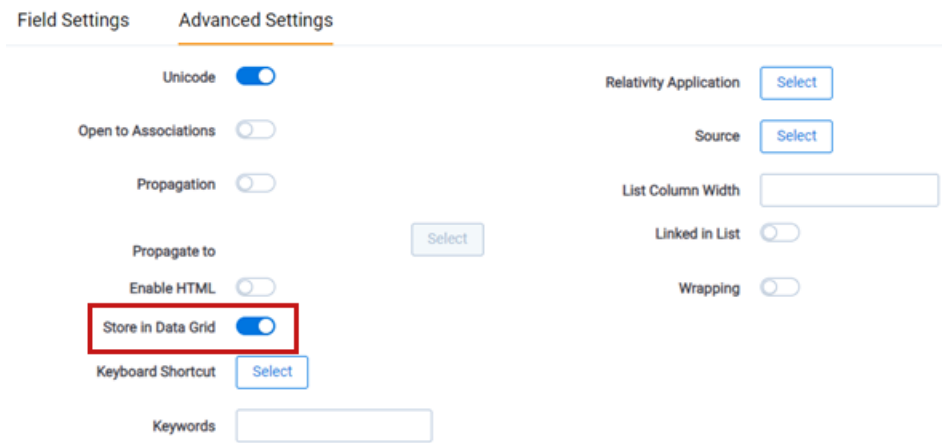
Note: If you run out of space in this repository, you can specify a new repository. Data Grid will continue to read from the old repository as well as the new repository.

4. Click **Save**.

To enable the extracted text field for Data Grid, perform the following steps:

1. Navigate to the **Fields** tab.
2. Locate the extracted text field and click the **Edit** link next to it.

3. Enable the **Store in Data Grid** field under the Advanced Settings tab.



Note: If you are storing extracted text in Data Grid, the **Include in Text Index** field is set to No because there is no SQL text index. If you want to search using dtSearch, you must follow best practice of creating a saved search of fields you want to index.

4. Click **Save**.

Note: Enabling extracted text fields for Data Grid works for new workspaces only. You can't enable Data Grid for fields that already have text in SQL. If you want to migrate fields from SQL to Data Grid, you must use the Data Grid Text Migration application.

Now that you've enabled processing to Data Grid, you can proceed to running a processing set the way you normally would.

Note: The processing engine and Data Grid no longer communicate directly with each other when you process data to Data Grid. Because of this, the Write-to-Grid phase of processing has been deprecated. Instead of writing directly to the grid, the processing engine sends data to the Import API. The Import API receives the data and looks to see whether the workspace is enabled for Data Grid. If the workspace is not Data Grid-enabled, then the Import API sends all of the data to SQL. If the workspace is Data Grid-enabled, then the Import API looks at each field to see which fields are Data Grid-enabled. If the field is Data Grid-enabled, then the Import API sends that data to Data Grid. If the field is not Data Grid-enabled, then it sends that data to SQL.

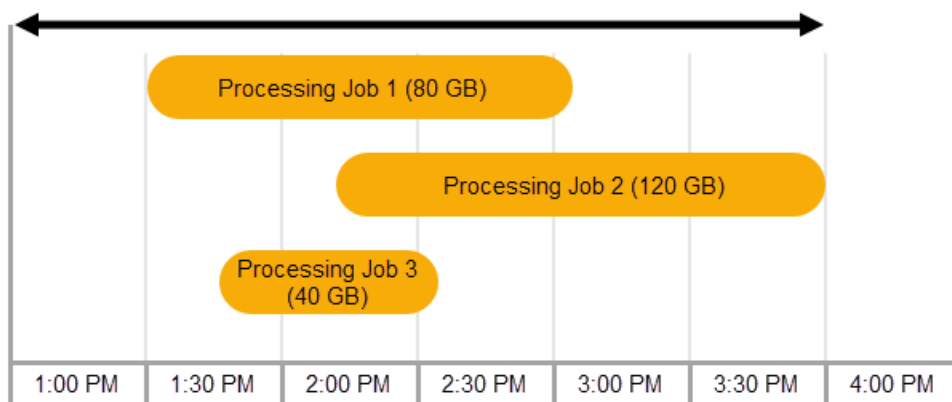
30 Factors that affect processing throughput

This topic explains several factors that may affect processing throughput, or performance for processing discover and publish jobs. If you consistently experience lower-than-usual throughput, check to see if any of the following factors are impacting job performance.

30.1 Calculating processing discover throughput

Processing discover throughput is defined as the amount of concurrent processing work performed on a customer's instance in a given time period. Processing discovery throughput is expressed in gigabytes per hour (GB/hr) and is calculated by adding the size of the data discovered from all of the processing jobs running on the instance, then dividing by the amount of time for the data to process, when at least one processing job was running.

For instance, in the image below, between 1 PM and 4 PM, there are three processing jobs running and 240 GB (80+120+40) of data is discovered in 2.5 hours (3 hours - 0.5 hours when no processing job was running). So, the processing throughput is $240/2.5 = 96$ GB/hr.



30.2 Discover job performance

The following content discusses common factors that may affect discover job throughput.

30.2.1 Factors related to the processing discover job

Discovered size of the dataset versus the raw data size

The discovered size of the dataset can be larger than the raw data size for compressed files.

Compressed files such as .pst, .zip, .mbox, Google Workspace, or Lotus Notes have a discovered dataset size that is larger than the raw dataset after processing discovery completes. In general, e-discovery practitioners use an estimated 1.5 - 2 times the rate for data expansion. So, a 50 GB .pst files which processed at a speed of 50 GB/hr could expand to 100 GB after discovery. Throughput speeds can vary accordingly.

The discovered dataset size depends on the compression type used during data collection and preparation. You can change the compression method for many of these file types, so discovered sizes will differ greatly. To see a table of supported container types, see [Supported container files types](#).

The data within the raw dataset can affect the discovered data size. Multimedia audio, video files, large text files, and large images will compress well and then expand later leading to higher than 2 times the discovered data size.

Number of layers of compression

Customers can compress an already compressed container into a zip and have a multiple-level deep compressed file that is easier to upload (a single large file instead of many small files).

During discovery, multiple levels of a deep compressed file can reduce processing speed as processing resources are spent to get to the files, rather than discovering them. You should balance the need to zip files with processing speed, and only have one-level deep containers when possible, or decompress files as much as possible before discovering.

Number of passwords in the password bank

You should limit the number of passwords stored in the password bank to 100. Relativity attempts each password for every encrypted document, which ultimately reduces discovery speed. For example, it takes approximately 0.4 seconds per password per file to decrypt and with 4000 passwords, your processing time can increase by more than 26 minutes. It takes 10 times as long for Relativity to try 1000 passwords then for 100 passwords.

Additionally, the type of document in the dataset can also affect a processing job with too many stored passwords. For example, a large encrypted Excel file takes longer to process than encrypted Word files.

Number of processing jobs running concurrently on an instance

Relativity distributes processing compute resources equally across all of the jobs in a workspace and across all of the workspaces in an instance. Individual job throughput can be affected when multiple jobs are running.

Priority of jobs in the processing queue

You can manually change the priority of a processing job from the Imaging and Processing Queue tab. Processing resources shift to the priority job.

When running multiple processing jobs concurrently, changing the priority level to a lower number (higher priority) results in all available resources focusing solely on that task until its completion, potentially delaying the progress of

other tasks. Additionally, job priority is at the instance level. A higher-priority job in one workspace can affect the completion of a lower-priority job in another workspace.

For example, if you have the following:

Processing Set A: Priority 10
Processing Set B: Priority 10
Processing Set C: Priority 20

Processing capacity is spread equally on jobs A and B. Job C waits on these sets to complete before starting.

30.2.1.1 Factors related to the files in the processing discover job

Number of large files versus small files in a dataset

Large files over 10 GB process faster than smaller files less than 10 GB as there is more resource startup time involved with the processing of small files. If a dataset has a large number of files, the startup time on the files increases, making the dataset slower. You will see higher GB/hr throughput processing 10 files that are 10 GB each versus 10 files that are 10 KB each.

Number of files in a set dataset (file count per GB of data)

Generally speaking, a dataset with a low file count per GB of data is faster than a dataset with high file count per GB. For example, you will see higher GB/hr throughput processing 10 files that are 10 GB each versus 100 files that are 1 GB each. You can affect throughput of a processing job depending on how you bring in data within the dataset.

Specific file types within a dataset

Certain file types such as large encrypted Excel files, deeply nested emails, or large 7-zips greater than 350 MB, .tar, .e01 and .i01 take longer to ingest data, compared to simpler file types such as email documents, thus affecting processing throughput. Similarly, text extraction is significantly slower with datasets containing large .pdf files needing OCR (when the system is performing OCR on them). Even 1-2 MB .pdf files, .png, .tiff, .jpeg, .cad, and other graphical image files can take several minutes to OCR if the embedded images have high resolution.

30.2.2 Processing profile settings that may impact discover job performance

Text extraction method and OCR accuracy field settings

- **Text Extraction Method**—the type of text extraction method used on the processing profile. Relativity internal performance testing has shown that Relativity text extraction method to be at least two times faster than the native method on Excel, Word, and PowerPoint file types. The method is also less error-prone than native.
- **OCR Accuracy**—the processing job performance is heavily impacted based on the OCR Accuracy setting chosen on the processing profile. A higher accuracy OCR can result in better text for scanned documents and images but at the cost of performance. Do not use the high OCR accuracy unless needed, as it impacts the performance of processing jobs. You can achieve good OCR results with low and medium accuracy levels.

30.3 Publishing job performance

The following content discusses common factors that may affect publish job throughput.

30.3.1 Factors related to the processing publish job

Timing of other jobs running concurrently on an instance

- **Other publish processing job**—a workspace can only have one publish job running deduplication and control number generation phases of publish at a time. If you have multiple publish jobs running on the workspace at the same time, the publish job in the queue waits for the running publish job to finish the deduplication process and control number generation phases before the publish job in the queue can start.
- **Post-publish delete job**—post-publish document deletion jobs have lower priority than other publish jobs in the workspaces and are deferred till all publish jobs finish. If a delete job is in progress and a publish job is sub-

mitted, the delete job will safely pause after a few moments for the publish job to start. The delete job resumes after the publish job completes.

Number of fields mapped on the workspace

Be aware of the number of mapped fields in your workspace. Having more than 100 fields mapped may affect the processing publish job performance.

Mapping field types to processing metadata fields

Be sure that processing metadata fields are mapped to the right field type in Relativity. Mapping a metadata field such as Email Participants to a long text field or a choice field, when the field can have a large number of choices or values, will negatively impact publish performance.

30.3.2 Processing profile settings that may impact publish job performance

Propagate deduplication data and source folder structure field settings

- **Propagate deduplication data**—setting this field to Yes applies the data from the mapped deduplication fields (such as the deduped custodians, deduped paths, all custodians, and all paths fields) from the parent to the children documents so that you can meet production specifications and perform searches on those fields without having to include family or manually overlay those fields. Setting this field to Yes is not recommended unless your use case needs it, as the feature reduces performance when turned on due to the time it takes to overlay the fields onto the children.
- **Do you want to use source folder structure**—setting this field to Yes keeps the folder structure of the source of the files you process when you bring the files into Relativity. This field slows job time if enabled as it reconstructs the folder and inbox structure in the documents tab.

31 Frequently asked questions

Did you know the fastest way to get answers to your questions is through Relativity's AI chatbot? Click the **Ask AI** button located in the upper right corner of your Relativity screen, and enter a prompt. The chatbot will search both the help and Community sites right from your Relativity screen!



31.1 Calendar appointments

Why does a single recurring Outlook appointment produce many attachments during processing?

A recurring appointment is stored as a series master. Each time an occurrence in the series is changed, such as when it is canceled, rescheduled, or edited, Outlook saves a separate snapshot of that occurrence in the PST file. Outlook displays only the most recent version of the occurrence on the calendar. However, Relativity extracts each stored snapshot as a separate attachment, or child document, linked to the series master. As a result, a recurring series that has been modified many times can produce dozens or even hundreds of attachments.

31.2 Date and time

How does Relativity handle calendar metadata?

The processing engine captures all the dates of calendar items. If there is not a field for it in Relativity, this data will end up in the "OtherProps" field.

How does processing handle time zones?

During Discovery, native date and time information is extracted and normalized to Coordinated Universal Time (UTC) if not already stored in this format. Processing then uses the time zone configured in the Processing Profile (and corresponding Data Sources that inherit this setting) to apply the desired UTC offset to date and time metadata fields. These adjusted values are also reflected in the extracted text, for example, the Sent Date/Time displayed in the

header portion of the email text.

In the case of emails, the Sent Date/Time value from the message header is used in the generation of the processing deduplication hash. However, this approach utilizes pre-adjusted date and time values, allowing for deduplication of emails sent and received in different time zones.

For Daylight Saving Time (DST), Relativity maintains a table that stores current and historical DST rules by location. In this way, the appropriate DST adjustment is always applied on top of the UTC offset, given a document's specific date in combination with the selected time zone. For example, a change to how the United States observes DST went into effect in 2007. Relativity applies the appropriate DST adjustment, which is different for documents authored before and after 2007, provided the configured time zone falls within the United States.

Processing also utilizes the configured time zone to populate two system fields that are automatically mapped and published to the workspace: Relativity Native Time Zone Offset and Time Zone Field. These are used downstream for Native Viewer file conversion and Imaging activities. For more information, see System-Mapped Fields.

Can the date and time format be modified?

Yes. Date/time formatting displayed in extracted text, imaging results, and the native viewer is influenced by the processing region configured for your Relativity instance. This setting typically corresponds to the geographical location of your Relativity instance, but it can be changed. Note that this setting does NOT change the display of dates and times within Document fields or elsewhere in the user interface; these are controlled by your web browser's language and localization settings. Contact your Customer Success Manager for questions regarding the processing region instance settings or browser-based localization settings.

31.3 DeNIST

What files display in the DeNIST report?

The DeNIST report displays only attachments. You can look at the INVXXXXX database in the DeNIST table to see the individual files.

Does Relativity allow the use of a custom NIST?

There's no official support for a custom NIST list.

31.4 Discover

Can file names be appended after discovery?

There's currently not a way to append a file name after discovery but prior to publishing.

31.5 Errors

Can images be reprocessed if they have errors?

As long as the set hasn't been published, if the image reports an error, you can retry the image and/or make corrections to the native and then retry the error.

How does processing work with errors?

If you publish a processing set, even documents that have an error associated with them will get a record entered in the Document object/tab, with the Control Number field populated at the very least, even if Relativity was unable to determine anything else about the document. In other words, just because a document has an error during processing doesn't mean that it won't be displayed in the Document list with a control number when you publish the processing set. The only way this doesn't happen is if the error occurs during ingestion and either Relativity is unable to extract the document from a container or the source media is corrupt.

31.6 Media processing

How does Relativity process audio and video?

Audio and video files are identified, but no metadata (other than basic metadata) or text is extracted from them. They will be marked as unprocessable.

31.7 Other

How does processing handle regional setting changes?

When a user submits a ticket to change the processing region setting for the Relativity instance, it is best to have no processing jobs running. This setting takes effect for new jobs and impacts the results of zip file/folder name extraction, text extraction, imaging, and native viewer conversion. Contact your Customer Success Manager for more details on how changing this setting can impact your review experience.

Is it possible to publish a subset of discovered files from a processing set?

Relativity supports publishing an entire processing set, excluding any files that are marked as deleted. This includes items like containers, files that meet one or more discover filter conditions, and deduplicated files. For best results, use discover filters to exclude specific documents from the publishing process. You can also apply filters during the inventory phase.

For more information, see [Inventory / Discovery settings](#). You can also configure and apply filters before discovery.

For more information, see [Filtering files](#).

Once files are published, are they deleted from the processing source location?

No, there is no alteration to the processing source location. Relativity reads from this intermediate location and copies the files to the Relativity workspace file repository.

31.8 Password-protected files

Does Relativity process password protected PST or OST files?

Passwords on PST and OST files are bypassed automatically by Relativity.

Proprietary Rights

This documentation (“**Documentation**”) and the software to which it relates (“**Software**”) belongs to Relativity ODA LLC and/or Relativity’s third party software vendors. Relativity grants written license agreements which contain restrictions. All parties accessing the Documentation or Software must: respect proprietary rights of Relativity and third parties; comply with your organization’s license agreement, including but not limited to license restrictions on use, copying, modifications, reverse engineering, and derivative products; and refrain from any misuse or misappropriation of this Documentation or Software in whole or in part. The Software and Documentation is protected by the **Copyright Act of 1976**, as amended, and the Software code is protected by the **Illinois Trade Secrets Act**. Violations can involve substantial civil liabilities, exemplary damages, and criminal penalties, including fines and possible imprisonment.

©2026. Relativity ODA LLC. All rights reserved. Relativity® is a registered trademark of Relativity ODA LLC.