Relativity one

Processing User Guide

March 12, 2024

For the most recent version of this document, visit our **documentation website**.

Table of Contents

1 Processing 10

1.1 Basic processing workflow 10

2 Processing to Data Grid 12

2.1 Enabling processing to Data Grid 12

3 Supported file types for processing 14

3.1 Supported file types 14

3.2 Notable unsupported file types 17

3.3 Supported container file types 20

3.3.1 Container file types supported for the password bank 21

3.3.2 Non-container file types supported for the password bank 21

4 Short message conversion for Slack 23

4.1 Required file structure 23

4.2 Troubleshooting 24

4.3 Resources 24

5 Google Workspace 25

5.1 Collect 25

5.2 Google Vault 25

5.3 Google Workspace metadata field lists 26

5.3.1 Google Workspace fields 26

6 Microsoft 30

6.1 Outlook message item (.msg) to MIME encapsulation (.mht) conversion 30

6.1.1 Email image extraction support 30

6.1.2 Microsoft Office child extraction support 31

6.2 Excel files 32

7 Other software support 33

7.1 Multi-part forensic file considerations 33

7.2 Native text extraction and OCR 33

7.3 Password-protected .rar files 33

7.3.1 Lotus Notes 34

7.4 Multi-part container considerations 37

7.5 Calendar file, vCard file considerations 38

8 Password bank 41

8.1 Password bank in processing workflow 41

8.2 Password Bank in imaging workflow 42

8.3 Creating or deleting a Password Bank entry 43

8.3.1 Fields 43

8.3.2 Example password 46

8.4 Validations, errors, and exceptions 46

8.5 Viewing audits 47

9 Mapping processing fields 48

9.1 Mapping fields 48

9.1.1 Relativity system field considerations 49

9.1.2 Field mapping validations 50

9.2 System-mapped fields 50

9.3 Optional fields 51

9.4 Email Store Name field 76

9.5 Virtual path 78

9.6 Processing folder path 79

9.7 Email folder path 79

9.8 Source path 79

9.9 Message ID considerations 80

9.10 Comments considerations 80

9.11 De-duped custodian and path considerations 81

10 Processing profiles 82

10.1 Creating or editing a processing profile 82

10.1.1 Fields 82

10.1.2 Inventory / discovery settings 91

10.1.3 Extraction settings 94

10.1.4 Short message conversion settings 99

10.1.5 Deduplication settings 100

10.1.6 Publish settings 102

10.1.7 Other considerations 104

11 Deduplication considerations 111
11.1 Global deduplication 111
11.2 Custodial deduplication 112
11.3 No deduplication 113
11.4 Global deduplication with attachments 114
11.5 Global deduplication with document-level errors 115
11.6 Technical notes for deduplication 116
11.6.1 Calculating MD5/SHA1/SHA256 hashes 116
11.6.2 Calculating deduplication hashes for emails 117
11.6.3 Calculating the Relativity deduplication hash 118
12 Processing sets 120
12.1 Processing sets default view 120
12.2 Creating a processing set 122
12.3 Processing set fields 123
12.4 Adding a data source 125
12.5 Data source fields 127
12.5.1 Order considerations 133
12.5.2 Edit considerations for data sources 134
12.5.3 Processing data source view 134
12.5.4 Job Errors view 136
12.6 Processing Data Sources tab 137
12.7 Deleting a processing set 141
12.8 Avoiding data loss across sets 144
12.9 Copying natives during processing 144
13 Processing Set Details layout (preview) 148
13.1 What’s new? 148
13.2 Accessing the preview page 148
13.2.1 Turning off the preview page 148
13.2.2 Providing feedback 148
13.3 Viewing job details 149
13.3.1 Inventory 149
13.3.2 Discover 150

13.3.3 Publish 150
13.3.4 Resources 151
14 Inventory overview 152
14.1 Inventory process 153
14.2 Monitoring inventory status 155
14.3 Inventory progress 156
14.4 Discovering files from Inventory 158
15 Inventory processing 159
15.1 Running inventory 159
15.1.1 Canceling inventory 160
15.2 Filtering files 160
15.2.1 Applying a Date range filter 164
15.2.2 Applying a File Size filter 166
15.2.3 Applying a deNIST filter 167
15.2.4 Applying a Location filter 167
15.2.5 Applying a File Type filter 168
15.2.6 Applying a Sender Domain filter 169
15.3 Removing filters 173
15.4 Inventory errors 173
15.4.1 Inventory error scenarios 173
15.5 Re-inventory 174
16 Discovering files 176
16.1 Running file discovery 177
16.1.1 Discovery process 179
16.1.2 Container extraction 180
16.2 Special considerations - OCR and text extraction 181
16.3 Monitoring discovery status 181
16.4 Viewing text extraction progress in processing sets 184
16.5 Canceling discovery 184
17 Files tab 187
17.1 Views on the Files tab 187
17.1.1 All Files view 188

17.1.2 Deleted Documents view 189
17.1.3 Current Errored Files view 190
17.1.4 All Errored Files view 191
17.2 Details modal 193
17.3 Retrying delete errors 197
17.4 Republishing files from the Files tab 198
17.4.1 Republishing files where job errors are present 200
17.4.2 Common use cases for using the Republish mass operation 201
17.5 Retrying errors 201
17.6 Saved filters 203
17.6.1 Right-click options 206
17.7 Single Download / Replace 208
17.8 Mass download errored files 210
18 Publishing files 213
18.1 Running file publish 214
18.1.1 Publish process 217
18.2 Monitoring publish status 219
18.3 Canceling publishing 220
18.4 Republishing a processing set 221
18.5 Retrying errors after publish 223
19 Post-publish delete 224
19.1 Overview 224
19.2 Running post-publish delete 224
19.3 Interacting with deleted documents 227
19.4 Retrying delete errors 227
20 Error overview 229
20.1 Types of processing errors 229
20.2 Processing phases where errors may occur 229
20.3 Processing error tabs 230
20.3.1 Files tab 230
20.3.2 Job Errors tab 239
20.4 Viewing errors from a processing set 242

20.5 Required security permissions 242

21 Error resolution 243

21.1 Types of processing errors 243

21.2 Processing phases where errors may occur 243

21.3 File errors 244

21.3.1 File error actions 244

21.4 Resolving file errors 246

21.4.1 Retrying an error from the processing set 247

21.4.2 Retry file errors mass action 248

21.4.3 Replacing a corrupted file 249

21.4.4 Resolving a password-protected file error 250

21.5 Job errors 250

21.5.1 Job error workflow 250

21.5.2 Resolving job errors 251

22 Reports 254

22.1 Generating a processing report 254

22.2 Data Migration 255

22.2.1 Summary Statistics: Data Migration 255

22.2.2 Excluded Files 256

22.2.3 Processing Sets 256

22.3 Master Document Replacement Summary 256

22.3.1 Deleted Master Documents 256

22.3.2 Replacements Master Documents 257

22.3.3 Processing Sets 257

22.4 Discovery Exclusion Results 257

22.4.1 Discover Filter Settings 257

22.4.2 File Type | File Size | Excluded File Count 257

22.4.3 Processing Sets 258

22.5 Discovered Files by Custodian 258

22.5.1 Discovered Files by Custodian 258

22.5.2 File Types Discovered - Processable 258

22.5.3 File Types Discovered - Processable (By Custodian) 258

22.5.4 File Types Discovered - Unprocessable 258
22.5.5 File Types Discovered - Unprocessable (by Custodian) 258
22.5.6 Processing Sets 259
22.6 Discovered Files by File Type 259
22.6.1 Discovered Files by Custodian 259
22.6.2 File Types Discovered - Processable 259
22.6.3 File Types Discovered - Processable (By File Type) 259
22.6.4 File Types Discovered - Unprocessable 259
22.6.5 File Types Discovered - Unprocessable (By File Type) 260
22.6.6 Processing Sets 260
22.7 Document Exception 260
22.7.1 Document Level Errors - Discovery 260
22.7.2 Document Level Errors - Publishing 260
22.7.3 Processing Sets 261
22.8 File Size Summary 261
22.8.1 Pre-Processed File Size 261
22.8.2 Processed File Size 261
22.8.3 Published File Size 261
22.9 Inventory Details 261
22.9.1 Inventory Filter Settings 262
22.9.2 Excluded by File Type Filter | Excluded File Count 262
22.9.3 Excluded by Location Filter | Excluded File Count 262
22.9.4 Excluded by Sender Domain Filter | Excluded File Count 262
22.9.5 Processing Sets 262
22.10 Inventory Details by Custodian 262
22.10.1 Inventory Filter Settings 262
22.10.2 Custodian | Excluded by File Type Filter | Excluded File Count 263
22.10.3 Custodian | Excluded by File Location Filter | Excluded File Count 263
22.10.4 Custodian | Excluded by Sender Domain | Excluded File Count 263
22.10.5 Processing Sets 263
22.11 Inventory Exclusion Results 263
22.11.1 Inventory Filter Settings 263

22.11.2 File Type | Excluded File Count 264
22.11.3 Location | Excluded File Count 264
22.11.4 Sender Domain | Excluded File Count 264
22.11.5 Processing Sets 264
22.12 Inventory Exclusion Results by Custodian 264
22.12.1 Custodian | Excluded by File Type Filter | Excluded File Count 264
22.12.2 Custodian | Excluded by File Location Filter | Excluded File Count 264
22.12.3 Custodian | Excluded by Sender Domain | Excluded File Count 264
22.12.4 Processing Sets 265
22.13 Inventory Summary 265
22.13.1 Initial Inventory Results 265
22.13.2 Filtering Summary 265
22.13.3 Final Inventory Results 266
22.13.4 Processing Sets 266
22.14 Job Exception 266
22.14.1 Job Level Errors 266
22.14.2 Processing Sets 266
22.15 Text Extraction 266
22.15.1 Text Extraction by Custodian 267
22.15.2 Text Extraction by File Type 267
22.15.3 Breakdown by Error Message 267
22.15.4 Processing Sets 267
23 Processing administration 269
23.1 Security considerations for processing administration 269
23.2 Monitoring active jobs 271
23.2.1 Active jobs mass operations 274
23.3 Using the Processing History sub-tab 275
23.3.1 Auto refresh options for processing history 277
24 Frequently asked questions 279

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.

Note: Your environment has been enabled to dynamically scale your Invariant worker servers dependent on load. Sustained activity is automatically detected by the system, and Relativity will add workers to handle this work. Once the work is done, they will automatically scale back down. This feature is continually being improved to be smarter about when we add workers and how many we add.

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.

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

Using processing

You are a litigation support specialist, and the lead attorney hands you a CD containing data on a key custodian. There are about 200,000 files on the disc, and he is only looking for files from an 18-month period.

You use Relativity's processing feature to bring in that custodian’s data to Relativity and then to filter it based on what the lead attorney is looking for in this case. To do this, you first save the files into a folder and create a new custodian, Joe Smith.

Then you create a new processing set, to which you add a data source that has only Joe Smith associated with it. This data source includes a source path that is the folder in which you saved the custodian's files.

Once you save the processing set, you can inventory that custodian's data and eliminate all the files that fall outside of the 18-month period you are dealing with. Once you narrow down the data set to the most relevant files, you can discover them and give the lead attorney a reviewable set of documents.

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.

2 Processing to Data Grid

By processing directly into Data Grid™, you have the opportunity to improve your publishing speeds.

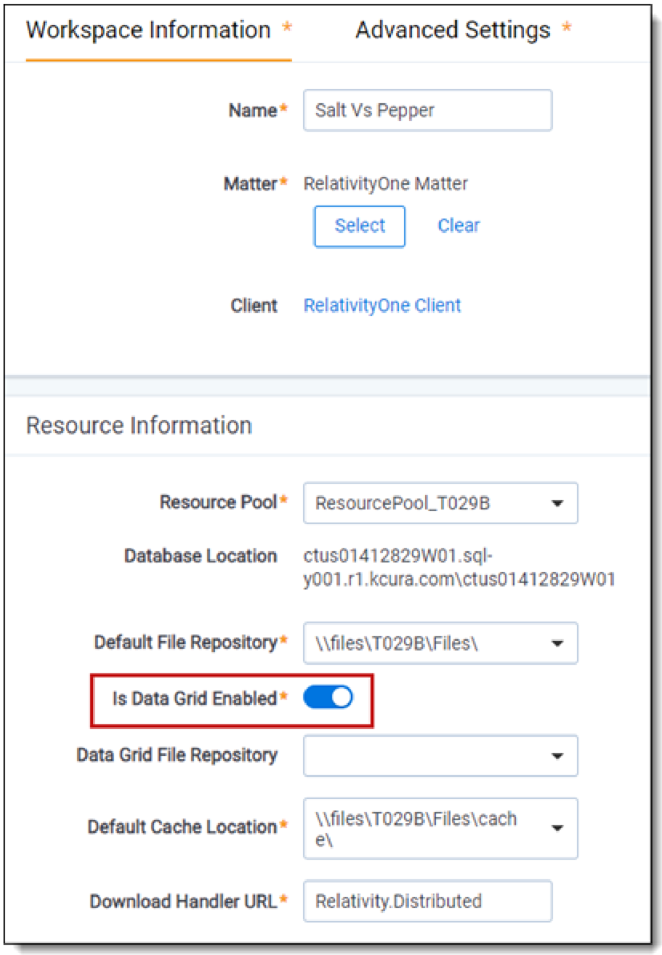
2.1 Enabling processing to Data Grid

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.



Name* Salt Vs Pepper

Matter* RelativityOne Matter

Client RelativityOne Client

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

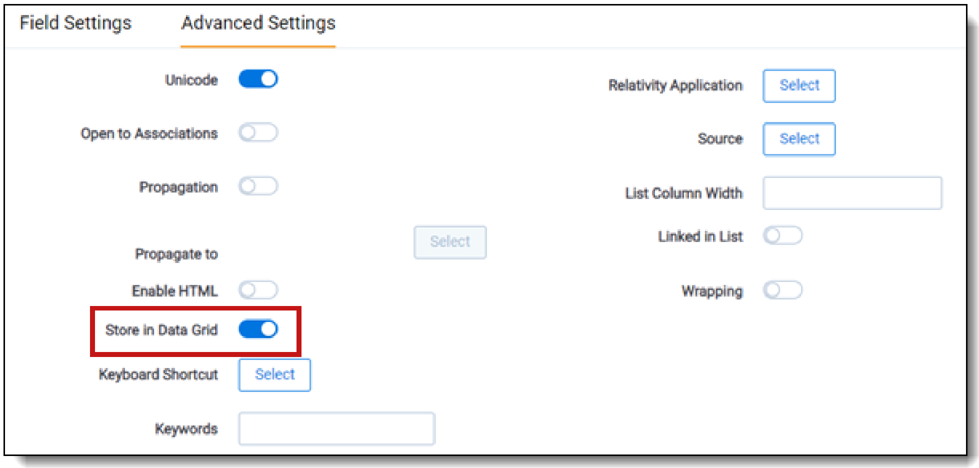
3. (Optional) Next to **Data Grid File Repository**, select the path for the physical location of the text files used by Data Grid. If no file repository is specified for this field, and Data Grid is enabled, Data Grid stores text in the default file repository.

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.



Unicode ●
Open to Associations ○
Propagation ○
Propagate to
Enable HTML ○
Store in Data Grid ●
Keyboard Shortcut
Keywords
Relativity Application
Source
List Column Width
Linked in List ○
Wrapping ○

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.

3 Supported file types for processing

Relativity supports many file types for processing. There are also some file types that are incompatible with the processing engine. Before processing your data, note the supported and unsupported file types, as well as any caveats involved with processing those file types.

Note: 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.

This page covers supported and unsupported file types. You can find additional supported file type details and considerations on the following pages:

- Google Workspace
- Microsoft
- Short message conversion for Slack
- Other software support

3.1 Supported file types

Relativity supports the following file types and extensions for processing.

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.

File type	Extensions
Adobe files	.pdf, .fm, .ps, .eps ▪ Processing support for .xfa pdf, .pdf web forms, includes extraction of text, metadata, and imaging. Some 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. ▪ Relativity performs Optical Character Recognition (OCR) on .pdf files during processing. Relativity handles a .pdf portfolio, which is an integrated .pdf unit containing multiple files, by extracting the metadata and associating it with the files contained in the portfolio.
AppleDouble	AppleDouble-encoded attachments in e-mails.
CAD files	.dxf, .dwg, .slddrw, .sldprt, .3dxml, .sldasm, .prtdot, .asmdot, .drwdot, .stl, . eprt, .easm, .edrw, .eprtx, .edrwx, .easmx

File type	Extensions
	■ The OCR output for processed CAD files can vary significantly.
Compressed files	.7z, .zip, .tar, .gz, .bz2, .rar, .z, .cab, .alzip Zip file containers do not store time zone information for CreatedOn, LastModified, and LastAccessed fields. When extracting files, time stamps are only meaningful if the time zone that the zip file container was created in is known. 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 local time zone. LastModified and LastAccessed fields will usually match. **Note:** Relativity does not support multi-part .zip, .tar, or .7z files.
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.
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
EnCase	e01, .ex01, .l01, .lx01 ■ Processing supports .e01 and .ex01 files for the following operating and file systems: ○ Windows—NTFS, FAT, ExFAT ○ Mac—HFS+ ○ Linux (Ubuntu)—EXT2, EXT3, EXT4 ■ Deleted files that exist on an .e01 and .ex01 (disk) image file are skipped

File type	Extensions
	during processing, with the exception of recycle bin items, which are processed with limited metadata. ■ Encrypted EnCase files are not supported. You must decrypt EnCase files prior to processing them. For details on .e01 file type handling, see Multi-part forensic files.
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. **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.
Google Workspace	For details, see Supported file types - Google Workspace.
Hangul	.hwp Hangul word processing files 1997 up to 2010 are supported.
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.
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, .exf, . exif, .webp, .wdp
JungUm Global	.gul
OneNote	.one ■ Relativity uses Microsoft connectors to extract information from OneNote files at the section, or tab, level. All pages within a section are extracted as one file. During ingestion, Relativity extracts embedded items from OneNote files, and for some object types, generates them as .pdf or .tiff files natively. ■ The Password Bank does not support OneNote files. ■ RelativityOne does not support OneNote 2003 files.
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. ■ 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.
Publisher	.pub

File type	Extensions
Project	.mpp, .mpt, .mpd, .mpx **Note:** The text extracted from Project files is from the Gantt chart view and includes Task Notes.
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.
Text files	Such as .txt or .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.
Vector files	.svg, .svgz, .wmf, .plt, .emf, .snp, .hpgl, .hpg, .plo, .prn, .emz, .wmz
Visio	.vsd, .vdx, .vss, .vsx, .vst, .vsw, .vsdx, .vsdm - Visio is a separate installation per the Worker Manager server page. - You must have Office 2013 or Office 2016 installed in order to process .vsdx and .vsdm file extensions.
Word	.docx, .docm, .dotx, .dotm, .doc, .dot, .rtf Word 2.0 through the current product version is supported, including templates.
WordPerfect	.wpd, .wps

Note: 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.

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.

3.2 Notable unsupported file types

Processing does not support files created with the following programs and versions:

Product category	Product name and version
DOS Word Processors	- 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

Product category	Product name and version
	■ Enable 3.0, 4.0, and 4.5 ■ 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
Windows Word Processors	■ 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

Product category	Product name and version
	■ 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
Mac Word Processors	MacWrite II Version 1.1
Disk Images	Symantec Ghost
Encryption	Pretty Good Privacy (PGP)
HEIC	High Efficiency Image Container
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 ■ Quattro Pro (Windows) Versions through 12.0, X3 ■ SmartWare II Version 1.02 ■ SuperCalc 5 Version 4.0 ■ VP Planner 3D Version 1.0

In addition, Processing does not support the following files:

- Self-extracting .rar files
- Private mail certificate (.pem) files

- Apple i-Works suite (Pages, Numbers, Keynote)
- Apple Mail:
  - .emlxpart
  - .partial.emlx

  Note: The .emlxpart and .partial.emlx are distinct from the .emlx file extension, which is supported by processing.

- Audio/Video files
  - .wav
- iCloud backup files
- Microsoft Access
- Microsoft Works
- Raw partition files:
  - ISO
  - NTFS
  - HFS

3.3 Supported container file types

The following file types can act as containers:

File type	Extensions
Bloomberg	.xml Relativity does not support Instant Bloomberg .xml files.
Cabinet	.cab Relativity does not support multi-part .cab files. Relativity does not support Password Protected .cab files.
Compressed files	.7z, .zip, .tar, .gz, .bz2, .rar, .z, .cab, .alzip 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. Inventory, however, only extracts data from first-level documents. For example, 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. **Note:** Relativity does not support multi-part .zip, .tar, or .7z files. **Note:** 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.
EnCase	.e01, .l01, .lx01, .ex01
AccessData Logical	.ad1

File type	Extensions
Image	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 for more information.
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 doe not apply to the calendar file itself.
Lotus Notes Database	.nsf 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.
Outlook Offline Storage	.ost
Outlook Mail Folder	.pst Relativity assigns duplicate hash values to calendar invites, as it does with email messages and other documents.
Outlook Express Mail Folder	.dbx
PDF Portfolio	.pdf
RAR	.rar You do not need to combine multi-part .rar files before processing them. See Password-protected .rar files for more information.
ZIP	See Compressed files.

Note: Container files do not store encoding information. Because of this, you may see garbled characters in the file names of children files processed from a container file (such as a .zip file) if the originating locale differs from the processed locale. For example, if the originating container file's locale is set to Russian, then processed on an instance set to US, the container's children files may have garbled characters.

3.3.1 Container file types supported for the password bank

The following container file types are supported by Relativity for Password Bank in Inventory.

File type	Extensions
Compressed files	.7z, .alzip, .zip, .z, .bz2, .gz
Lotus Notes Database	.nsf
PDF Portfolio	.pdf
PST	.pst
RAR	.rar

3.3.2 Non-container file types supported for the password bank

The Password Bank also supports the following non-container formats:

- .pdf
- Excel*
- Word*
- PowerPoint*
- S/MIME
- .p7m

* Except .drm files or custom encryption

4 Short message conversion for Slack

This page covers information for Slack conversion. To view the Relativity supported and unsupported file type tables, see Supported file types for processing.

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.

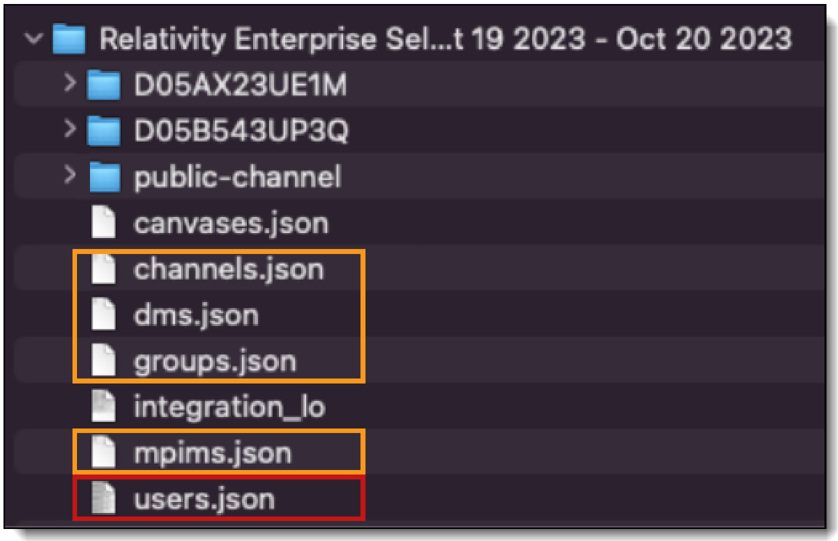
4.1 Required file structure

Processing supports original structure of zip containers coming from Slack. Do not change the contents of the zip file or you may see corrupt or incomplete results.

Note: Slack exports greater than 10 GB are not supported and may be unable to process. You must repackage the contents into smaller packages if you have a Slack export over 10 GB. For example, you might split different conversation types across different zips.

You must have at least two of the following files at the root level of your zip container:

- users.json (required)
- Plus, one of the following:
  - channels.json
  - dms.json
  - groups.json
  - mpims.json



4.2 Troubleshooting

- Processing supports workspace-level exports containing exports from a single workspace. Organizational-level (org-level) exports are not supported.

  Note: If your directory structure includes the file org_users.json instead of users.json, you have an org-level export package, which is not supported. To process the data, you must repackage the directories found in the teams directory into standalone workspace-level zip containers.

- Inventory for Slack export containers is not supported.
- 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.

4.3 Resources

For additional information, see:

- 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.

5 Google Workspace

This page covers additional supported file type (for processing) information for Google Workspace. To view the Relativity supported and unsupported file type tables, see Supported file types for processing.

When you gather Google Workspace data, Google provides a supplemental metadata file. For Google Mail (Gmail), the metadata file is a .csv file. For Google Drive, the metadata file is an .xml file.

Processing automatically identifies when the supplemental metadata file is present in your file share, links the metadata fields to the processed files, and makes them available as mappable fields.

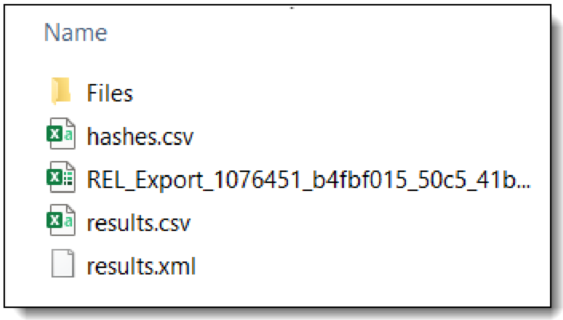
There are two methods for collecting Google Workspace data:

- Collect in Relativity. For more information, see the Google Workspacedocumentation.
- Manual download via Google Vault.

5.1 Collect

When using Collect to gather Gmail data, the metadata file is provided as a .csv file, an .xml file is also provided but not used. When using Collect to gather Google Drive data, the metadata file is provided as an .xml file.

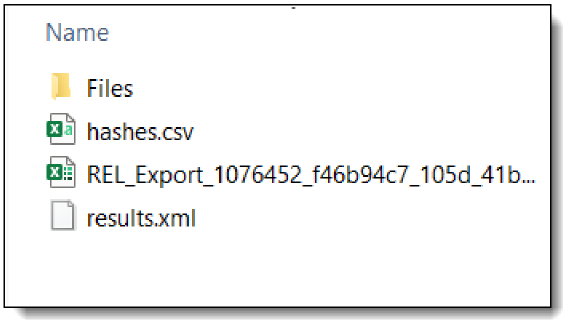
Relativity Collect: Gmail



Name

Files
hashes.csv
REL_Export_1076451_b4fbf015_50c5_41b...
results.csv
results.xml

Relativity Collect: Google Drive



Name

Files
hashes.csv
REL_Export_1076452_f46b94c7_105d_41b...
results.xml

Note: Collect places additional files in the root of the data source. They are processed as loose individual files if they are not removed prior to processing.

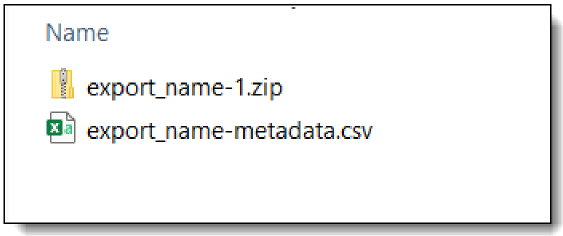
Other notes

- Google Chat data is converted to Relativity's short message format (RSMF) for processing when using Collect to gather data.
- Do not edit the supplemental metadata files or Processing will not recognize them.

5.2 Google Vault

When using Google Vault to gather Gmail data, download the [EXPORT_NAME].zip, ;and [EXPORT_NAME]-metadata.csv files. For Google Drive data, download the [EXPORT_NAME].zip and [EXPORT_NAME]-metadata.xml files. When collecting data manually from Google Vault, you must export Gmail data in MBOX format. Relativity does not sync additional metadata if exported as a .pst file.

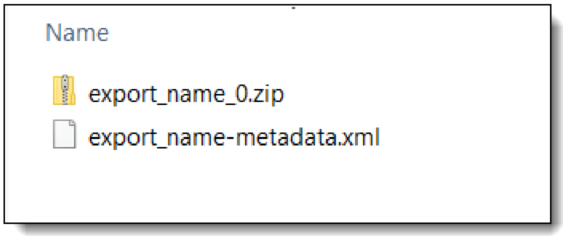
Google Vault: Gmail



export_name-1.zip

export_name-metadata.csv

Google Vault: Google Drive



export_name_0.zip

export_name-metadata.xml

Note: If you have files saved in your root folder other than the supplemental metadata files, they are processed as loose individual files.

Other notes

- Google chat data must be converted to Relativity's short message format (RSMF) prior to processing when using Google Vault. Google chat data does not need to be converted to RSMF when using Collect.
- Do not edit the supplemental metadata files or Processing will not recognize them.

5.3 Google Workspace metadata field lists

The Google supplemental metadata files contain the fields listed below, all of which are available for mapping to Document object fields.

Note: Google may add, remove, or edit fields at any time. You can find the most current lists on their Vault export contents web page.

Mapping Google Drive fields with deduplication turned off

This section only applies when processing Google Drive data with deduplication turned off.

Relativity uses MD5 hash for deduplication when processing, and with deduplication turned off, two files with differing cloud-based metadata (sidecar metadata) can have the same physical MD5 hash. Since Relativity associates sets of metadata with a physical file hash, files with identical file hashes can be populated with identical sidecar metadata. To resolve this issue, Relativity has developed a set of source fields specific to Google Drive. When processing with deduplication turned off, this forces files with the same hash to have their own unique set of metadata.

5.3.1 Google Workspace fields

The following table lists the metadata fields found in the Google Drive .xml file. Use these fields when processing Google Drive files with deduplication turned off.

Google drive .xml field	Relativity source field name	Description
DocID	GoogleDrive/DocID	A unique identifier for the file. For site exports, the value is the page ID.
#Author	GoogleDrive/Author	The email address of the person who owns the file in Drive. For a shared drive file, it shows the shared

Google drive .xml field	Relativity source field name	Description
		drive name.
Collaborators	GoogleDrive/Collaborators	The accounts and groups that have direct permission to edit the file or add comments. Also includes users with indirect access to the file if you chose this option during export.
Viewers	Google/Viewers	The accounts and groups that have direct permission to view the file. Also includes users with indirect access to the file if you chose this option during export.
#DateCreated	GoogleDrive/DateCreated	The date a Google file was created in Drive. For non-Google files, the date the file was uploaded to Drive.
#DateModified	GoogleDrive/DateModified	The date the file was last modified.
#Title	Google/Title	The file name as assigned by the user. Because some operating systems cannot expand zip files with extremely long file names, Vault truncates the file name at 128 characters during export. The value shown by the #Title tag isn't truncated.
DocumentType	GoogleDrive/DocumentType	The file type for Google files. Possible values are: ■ DOCUMENT—a document created 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.
Others	Google/Others	The accounts from your query that have indirect access to the file if you opted to exclude access level information during export. May also include users for whom Vault could not determine permission levels at the time of export.
DocParentID	Google/DocParentID	For sites, the site ID.
SharedDriveID	Google/SharedDriveID	The identifier of the shared drive that contains the file, if applicable.
SourceHash	Google/SourceHash	A unique hash value for each version of a file. Can be used to deduplicate file exports and verify the exported file is an exact copy of the source file. Supported by Google Docs, Sheets, and Slides files only.

Google drive .xml field	Relativity source field name	Description
FileName	GoogleDrive/FileName	The file name. Use this value to correlate the metadata with the file in the export ZIP file.
FileSize	GoogleDrive/FileSize	The size of the file in bytes.
Hash	GoogleDrive/FileHash	The MD5 hash of the file.
UserQuery	Google/UserQuery	The query submitted by the Vault user that retrieved the files included in this export.
TimeZone	Google/TimeZone	The time zone used for date-based searches.
Custodians	Google/Custodians	The email addresses of the users whose accounts were searched. If you searched for content rather than individual user accounts, there are no custodians listed here.
Labels	GoogleDrive/Labels	Labels applied to the message by Google Drive or the user.

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

Google mail .cvs field	Relativity source field name	Description	Notes
Rfc822MessageId	Google/Rfc822MessageId	A message ID that is the same for the receiver's and sender's 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	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	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	

Google mail .cvs field	Relativity source field name	Description	Notes
		value of Account is user-1@company.com.	
From	Google/From	The sender account.	
To	Google/To	The recipient account. Multiple recipients are comma-separated and the list is in double quotes.	Gmail only
CC	Google/CC	Accounts in the cc: field.	Gmail only
BCC	Google/BCC	Accounts in the bcc: field.	Gmail only
Subject	Google/Subject	The message subject.	Gmail only
Labels	Google/Labels	Labels applied to the message by Gmail or the user.	Gmail only
DateSent	Google/DateSent	The message send date in UTC, yyyy-MM-dd'T'HH:mm:ssZZZZ.	Gmail only
DateRecieved	Google/DateRecieved	The message received date, yyyy-MM-dd'T'HH:mm:ssZZZZ.	Gmail only
SubjectAtStart	Google/SubjectAtStart	The subject of the conversation when the first message was sent.	Classic Hangouts only
SubjectAtEnd	Google/SubjectAtEnd	The subject of the conversation when the last message was sent.	Classic Hangouts only
DateFirstMessageSent	Google/DateFirstMessageSent	The time stamp for when the first message in a conversation was sent.	Classic Hangouts only
DateLastMessageSent	Google/DateLastMessageSent	The time stamp for when the last message in a conversation was sent.	Classic Hangouts only
DateFirstMessageReceived	Google/DateFirstMessageReceived	The time stamp for when the first message in a conversation was received.	Classic Hangouts only
DateLastMessageReceived	Google/DateLastMessageReceived	The time stamp for when the last message in a conversation was received.	Classic Hangouts only
ThreadedMessageCount	Google/ThreadedMessageCount	The number of messages in the conversation.	Classic Hangouts only

6 Microsoft

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

6.1 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 encapsultation (.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 OutsideIn. 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.

6.1.1 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

Email format	Attachments that are inline, embedded, images
	▪ 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**.

6.1.1.1 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.

6.1.2 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	.xltx	√	√
Excel	.xltm	√	√
Excel	.xls	√	√*
Excel	.xlt	√	√*
Excel	.xla	√	√*
Excel	.xlm	√	√*
Excel	.xlw	√	√*
Excel	.uxdc		
Outlook	.msg	√	√
Word	.docx	√	√
Word	.docm	√	√

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

6.2 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.

7 Other software support

This page 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.

7.1 Multi-part forensic file considerations

When processing a multi-part forensic image, make sure the source location points to the root folder that contains all of the files that make up the image. If you select only the first file of the image, such as .e01, .l01, .ex01, .lx01, inventory and discovery fail with an unrecoverable error.

This is because inventory looks at where files reside in the processing source folder and does not copy them to the repository. If you only select the first file, that file only is copied to the repository during discovery. The workers attempt to extract from it and fail since the rest of the archive is not available.

When processing .e01 files, the following NTFS file system files are skipped:

- Unallocated space files
- Index $I30 files
- $TXF_DATE files

For details on EnCase file type support, see the EnCase entry on the supported files types table.

7.2 Native text extraction and OCR

Processing distinguishes between text and line art in the documents you process. For these documents, processing will only OCR the line art. This means that Relativity does not skip OCR if a page has electronic text.

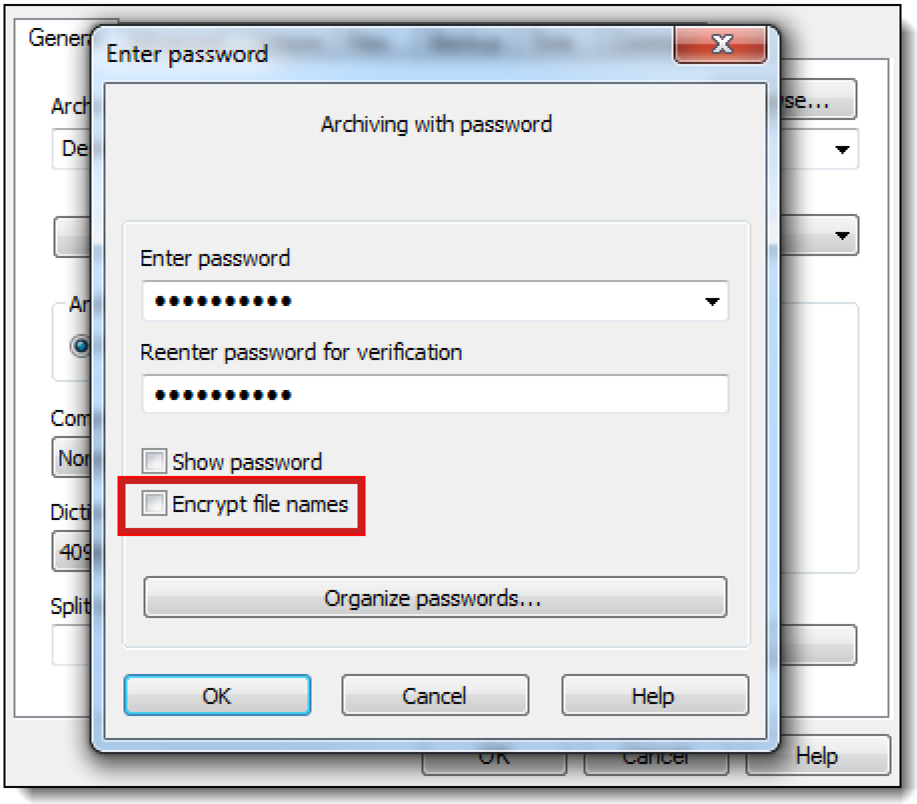
Accordingly, Relativity performs both native text extraction and OCR on the following file formats:

- All vector formats—.svg, CAD files, Metafiles [.wmf, .emf], Postscript, Encapsulated postscript
- .pdf, Visio, Publisher, MS Project, Hancom and JungUm files

All image formats, such as .tiff, .jpeg, .gif, .bmp, and .png, do not have native text, so only OCR is performed. If the file has electronic text and images, native text extraction and OCR is performed.

7.3 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.



Archiving with password

Enter password: ••••••••••

Reenter password for verification: ••••••••••

☐ Show password

☐ Encrypt file names

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	√		

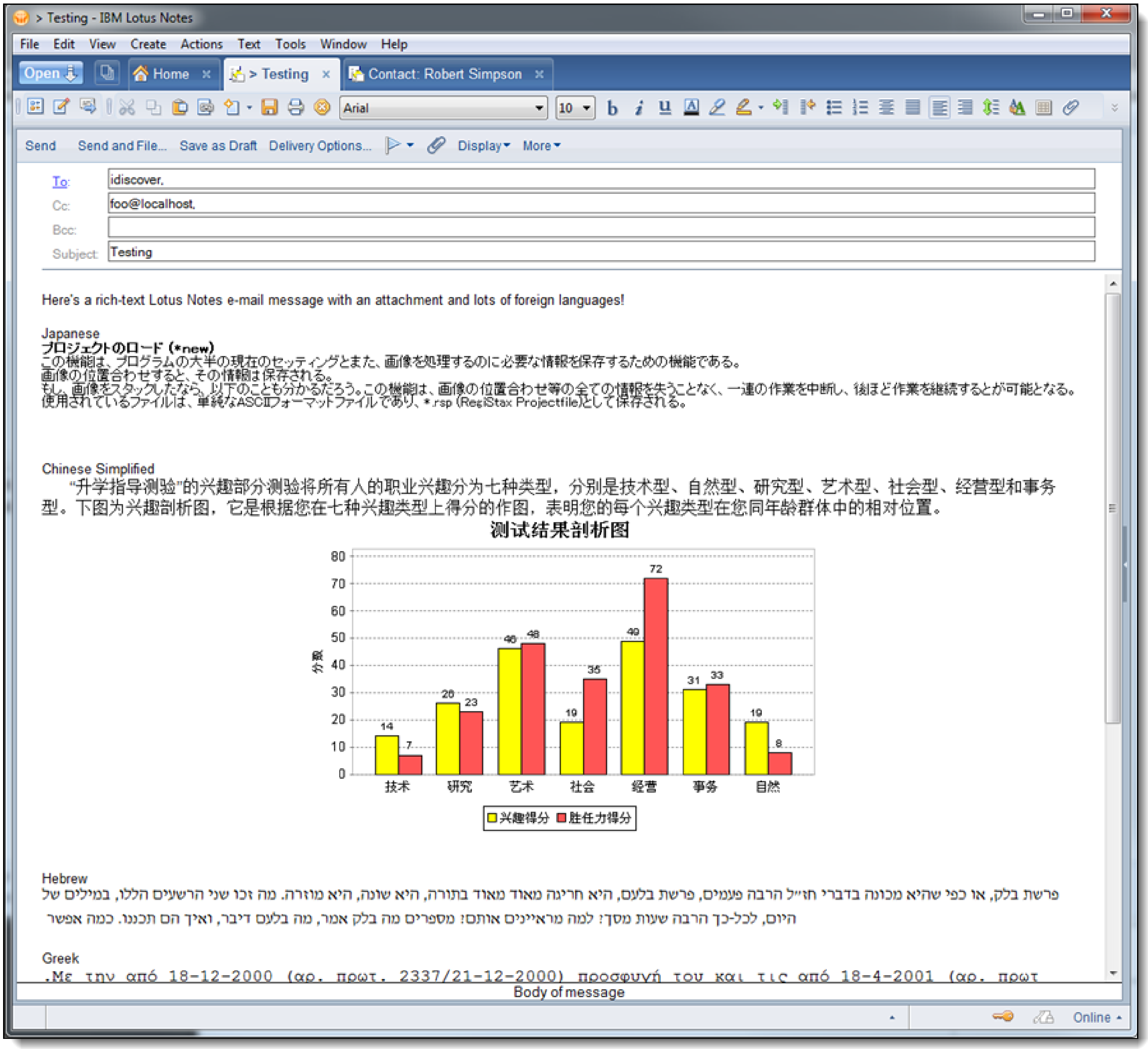
7.3.1 Lotus Notes

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 .dxl 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:


> Testing - IBM Lotus Notes
File
Edit
View
Create
Actions
Text
Tools
Window
Help
Open
Home
> Testing
Contact: Robert Simpson
Arial
10
Send
Send and File...
Save as Draft
Delivery Options...
Display
More
Body of message
Online

To: idiscover,
Cc: foo@localhost,
Bcc:
Subject: Testing

Here's a rich-text Lotus Notes e-mail message with an attachment and lots of foreign languages!

Japanese
プロジェクトのロード (*new)
この機能は、プログラムの大半の現在のセッティングとまた、画像を処理するのに必要な情報を保存するための機能である。
画像の位置合わせすると、その情報は保存される。
もし、画像をスタックしたなら、以下のことも分かるだろう。この機能は、画像の位置合わせ等の全ての情報を失うことなく、一連の作業を中断し、後ほど作業を継続するとが可能となる。
使用されているファイルは、単純なASCIIフォーマットファイルであり、*.rsp (RegiStax Projectfile)として保存される。

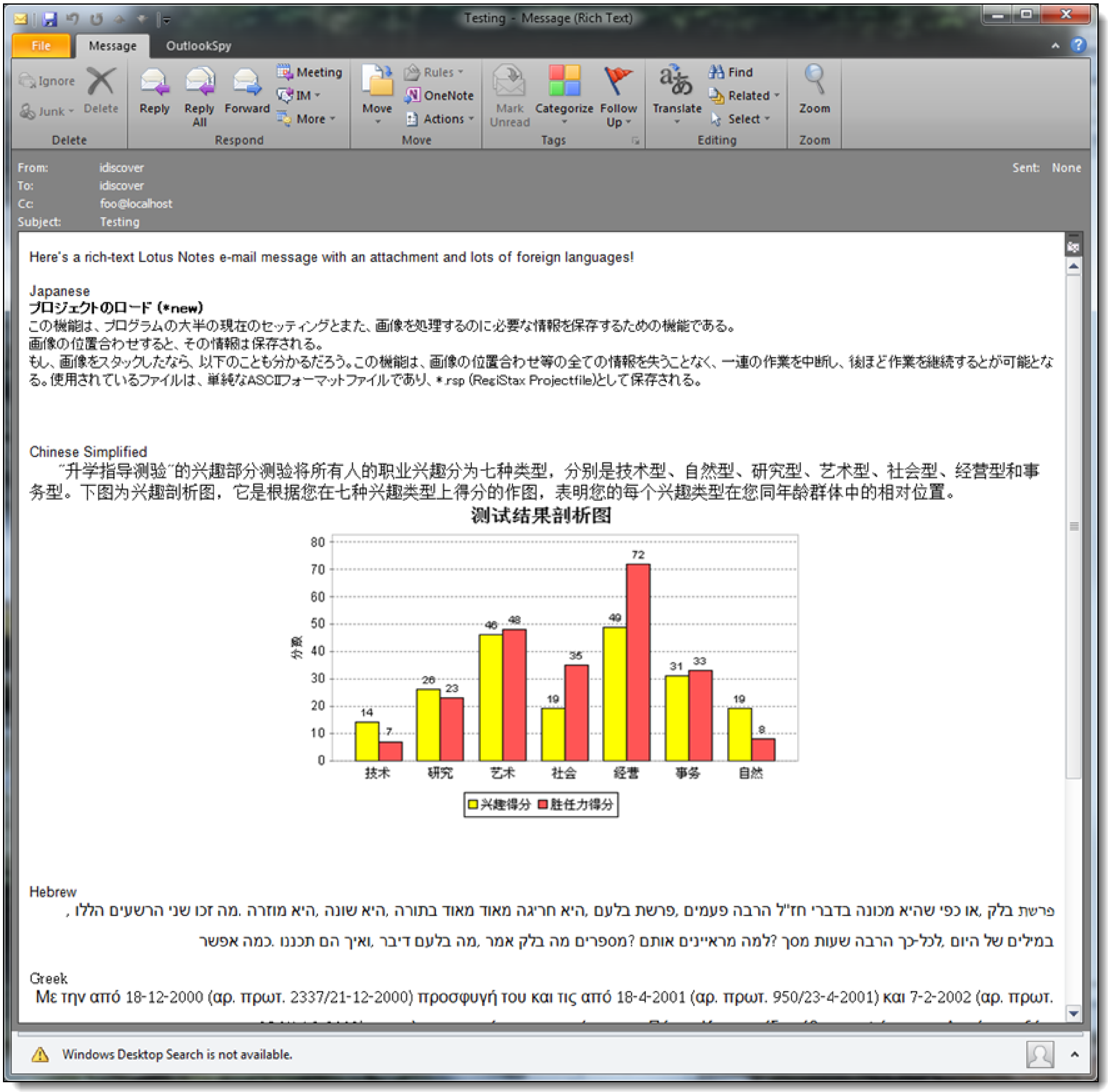
Chinese Simplified
“升学指导测验”的兴趣部分测验将所有人的职业兴趣分为七种类型，分别是技术型、自然型、研究型、艺术型、社会型、经营型和事务型。下图为兴趣剖析图，它是根据您在七种兴趣类型上得分的作图，表明您的每个兴趣类型在您同年龄群体中的相对位置。

测试结果剖析图
80
70
60
50
40
30
20
10
0
分数
14
7
26
23
46
48
19
35
49
72
31
33
19
8
技术
研究
艺术
社会
经营
事务
自然
□兴趣得分 ■胜任力得分

Hebrew
פרשת בלק, או כפי שהיא מכונה בדברי חז״ל הרבה פעמים, פרשת בלעם, היא חריגה מאוד מאוד בתורה, היא שונה, היא מוזרה. מה זכו שני הרשעים הללו, במילים של
היום, לכל-כך הרבה שעות מסך! למה מראיינים אותם! מספרים מה בלק אמר, מה בלעם דיבר, ואיך הם תכננו. כמה אפשר

Greek
.Με την από 18-12-2000 (αρ. πρωτ. 2337/21-12-2000) προσφυγή του και τις από 18-4-2001 (αρ. πρωτ

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


Testing - Message (Rich Text)
File
Message
OutlookSpy
Ignore
Junk
Delete
Delete
Reply
Reply All
Forward
Meeting
IM
More
Respond
Move
Rules
OneNote
Actions
Move
Mark Unread
Categorize
Follow Up
Tags
Translate
Find
Related
Select
Editing
Zoom
Zoom
测试结果剖析图
80
70
60
50
40
30
20
10
0
分数
14
7
26
23
46
48
19
35
49
72
31
33
19
8
技术
研究
艺术
社会
经营
事务
自然
🟨兴趣得分 🟥胜任力得分
Windows Desktop Search is not available.

From: idiscover
To: idiscover
Cc: foo@localhost
Subject: Testing
Sent: None

Here's a rich-text Lotus Notes e-mail message with an attachment and lots of foreign languages!

Japanese
プロジェクトのロード (*new)
この機能は、プログラムの大半の現在のセッティングとまた、画像を処理するのに必要な情報を保存するための機能である。
画像の位置合わせすると、その情報は保存される。
もし、画像をスタックしたなら、以下のことも分かるだろう。この機能は、画像の位置合わせ等の全ての情報を失うことなく、一連の作業を中断し、後ほど作業を継続するとが可能となる。使用されているファイルは、単純なASCIIフォーマットファイルであり、*.rsp (RegiStax Projectfile)として保存される。

Chinese Simplified
"升学指导测验"的兴趣部分测验将所有人的职业兴趣分为七种类型，分别是技术型、自然型、研究型、艺术型、社会型、经营型和事务型。下图为兴趣剖析图，它是根据您在七种兴趣类型上得分的作图，表明您的每个兴趣类型在您同年龄群体中的相对位置。

Hebrew
פרשת בלק, או כפי שהיא מכונה בדברי חז"ל הרבה פעמים, פרשת בלעם, היא חריגה מאוד מאוד בתורה, היא שונה, היא מוזרה. מה זכו שני הרשעים הללו ,
במילים של היום, לכל-כך הרבה שעות מסך? למה מראיינים אותם? מספרים מה בלק אמר, מה בלעם דיבר, ואיך הם תכננו. כמה אפשר

Greek
Με την από 18-12-2000 (αρ. πρωτ. 2337/21-12-2000) προσφυγή του και τις από 18-4-2001 (αρ. πρωτ. 950/23-4-2001) και 7-2-2002 (αρ. πρωτ.

7.3.1.1 Lotus Notes supported versions

Lotus Notes are supported through Version 10.

7.4 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.

7.5 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]
EmailBCCSmtp	BCC (SMTP Address)
EmailCC	[other metadata]
EmailCCName	[other metadata]

Processing engine property name	Relativity property name
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]
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]

Processing engine property name	Relativity property name
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]
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]

8 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, you can have Relativity 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 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.

Note: You can locate the Password Bank tab under both the Imaging and the Processing applications, if both are installed.

Using the password bank

Imagine you are a project manager, and you have been experiencing a high volume of files not making it into Relativity when you run a processing set because these files were unexpectedly password protected. As a result, the processing errors tab in your workspace is overrun and the data set you are able to inventory, discover, and publish is smaller than anticipated.

To deal with this, set up a password bank that includes the passwords for as many of the files you intend to process as you can locate. Then Relativity can decrypt these files and you can be sure that you are bringing the most relevant material into your workspace.

8.1 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 and/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 in the Password Bank, otherwise you will see an error.

The following scenario depicts the basic procedure by which you would address errors 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 errors 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 errored files.
5. Run retry on the files that previously resulted in password-protection errors.
6. From the Files tab, use the Republish mass action to retry to job.

Note: The PDF mass action does not work with the Password Bank. Specifically, the PDF mass action cannot connect with the password bank to grab passwords for encrypted documents. That connection is only available for native imaging and processing.

8.2 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. Click View Document Errors in the Imaging Set console after you run an imaging set.
2. Outside of Relativity, locate the passwords designated to unlock those files.
3. Return to Relativity, go to the Password Bank, and create entries for every password that corresponds with the errored files.

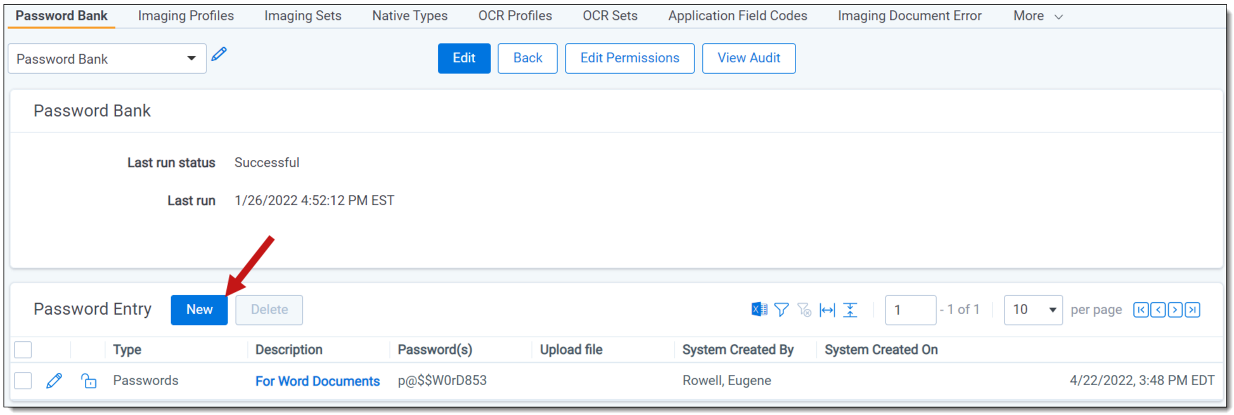
4. Click Retry Errors in the Imaging Set console to retry imaging the files that previously resulted in password-protection errors.

8.3 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.


Password Bank
Imaging Profiles
Imaging Sets
Native Types
OCR Profiles
OCR Sets
Application Field Codes
Imaging Document Error
More
Password Bank
Edit
Back
Edit Permissions
View Audit
Password Bank
Password Entry
New
Delete
1
- 1 of 1
10
per page

Last run status Successful

Last run 1/26/2022 4:52:12 PM EST

☐		Type	Description	Password(s)	Upload file	System Created By	System Created On
☐		Passwords	For Word Documents	p@$$W0rD853		Rowell, Eugene	4/22/2022, 3:48 PM EDT

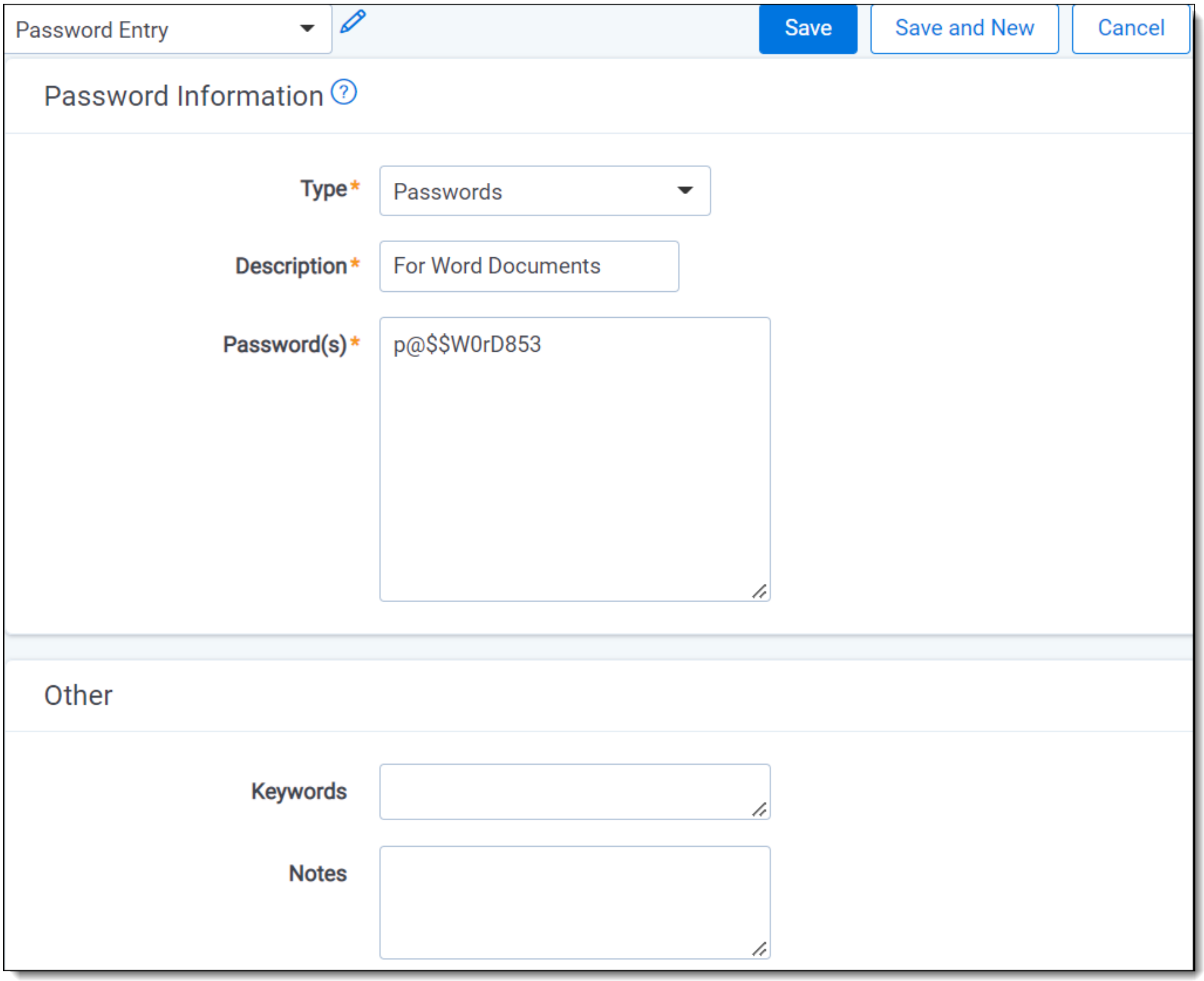
3. Complete the fields on the Password Entry Layout. See Fields for more information.
4. Click **Save**. The entry appears among the others under the Password Entry object.

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

Note: When you create a password entry and submit any job that syncs with the processing engine (imaging or processing), an entry is created in the engine for that password and that workspace. Even if you delete that password entry from the password bank, any future jobs continue to try that password.

8.3.1 Fields

The Password Bank layout contains the following fields:



Type* Passwords

Description* For Word Documents

Password(s)* p@$$W0rD853

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 error 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.

8.3.2 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.

8.4 Validations, errors, and exceptions

Note the following:

- Including a password that does not belong to a document in your data set does not throw an error 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 encounter 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 errors by supplying the correct password to the bank and then retrying those errors in their respective processing or imaging jobs.

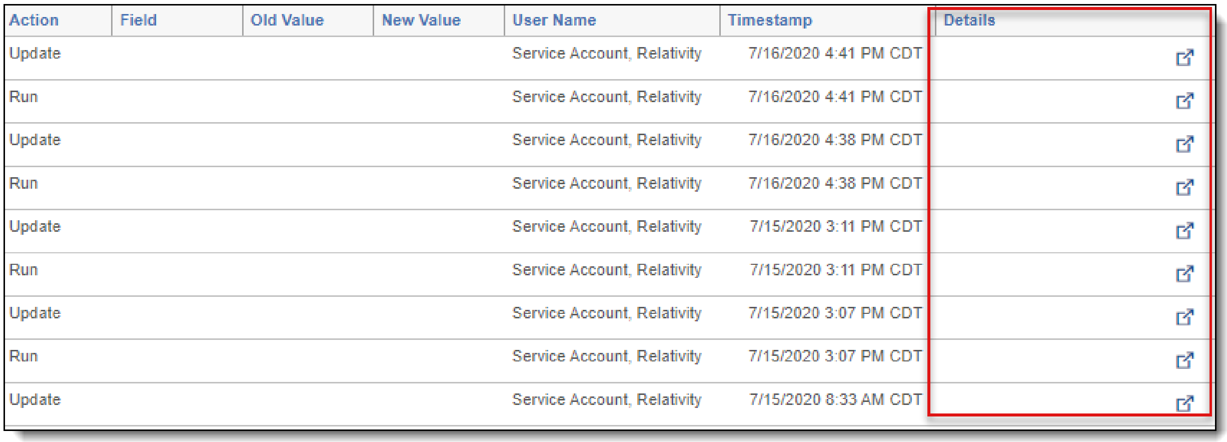
Note: When you supply a valid password to the password bank, the processing engine extracts metadata and extracted text from the document that the password unlocks. However, when you publish that document, its password security is not removed, in which case it technically remains in an encrypted state 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.

8.5 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, ands 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.
3. Click **Details** on the Password Bank history layout.



Action	Field	Old Value	New Value	User Name	Timestamp	Details
Update				Service Account, Relativity	7/16/2020 4:41 PM CDT	↗
Run				Service Account, Relativity	7/16/2020 4:41 PM CDT	↗
Update				Service Account, Relativity	7/16/2020 4:38 PM CDT	↗
Run				Service Account, Relativity	7/16/2020 4:38 PM CDT	↗
Update				Service Account, Relativity	7/15/2020 3:11 PM CDT	↗
Run				Service Account, Relativity	7/15/2020 3:11 PM CDT	↗
Update				Service Account, Relativity	7/15/2020 3:07 PM CDT	↗
Run				Service Account, Relativity	7/15/2020 3:07 PM CDT	↗
Update				Service Account, Relativity	7/15/2020 8:33 AM CDT	↗

4. 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.

9 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.

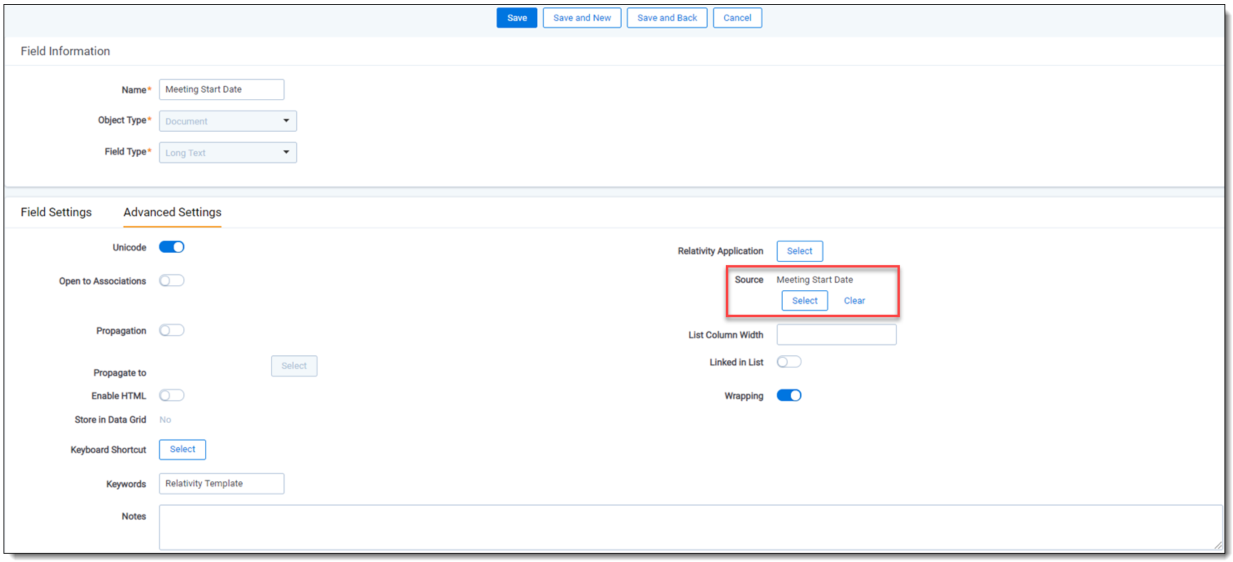
Sometimes, you will see a field stored in both Relativity system fields and optional fields, such as Date Created. In this case, Relativity determines which field to keep and which to ignore.

9.1 Mapping fields

To map processing fields, perform the following steps:

1. Open the **Fields** tab.
2. Click **New Field** or **Edit** on an existing field.
3. Provide a name in the **Name** field. We recommend that you give the field an identical name to the one you are mapping to.
4. 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.
5. In the Field Type field, select the type of field to set what type of data can be entered into the field.
6. 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.


Save
Save and New
Save and Back
Cancel
Field Information
Field Settings
Advanced Settings
Unicode
Open to Associations
Propagation
Propagate to
Select
Enable HTML
Keyboard Shortcut
Select
Notes
Relativity Application
Select
Select
Clear
List Column Width
Linked in List
Wrapping

Name* Meeting Start Date
Object Type* Document
Field Type* Long Text
Store in Data Grid No
Keywords Relativity Template
Source Meeting Start Date

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**.

Note: If the Processing application is not installed, you can still map fields as long as you have added the worker manager server to the resource pool.

9.1.1 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.

9.1.2 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 long text field to a 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."

9.2 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:

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.
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**.
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.

Processing Field Name	Source view	Field Type	Description
Originating Processing Data Source		Single Object	A single object field that refers to the processing data source.
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 on page 79 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 viewer for processed emails. This field will be populated with the UTC offset value of the time zone chosen in the processing profile. For example, documents processed to Central Standard Time (CST), would be populated with a value of "-6" because CST is UTC-6. 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.
Virtual Path	Other Fields	Long Text	Folder structure and path to file from the original location identified during processing. See Virtual path on page 78 for more information.

9.3 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:

- You can map one processing field to multiple Document object fields.
- You can only map a processing field to a Unicode-enabled field.

- 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 **All 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 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.	Lay, Kenneth; Doe, John
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 Folders\Notes inbox; Doe, John|\Doe, John\John_Doe_000_1_2_1.pst \Doe-J\John_Doe_Dec2000\Notes Folders\Discussion threads

Processing field/ source name	Source view	Field type	Description	Example value
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;KL000000↪0031.0002
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. This field has a maximum length of 50 alphanumeric characters.	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
CC		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	BA8F37866F59F269AE1D6 ↪2D962B887B65DE7474D1 ↪3679D9388B75C95EE7780FE

Processing field/ source name	Source view	Field type	Description	Example value
			Information Processing Standards cryptography) enabled for the worker manager server.	
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.	1989C1E539B5AE981820648623954872BEE3E483; 58D9E4B4A3068DA6E9BCDD969523288CF38F9FB3
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.	7848EEFC40C40F868929600BF033617642E0D37C2F5FA444C7EF83350AE19883;628B6233DD6E0C89F32D6EFF2885F26917F144B19F3678265BEBAC7E9ACAAF5B
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 document has additional files embedded in it.	Yes
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 alpha-numeric 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	KL0000000001.0002

Processing field/ source name	Source view	Field type	Description	Example value
			populated for documents with no family members. This field has a maximum length of 50 alphanumeric characters.	
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.	12/24/2015
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	"12/24/2015 11:59 PM"

Processing field/ source name	Source view	Field type	Description	Example value
			value, this field will display the document metadata date created value.	
Created Time	Standard Fields	Long Text	The time at which a file was created.	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. You are not able to retroactively populate this field with custodian information.	2
DeDuped Custodians	Standard Fields	Multiple Object	The custodians associated with the de-duped 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. You are not able to retroactively populate this field with custodian information. The All Custodians field is mapped to a document and is updated only.	Lay, Kenneth;Doe, John
DeDuped Paths	Standard Fields	Long Text	The virtual paths of	Lay, Kenneth|\Lay, Kenneth\kenneth_lay_

Processing field/ source name	Source view	Field type	Description	Example value
			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. You are not able to retroactively populate this field with path information.	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
Discover Errors on Child Documents	Standard Fields	Multiple Object	Identifier of the file that contains the parent document on which the error occurred.	
Document Subject		Long Text	Subject of the document extracted from the properties of the native file.	RE: Our trip to Washington
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 on page 79 for more information.	Inbox\New Business
Email Format	Standard Fields	Single Choice	The indicator of whether an email is HTML, Rich Text, or Plain Text.	HTML

Processing field/ source name	Source view	Field type	Description	Example value
Email Has Attachments	Standard Fields	Yes/No	The yes/no indicator of whether an email has children, attachments.	Yes
Email In Reply To ID	Standard Fields	Long Text	The internal metadata value within an email for the reply-to ID.	<F9B1A278195DF640A4CC6E↪C973DFF0C85FBBEDEB↪@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		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 field on page 76. 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
Error Category	Standard Fields	Fixed-Length Text (50)	The category assigned by the system to a processing error.	Password Protected Container

Processing field/ source name	Source view	Field type	Description	Example value
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	Fixed-Length Text (50)	The phase of processing in which the error occurred—discovery, text extraction, publish, file deletion.	Discovery
Error Status	Standard Fields	Fixed-Length Text (200)	The status of the error—undetermined, ready to retry, retried, 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 worksheets.	No
Excel Pivot Tables	Standard Fields	Yes/No	The yes/no indicator of whether an Excel file contains pivot tables.	Yes
Extracted Text Size in KB	Other Fields	Decimal	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 alphanumeric characters.	KL0000000002
File Extension	Standard	Fixed-	The extension of the file,	MSG

Processing field/ source name	Source view	Field type	Description	Example value
	Fields	Length Text (25)	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.	
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	Decimal	Generally a decimal number indicating the size in bytes of a file.	15896
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	Capellas Michael D. [Michael.Capellas@COMPAQ.com]

Processing field/ source name	Source view	Field type	Description	Example value
			255 alpha-numeric characters.	
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 definitively state that a document contained no hidden data just because the system could not detect it.	Yes
Has OCR Text	Standard Fields	Yes/No	The yes/no indicator of whether the extracted text field contains OCR text.	Yes
HiddenAttachment		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"

Processing field/ source name	Source view	Field type	Description	Example value
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
Keywords	Standard Fields	Long Text	The internal value entered for keywords associated with a Microsoft Office document.	Enron, Security Agreement
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 changes to a file were last saved.	12/24/2015
Last Modified Date/Time	Standard Fields	Date	The date and time at which changes to a file were last saved.	"12/24/2015 11:59 PM"
Last Modified Time	Standard Fields	Long Text	The time at which changes to a file were	11:59 PM

Processing field/ source name	Source view	Field type	Description	Example value
			last saved.	
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 Text (255)	The internal value indicating the last user to save a document. This field has a maximum length of 255 alphanumeric characters.	ymendez
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 deduplication and authentication generated using the MD5 hash algorithm. Relativity cannot calculate this value if you have FIPS (Federal Information Processing Standards cryptography) enabled for the worker manager server. This	21A74B494A1BFC2FE217 ↪CC274980E915

Processing field/ source name	Source view	Field type	Description	Example value
			field has a maximum length of 40 alpha-numeric characters.	
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-numeric characters.	72
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

Processing field/ source name	Source view	Field type	Description	Example value
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 Text (255)	The message number created by an email application and extracted from the email's metadata. For more information, see [Message ID considerations on page 80](). This field has a maximum length of 255 alpha-numeric characters.	<PLSRGLMRNQWEDFYPJL5 ↪ZJFF41USDEIQHB@zlsvr22>
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.MSG
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	Jane Doe

Processing field/ source name	Source view	Field type	Description	Example value
			email. This field has a maximum length of 50 alpha-numeric characters.	
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 field has a maximum length of 25 alpha-numeric characters.	DOC
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	The indicator of which flag, if any, an Outlook item has assigned to it—NoFlag, FlagMarked, or FlagComplete.	NoFlag
Parent Document ID	Standard Fields	Fixed-Length Text	Document ID of the parent document. This field is only available on child items.	EN11, 12345

Processing field/ source name	Source view	Field type	Description	Example value
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 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**.	Encrypted
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"
Processing Errors	Standard Fields	Multiple Object	Associated errors that occurred on the document during processing. This field is a link to the associated Processing Errors record.	
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	"12/24/2015 11:59 PM"

Processing field/ source name	Source view	Field type	Description	Example value
			was received.	
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
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 recipients of an email message.	Jane Doe
Record Type		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	Yes

Processing field/ source name	Source view	Field type	Description	Example value
			exceeding 2GB. To locate this field, navigate to the All Fields view.	
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.	Rolll 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	Mea Ftad-atields	Long Text	This is used to identify source of the data, which is intended to be ambiguous. For example, it could be the application of the data contained in the RSMF file.	Slack
*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

Processing field/ source name	Source view	Field type	Description	Example value
*RSMF Event 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.	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
SHA1 Hash	Standard Fields	Fixed-Length Text (50)	Identifying value of an electronic record that can be used for deduplication and authentication generated using the SHA1 hash algorithm. This field has a maximum length of 50 alpha-numeric characters.	D4C4EB398980E82B4B3064 ↪CC2005F04D04BBAAE6
SHA256 Hash	Standard Fields	Fixed-Length Text (70)	Identifying value of an electronic record that can be used for deduplication and authentication generated using the SHA256 hash algorithm. This field has a maximum length of 70 alpha-numeric characters.	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	Long	The date on which an	12/24/2015

Processing field/ source name	Source view	Field type	Description	Example value
	Fields	Text	email message was sent.	
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 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, 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:	"12/24/2015 11:59 PM"

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, and Primary 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 on page 79 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

Processing field/ source name	Source view	Field type	Description	Example value
Suspect File Extension	Standard Fields	Yes/No	The yes/no indicator if whether the extension of a file does not correspond to the actual type of the file. For example. XLS for a Word document.	Yes
Text Extraction Method	Standard Fields	Single Choice	The method used to run text extraction.	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 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 look-	Yes

Processing field/ source name	Source view	Field type	Description	Example value
			ing 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 TrackedChangesCombined Invariant field. This will be Yes if either of the following are true:	

Processing field/ source name	Source view	Field type	Description	Example value
			■ 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

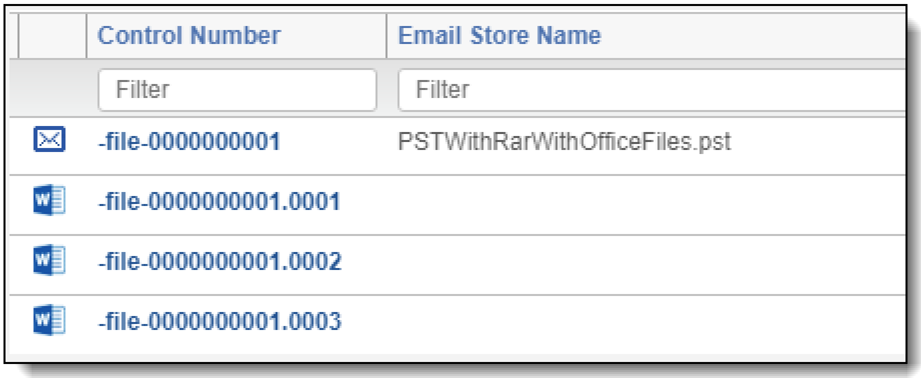
Processing field/ source name	Source view	Field type	Description	Example value
			ging/OCRing that document type. For example, it is not set for a corrupt file for which we cannot extract text, such as a corrupt Word document that logs an error during data extraction. ▪ Unprocessable documents do not have errors associated with them because they never reach a point at which they can register a processing error.	

Note: You can track which passwords successfully decrypted published documents by mapping the Password field found in the All Fields view. Specifically, you can find this Password field by clicking Source on the field layout, selecting the All Fields view, and locating the source field name of Password with a field type of Long Text.

9.4 Email Store Name field

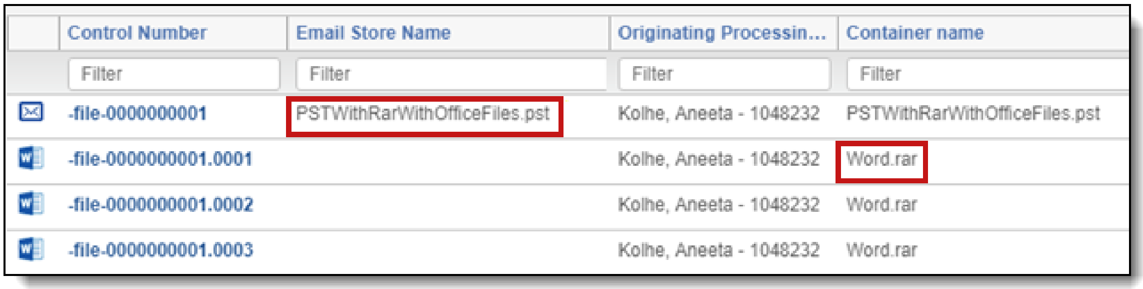
To better understanding how the Email Store Name field works, consider the following examples:

- 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.



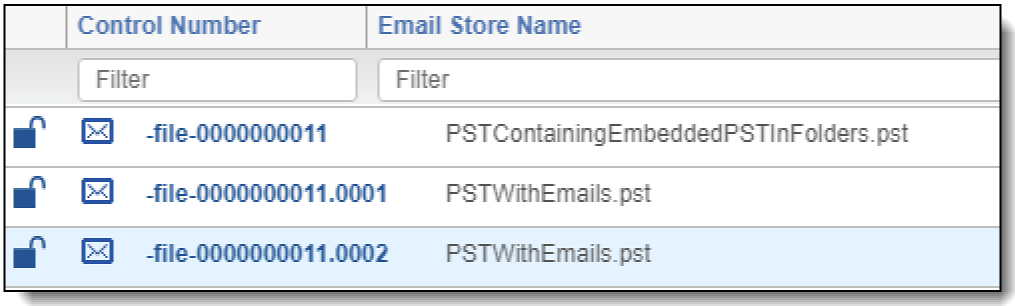
	Control Number	Email Store Name
	Filter	Filter
✉	-file-0000000001	PSTWithRarWithOfficeFiles.pst
W	-file-0000000001.0001	
W	-file-0000000001.0002	
W	-file-0000000001.0003	

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



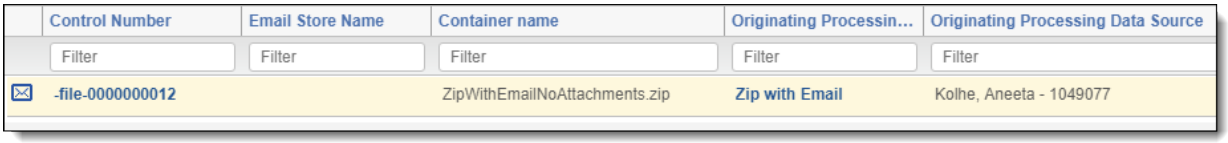
	Control Number	Email Store Name	Originating Processin...	Container name
	Filter	Filter	Filter	Filter
✉	-file-0000000001	PSTWithRarWithOfficeFiles.pst	Kolhe, Aneeta - 1048232	PSTWithRarWithOfficeFiles.pst
W	-file-0000000001.0001		Kolhe, Aneeta - 1048232	Word.rar
W	-file-0000000001.0002		Kolhe, Aneeta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

- For an email taken from a zip folder, the Email Store Name field is blank.

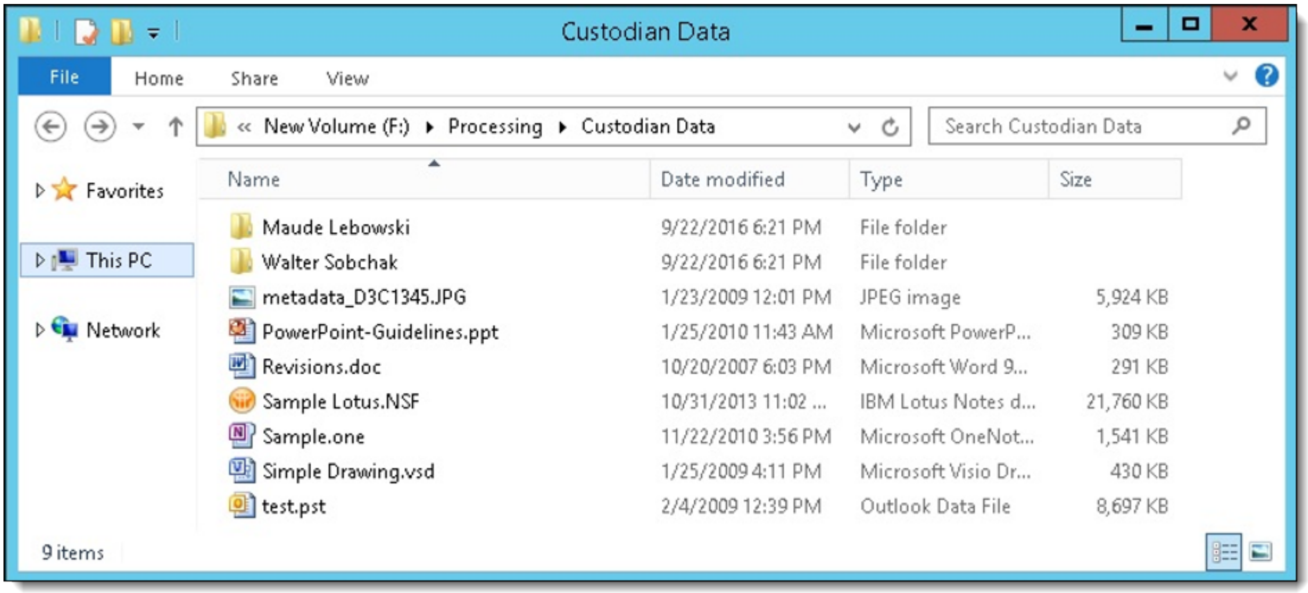


	Control Number	Email Store Name	Container name	Originating Processin...	Originating Processing Data Source
	Filter	Filter	Filter	Filter	Filter
✉	-file-0000000012		ZipWithEmailNoAttachments.zip	Zip with Email	Kolhe, Aneeta - 1049077

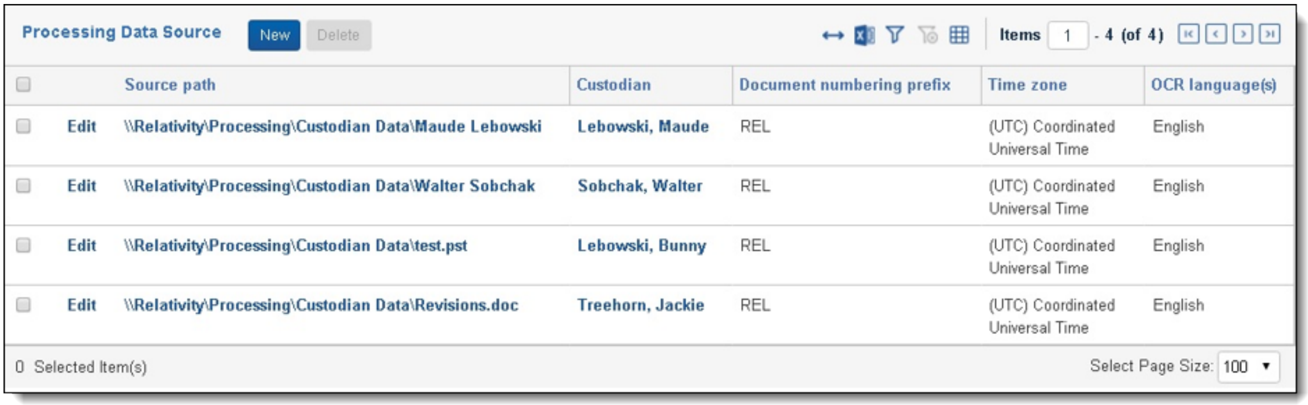
9.5 Virtual path

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.


Custodian Data
File
Home
Share
View
« New Volume (F:) ▸ Processing ▸ Custodian Data
Search Custodian Data
Favorites
This PC
Network
9 items

Name	Date modified	Type	Size
Maude Lebowski	9/22/2016 6:21 PM	File folder	
Walter Sobchak	9/22/2016 6:21 PM	File folder	
metadata_D3C1345.JPG	1/23/2009 12:01 PM	JPEG image	5,924 KB
PowerPoint-Guidelines.ppt	1/25/2010 11:43 AM	Microsoft PowerP...	309 KB
Revisions.doc	10/20/2007 6:03 PM	Microsoft Word 9...	291 KB
Sample Lotus.NSF	10/31/2013 11:02 ...	IBM Lotus Notes d...	21,760 KB
Sample.one	11/22/2010 3:56 PM	Microsoft OneNot...	1,541 KB
Simple Drawing.vsd	1/25/2009 4:11 PM	Microsoft Visio Dr...	430 KB
test.pst	2/4/2009 12:39 PM	Outlook Data File	8,697 KB


Processing Data Source
New
Delete
Items 1 - 4 (of 4)
0 Selected Item(s)
Select Page Size: 100

☐		Source path	Custodian	Document numbering prefix	Time zone	OCR language(s)
☐	Edit	\\Relativity\Processing\Custodian Data\Maude Lebowski	Lebowski, Maude	REL	(UTC) Coordinated Universal Time	English
☐	Edit	\\Relativity\Processing\Custodian Data\Walter Sobchak	Sobchak, Walter	REL	(UTC) Coordinated Universal Time	English
☐	Edit	\\Relativity\Processing\Custodian Data\test.pst	Lebowski, Bunny	REL	(UTC) Coordinated Universal Time	English
☐	Edit	\\Relativity\Processing\Custodian Data\Revisions.doc	Treehorn, Jackie	REL	(UTC) Coordinated Universal Time	English

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.

9.6 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.

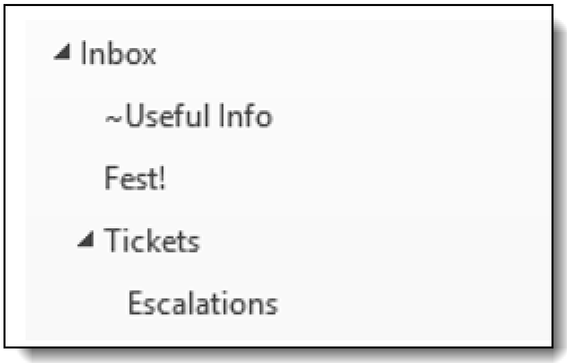
The following are examples of virtual paths and corresponding processing folder paths.

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

9.7 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."



9.8 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.

9.9 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.

9.10 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.

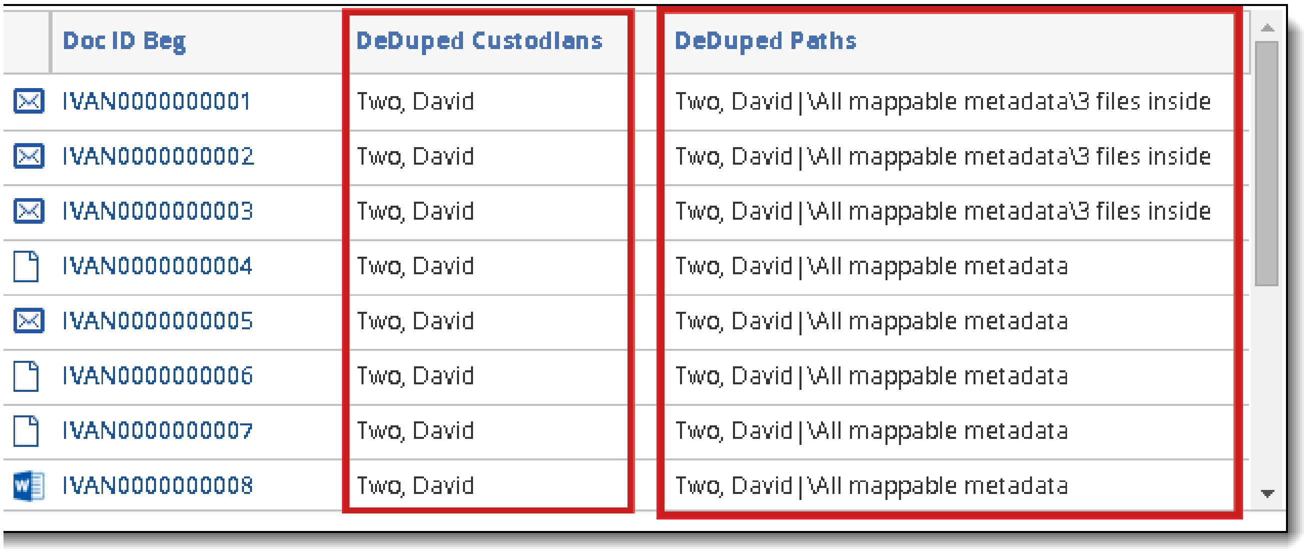
9.11 De-duped custodian and path considerations

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	DeDuped Custodians	DeDuped Paths
✉	IVAN0000000001	Two, David	Two, David | \All mappable metadata\3 files inside
✉	IVAN0000000002	Two, David	Two, David | \All mappable metadata\3 files inside
✉	IVAN0000000003	Two, David	Two, David | \All mappable metadata\3 files inside
🗋	IVAN0000000004	Two, David	Two, David | \All mappable metadata
✉	IVAN0000000005	Two, David	Two, David | \All mappable metadata
🗋	IVAN0000000006	Two, David	Two, David | \All mappable metadata
🗋	IVAN0000000007	Two, David	Two, David | \All mappable metadata
W	IVAN0000000008	Two, David	Two, David | \All mappable metadata

Note: When Relativity populates the Deduped Custodians and Deduped 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.

10 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.

Relativity provides a Default profile upon installation of processing.

Using Processing profiles

You're a litigation support specialist, and your firm has requested you to bring a custodian's data into Relativity without bringing in any embedded Microsoft office objects or images. You have to create a new processing profile for this because none of the profiles in the workspace have specified to exclude embedded images or objects when extracting children from a data set.

To do this, you simply create a new profile with those specifications and select that profile when creating the processing set that you want to use to bring the data into Relativity.

10.1 Creating or editing a processing profile

To create or edit a processing profile:

1. Go 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.

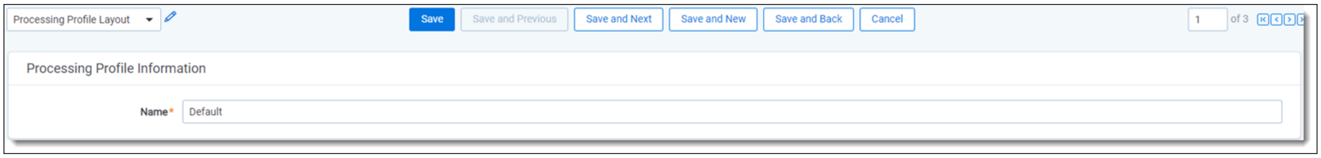
Note: You can't delete the Default processing profile. If you delete a profile that is associated with a processing set you've already started, the in-progress processing phase will continue with the original profile settings you applied when you submitted the job, but you won't be able to proceed to 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.

10.1.1 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.

10.1.1.1 Processing profile information

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



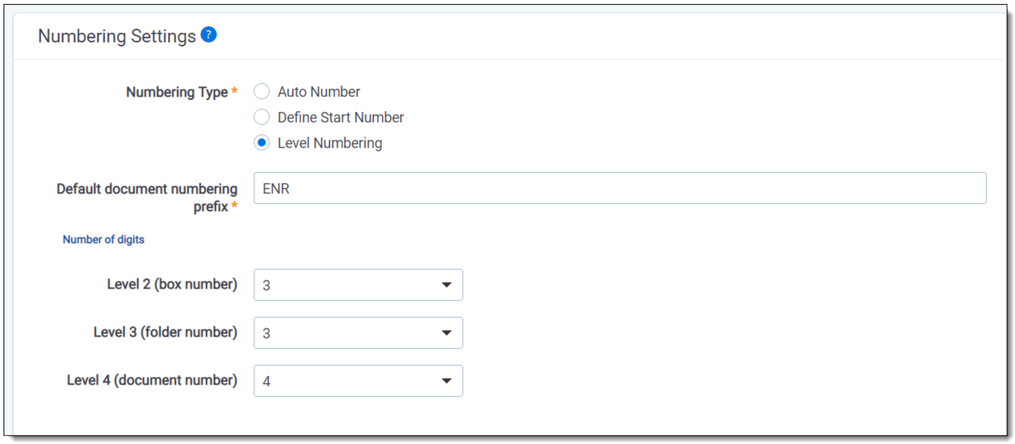
Processing Profile Information

Name* Default

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

10.1.1.2 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 * ENR

Number of digits

Level 2 (box number) 3

Level 3 (folder number) 3

Level 4 (document number) 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’.

- **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 Numbering**—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.
- **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 on page 1.
    - **Suffix Always**—arranges for child documents to be appended to their parent with a delimiter.
    - **Continuous Always**—arranges for child documents to receive a sequential control number after their parent.
    - **Continuous, Suffix on Retry**—arranges for child documents to receive a sequential control number after their parent except for child documents that weren't published to the workspace. When these unpublished child documents are retried and published, they will receive the parent's number with a suffix. If you resolve the error post-publish, the control number doesn’t change.

      Note: It's 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.

- **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.

Note: Level numbering cannot be used with Quick-Create Set(s).

Note: 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.

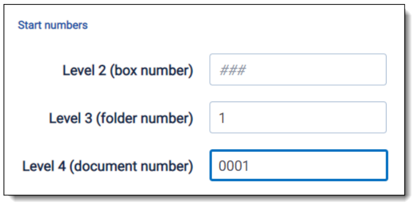
Level 2 (box number)—corresponds to the BBB level in the PPP.BBB.FFF.NNNN format. Default value is 3 digits.

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

10.1.1.3 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:

REL.001.0001.9999	Excel document
REL.001.0002.0001	Word document
REL.001.0002.0002	Word document

9,997 documents later

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:

REL.001.0001.9999 – excel document
REL.001.0002.0001 – word document
REL.001.0002.0002 – word document

9,995 documents later

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.001.0003.0002
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:

```
REL.001.0003.0001 – email with 10,000 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
[...]
REL.001.0003.9999 – attachment 9998
REL.001.0003.0001_0001 – attachment 9999
REL.001.0003.0001_0002 – attachment 10,000
```

10.1.1.4 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:

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

Republish:

```
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:

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

Republish:

```
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:

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

Republish:

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:

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

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

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:

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.

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.

10.1.1.5 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 souce 2: REL.001.001.0011- REL.001.001.0020

Data souce 2: REL.001.001.0021 - REL.001.001.0030

10.1.1.6 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:

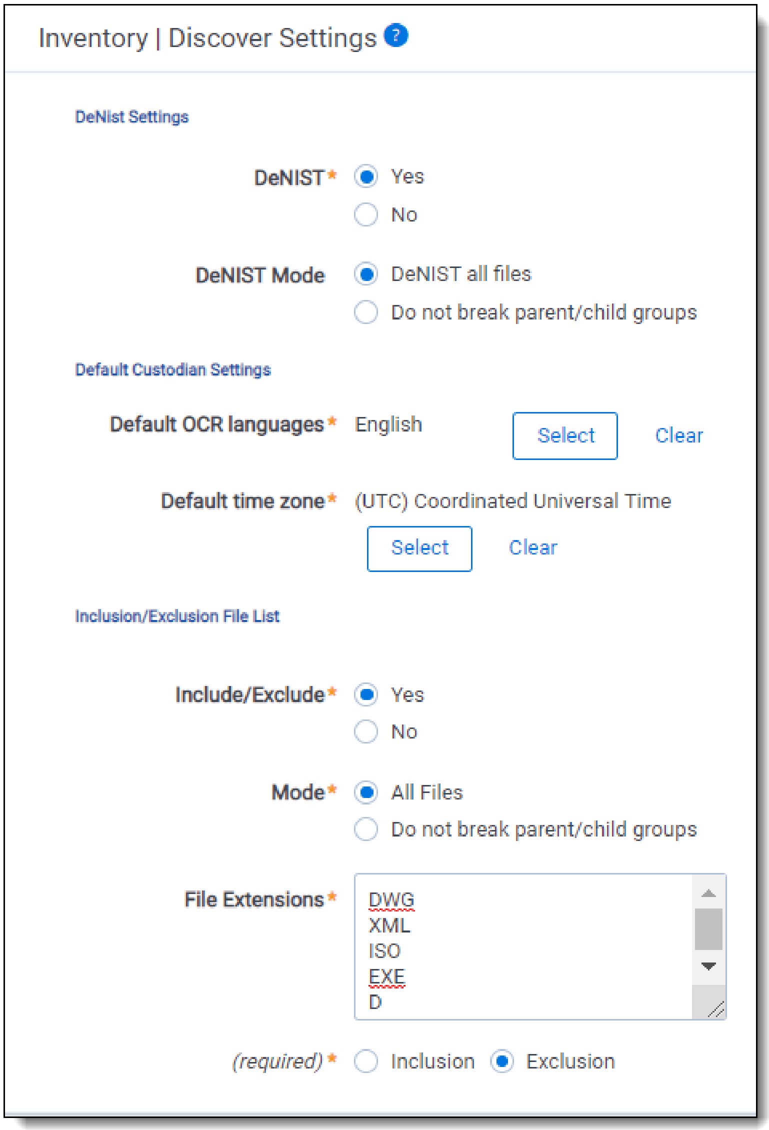
REL.001.001.9998
REL.001.001.9999 (Password-protected file)
REL.001.002.0001
REL.001.002.0002

Republish:

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	

10.1.2 Inventory / discovery settings

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



Inventory | Discover Settings

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*
DWG
XML
ISO
EXE
D

(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 because it is stored on the worker manager server. You should not edit the NIST list. Relativity makes new versions of the NIST list available shortly after the National Software Reference Library (NSRL) releases them quarterly. Login to the NIST Package Download webpage on the Relativity Community website to download the latest package and installer files.

- **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.

Note: The processing engine discovers all natives in UTC and then converts metadata dates and times into the value you enter for the Default Time Zone field. The engine needs the time zone at the time of text extraction to write the date/time into the extracted text and automatically applies the daylight saving time for each file based on its metadata during the publishing stage.

- **Include/Exclude**—enables the toggle for the inclusion/exclusion fields. The Inclusion/Exclusion File List allows you to upload 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.

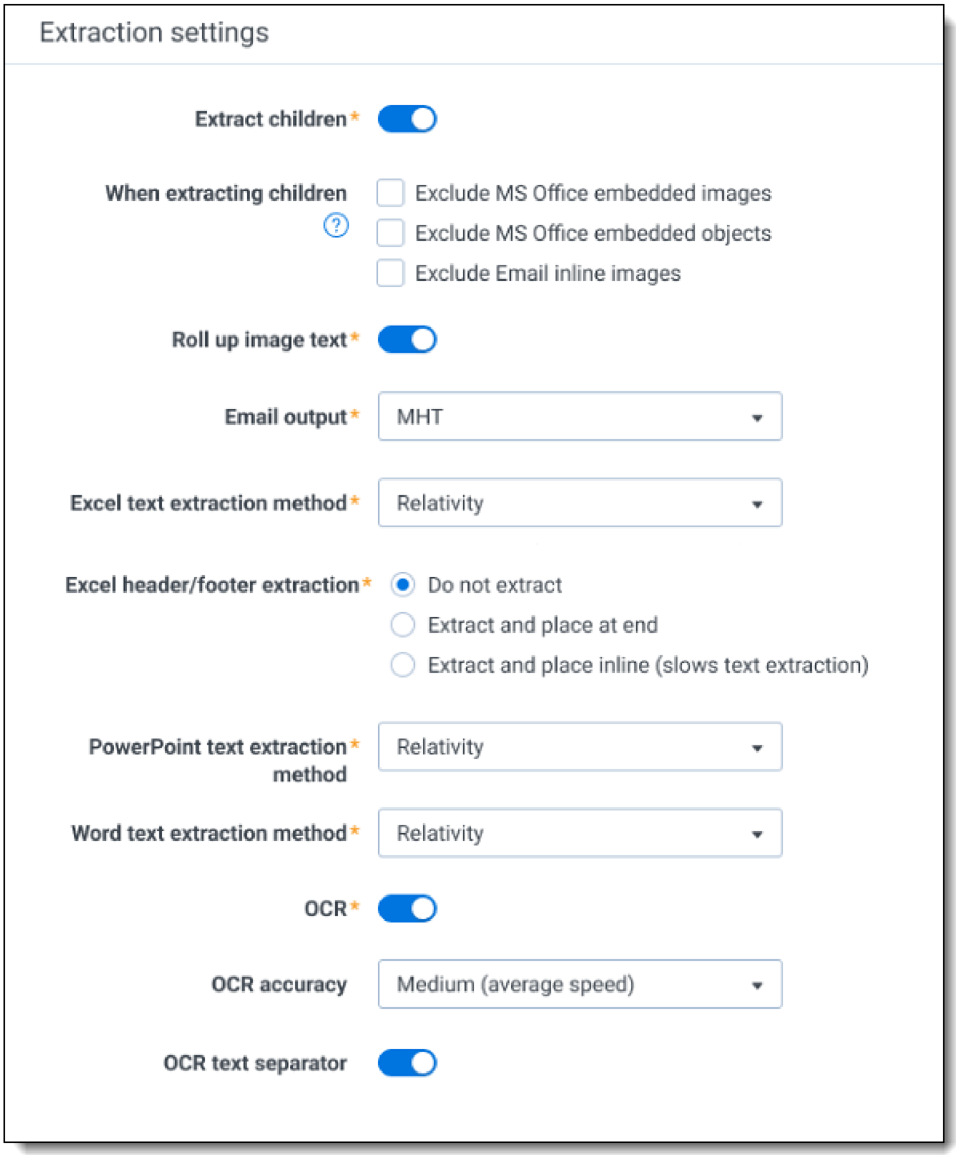
  Note: File extensions must be separated with a hard return in order to be filed as a new extension. Extensions are case insensitive and should be entered as just the name of the extension (i.e., EXE versus .EXE).

- **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.

10.1.3 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* ●

When extracting children
☐ Exclude MS Office embedded images
☐ Exclude MS Office embedded objects
☐ Exclude Email inline images

Roll up image text* ●

Email output* MHT

Excel text extraction method* Relativity

Excel header/footer extraction*
● Do not extract
○ Extract and place at end
○ Extract and place inline (slows text extraction)

PowerPoint text extraction method* Relativity

Word text extraction method* Relativity

OCR* ●

OCR accuracy Medium (average speed)

OCR text separator ●

- **Extract children**—Choose whether or not to extract child items during discovery, including 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 removes the options for embedded images and objects, and rolled up image text.

    Note: You do not need to set the Extract children field to Yes to have the files within PST and other container files extracted and processed. This is because Relativity breaks down container files by default without the need to specify to extract children.

- **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 embedded images**—selecting this option excludes images of various file types found inside Microsoft Office files (such as .jpg, .bmp, or .png in a Word file) from discovery. Embedded images are not published separately in Relativity.
  - **MS Office embedded objects**—selecting this option excludes objects of various file types found inside Microsoft Office files (such as an Excel spreadsheet inside a Word file, or a PDF file inside in a Word file) from discovery. Embedded objects are not published separately in Relativity. MS Office embedded objects do not have text extracted and are not searchable.
  - **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. Inline images are not published separately in Relativity.
- **Roll up image text**—if you include **MS Office embedded images** or **Email inline images**, you can append the image's text to the end of the parent document. The image itself is not published. The benefit of this feature is cost savings due to reduced file count and size in the hosted workspace.

  Note: Relativity rolls up all MS Office embedded images and Email inline images to their parent documents, whether or not the image has text. The Rolled up image text option allows you to select whether or not to append the image's extracted text (if there is text) to the parent document.

  - If the **Extract children** field is set to no, the **Roll up image text** field is not displayed.
  - You must include **MS Office embedded images** or **Email inline images** (one or both) to use the roll up text feature.
  - If the **Roll up image text** field is set to yes, Relativity appends the text from **MS Office embedded images** (if included) and **Email inline images** (if included) to the end of their respective parent documents.
  - Rolled up text is visible at the end of the parent's extracted text, with a text separator.

  Note: See Microsoft Office child extraction support for information on what MS Office documents can have embedded images extracted.

  Note: See Email image extraction support for information on what type of emails can have inline images extracted.

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**

    Note: This option affects the following file types: Outlook files, Lotus Notes files, Bloomberg files

Note: 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.
  - **Relativity**—tells Relativity to use Relativity's built-in engine to extract text from Excel files. This setting is recommended for performance and accuracy.
  - **Native**—tells Relativity to use Excel to extract text from Excel files.
  - **Native (failover to dtSearch)—**tells Relativity to use Excel to extract text from Excel files with dtSearch as a backup text extraction method if extraction fails.
  - **dtSearch (failover to Native)**—tells Relativity to use 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.
- **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. This field isn't available if you selected dtSearch for the Excel Text Extraction Method field above because dtSearch automatically extracts header and footer information and places it at the end of the text; if you selected a value for this field and then select dtSearch above,

your selection here is nullified. The 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.
  - **Relativity**—tells Relativity to use Relativity's built-in engine to extract text from PowerPoint files. This setting is recommended for performance and accuracy.
  - **Native**—tells Relativity to use PowerPoint to extract text from PowerPoint files.
  - **Native (failover to dtSearch)**—tells Relativity to use PowerPoint to extract text from PowerPoint files with dtSearch as a backup text extraction method if extraction fails.
  - **dtSearch (failover to Native)**—tells Relativity to use 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.
  - **Relativity** -tells Relativity to use Relativity's built-in engine to extract text from Word files. This setting is recommended for performance and accuracy.
  - **Native**—tells Relativity to use Word to extract text from Word files.
  - **Native (failover to dtSearch)**—tells Relativity to use Word to extract text from Word files with dtSearch as a backup text extraction method if extraction fails.
  - **dtSearch (failover to Native)**—tells Relativity to use 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.

Note: If OCR isn't essential to your processing job, it's recommended to disable the OCR field on your processing profile, as doing so can significantly reduce processing time and prevent irrelevant documents from having OCR performed on them. You can then perform OCR on only relevant documents outside of the processing job.

- **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.
    - **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.
- **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.

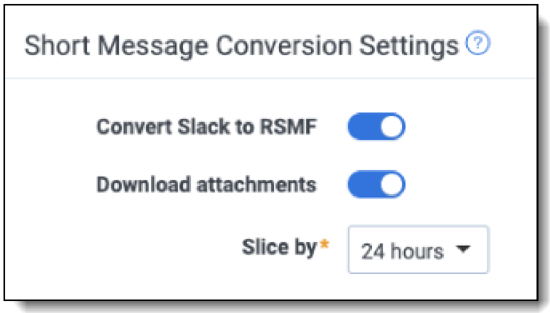
Note: When you process files with both the OCR and the OCR Text Separator fields enabled, any section of a document that required OCR will include text that says **OCR from Image**. This can then pollute a dtSearch index because that index is typically built off of the extracted text field, and OCR from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

10.1.4 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.

Note: To view information on supported file types for short messages, see Short message conversion for Slack. For short message mapping considerations, see Relativitity's short message format.

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



Short Message Conversion Settings

Convert Slack to RSMF ●

Download attachments ●

Slice by* 24 hours

- **Convert Slack to RSMF**—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.

    Note: Downloading attachments can lead to a significant increase in the size of your import container. Make sure you have the necessary storage resources to accommodate the additional file volume.

- **Slice by**—determines how Relativity splits conversations into RSMF files in terms of time.

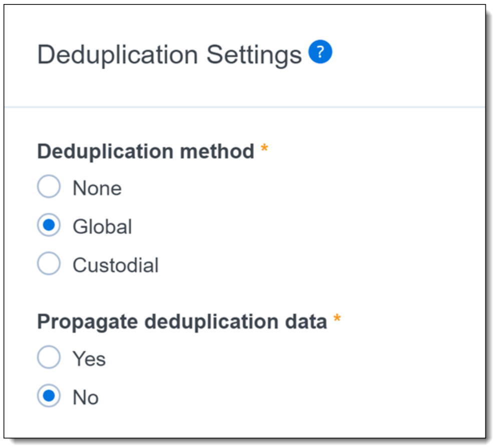
  Note: There are a few limitations when creating RSMF files. The first, a Viewer condition, limits the number of events to 10,000. The second, a Processing condition, limits the RSMF file size to 2 GB. Relativity splits the RSMF output into multiple smaller files if it encounters either condition for a select time frame.

  Note: Relativity uses the UTC time zone setting to calculate blocks 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.

10.1.5 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

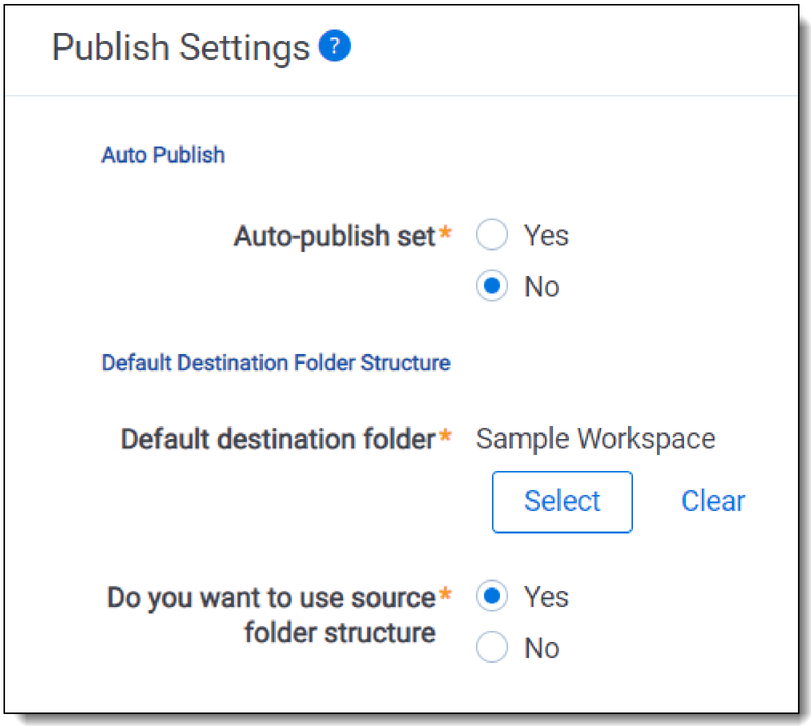
Propagate deduplication data field below visible and required.

> **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.

10.1.6 Publish settings

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



Publish Settings

Auto Publish

Auto-publish set* ○ Yes ● No

Default Destination Folder Structure

Default destination folder* Sample Workspace

Do you want to use source folder structure* ● Yes ○ 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.

10.1.7 Other considerations

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

10.1.7.1 Parent/child numbering type examples

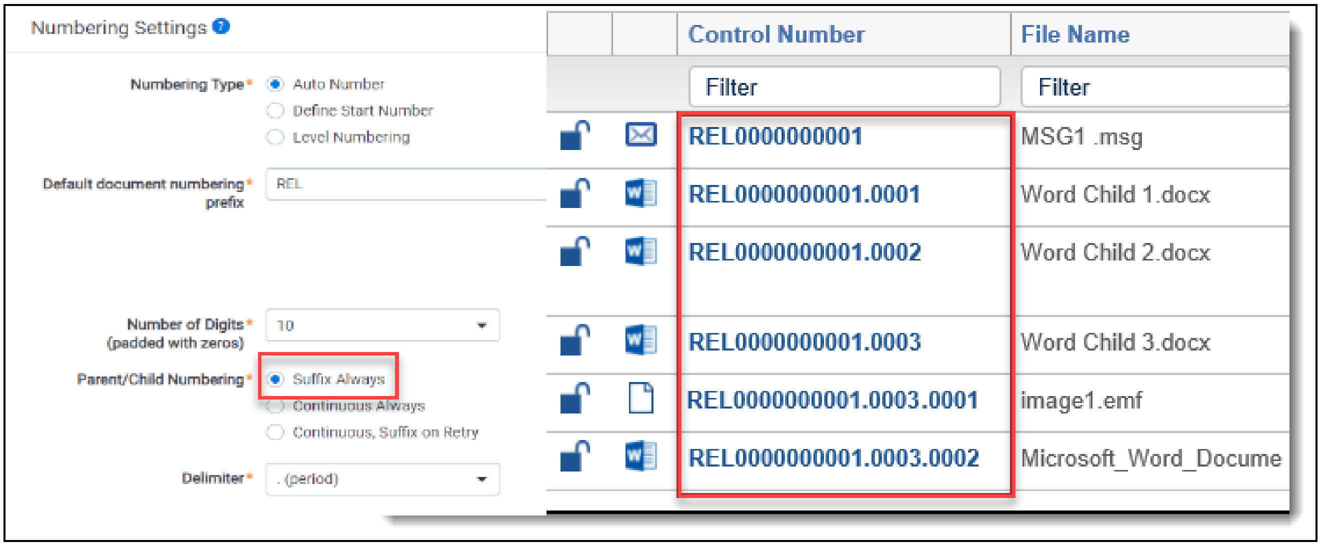
To better understand how each parent/child numbering option appears for published documents, consider the following scenario.

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'd chosen Suffix 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

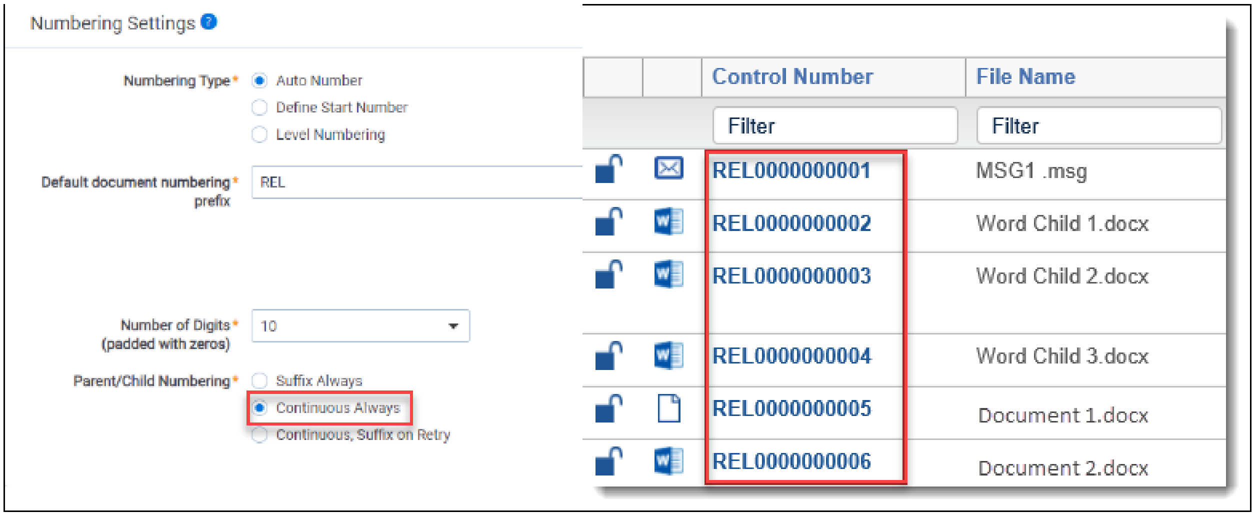
Number of Digits (padded with zeros)*: 10

Parent/Child Numbering*: ● Suffix Always ○ Continuous Always ○ Continuous, Suffix on Retry

Delimiter*: . (period)

Control Number	File Name
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'd chosen 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

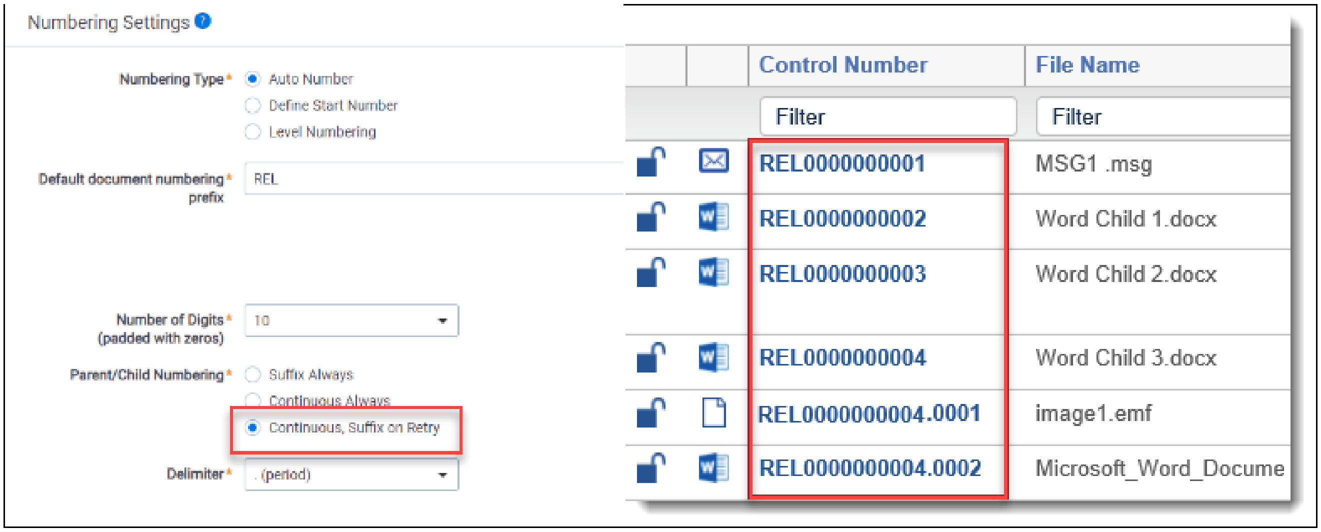
Number of Digits (padded with zeros)*: 10

Parent/Child Numbering*: ○ Suffix Always ● Continuous Always ○ Continuous, Suffix on Retry

Control Number	File Name
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'd chosen 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 haven't 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.

10.1.7.2 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

10.1.7.3 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*.
- Note the following differences between unpublished documents and published documents with errors:
    - If a file is unpublished, and Continuous Always is the numbering option on the profile, Relativity will not add a suffix
    - If a file is unpublished, and Suffix Always is the numbering option on the profile, Relativity will add a suffix to it.
    - If a file has an error and is published, and Continuous, Suffix on Retry is the numbering option on the profile, Relativity will add a suffix to it.
- It's 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.

10.1.7.4 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.

10.1.7.5 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:

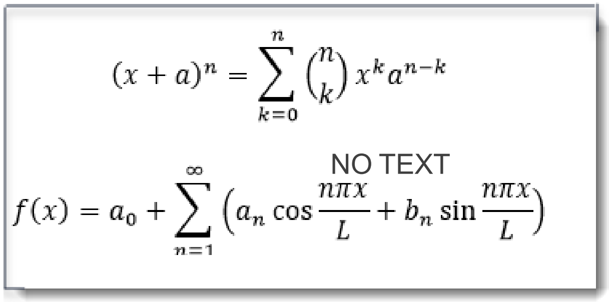
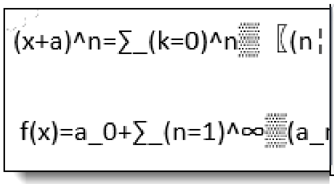
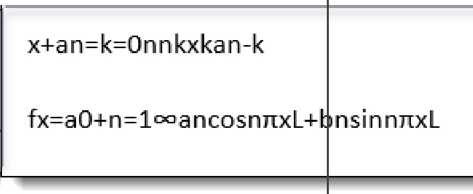
	Relativity			Native			dtSearch		
FEATURES	Excel Features Supported	Word Features Supported	Power Point Features Supported	Excel Features Supported	Word Features Supported	Power Point Features Supported	Excel Features Supported	Word Features Supported	Power Point Features Supported
FEATURE DIFFERENCES									
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	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A

(column level)									
Password protected (file level)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Password protected (row level)	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Password protected (sheet / page level)	✓	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
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

10.1.7.6 Math equations

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

		Text Extraction Method	
Original Document	Relativity	Native	dtSearch
 $(x+a)^n = \sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k}$ $f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L}\right)$	NO TEXT	 (x+a)^n=∑_(k=0)^n▒ 〖(n¦ f(x)=a_0+∑_(n=1)^∞▒(a_	 x+an=k=0nnkxkan-k fx=a0+n=1∞ancosnπxL+bnsinnπxL

11 Deduplication considerations

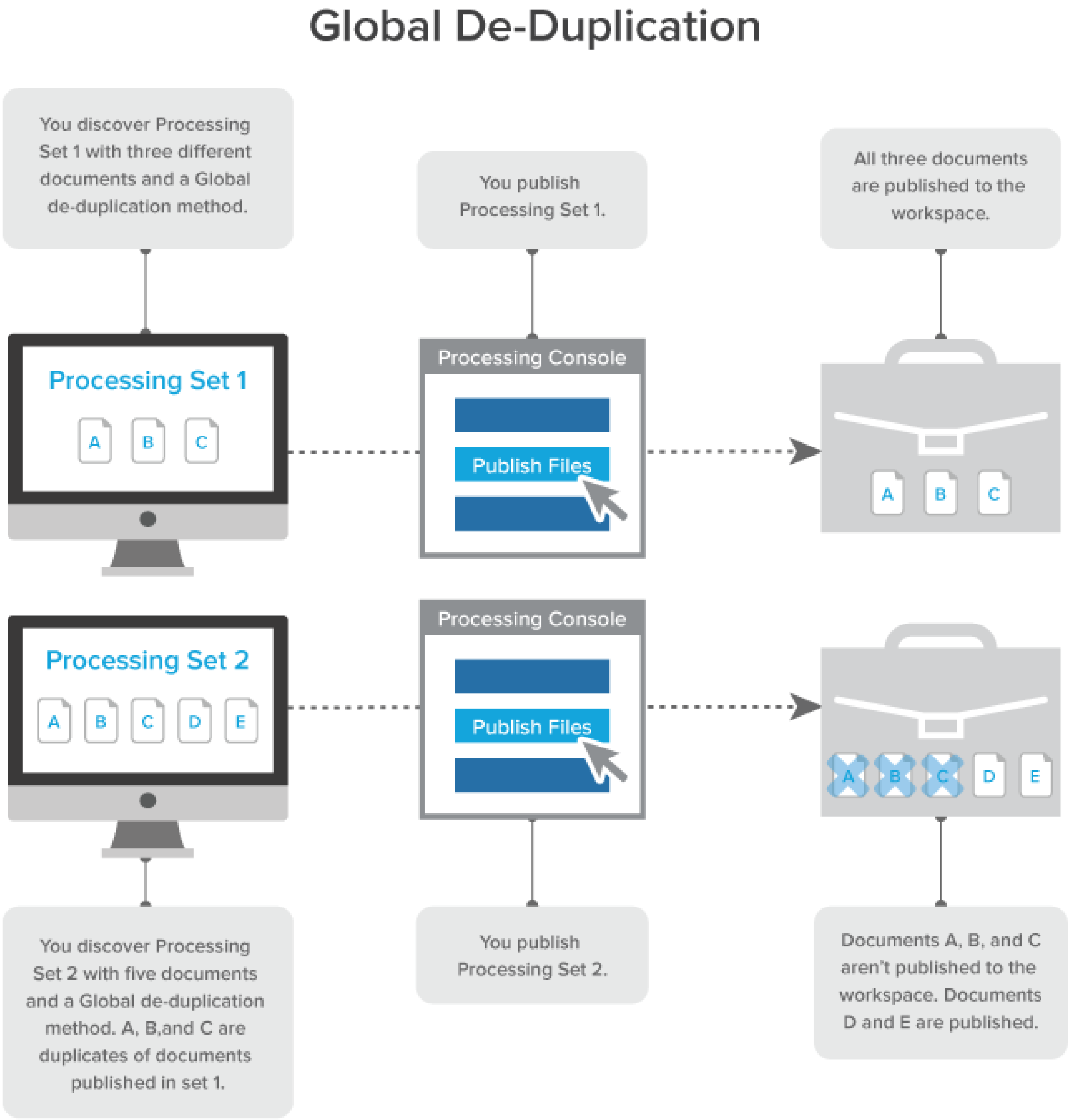
The following scenarios depict what happens when you publish a processing set using each of the deduplication methods available on the processing profile.

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.
- When deduplication runs as part of publish, it doesn't occur inside the Relativity database or the Relativity SQL Server and thus has no effect on review. As a result, there is no database lockout tool available, and reviewers are able to access a workspace and perform document review inside it while publish and deduplication are in progress.
- 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.
- If you change the deduplication method between publications of the same data, even if you're using different processing sets, you may encounter unintended behavior. For example, if you publish a processing set with None selected for the deduplication method on the profile and then make a new set with the same data and publish it with Global selected, Relativity won't publish any new documents because they will all be considered duplicates. In addition, the All Custodians field will display unexpected data. This is because the second publish operation assumed that all previous publications were completed with the same deduplication settings.

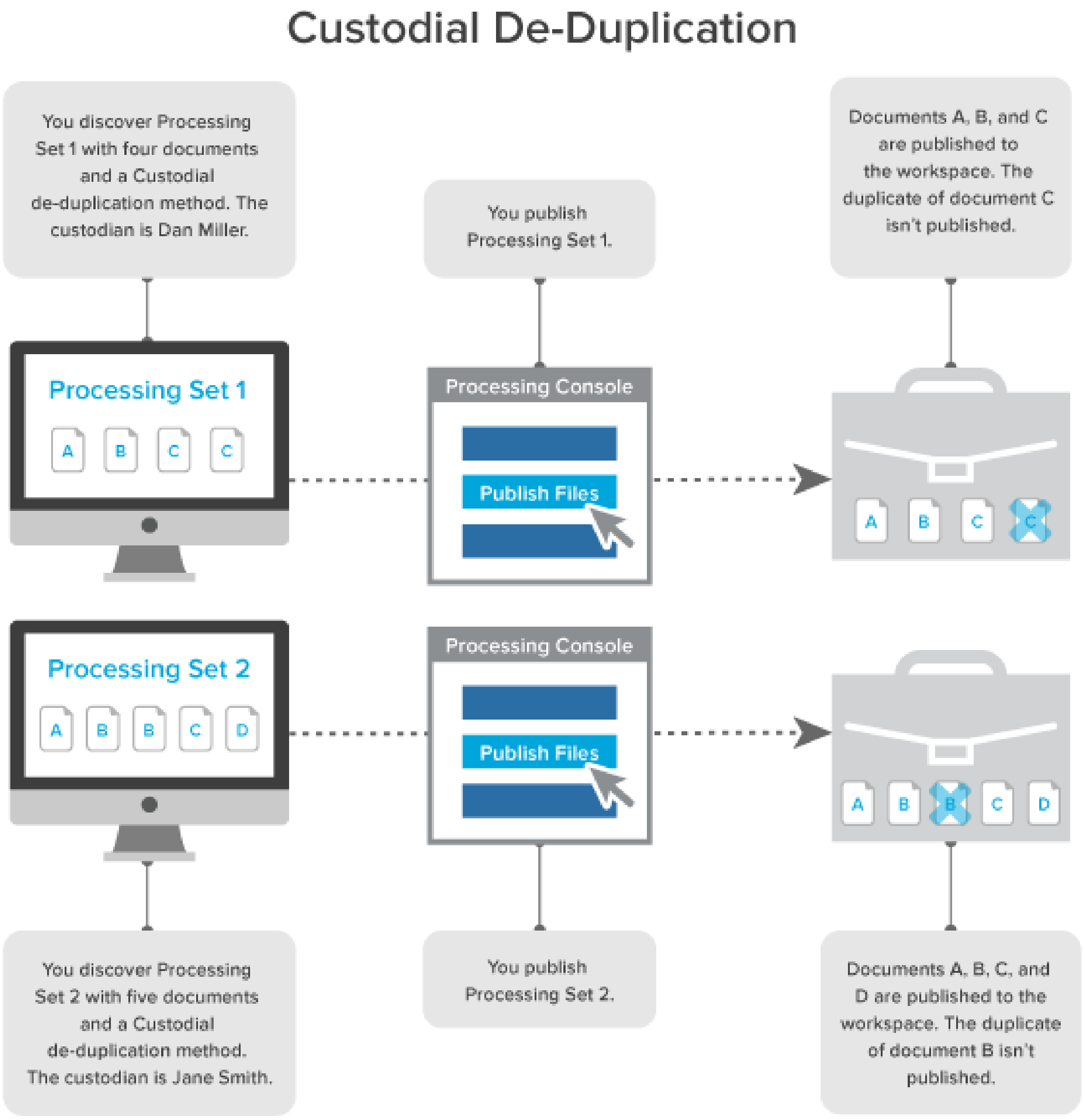
11.1 Global deduplication

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
You discover Processing Set 1 with three different documents and a Global de-duplication method.
You publish Processing Set 1.
All three documents are published to the workspace.
Processing Set 1
A
B
C
Processing Console
Publish Files
A
B
C
Processing Set 2
A
B
C
D
E
Processing Console
Publish Files
A
B
C
D
E
You discover Processing Set 2 with five documents and a Global de-duplication method. A, B,and C are duplicates of documents published in set 1.
You publish Processing Set 2.
Documents A, B, and C aren't published to the workspace. Documents D and E are published.

11.2 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

You discover Processing Set 1 with four documents and a Custodial de-duplication method. The custodian is Dan Miller.

You publish Processing Set 1.

Documents A, B, and C are published to the workspace. The duplicate of document C isn't published.

You discover Processing Set 2 with five documents and a Custodial de-duplication method. The custodian is Jane Smith.

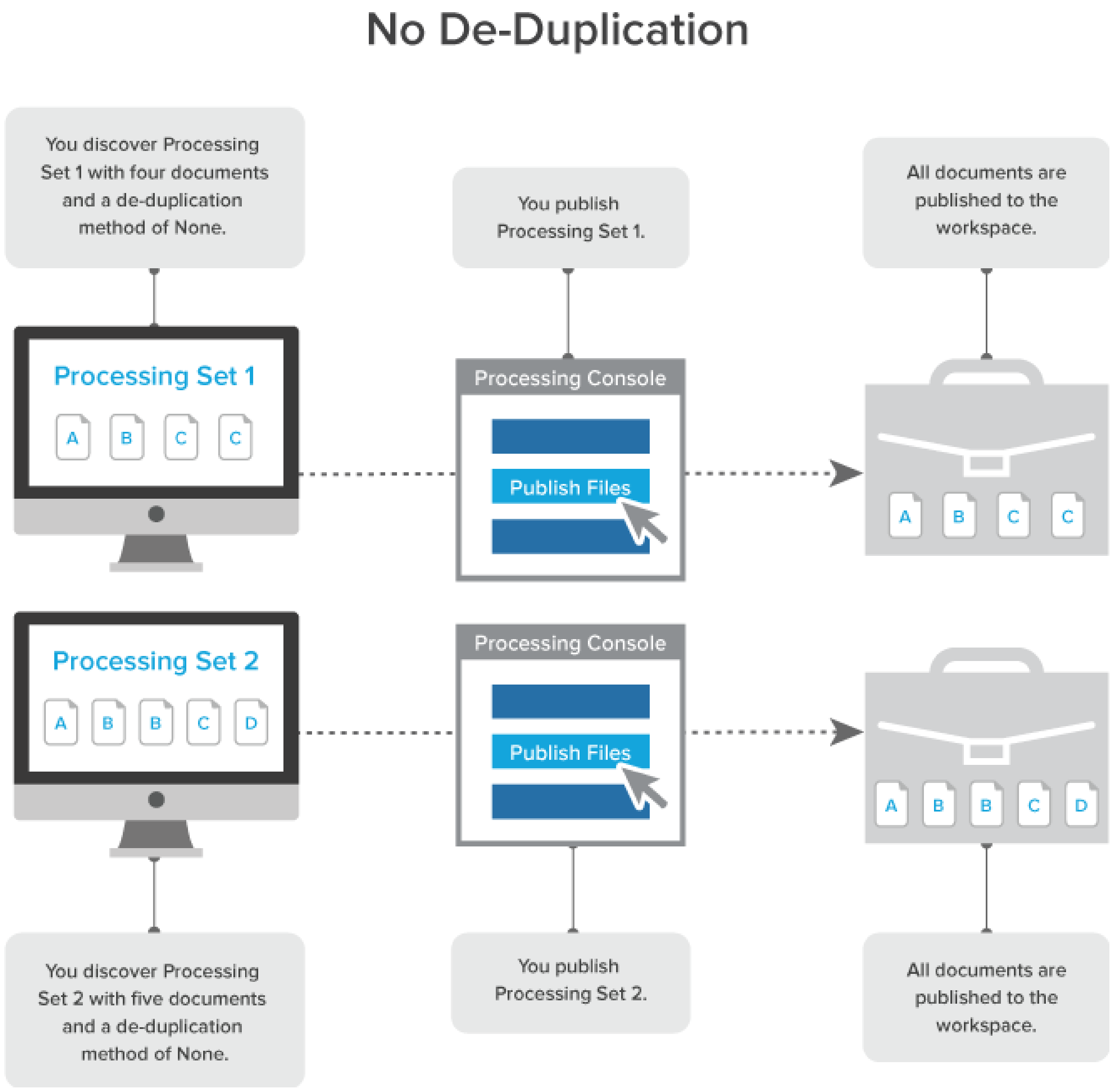
You publish Processing Set 2.

Documents A, B, C, and D are published to the workspace. The duplicate of document B isn't published.

Processing Set 1
A
B
C
C
Processing Console
Publish Files
A
B
C
C
Processing Set 2
A
B
B
C
D
Processing Console
Publish Files
A
B
B
C
D

11.3 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

You discover Processing Set 1 with four documents and a de-duplication method of None.

You publish Processing Set 1.

All documents are published to the workspace.

Processing Set 1
A
B
C
C
Processing Console
Publish Files
A
B
C
C
Processing Set 2
A
B
B
C
D
Processing Console
Publish Files
A
B
B
C
D

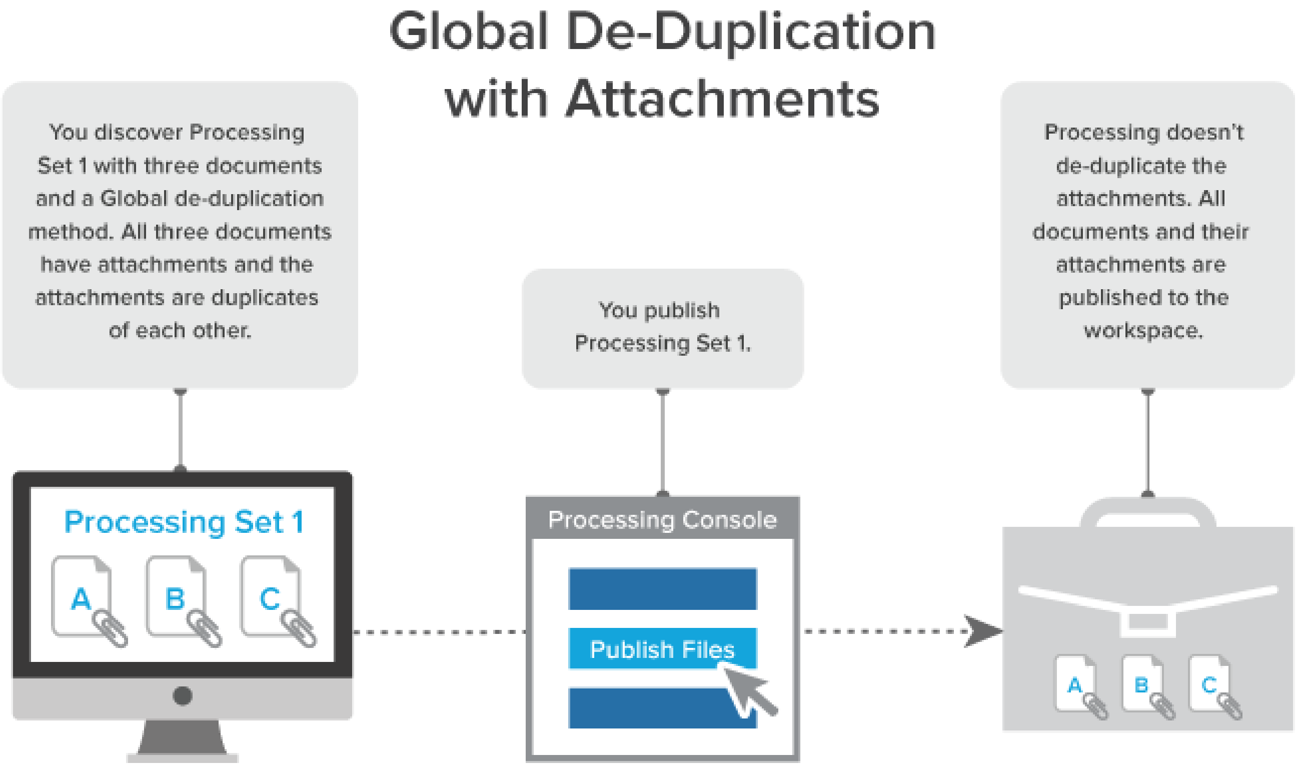
You discover Processing Set 2 with five documents and a de-duplication method of None.

You publish Processing Set 2.

All documents are published to the workspace.

11.4 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
You discover Processing Set 1 with three documents and a Global de-duplication method. All three documents have attachments and the attachments are duplicates of each other.
You publish Processing Set 1.
Processing doesn't de-duplicate the attachments. All documents and their attachments are published to the workspace.
Processing Set 1
A
B
C
Processing Console
Publish Files
A
B
C

11.5 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.

COPYING NATIVES DURING PROCESSING
Source
EDDS Repository
\\filerepo\EDDS\GUID\ABC.msg
Inventory
Discovery
Publish

11.6 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**—the hash used by processing in Relativity to de-duplicate files, which references a Unicode string of the header, body, attachment, and recipient hashes generated by processing. For loose files, the Processing Duplicate Hash is a hash of the file's SHA256 hash.

11.6.1 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 can't calculate the MD5 hash value if you have FIPS (Federal Information Processing Standards cryptography) enabled for the worker manager server.

11.6.2 Calculating deduplication hashes for emails

11.6.2.1 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.

11.6.2.2 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

11.6.2.3 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

11.6.2.4 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:
   80D03318867DB05E40E20CE10B7C8F511B1D0B9F336EF2C787CC3D51B9E26BC9974C9D2C-
   0EEC0F515C770B8282C87C1E8F957FAF34654504520A7ADC2E0E23EA

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.

11.6.3 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.
   -
     ```
     `6283cf-
     b34e4831c97e363a9247f1f01-
     beaaed01d-
     b3a65a47be310c27e3729a3ee05dce5acaec3696c681cd7e-
     b646a221a8fc376478b655c81214d-
     ca7419aa-
     bee6283cf-
     b34e4831c97e363a9247f1f01-
     beaaed01d-
     b3a65a47be310c27e3729a3ee3843222f1805623930029bad6f32a7604e2a7ac-
     c10db9126e34d7be289cf86e`
     ```

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.

Note: If two emails have an identical body, attachment, recipient, and header hash, they are duplicates.

Note: For loose files, the Processing Duplicate Hash is a hash of the file's SHA256 hash.

12 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.

Note: Your environment has been enabled to dynamically scale your Invariant worker servers dependent on load. Sustained activity is automatically detected by the system, and Relativity will add workers to handle this work. Once the work is done, they will automatically scale back down. This feature is continually being improved to be smarter about when we add workers and how many we add.

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.

Using a Processing set

Imagine that you’re a litigation support specialist. The firm you work for has put you in charge of setting the groundwork to bring data owned by two specific custodians into a Relativity workspace because those custodians have just been identified as maintaining material that is potentially relevant to the case.

To do this, you need to create a new processing set using the Default profile. Once you save the set, you need to attach those two specific custodians to the set via two separate processing data sources.

You can now bring only these two custodians' files through the processing phases of inventory, discovery, and publish.

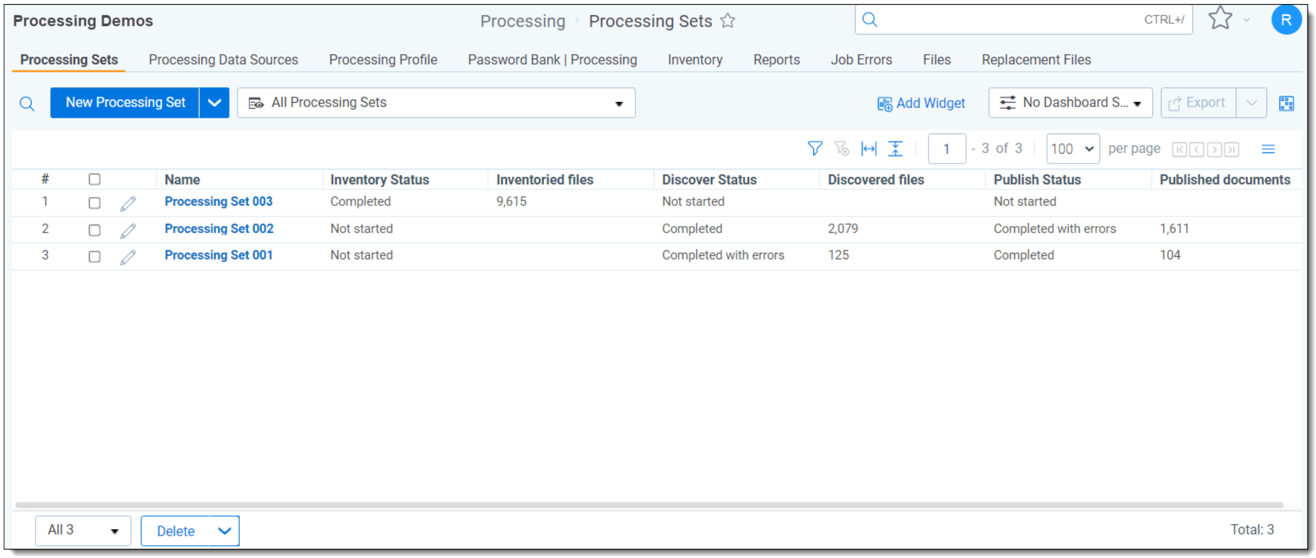
For the details of creating a processing set, see Creating a processing set on page 122.

12.1 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 Demos
Processing › Processing Sets ☆
CTRL+/
R
Processing Sets
Processing Data Sources
Processing Profile
Password Bank | Processing
Inventory
Reports
Job Errors
Files
Replacement Files
New Processing Set
All Processing Sets
Add Widget
No Dashboard S...
Export
1 - 3 of 3
100
per page
All 3
Delete
Total: 3

#	☐		Name	Inventory Status	Inventoried files	Discover Status	Discovered files	Publish Status	Published documents
1	☐		Processing Set 003	Completed	9,615	Not started		Not started	
2	☐		Processing Set 002	Not started		Completed	2,079	Completed with errors	1,611
3	☐		Processing Set 001	Not started		Completed with errors	125	Completed	104

- **Name** - the name of the processing set.
- **Inventory Status** - the current status of the inventory phase of the set. This field could display any of the following status values:
  - Not started
  - In progress
  - Completed
  - Completed with errors
  - Re-inventory required - Upgrade
  - Re-inventory required - Data sources modified
  - Canceled
  - Finalized failed
- **Inventoried files** - the number of files across all data sources on the set that have been inventoried.

  Note: Inventory populates only job level errors.
- **Discover Status** - the current status of the discovery phase of the set. This field could display any of the following status values:
  - Not started
  - In progress

  - Completed
  - Completed with errors
  - Canceled
- **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.

12.2 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:

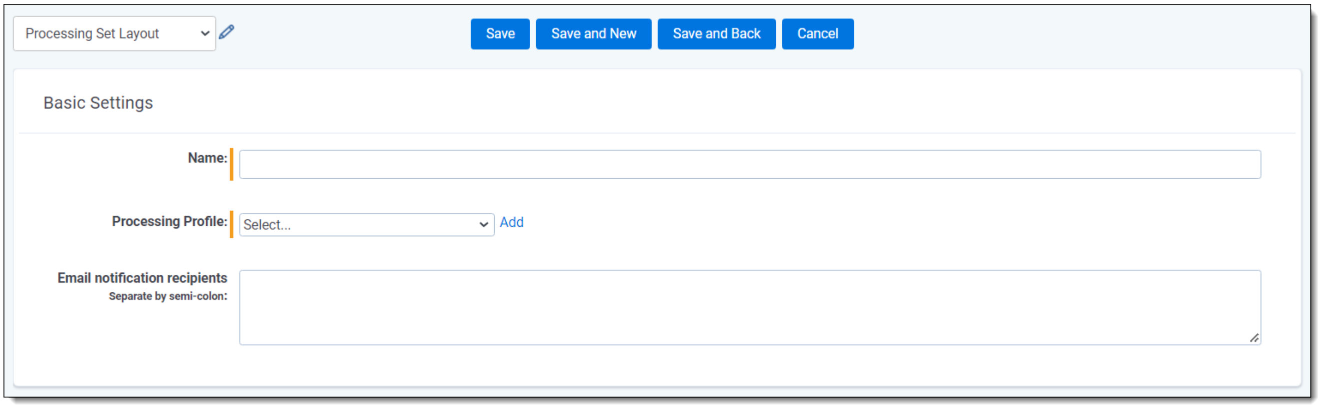
1. Navigate to the **Processing** tab and then click the **Processing Sets** sub-tab.
2. Click the **New Processing Set** button to display the Processing Set layout.
3. Complete the fields on the Processing Set layout. See Processing set fields on the next page.
4. Click **Save**.

5. Add as many Processing Data Sources to the set as you need. See Adding a data source on page 125.

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.

12.3 Processing set fields

To create a processing set, complete the following fields:

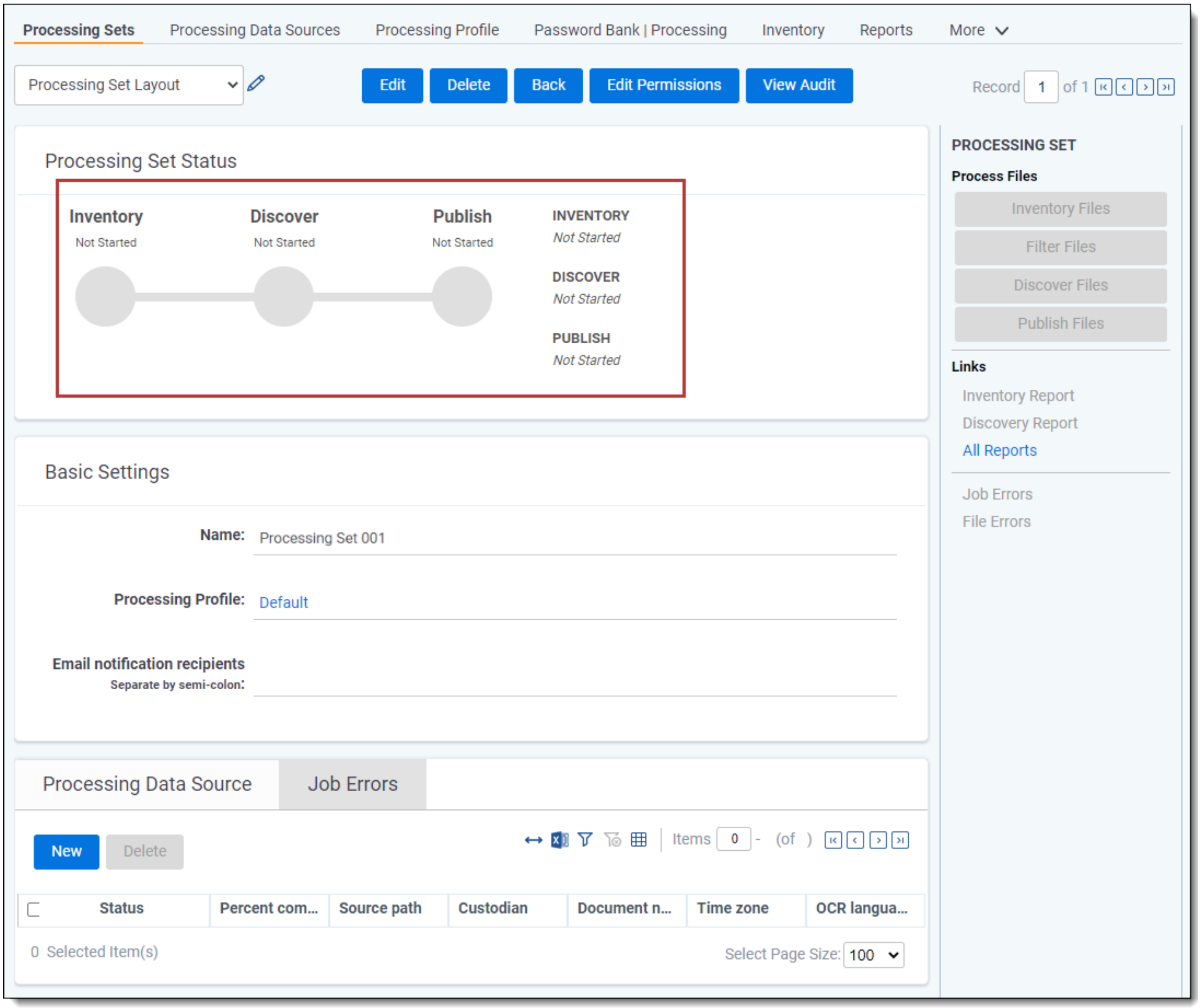

Processing Set Layout
Save
Save and New
Save and Back
Cancel
Basic Settings
Name:
Processing Profile:
Select...
Add
Email notification recipients
Separate by semi-colon:

- **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 on page 82.
- **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.


Processing Sets
Processing Data Sources
Processing Profile
Password Bank | Processing
Inventory
Reports
More
Processing Set Layout
Edit
Delete
Back
Edit Permissions
View Audit
Record 1 of 1
PROCESSING SET
Process Files
Inventory Files
Filter Files
Discover Files
Publish Files
Links
Inventory Report
Discovery Report
All Reports
Job Errors
File Errors
Processing Data Source
Job Errors
New
Delete
Items 0 - (of)
Select Page Size: 100

Processing Set Status

Inventory Not Started
Discover Not Started
Publish Not Started

INVENTORY *Not Started*

DISCOVER *Not Started*

PUBLISH *Not Started*

Basic Settings

Name: Processing Set 001

Processing Profile: Default

Email notification recipients Separate by semi-colon:

☐	Status	Percent com...	Source path	Custodian	Document n...	Time zone	OCR langua...

0 Selected Item(s)

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.

12.4 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.

Note:
Be sure your data sources have unique names. Data sources are not case sensitive. 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]

Note: 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.

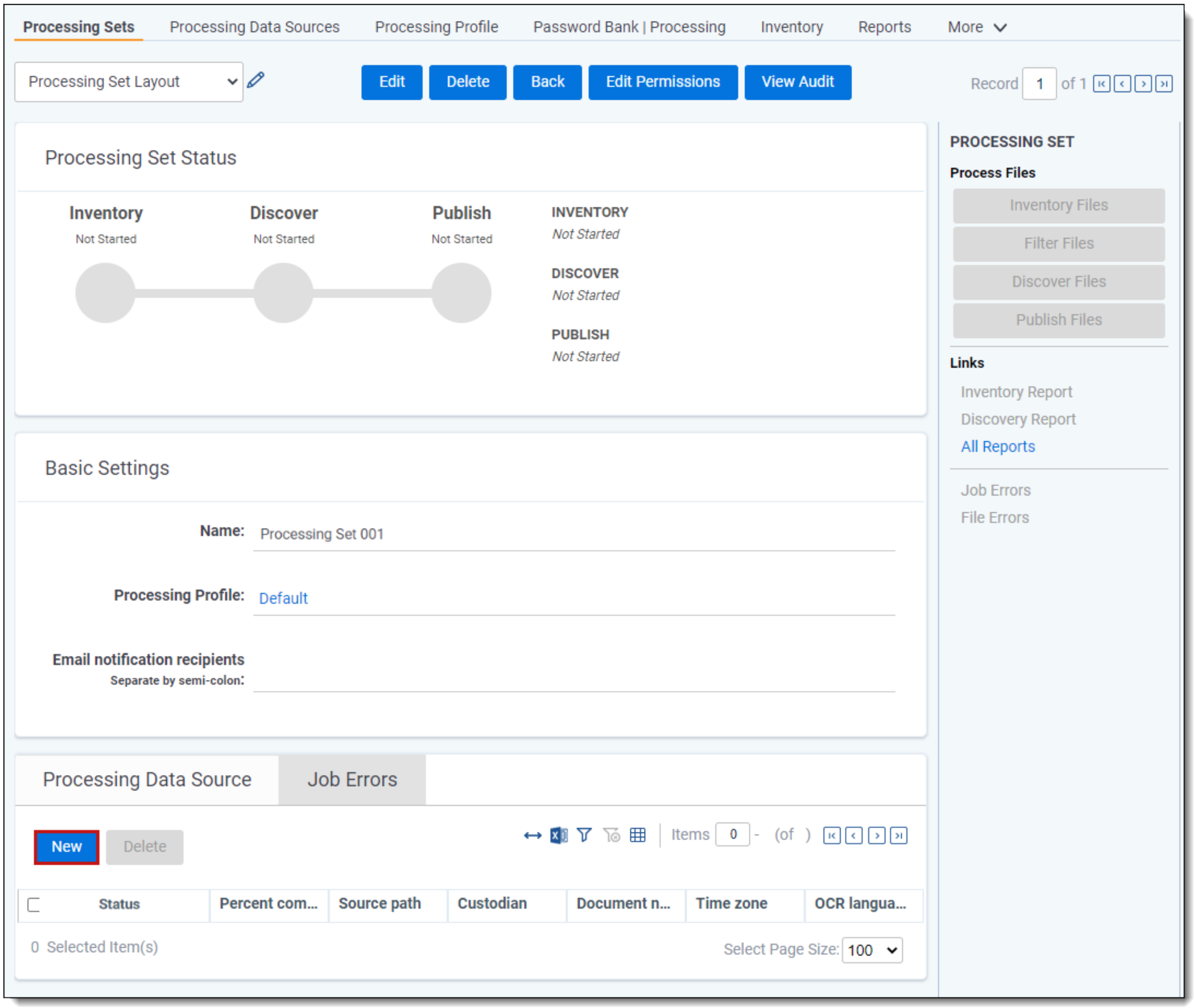
You can add multiple data sources to a single processing set, which means that you can process data for multiple custodians through a single set. There is no limit to the number of data sources you can add to a set; however, most sets contain ten or fewer.

Note: 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 on page 122.

2. On the **Processing Data Source** object of the processing set click **New**.


Processing Sets
Processing Data Sources
Processing Profile
Password Bank | Processing
Inventory
Reports
More
Processing Set Layout
Edit
Delete
Back
Edit Permissions
View Audit
Record 1 of 1
Processing Set Status
Inventory
Not Started
Discover
Not Started
Publish
Not Started
INVENTORY
Not Started
DISCOVER
Not Started
PUBLISH
Not Started
PROCESSING SET
Process Files
Inventory Files
Filter Files
Discover Files
Publish Files
Links
Inventory Report
Discovery Report
All Reports
Job Errors
File Errors
Basic Settings
Processing Data Source
Job Errors
New
Delete
Items 0 - (of)
Status
Percent com...
Source path
Custodian
Document n...
Time zone
OCR langua...
0 Selected Item(s)
Select Page Size: 100

Name: Processing Set 001

Processing Profile: Default

Email notification recipients Separate by semi-colon:

3. Complete the fields on the **Add Processing Data Source** layout. See Data source fields below.

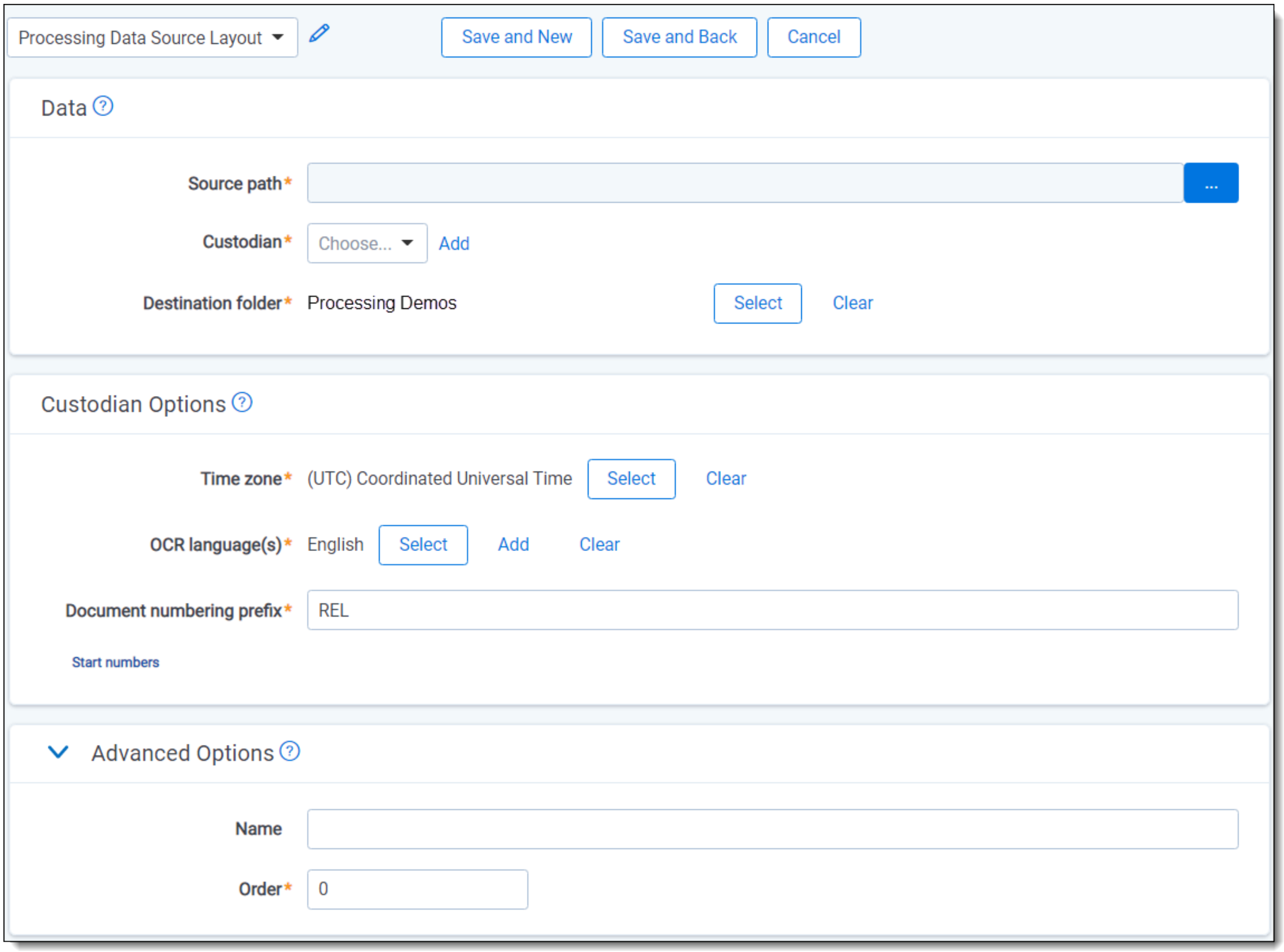
4. Click **Save**. When you save the data source, it becomes associated with the processing set and the console on the right side is enabled for inventory and file discovery.

For details on what information is displayed in the data source view while the processing set is running, see Processing data source view on page 134.

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.

12.5 Data source fields

To add a data source, complete the following fields:


Processing Data Source Layout
Save and New
Save and Back
Cancel
Data
...
Add
Select
Clear
Custodian Options
Select
Clear
Select
Add
Clear
Start numbers
Advanced Options

Source path*

Custodian* Choose...

Destination folder* Processing Demos

Time zone* (UTC) Coordinated Universal Time

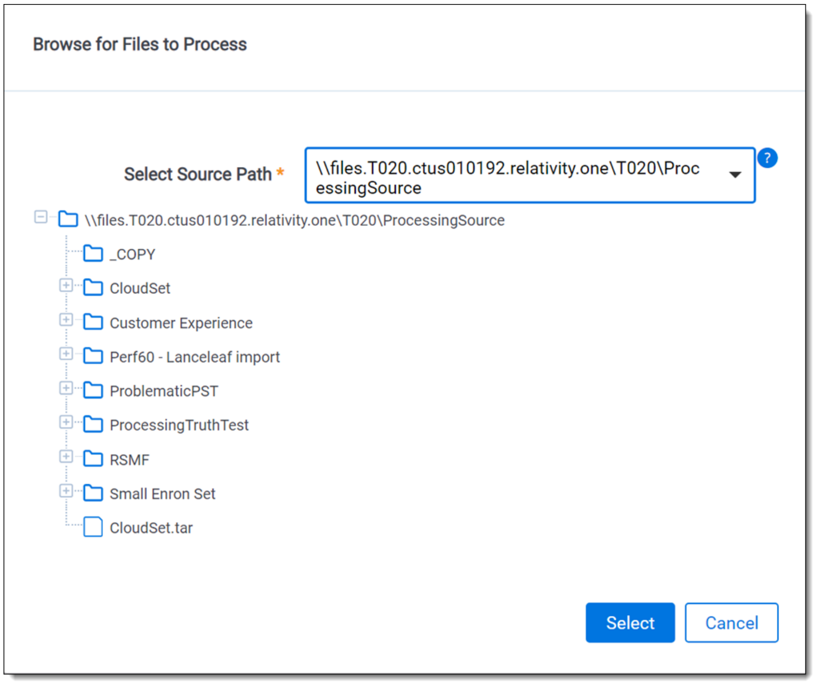
OCR language(s)* English

Document numbering prefix* REL

Name

Order* 0

- **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.



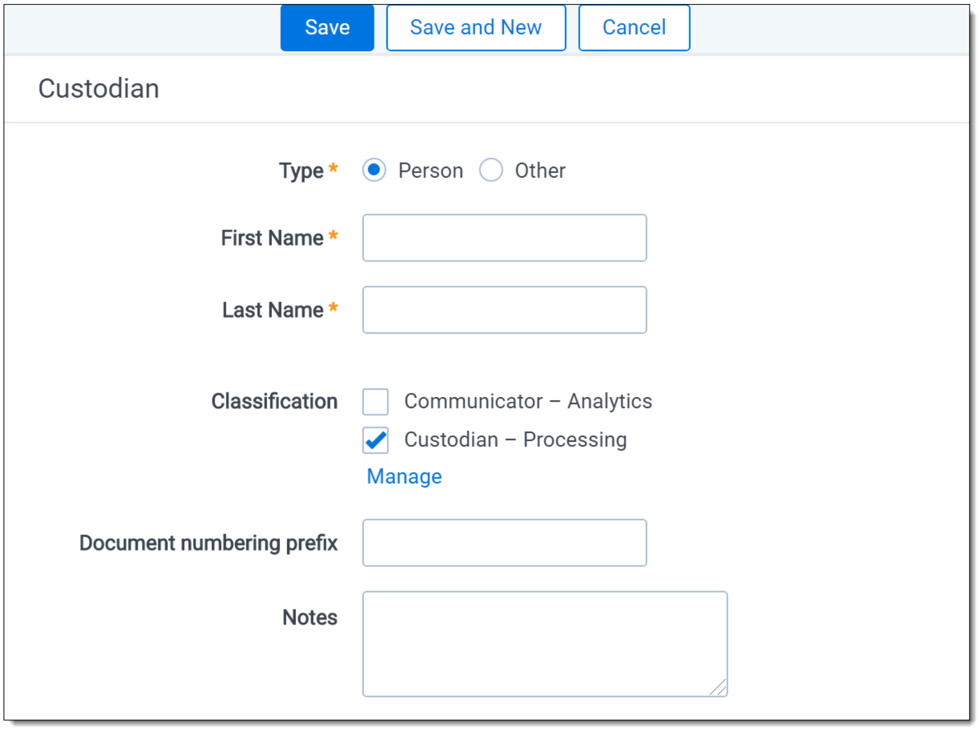
Browse for Files to Process
Select Source Path *
\\files.T020.ctus010192.relativity.one\T020\ProcessingSource
\\files.T020.ctus010192.relativity.one\T020\ProcessingSource
_COPY
CloudSet
Customer Experience
Perf60 - Lanceleaf import
ProblematicPST
ProcessingTruthTest
RSMF
Small Enron Set
CloudSet.tar
Select
Cancel

  - 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\SambaShare\Samba*, you must enter this exactly and not as “\\files\SambaShare\samba” or “\\files\sambashare\Samba”, 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 on page 144.

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.



Custodian

Type * ● Person ○ Other

First Name *

Last Name *

Classification ☐ Communicator – Analytics
☑ Custodian – Processing

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.

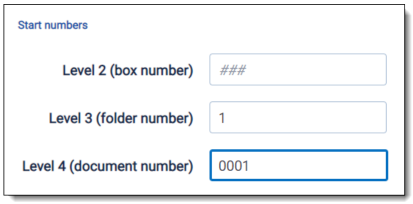
- **Destination folder** - the folder in Relativity where the processed data is published. This default value of this field is pulled from the processing profile. If you edit this field to a different destination folder location, the processing engine reads this value and not the folder specified on 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.
- **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.
    - The OCR settings used during processing are the same as those used during standard OCR.

  - 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>xxxxxxxxxx - 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 is 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).

- **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 below.

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.

12.5.1 Order considerations

The Order field determines:

- The job priority of the data source within a given processing set when the set is submitted to the queue (e.g., for discovery or publication). For example, a data source with a lower order number assigned is discovered and/or published before a data source with a higher order number assigned in a given set.
- Changing the order of a data source has no effect on the priority of the processing set. This means that if you set the order of a data source in one processing set to a higher priority than all of the data sources in another processing set, the priorities of the processing sets won't be modified.
- The priority of deduplication if you select a deduplication method other than None. For example, if Global deduplication is specified for a processing set, the data source with the lowest order number assigned would be designated as the primary data source within that processing set. This means that all duplicate files in higher-ordered data sources within that processing set would be deduplicated out against the files in the “primary” source. Any files in the source with the lowest order number assigned would not be removed via deduplication.

Note the following about the Order field:

- It isn't editable after you publish the files in this data source.
- If two data sources have the same order, or if you don't specify an order, Relativity sorts them by their system-assigned artifact ID number. At the time of publish, if two data sources have the same order, or if you don't specify an order, deduplication order is also determined by Artifact ID.
- 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. When you change the priority of a publish or republish job, Relativity respects the deduplication method used by the processing set containing the modified data sources.

- This value should always be lower than the maximum allowable integer of 2,147,483,647. If this is at or higher, subsequent data sources will have a negative order value.

12.5.2 Edit considerations for data sources

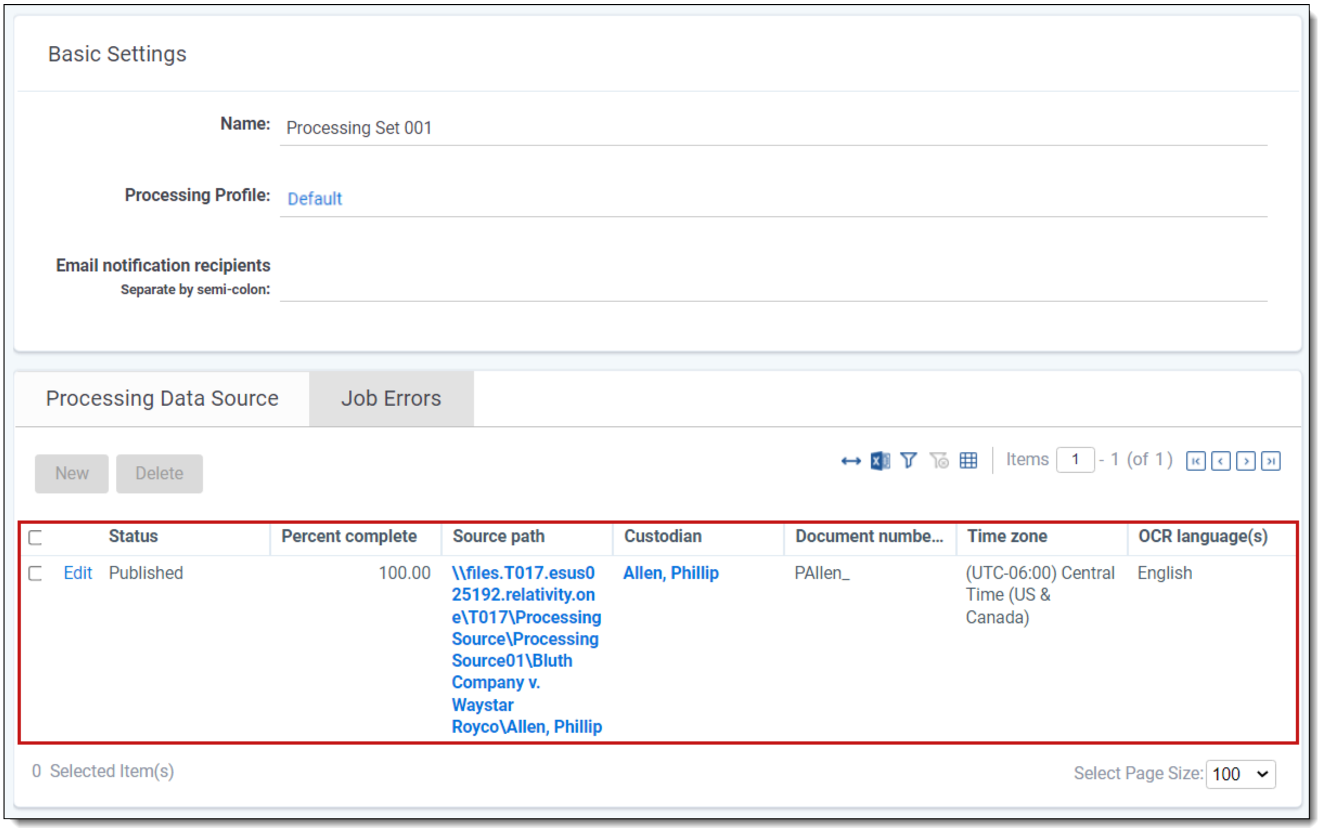
Note the following guidelines for modifying data sources:

- 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.

12.5.3 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.



Basic Settings

Name: Processing Set 001

Processing Profile: Default

Email notification recipients
Separate by semi-colon:

Processing Data Source | Job Errors

Items 1 - 1 (of 1)

☐	Status	Percent complete	Source path	Custodian	Document numbe...	Time zone	OCR language(s)
☐ Edit	Published	100.00	\\files.T017.esus025192.relativity.one\T017\Processing Source\Processing Source01\Bluth Company v. Waystar Royco\Allen, Phillip	Allen, Phillip	PAllen_	(UTC-06:00) Central Time (US & Canada)	English

0 Selected Item(s)

Select Page Size: 100

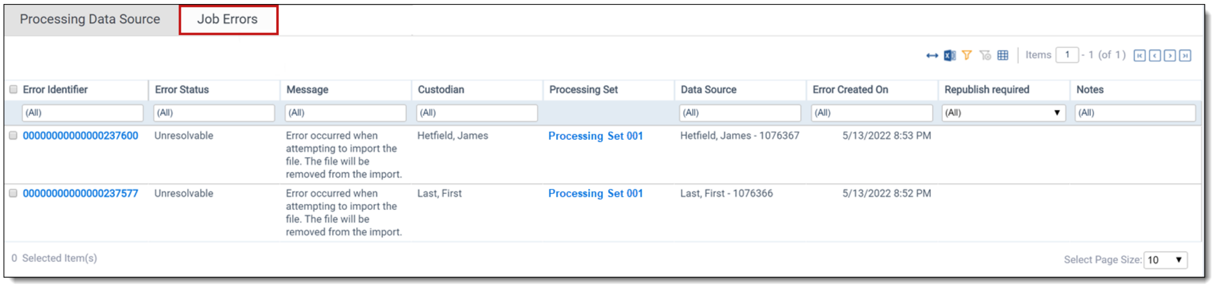
This view provides the following fields:

- **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. This status was added in July 2017 as part of the distributed publish enhancement.
    - **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. This status was added in July 2017 as part of the distributed publish enhancement. 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.

12.5.4 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.



☐ Error Identifier	Error Status	Message	Custodian	Processing Set	Data Source	Error Created On	Republish required	Notes
(All)	(All)	(All)	(All)		(All)	(All)	(All)	(All)
☐ 00000000000000237600	Unresolvable	Error occurred when attempting to import the file. The file will be removed from the import.	Hetfield, James	Processing Set 001	Hetfield, James - 1076367	5/13/2022 8:53 PM		
☐ 00000000000000237577	Unresolvable	Error occurred when attempting to import the file. The file will be removed from the import.	Last, First	Processing Set 001	Last, First - 1076366	5/13/2022 8:52 PM		

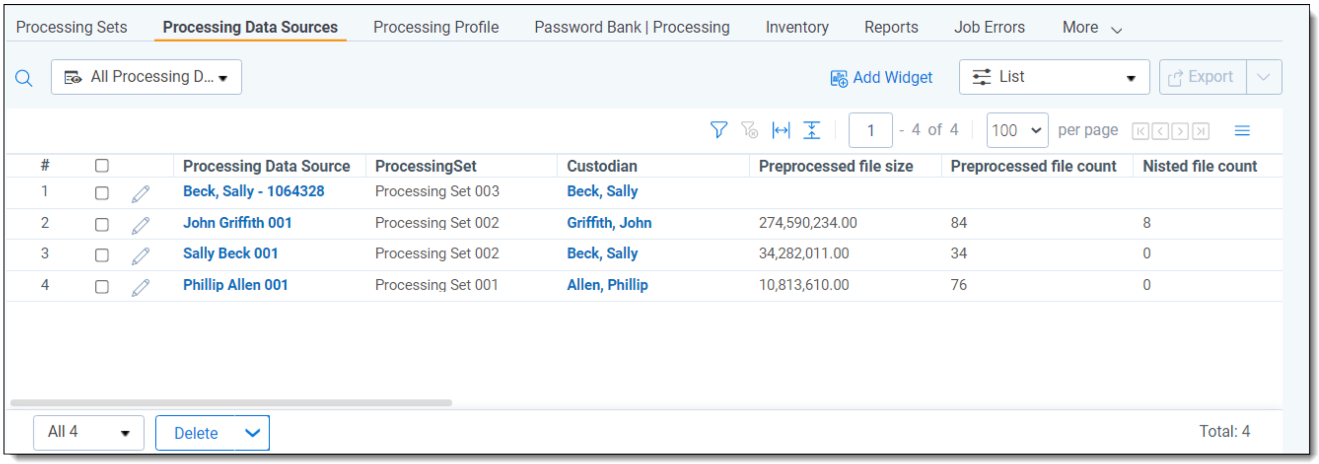
- **Error Identifier**—the unique identifier of the error as it occurs in the database. When you click this message, you are taken to the error details layout, where you can view the stack trace and other information. Note that for Unresolvable errors, the console is disabled because you cannot take 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 deduplication 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.

12.6 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 Data Source	ProcessingSet	Custodian	Preprocessed file size	Preprocessed file count	Nisted file count
1	☐	Beck, Sally - 1064328	Processing Set 003	Beck, Sally			
2	☐	John Griffith 001	Processing Set 002	Griffith, John	274,590,234.00	84	8
3	☐	Sally Beck 001	Processing Set 002	Beck, Sally	34,282,011.00	34	0
4	☐	Phillip Allen 001	Processing Set 001	Allen, Phillip	10,813,610.00	76	0

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.
- **Processing Set** - 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.
- **Preexpansion file count** - the number of files in the data source for all non-container files at the first level after initial expansion.
- **Nisted file count** - the number of files from the data source that were then removed, per the de-NIST setting.
- **Inventoried files** - the number of files from the data source that were inventoried.
- **Preexpansion file size** - the total size, in bytes, of all the files in the data source for all non-container files at the first level after initial expansion.
- **Preprocessed file count** - the number of files in the data source before you started the processing 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.

- **Filtered file count** - the number of files from the data source that were filtered out before discovery.
- **Filtered file size** - the total size, in bytes, of the 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 GB, 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** -
- **Destination folder** - the folder in Relativity into which documents are placed once they're published to the workspace.
- **Discover time complete** -

- **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 GB, 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 sum of all the file sizes, in GB, excluded from discovery by way of an exclusion filter after inventory.
- **Inventoried files** - the count of all files found during an inventory run.
- **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 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 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.
- **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 GB, 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.

12.7 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
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.

The following security permissions are required to delete a processing set:

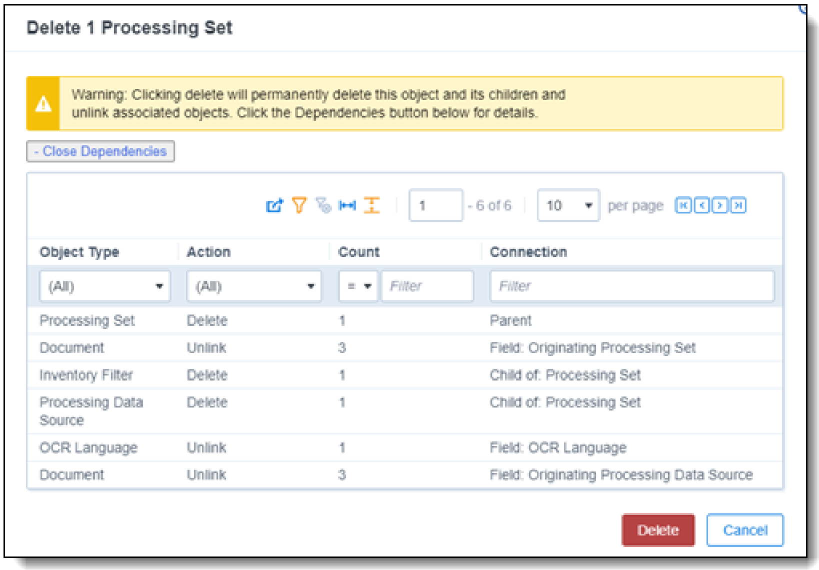
- Tab Visibility - Processing Application. (Processing and Processing Sets at minimum.)
- Other Settings - Delete Object Dependencies. This is required to delete the processing set's child objects and linked associated objects.
- Object Security
  - Edit permissions for Field, with the Add Field Choice By Link setting checked
  - (Optional) Delete permissions for OCR Language
  - Delete permissions for Processing Data Source, Processing Error, Processing Field, and Processing Set

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 its children and unlink associated objects. Click the Dependencies button below for details.

Object Type	Action	Count	Connection
(All)	(All)	=	Filter
Processing Set	Delete	1	Parent
Document	Unlink	3	Field: Originating Processing Set
Inventory Filter	Delete	1	Child of: Processing Set
Processing Data Source	Delete	1	Child of: Processing Set
OCR Language	Unlink	1	Field: OCR Language
Document	Unlink	3	Field: Originating Processing Data Source

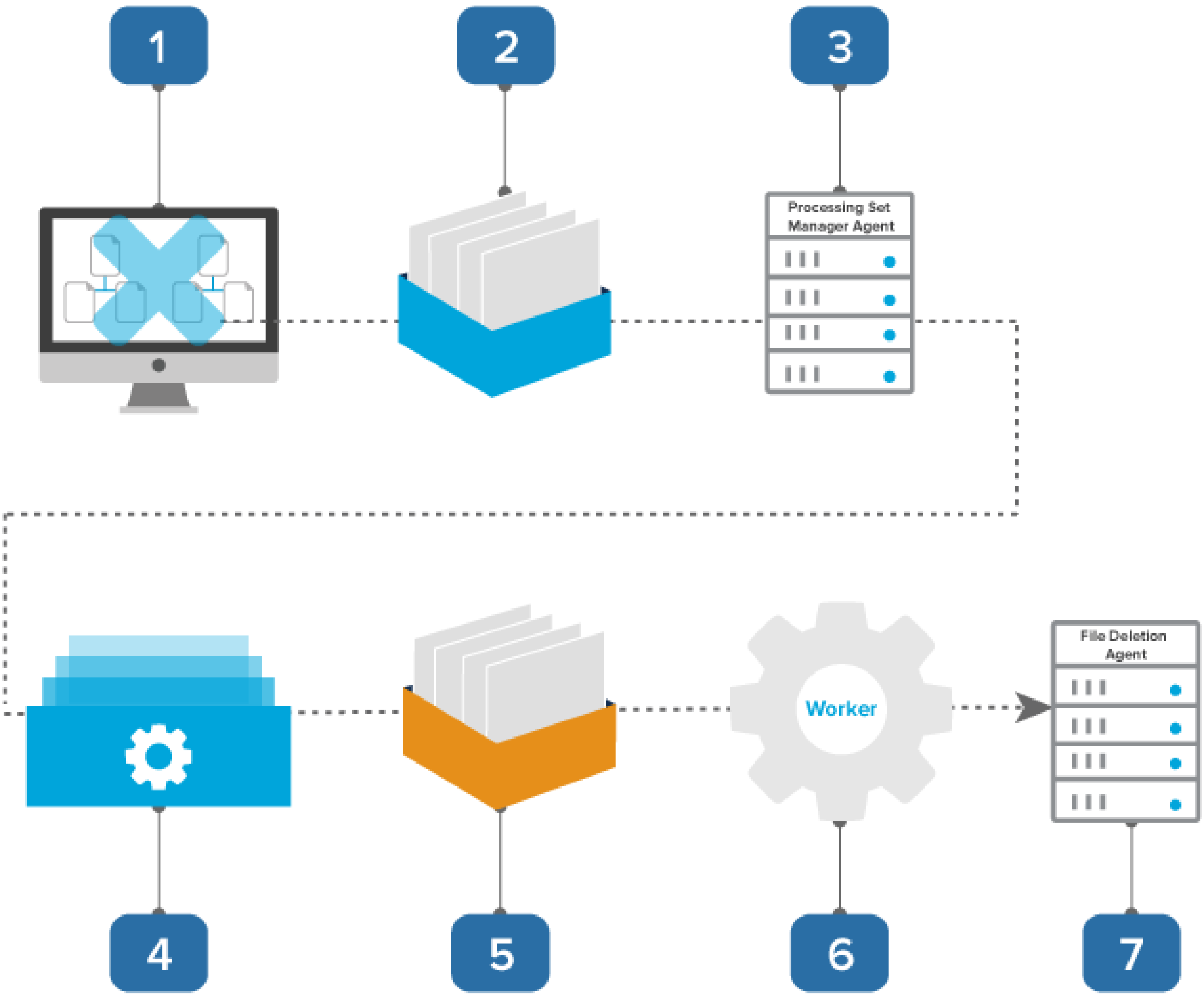
3. Click **Delete** on the confirmation window. When you proceed, you permanently delete the processing set object, its children, and its processing errors, and you unlink all associated objects.

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 subdirectories. 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.

12.8 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:

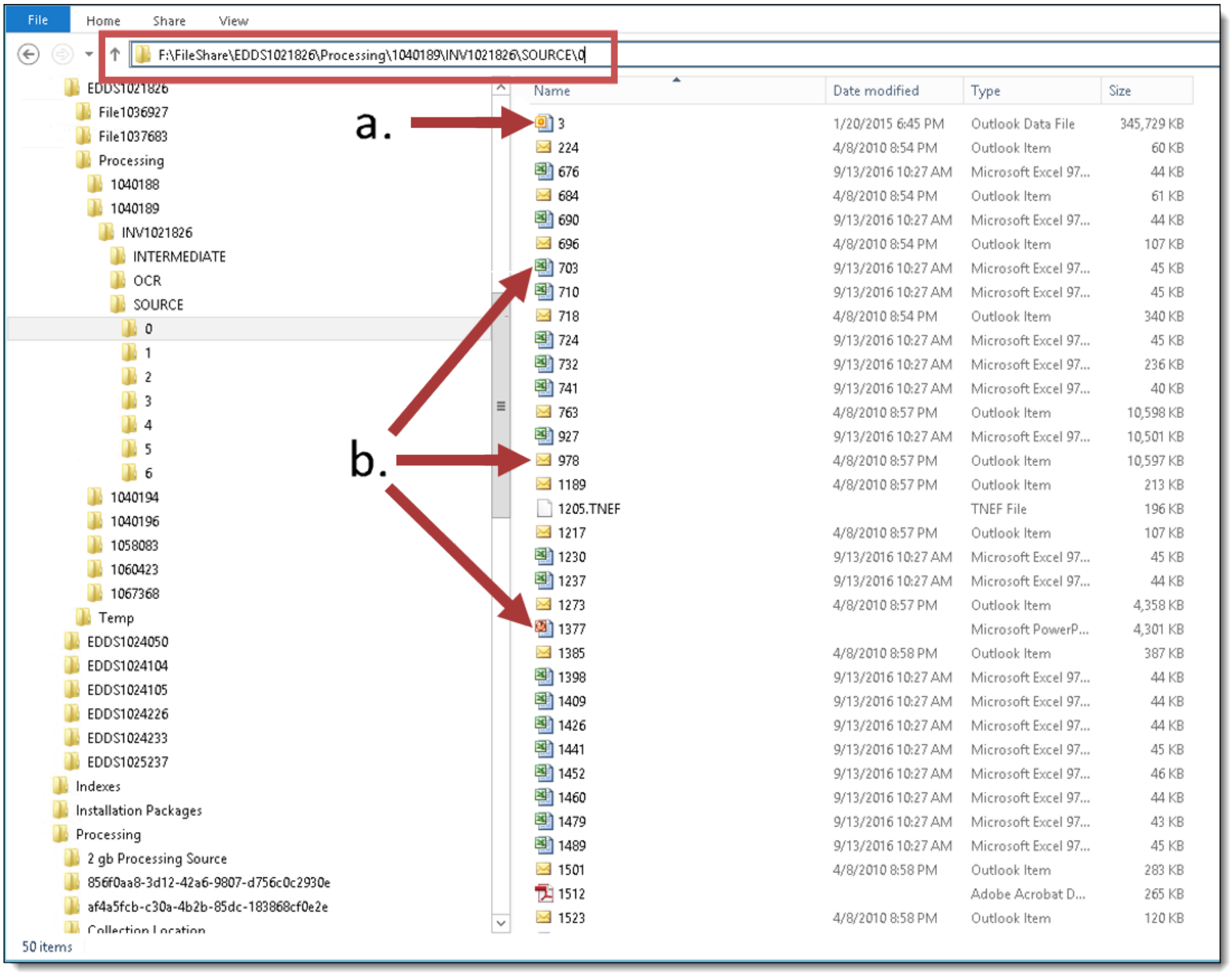
1. Run a processing set.
2. After the processing set is complete, import a small amount of data using Import/Export so that you can keep one steady stream of control numbers and pick up manually where the previous processing set left off.
3. After importing data through Import/Export, run another processing 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.

12.9 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.


File
Home
Share
View
a.
b.
EDDS1021826
File1036927
File1037683
Processing
1040188
1040189
INV1021826
INTERMEDIATE
OCR
SOURCE
0
1
2
3
4
5
6
1040194
1040196
1058083
1060423
1067368
Temp
EDDS1024050
EDDS1024104
EDDS1024105
EDDS1024226
EDDS1024233
EDDS1025237
Indexes
Installation Packages
Processing
2 gb Processing Source
856f0aa8-3d12-42a6-9807-d756c0c2930e
af4a5fcb-c30a-4b2b-85dc-183868cf0e2e
Collection Location
50 items

F:\FileShare\EDDS1021826\Processing\1040189\INV1021826\SOURCE\0

Name	Date modified	Type	Size
3	1/20/2015 6:45 PM	Outlook Data File	345,729 KB
224	4/8/2010 8:54 PM	Outlook Item	60 KB
676	9/13/2016 10:27 AM	Microsoft Excel 97...	44 KB
684	4/8/2010 8:54 PM	Outlook Item	61 KB
690	9/13/2016 10:27 AM	Microsoft Excel 97...	44 KB
696	4/8/2010 8:54 PM	Outlook Item	107 KB
703	9/13/2016 10:27 AM	Microsoft Excel 97...	45 KB
710	9/13/2016 10:27 AM	Microsoft Excel 97...	45 KB
718	4/8/2010 8:54 PM	Outlook Item	340 KB
724	9/13/2016 10:27 AM	Microsoft Excel 97...	45 KB
732	9/13/2016 10:27 AM	Microsoft Excel 97...	236 KB
741	9/13/2016 10:27 AM	Microsoft Excel 97...	40 KB
763	4/8/2010 8:57 PM	Outlook Item	10,598 KB
927	9/13/2016 10:27 AM	Microsoft Excel 97...	10,501 KB
978	4/8/2010 8:57 PM	Outlook Item	10,597 KB
1189	4/8/2010 8:57 PM	Outlook Item	213 KB
1205.TNEF		TNEF File	196 KB
1217	4/8/2010 8:57 PM	Outlook Item	107 KB
1230	9/13/2016 10:27 AM	Microsoft Excel 97...	45 KB
1237	9/13/2016 10:27 AM	Microsoft Excel 97...	44 KB
1273	4/8/2010 8:57 PM	Outlook Item	4,358 KB
1377		Microsoft PowerP...	4,301 KB
1385	4/8/2010 8:58 PM	Outlook Item	387 KB
1398	9/13/2016 10:27 AM	Microsoft Excel 97...	44 KB
1409	9/13/2016 10:27 AM	Microsoft Excel 97...	44 KB
1426	9/13/2016 10:27 AM	Microsoft Excel 97...	44 KB
1441	9/13/2016 10:27 AM	Microsoft Excel 97...	45 KB
1452	9/13/2016 10:27 AM	Microsoft Excel 97...	46 KB
1460	9/13/2016 10:27 AM	Microsoft Excel 97...	44 KB
1479	9/13/2016 10:27 AM	Microsoft Excel 97...	43 KB
1489	9/13/2016 10:27 AM	Microsoft Excel 97...	45 KB
1501	4/8/2010 8:58 PM	Outlook Item	283 KB
1512		Adobe Acrobat D...	265 KB
1523	4/8/2010 8:58 PM	Outlook Item	120 KB

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.

---

   a. The **EDDS12345\Processing\ProcessingSetArtifactID\INV12345\Source\0** folder displays as the original PST.

   b. 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:

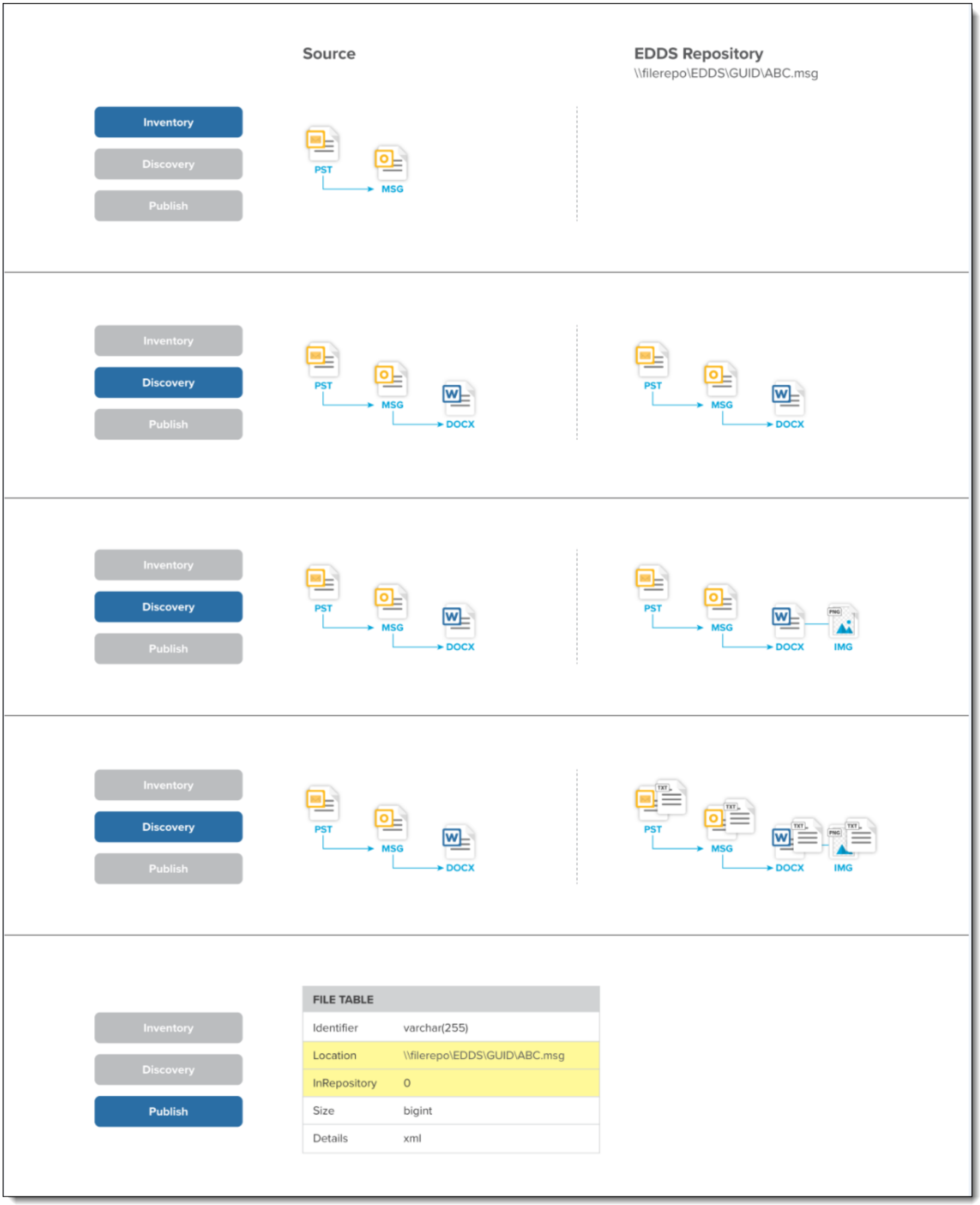
   - You can delete the processing source PST.
   - You can delete the PST file in the EDDS folder, assuming there are no errors.

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
\\filerepo\EDDS\GUID\ABC.msg
Inventory
Discovery
Publish
PST
MSG
DOCX
IMG
TXT
PNG

FILE TABLE	
Identifier	varchar(255)
Location	\\filerepo\EDDS\GUID\ABC.msg
InRepository	0
Size	bigint
Details	xml

13 Processing Set Details layout (preview)

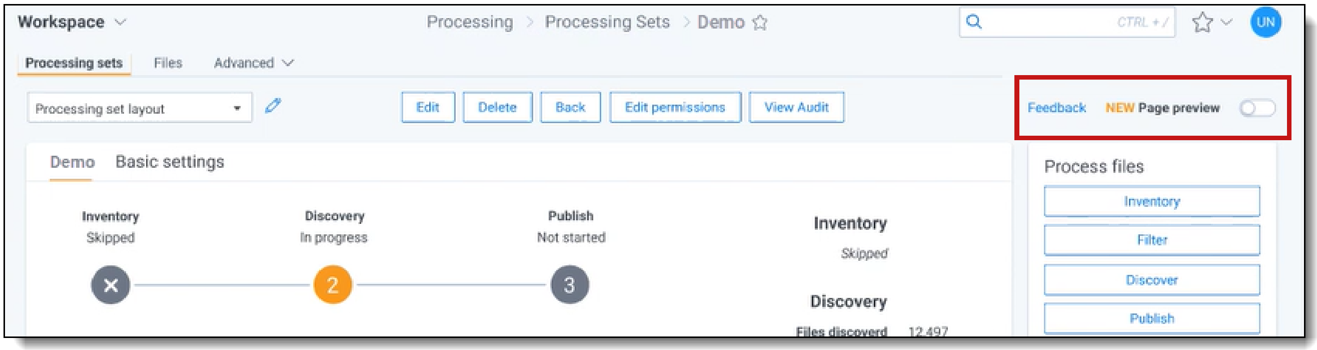
We welcome you to preview the new layout and features of the next-generation Processing Set Details page! The new layout will launch to all users in spring 2024 and features dynamic progress bars and annotations that take a deeper dive into processing job and sub-job progression.

13.1 What’s new?

Based on customer feedback, we have redesigned the Processing Set Details page so that you can see the progression of each processing job and associated sub-jobs. The three processing phases (Inventory, Discover, Publish) appear independently and include sub-job details to give you a clearer picture of the job's status. Each section has a dashboard with information on the number of files processed and file and job errors encountered for that phase.

13.2 Accessing the preview page

To access the preview page, toggle on the **Page preview** button near the upper right corner of your screen. Click **Turn on** when you see the confirmation modal.

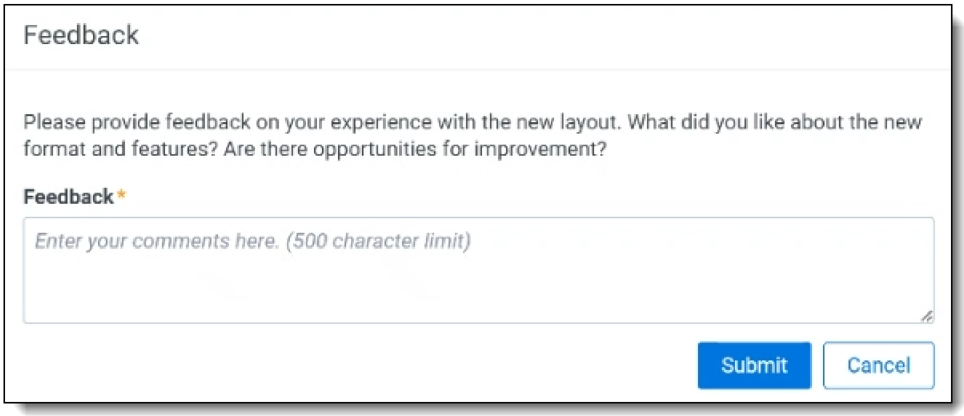

Workspace
Processing > Processing Sets > Demo
CTRL + /
UN
Processing sets
Files
Advanced
Processing set layout
Edit
Delete
Back
Edit permissions
View Audit
Feedback
NEW Page preview
Demo
Basic settings
Inventory
Skipped
Discovery
In progress
Publish
Not started
2
3
Inventory
Skipped
Discovery
Files discoverd 12,497
Process files
Inventory
Filter
Discover
Publish

13.2.1 Turning off the preview page

You can leave the preview page anytime and return to the classic layout. To leave the preview page, toggle off the **Page preview** button. A confirmation modal appears where you can leave feedback on your experience. Click **Turn off** to exit the preview layout.

13.2.2 Providing feedback

During the preview phase, we invite you to provide feedback on your experience with the new layout. You can submit feedback at any time. Click the **Feedback** link next to the **Page preview** toggle to submit your comments and questions.


Submit
Cancel

Feedback

Please provide feedback on your experience with the new layout. What did you like about the new format and features? Are there opportunities for improvement?

Feedback*

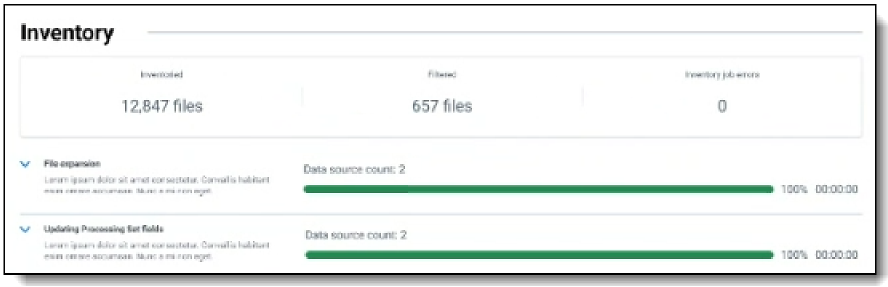
Enter your comments here. (500 character limit)

13.3 Viewing job details

Each processing phase (Inventory, Discover, and Publish) displays the progress of their associated sub-jobs. Each sub-job has a progress bar that progresses in real-time as the job moves to completion. In addition, each phase has a dashboard that displays file counts for the stage, along with file and job errors encountered.

The processing phases and their sub-jobs include:

13.3.1 Inventory


Inventory
Inventoried
12,847 files
Filtered
657 files
Inventory job errors
0
File expansion
Data source count: 2
100% 00:00:00
Updating Processing Set fields
Data source count: 2
100% 00:00:00

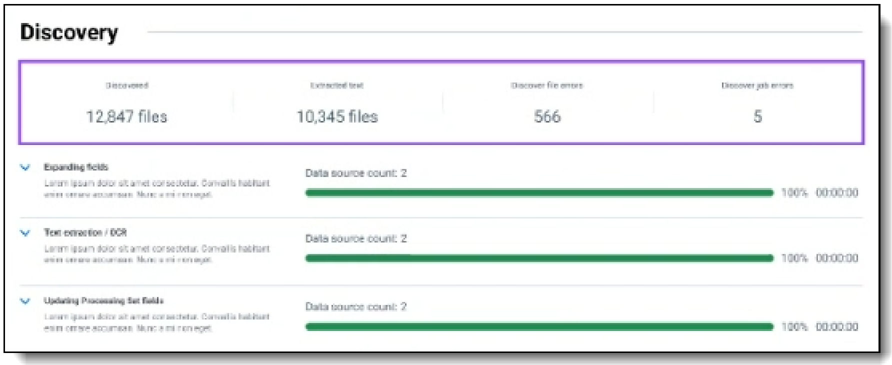
- **File expansion**—Relativity expands the data sources to locate files during this time.
- **Updating processing set fields**—Relativity updates associated fields in the workspace with the job results.

  Data source count—the number of data sources scheduled for inventory.

Dashboard numbers

- **Inventoried**—the number of files that Relativity has inventoried.
- **File errors**—the number of file errors encountered.
- **Inventory job errors**—the number of job errors encountered.

13.3.2 Discover



Discovery

Discovered: 12,847 files
Extracted text: 10,345 files
Discover file errors: 566
Discover job errors: 5

Expanding fields
Data source count: 2
100% 00:00:00

Text extraction / OCR
Data source count: 2
100% 00:00:00

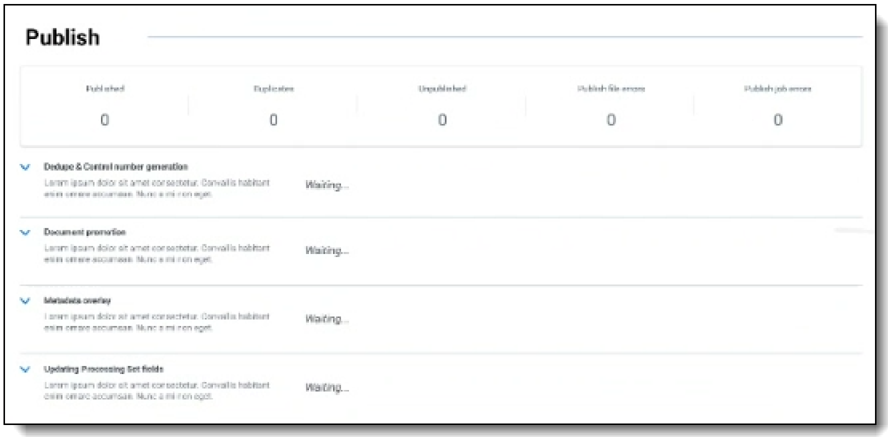
Updating Processing Set fields
Data source count: 2
100% 00:00:00

- **Expanding files**—during this time, Relativity expands the data sources to locate files and then extracts 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 Discover job time.
- **Updating processing set fields**—during this time, Relativity updates associated fields in the workspace with the job results.

Dashboard numbers

- **Discovered**—the number of non-container files that Relativity extracts from the data source.
- **Extracted files**—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.
- **Discover file errors**—the number of files with errors.
- **Discover job errors**—the number of job errors.

13.3.3 Publish



Publish

Published: 0
Duplicates: 0
Unpublished: 0
Publish file errors: 0
Publish job errors: 0

Dedupe & Control number generation
Waiting...

Document promotion
Waiting...

Metadata overlay
Waiting...

Updating Processing Set fields
Waiting...

- **Dedup & control number generation**—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**—primary documents and their families are published to the Documents tab.
- **Metadata overlay**—information (such as dedupe custodians and depupe paths) is overlayed from duplicates to their primary documents in the workspace.
- **Updating processing set fields**—during this time, Relativity updates associated fields in the workspace with the job results.

Dashboard numbers

- **Published**—the number of files published.
- **Duplicates**—the number of duplicate files encountered.
- **Unpublished**—the number of files that have not been published.
- **Publish file errors**—the number of files with errors.
- **Publish job errors**—the number of jobs with errors.

13.3.4 Resources

Visit these topics for additional information on the processing phases and errors.

- Inventory
- Discovering files
- Publishing files
- Processing error overview

14 Inventory overview

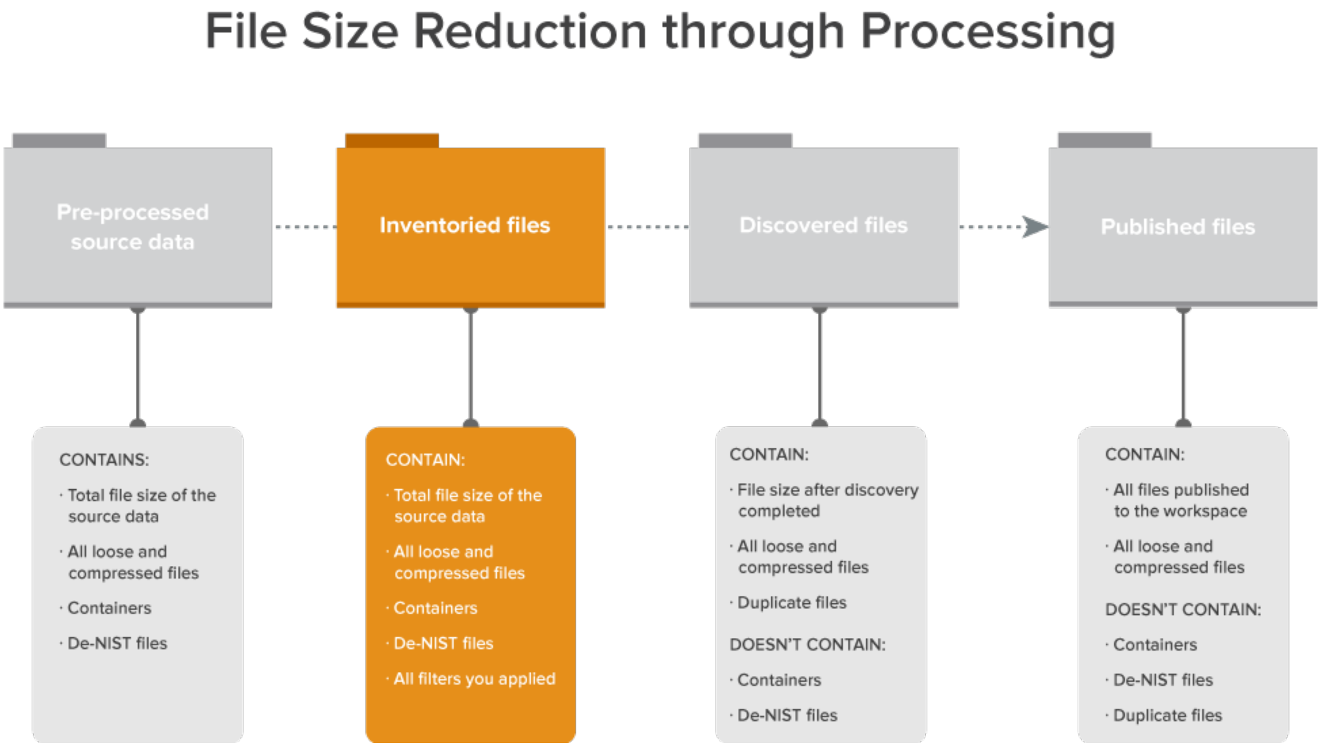
Use Inventory to narrow down your files before discovering them by eliminating irrelevant raw data from the discovery process through a variety of preliminary filters. With inventory you can exclude certain file types, file locations, file sizes, NIST files, date ranges, and sender domains. Doing this gives you a less-cluttered data set when you begin to discover your files.

Note: Your environment has been enabled to dynamically scale your Invariant worker servers dependent on load. Sustained activity is automatically detected by the system, and Relativity will add workers to handle this work. Once the work is done, they will automatically scale back down. This feature is continually being improved to be smarter about when we add workers and how many we add.

This page provides an overview of the inventory process. For information on running an inventory job, applying filters, and managing inventory errors, see Inventory processing.

Note: Your environment has been enabled to dynamically scale your Invariant worker servers dependent on load. Sustained activity is automatically detected by the system, and Relativity will add workers to handle this work. Once the work is done, they will automatically scale back down. This feature is continually being improved to be smarter about when we add workers and how many we add.

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 de-NIST and deduplication.


File Size Reduction through Processing
Pre-processed source data
Inventoried files
Discovered files
Published files
CONTAINS:
· Total file size of the source data
· All loose and compressed files
· Containers
· De-NIST files
CONTAIN:
· Total file size of the source data
· All loose and compressed files
· Containers
· De-NIST files
· All filters you applied
CONTAIN:
· File size after discovery completed
· All loose and compressed files
· Duplicate files
DOESN'T CONTAIN:
· Containers
· De-NIST files
CONTAIN:
· All files published to the workspace
· All loose and compressed files
DOESN'T CONTAIN:
· Containers
· De-NIST files
· Duplicate files

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.

Read an inventory scenario

Using Inventory and file filters

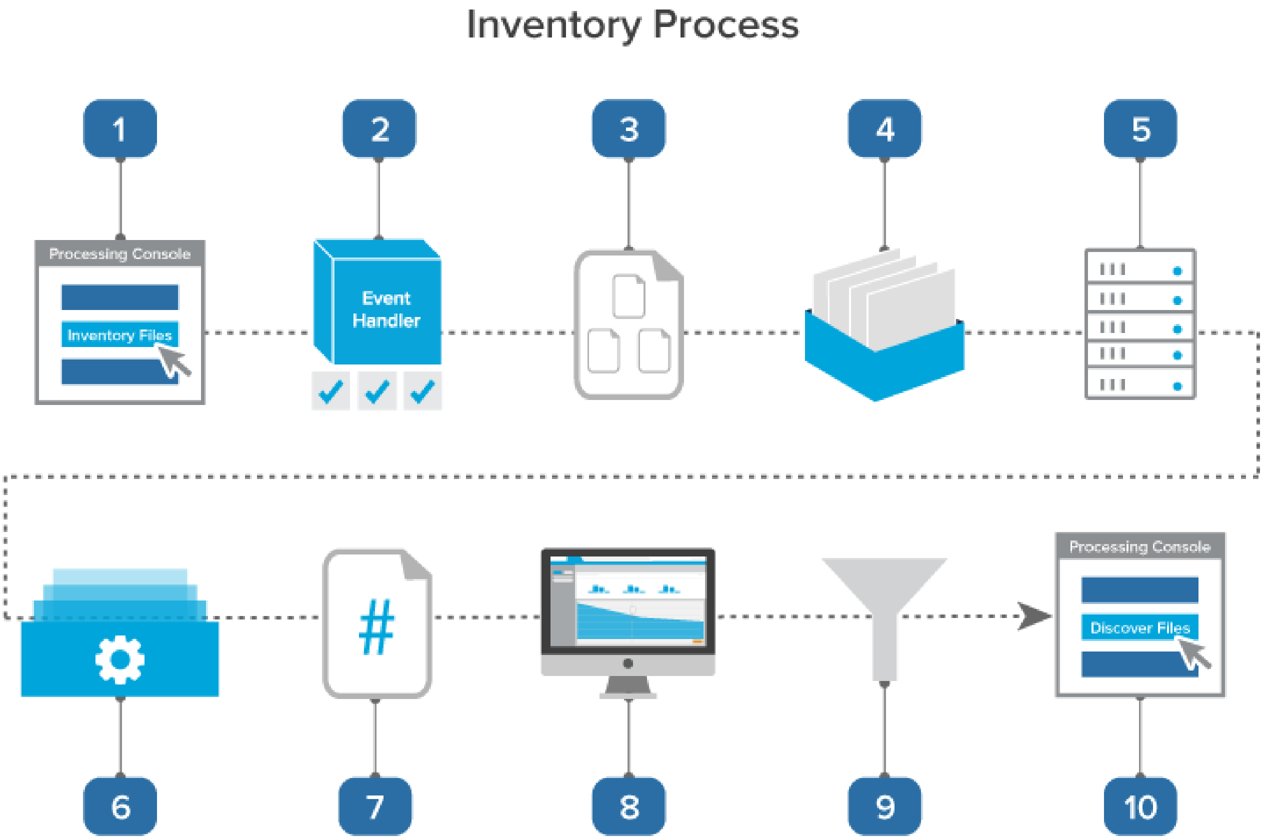
You are a project manager, and your firm requests that you create a processing set that includes a purposefully large data set from a custodian, with loose files and multiple email PST files. They then want you to eliminate all emails from a repository created in 2012, because those pre-date the case and are not admissible.

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.

14.1 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.

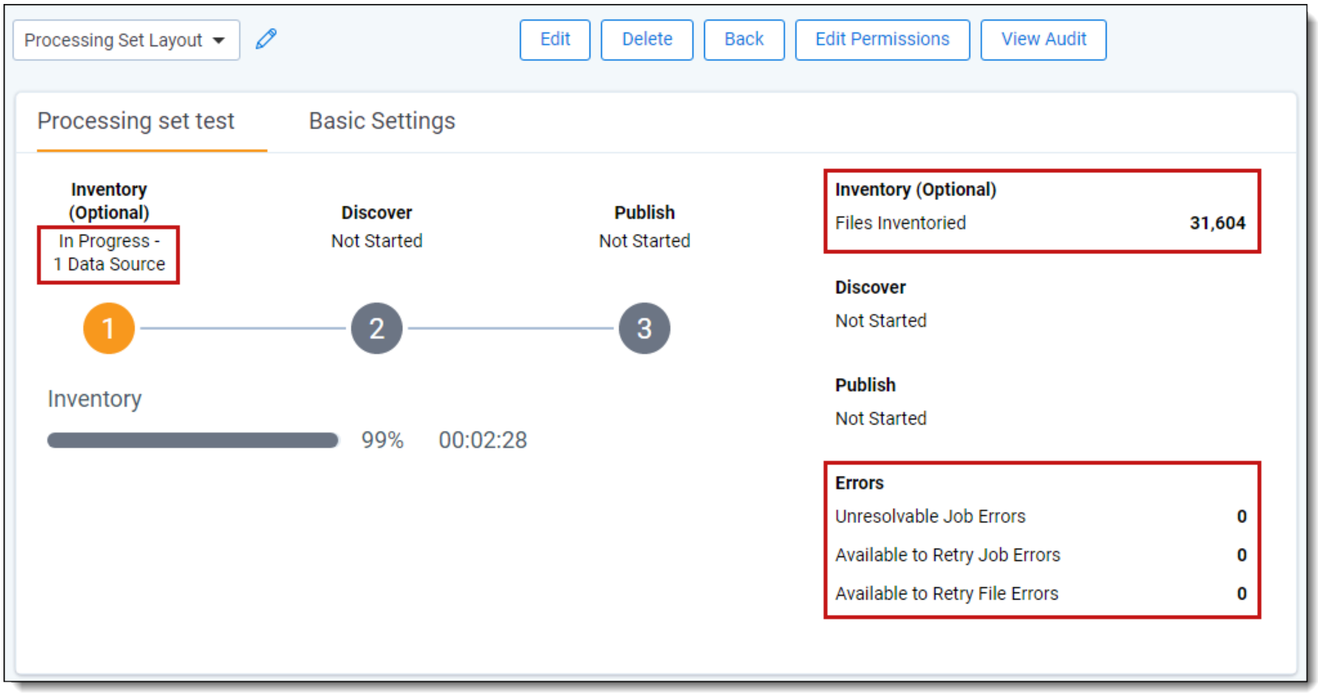


1. You click Inventory Files on the processing console.
2. A console event handler checks to make sure the processing set is valid and ready to proceed.
3. The event handler inserts the data sources to the processing queue.
4. The data sources wait in the queue to be picked up by an agent, during which 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.
6. 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.
7. The processing engine sends a count of all inventoried files to Relativity.
8. 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.

9. You apply the desired filters to your inventoried files to further narrow down the data set.
10. Once you've applied all desired filters, you move on to discovery.

14.2 Monitoring inventory status

You can monitor the progress of the inventory job through the information provided in the Processing Set Status display on the set layout.



Inventory (Optional)
In Progress -
1 Data Source

Discover
Not Started

Publish
Not Started

Inventory
99% 00:02:28

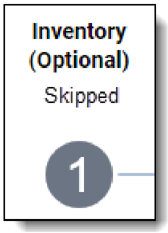
Inventory (Optional)	
Files Inventoried	**31,604**
Discover	
Not Started	
Publish	
Not Started	
Errors	
Unresolvable Job Errors	**0**
Available to Retry Job Errors	**0**
Available to Retry File Errors	**0**

Through this display, you can monitor the following:

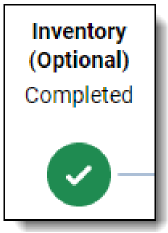
- **# of Data Sources**—the number of data sources currently in the processing queue.
- **Inventory | Files Inventoried**—the number of files across all data sources submitted that the processing engine has inventoried.
- **Errors**—the number of errors that have occurred across all data sources submitted, which fall into the following categories:
  - **Unresolvable Job Errors**—errors that you cannot retry.
  - **Available to Retry Job Errors**—job errors that are available for retry.
  - **Available to Retry File Errors**—file errors that are available for retry.

See Processing error resolution for details.

If you skip inventory, the status section displays a Skipped status throughout the life of the processing set.

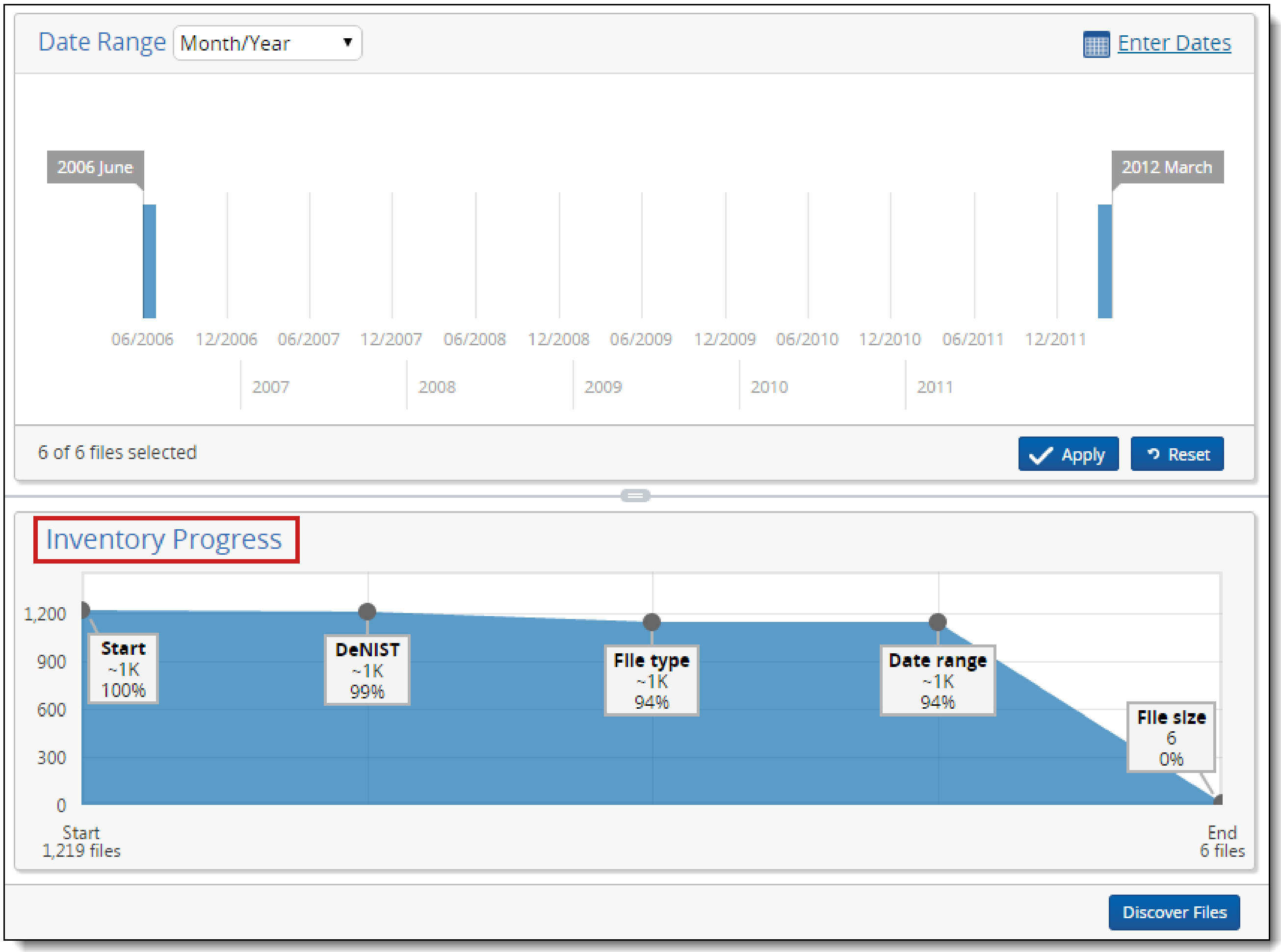


Once inventory is complete, the status section displays a Complete status, indicating that you can move on to either filtering or discovering your files. For more information, see Filtering files and Discovering files.



14.3 Inventory progress

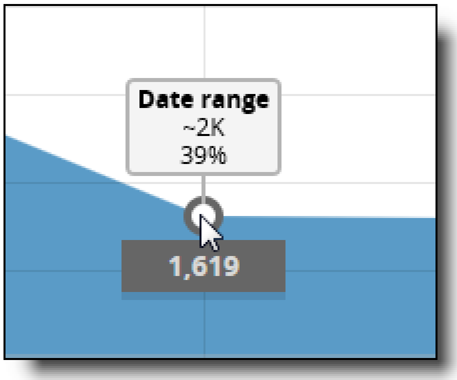
The graph in the Inventory Progress pane reflects all the filters you have applied to the processing set. This graph updates automatically as the inventory job progresses, and provides information on up to six different filters. The vertical axis contains the number of files. The horizontal axis contains the filters.


Date Range
Month/Year
Enter Dates
2006 June
2012 March
06/2006
12/2006
06/2007
12/2007
06/2008
12/2008
06/2009
12/2009
06/2010
12/2010
06/2011
12/2011
2007
2008
2009
2010
2011
6 of 6 files selected
Apply
Reset
Inventory Progress
1,200
900
600
300
0
Start
~1K
100%
DeNIST
~1K
99%
File type
~1K
94%
Date range
~1K
94%
File size
6
0%
Start
1,219 files
End
6 files
Discover Files

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.


Date range
~2K
39%
1,619

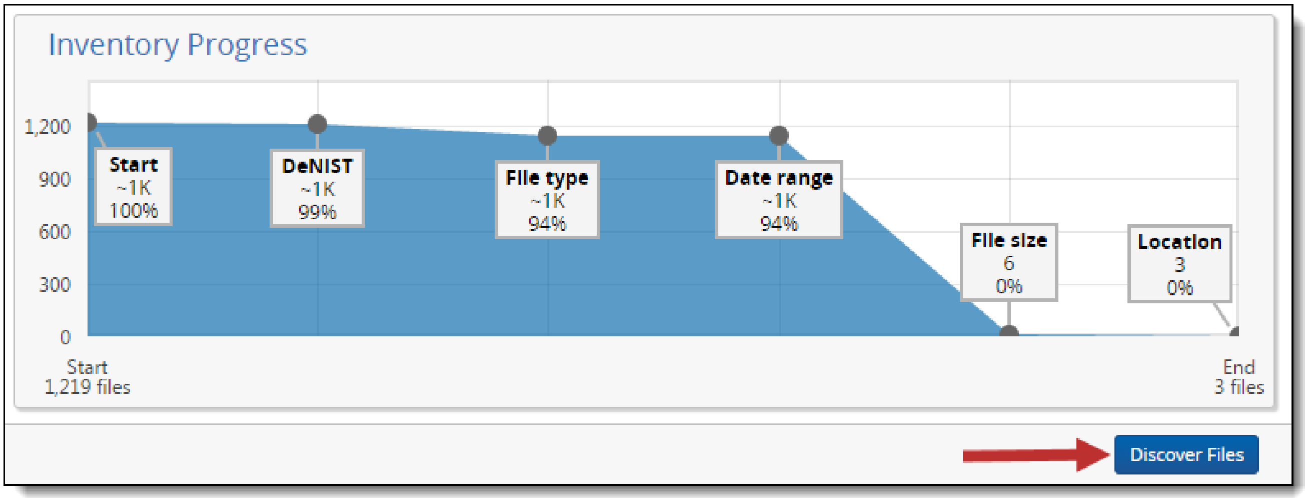
At any time before you discover the files reflected in the Inventory Progress pane, you can reset or delete any filters you already applied.

Once you determine that the filters you have applied have reduced the data set appropriately, you can discover the remaining files.

14.4 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.


Inventory Progress
1,200
900
600
300
0
Start
~1K
100%
DeNIST
~1K
99%
File type
~1K
94%
Date range
~1K
94%
File size
6
0%
Location
3
0%
Start
1,219 files
End
3 files
Discover Files

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.

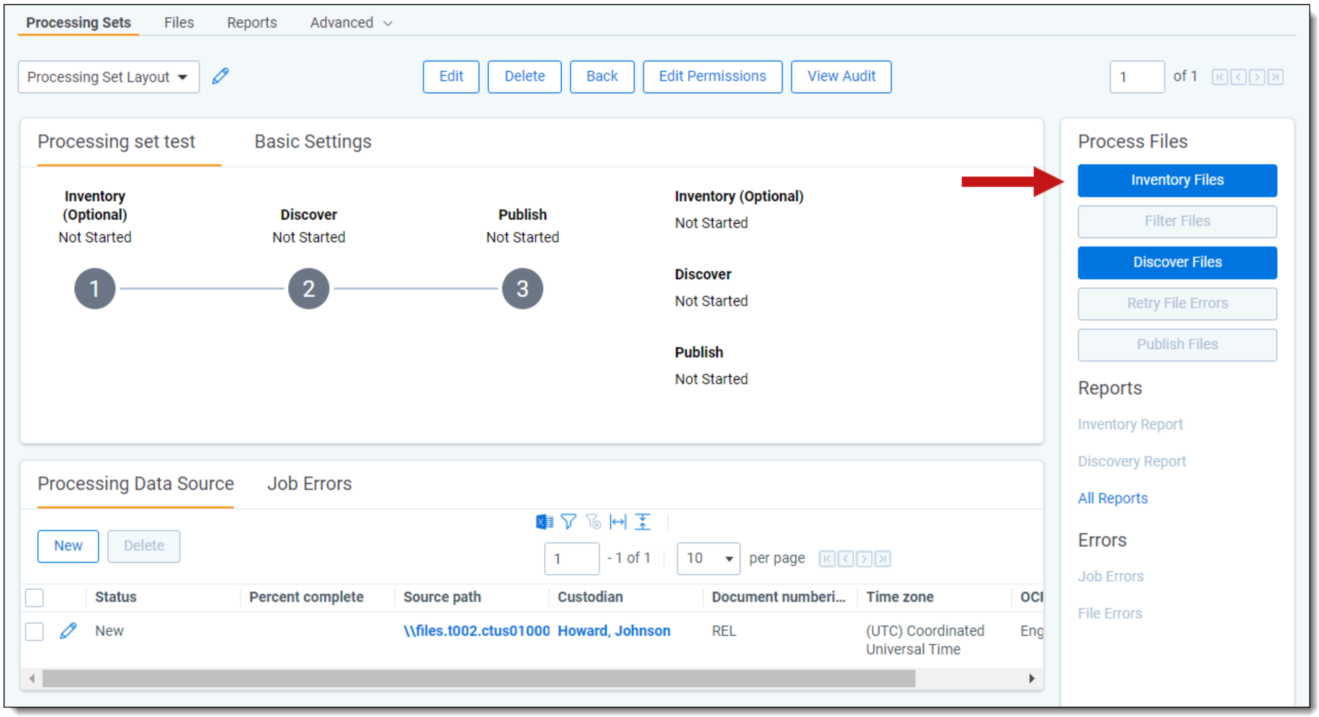
The same validations that apply when you start discovery from the processing set layout apply when discovering from the Inventory tab.

15 Inventory processing

This page provides details on inventory processing, applying filters, and managing inventory errors. For an overview of the inventory process, see Inventory overview.

15.1 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
Files
Reports
Advanced
Processing Set Layout
Edit
Delete
Back
Edit Permissions
View Audit
1
of 1
Processing set test
Basic Settings
Inventory (Optional)
Discover
Publish
1
2
3
Process Files
Inventory Files
Filter Files
Discover Files
Retry File Errors
Publish Files
Reports
Inventory Report
Discovery Report
All Reports
Errors
Job Errors
File Errors
Processing Data Source
Job Errors
New
Delete
1
- 1 of 1
10
per page

Inventory (Optional)
Not Started

Discover
Not Started

Publish
Not Started

☐		Status	Percent complete	Source path	Custodian	Document numberi...	Time zone	OCI
☐		New		\\files.t002.ctus01000	Howard, Johnson	REL	(UTC) Coordinated Universal Time	Eng

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.

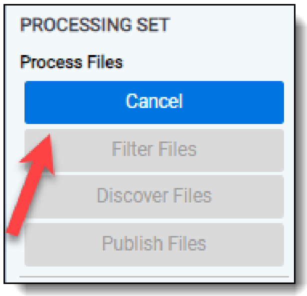
Note: The processing set manager agent sends the password bank to the processing engine when you start inventory. If a custodian is associated with a Lotus Notes password bank, that custodian's information is sent with the inventory job.

You can re-inventory files any time after the previous inventory job is complete. For more information, see Re-inventory.

15.1.1 Canceling inventory

If the need arises, you can cancel inventory before the job encounters its first error or before it is complete.

To cancel discovery, click **Cancel**.


PROCESSING SET
Process Files
Cancel
Filter Files
Discover Files
Publish Files

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.

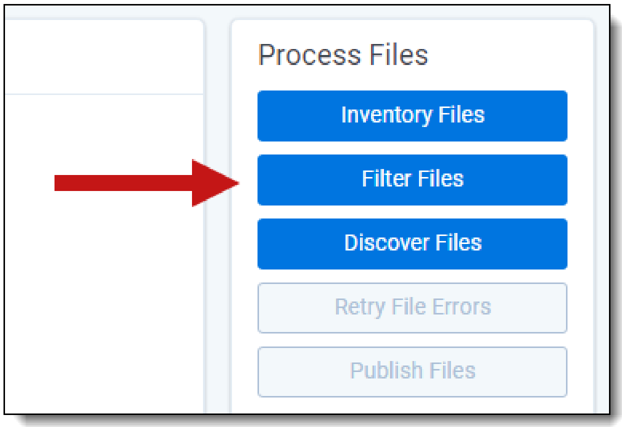
15.2 Filtering files

When inventory is complete you have the option of filtering your files in the Inventory tab before moving to discovery.

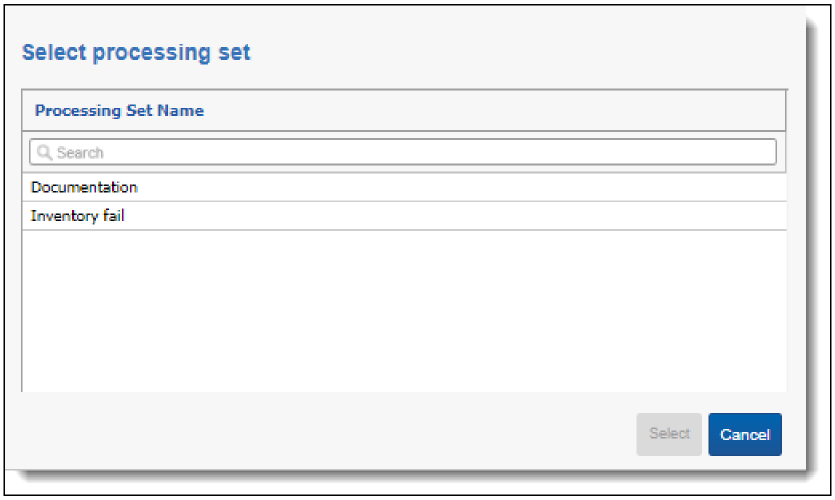
Note that Relativity only filters on the files that you have inventoried. Everything else that cascades down from the files that were discovered is not subject to the inventory filters that you set.

Note: Filtering is not supported for Relativity Short Message Format (RSMF) metadata.

To do this, click **Filter Files** on the console.


Process Files
Inventory Files
Filter Files
Discover Files
Retry File Errors
Publish Files

After clicking the Filter Files button, you are redirected to the Inventory tab. You see the current Processing set name, along with an option to change the set. If you navigated to the Inventory tab from another location in Relativity, you see a list of Processing sets that are eligible for filtering. Click a set and click **Select** to load the set on the Inventory layout.



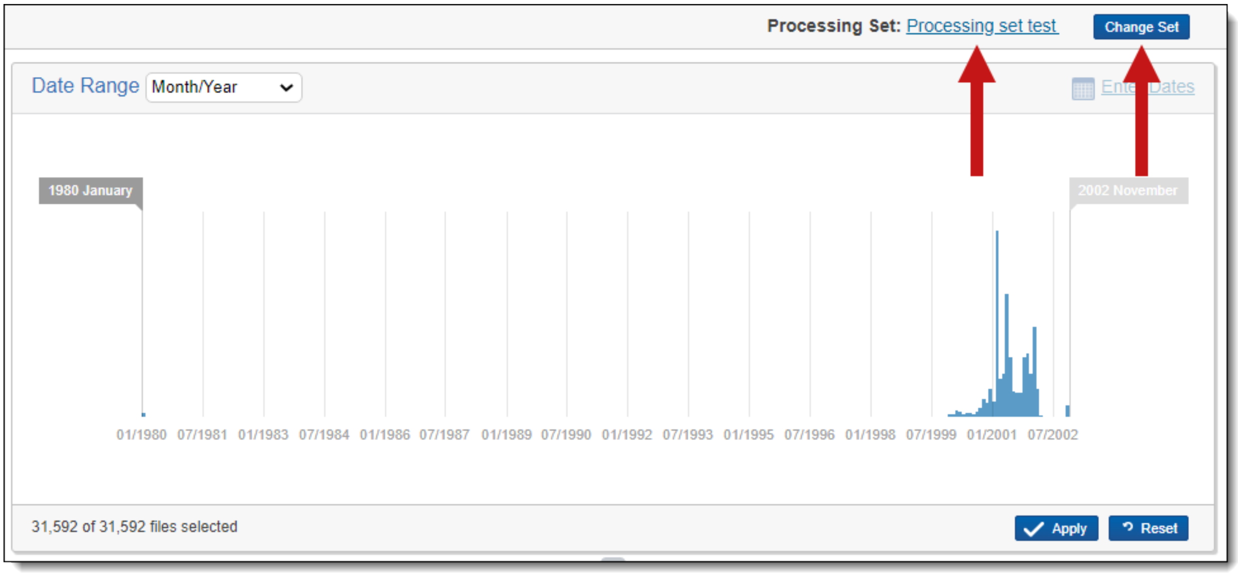
A processing set is not eligible for use in the Inventory tab if:

- 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.

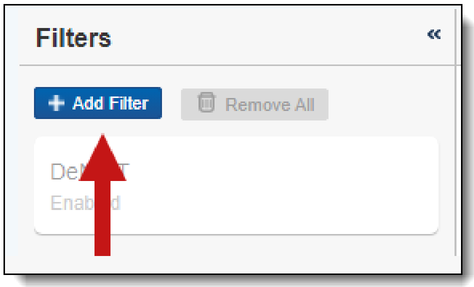

Processing Set: Processing set test
Change Set
Date Range
Month/Year
Enter Dates
1980 January
2002 November
01/1980
07/1981
01/1983
07/1984
01/1986
07/1987
01/1989
07/1990
01/1992
07/1993
01/1995
07/1996
01/1998
07/1999
01/2001
07/2002
31,592 of 31,592 files selected
Apply
Reset

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.

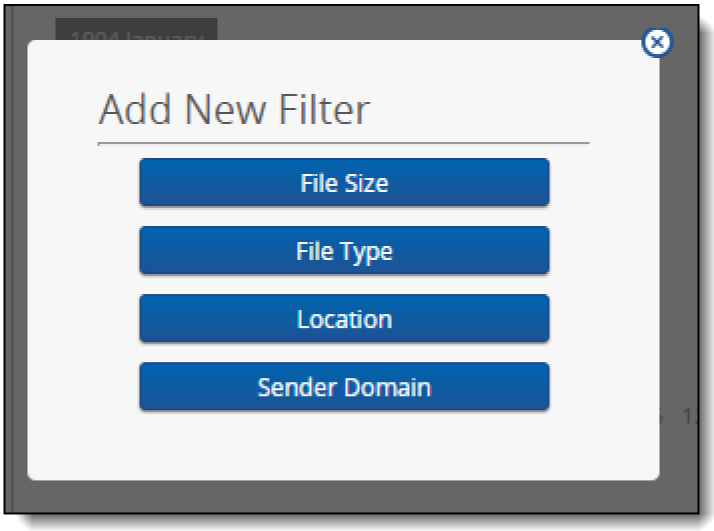
There are six different filters you can apply to a processing set. You can apply these filters in any order; however, you can only apply one filter of each type. This section describes how to apply file location, file size, file type, and sender domain filters. See Applying a Date range filter or Applying a DeNIST filter for instructions on filtering inventoried files by those properties.

To add a new filter, click **Add Filter**.


Filters
«
Add Filter
Remove All
DeNIST
Enabled

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:

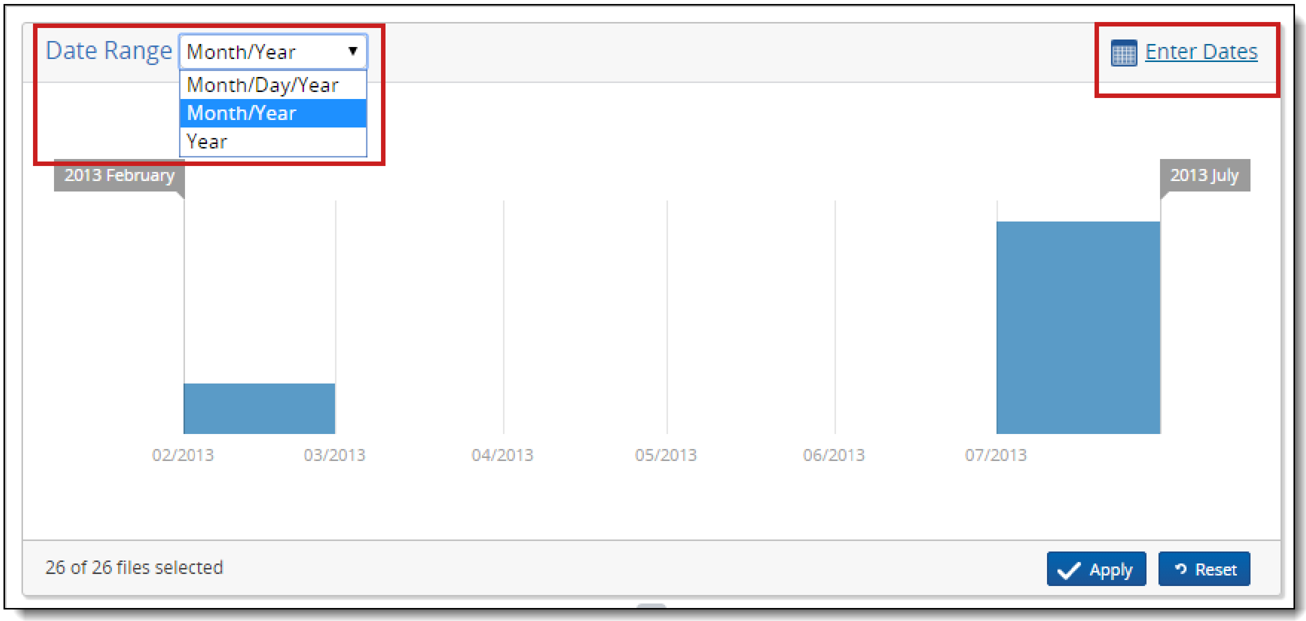
- If the applied filter conditions have excluded all files from the set, then there are no results for you to interact with and you cannot add or apply more filters.
- If a filter is already applied to the data, the corresponding button is disabled.
- The Inventory Progress graph displays the effect 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.

15.2.1 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.



26 of 26 files selected

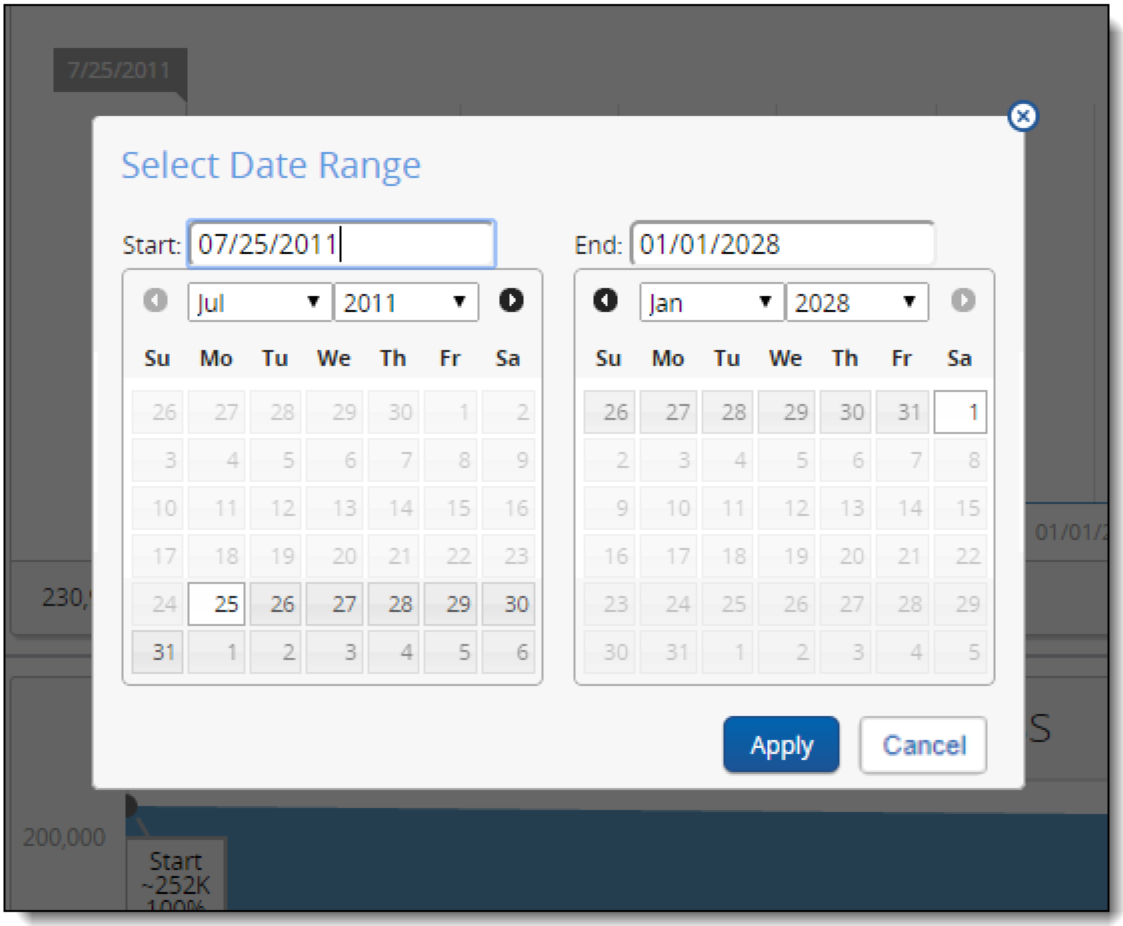
Note: When you filter for dates, you are filtering specifically on the Sort Date/Time field, which is taken from the file's Sent Date, Received Date, and Last Modified Date fields in that order of precedence. This happens on email messages repeated for the parent document and all child items to allow 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, and Primary 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.



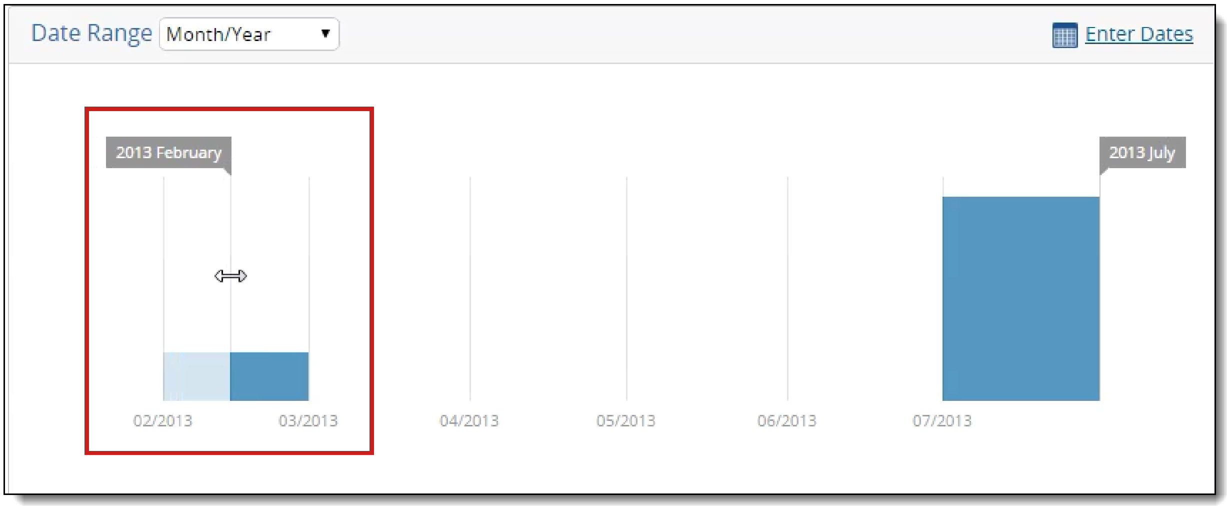
Select Date Range

Start: 07/25/2011

End: 01/01/2028

7/25/2011
Jul
2011
Jan
2028
Su Mo Tu We Th Fr Sa
Apply
Cancel
230,
01/01/2
200,000
Start
~252K

- 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
2013 February
2013 July
02/2013
03/2013
04/2013
05/2013
06/2013
07/2013

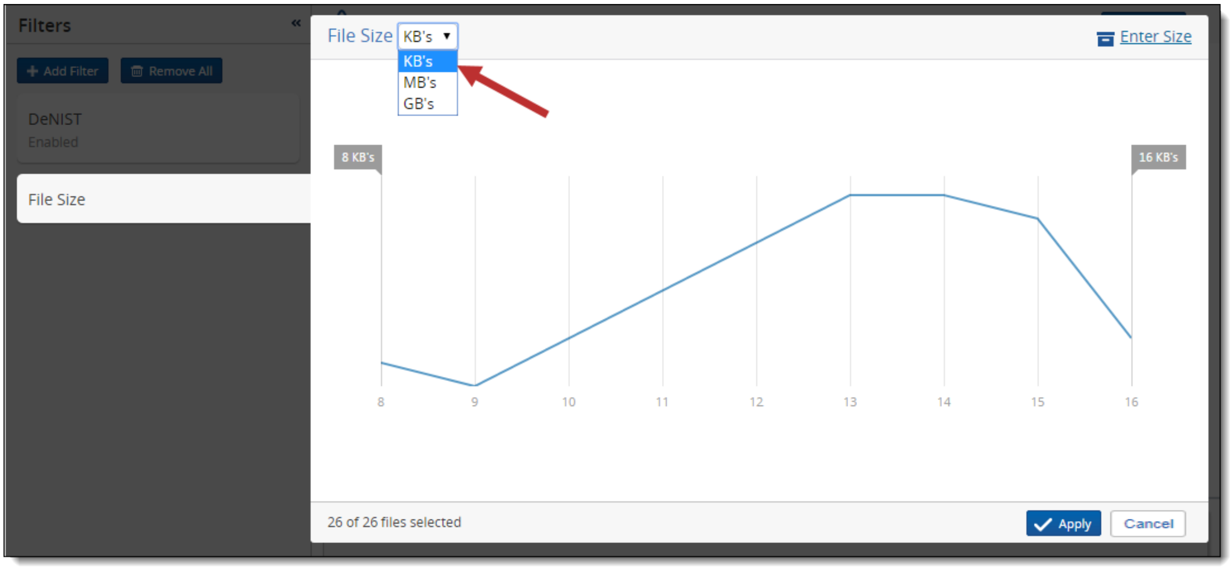
- To reset the parameters of a filter after you apply it, click **Reset**.

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.

15.2.2 Applying a File Size filter

To filter your processing set files by size:

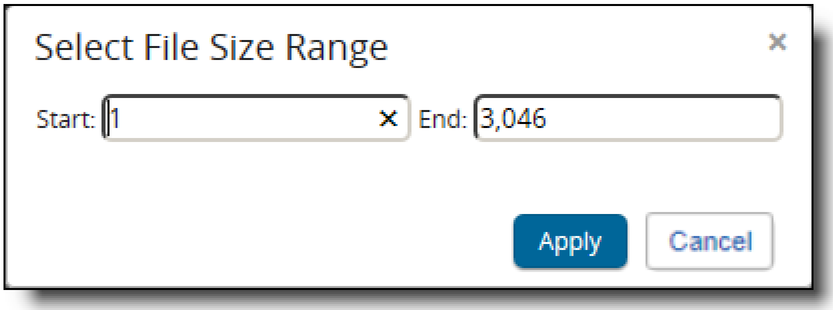
1. Click **Add Filter**.
2. Select **File Size** from the filter list.
3. Use the available options on the **File Size** range graph filter to specify how you want to apply the file size filter to your files.



26 of 26 files selected

- 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: 1

End: 3,046

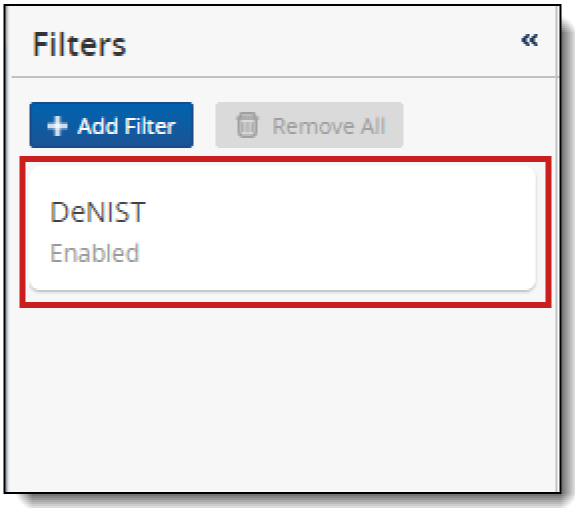
4. Click **Apply** once you have designated all the file sizes you want to exclude. The Inventory Progress pane reflects the addition of the file size filter, as well as the percentage and number of files that remain from the original data set. For more information, see Inventory progress.

Inventory reduces your processing set by the date parameters you defined. You can now apply additional filters to further reduce the data set, or you can discover the files.

15.2.3 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.


Filters

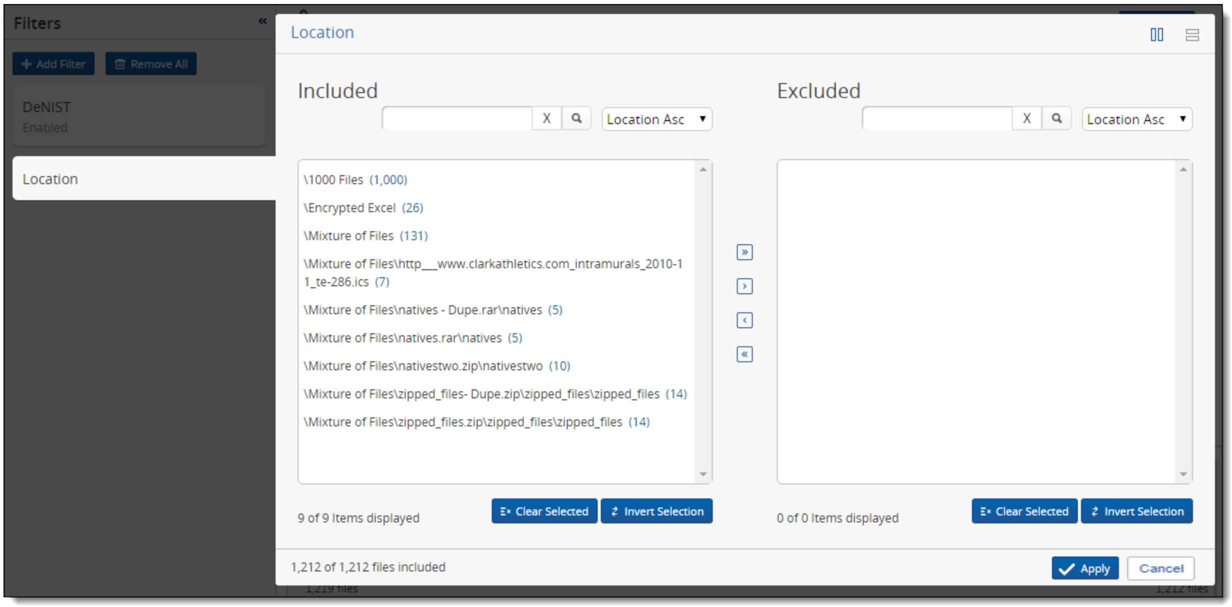
DeNIST
Enabled

15.2.4 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.


Filters
Add Filter
Remove All
DeNIST
Enabled
Location
Location
Included
Location Asc
Excluded
Location Asc
Clear Selected
Invert Selection
Clear Selected
Invert Selection
Apply
Cancel

\1000 Files (1,000)
\Encrypted Excel (26)
\Mixture of Files (131)
\Mixture of Files\http__www.clarkathletics.com_intramurals_2010-11_te-286.ics (7)
\Mixture of Files\natives - Dupe.rar\natives (5)
\Mixture of Files\natives.rar\natives (5)
\Mixture of Files\nativestwo.zip\nativestwo (10)
\Mixture of Files\zipped_files- Dupe.zip\zipped_files\zipped_files (14)
\Mixture of Files\zipped_files.zip\zipped_files\zipped_files (14)

9 of 9 Items displayed

0 of 0 Items displayed

1,212 of 1,212 files included

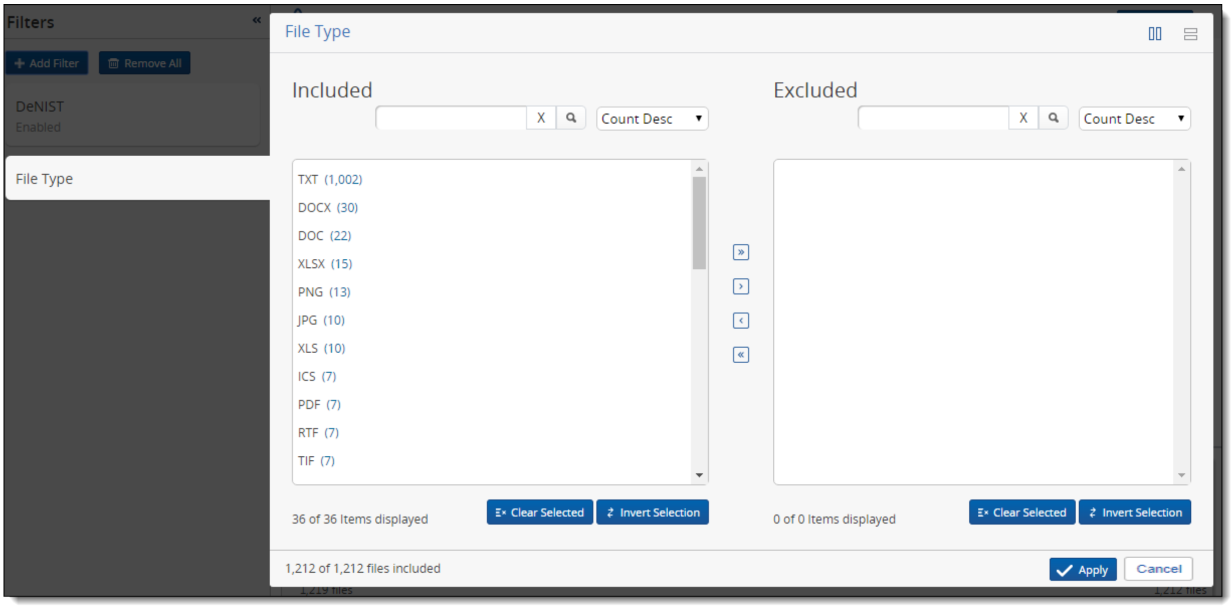
4. Click **Apply** once you have designated all the file locations you want to exclude. The Inventory Progress pane reflects the addition of the location filter, as well as the percentage and number of files that remain from the original data set. For more information, see Inventory progress.

You can now apply an additional filter to further reduce the data set, or you can discover the files.

15.2.5 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.


Filters
Add Filter
Remove All
DeNIST
Enabled
File Type
File Type
Included
X
Count Desc
TXT (1,002)
DOCX (30)
DOC (22)
XLSX (15)
PNG (13)
JPG (10)
XLS (10)
ICS (7)
PDF (7)
RTF (7)
TIF (7)
36 of 36 Items displayed
Clear Selected
Invert Selection
Excluded
X
Count Desc
0 of 0 Items displayed
Clear Selected
Invert Selection
1,212 of 1,212 files included
Apply
Cancel

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.

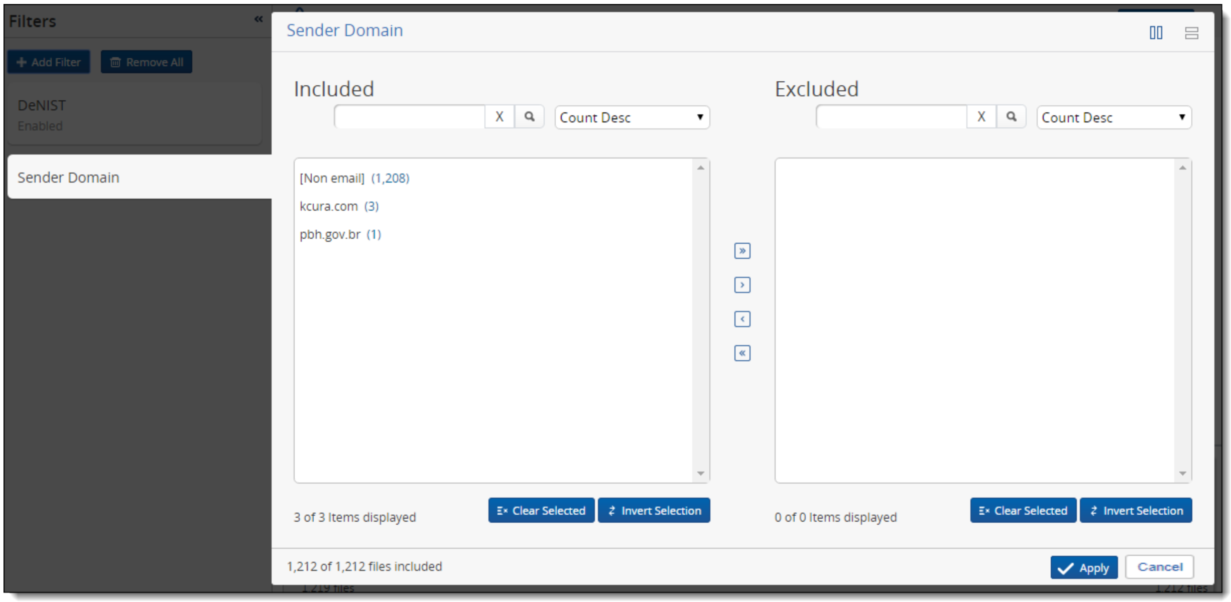
You can now apply an additional filter to further reduce the data set, or you can discover the files.

15.2.6 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.



[Non email] (1,208)
kcura.com (3)
pbh.gov.br (1)

3 of 3 Items displayed

0 of 0 Items displayed

1,212 of 1,212 files included

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.

15.2.6.1 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.

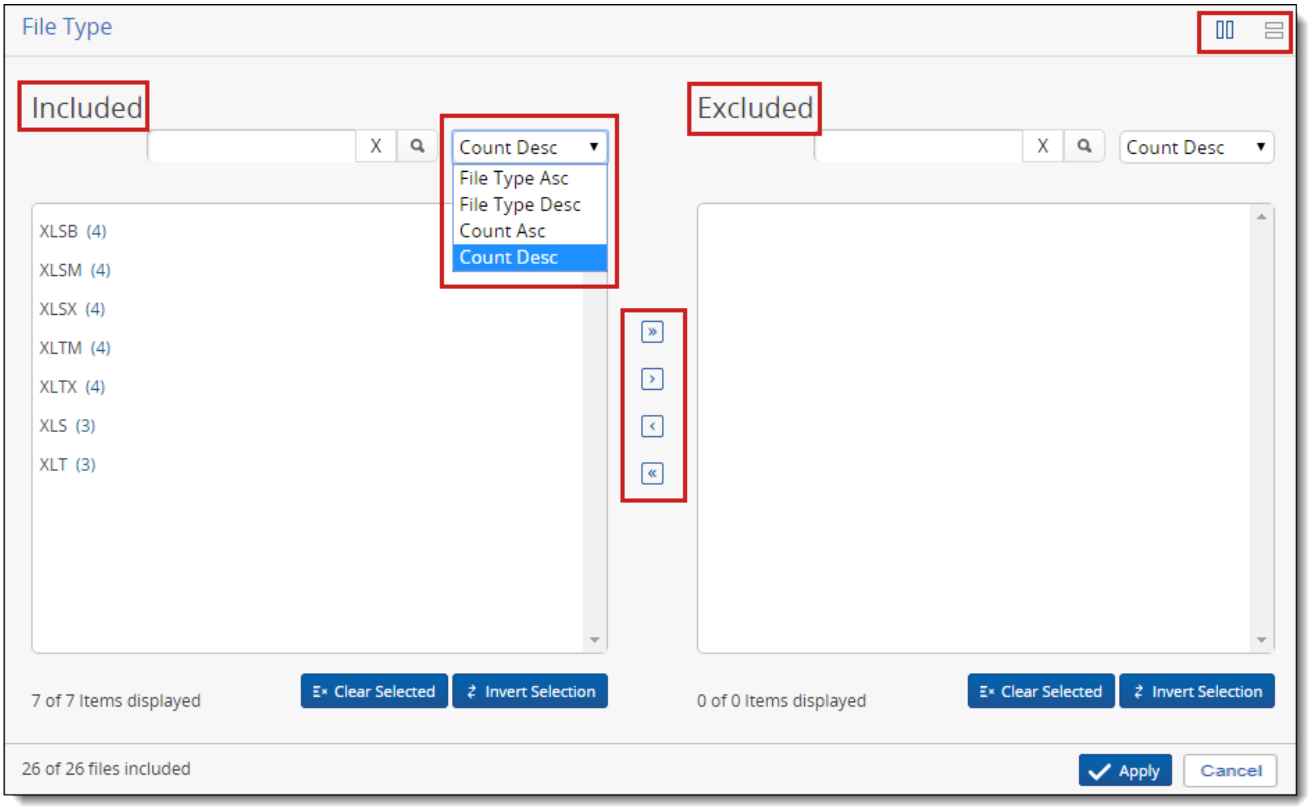
15.2.6.2 Applying a two-list filter

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).

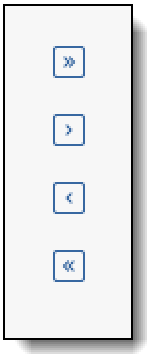
Note: If you add items from the Included list to the Excluded or vice versa, and these additions affect the sort and search criteria of the modified list, you can refresh the list to re-apply the sort and search.


File Type
Included
X
Count Desc
File Type Asc
File Type Desc
Count Asc
Count Desc
XLSB (4)
XLSM (4)
XLSX (4)
XLTM (4)
XLTX (4)
XLS (3)
XLT (3)
Excluded
X
Count Desc
7 of 7 Items displayed
Clear Selected
Invert Selection
0 of 0 Items displayed
Clear Selected
Invert Selection
26 of 26 files included
Apply
Cancel

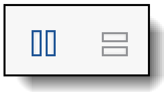
Note: Items removed from the data by edits to a previously applied filter are displayed in later filters with a value of (0) next to them. For example, if you apply the file type filter and then later narrow the date range to the extent that it filters out all files of the PDF type, then the next time 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:

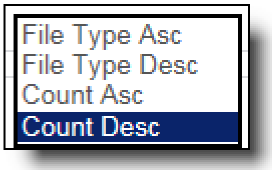
- Move over all items with double left and right arrows. Move over only the item(s) selected with the single left and right arrows.



- 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.

Clear Selected

- **Invert Selection**—marks the previously selected items in the Included or Excluded list as unselected while selecting the items that were not selected before.

Invert Selection

15.3 Removing filters

Clicking **Remove All** under Filter Controls removes all the filters from the menu on the left side of the menu.

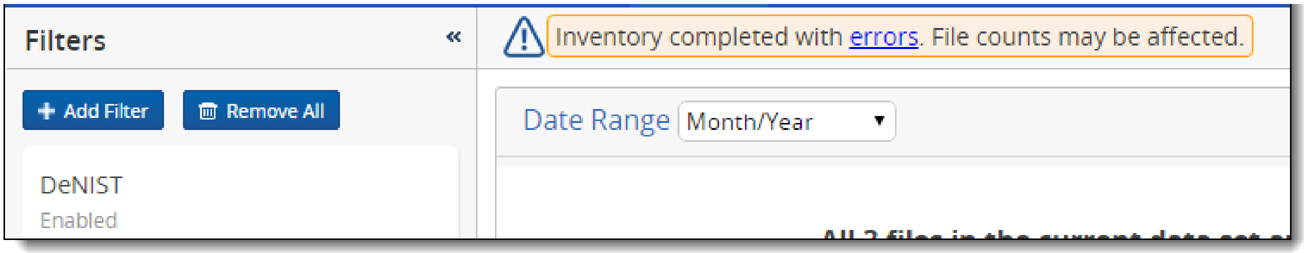
You can also remove filters individually by clicking the **X** on a single filter in the menu. You cannot delete a filter if you are currently working with it.

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

15.4 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.



Inventory completed with errors. File counts may be affected.

Date Range: Month/Year

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.

15.4.1 Inventory error scenarios

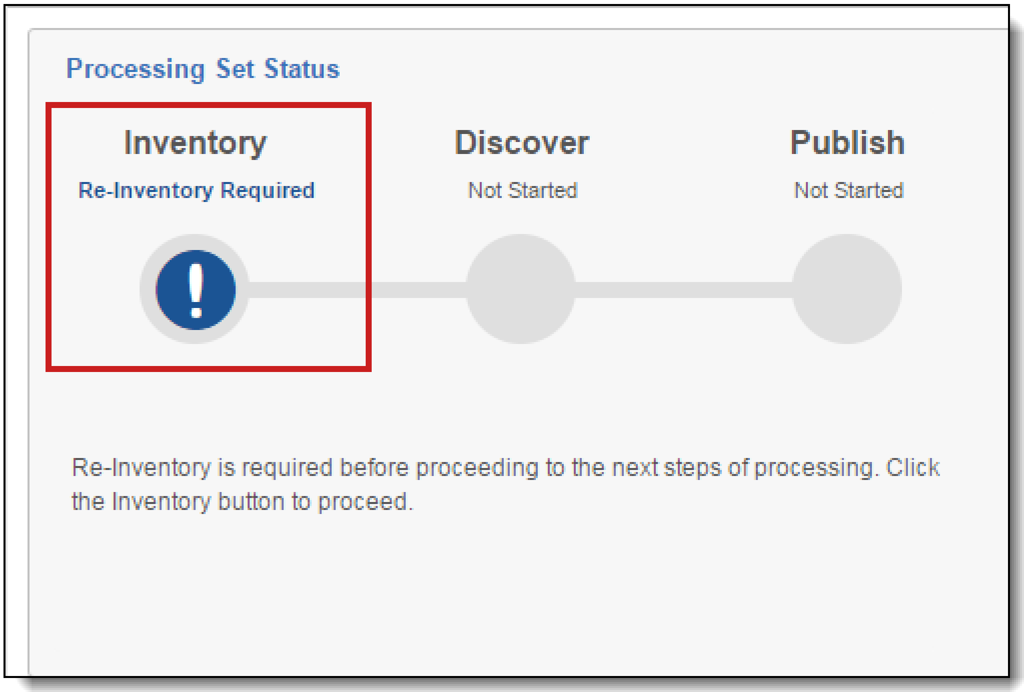
You receive an error when starting file inventory if any of the following scenarios occur:

- 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.

- There is no worker manager server associated with the workspace in which you are performing file inventory.
- The queue manager service is disabled.

15.5 Re-inventory

You may be prompted to run inventory again in the status display on the processing set layout.



Re-Inventory is required before proceeding to the next steps of processing. Click the Inventory button to proceed.

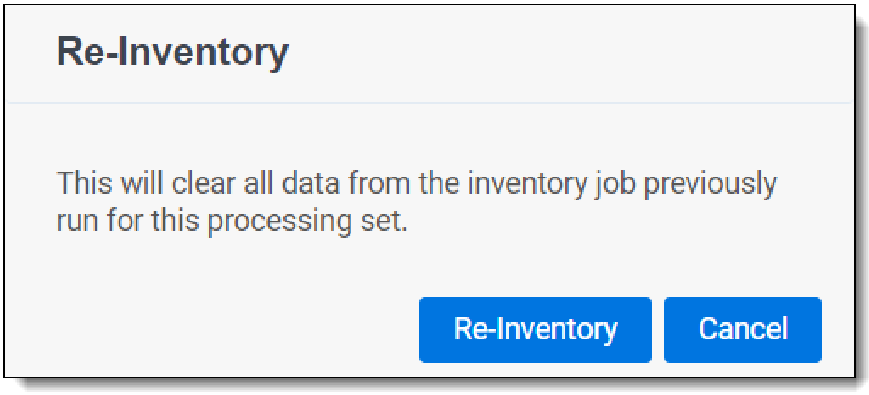
You must run inventory again on a processing set if:

- 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.

When you re-inventory files:

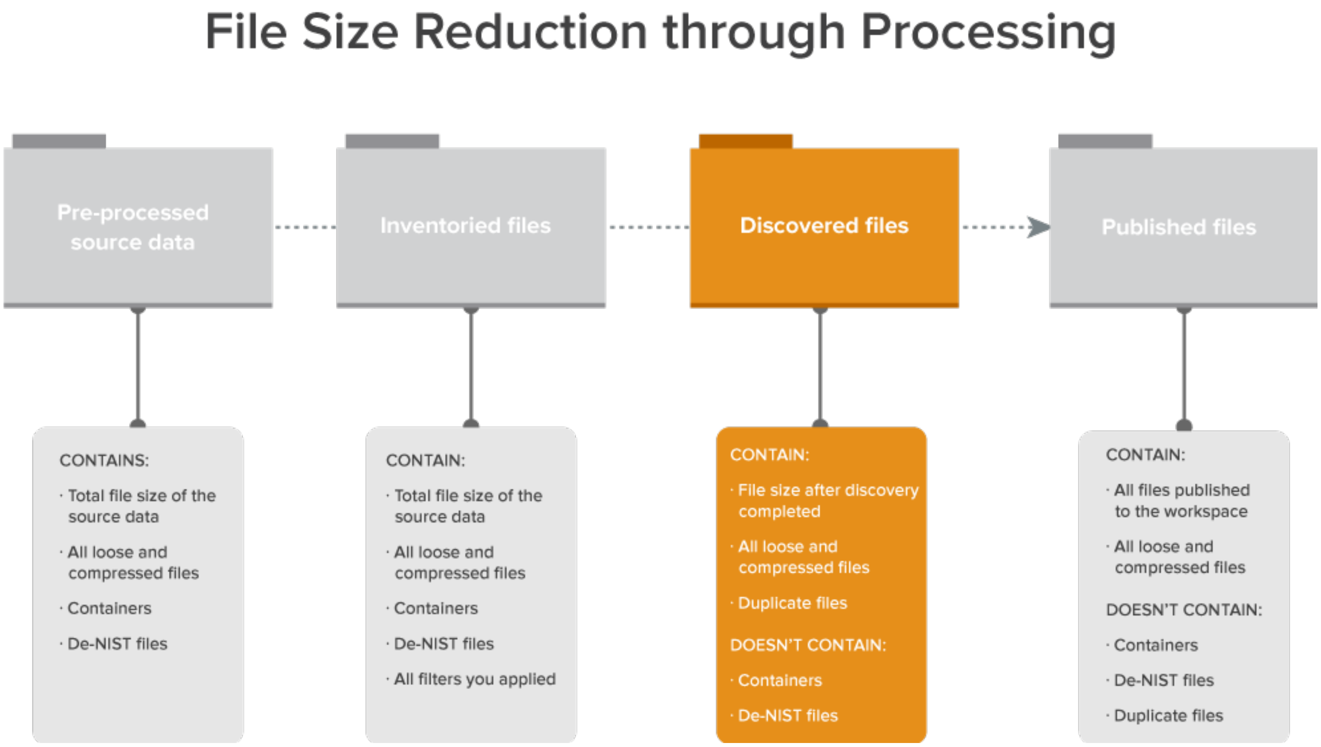
- Filters that you previously applied in the Inventory tab do not get cleared.
- Errors that you encountered in a previous Inventory job are cleared.

16 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.

Note: Your environment has been enabled to dynamically scale your Invariant worker servers dependent on load. Sustained activity is automatically detected by the system, and Relativity will add workers to handle this work. Once the work is done, they will automatically scale back down. This feature is continually being improved to be smarter about when we add workers and how many we add.

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
Pre-processed source data
Inventoried files
Discovered files
Published files
CONTAINS:
· Total file size of the source data
· All loose and compressed files
· Containers
· De-NIST files
CONTAIN:
· Total file size of the source data
· All loose and compressed files
· Containers
· De-NIST files
· All filters you applied
CONTAIN:
· File size after discovery completed
· All loose and compressed files
· Duplicate files
DOESN'T CONTAIN:
· Containers
· De-NIST files
CONTAIN:
· All files published to the workspace
· All loose and compressed files
DOESN'T CONTAIN:
· Containers
· De-NIST files
· Duplicate files

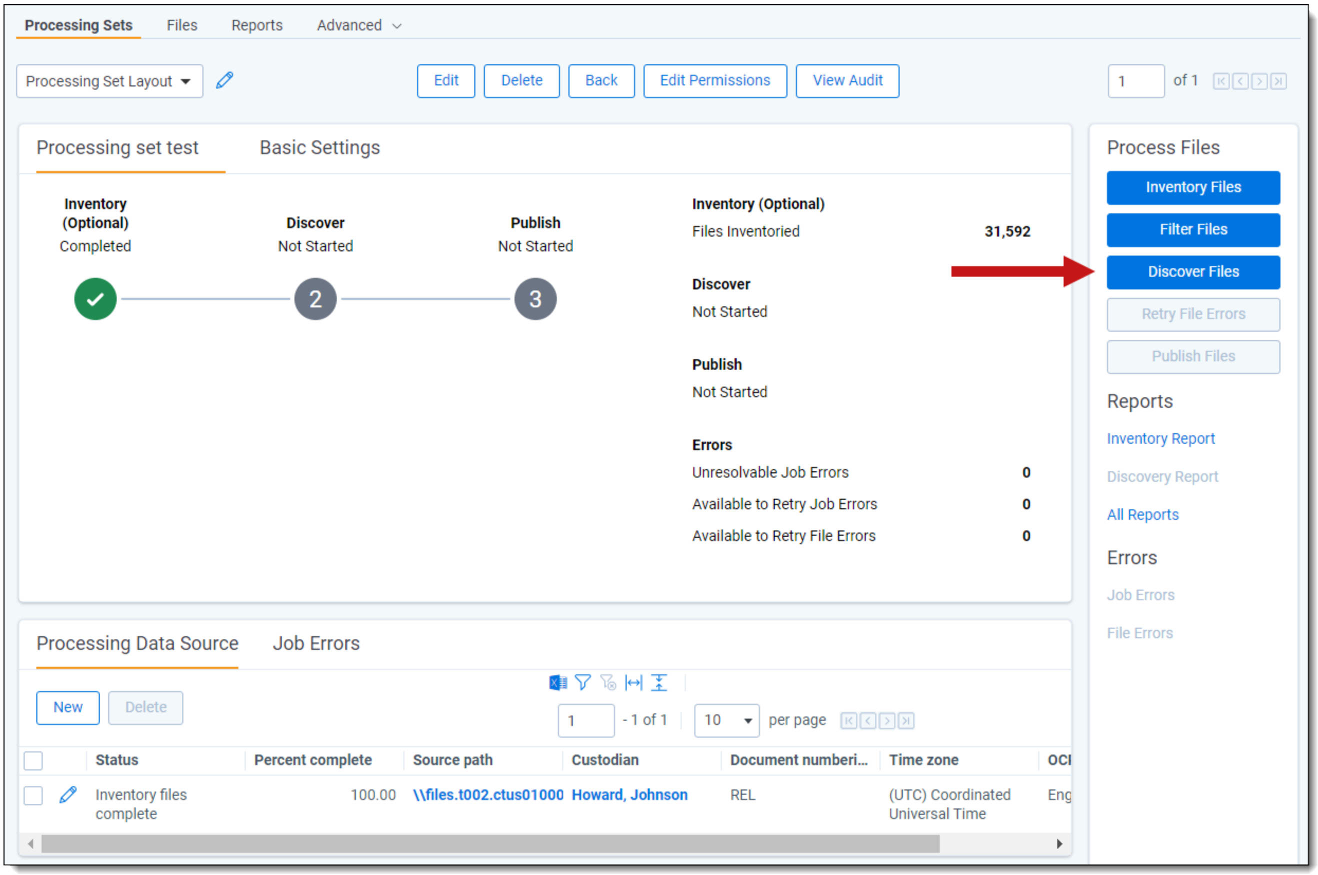
The following is a typical workflow that incorporates discovery:

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.

16.1 Running file discovery

To start discovery click **Discover Files** on the processing set console. You can click this 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, and Primary Date/Time. The null value is excluded and not represented in the filtered list.



Processing Sets
Files
Reports
Advanced
Processing Set Layout
Edit
Delete
Back
Edit Permissions
View Audit
1
of 1
Processing set test
Basic Settings
Process Files
Inventory Files
Filter Files
Discover Files
Retry File Errors
Publish Files
Reports
Inventory Report
Discovery Report
All Reports
Errors
Job Errors
File Errors
Processing Data Source
Job Errors
New
Delete
1
- 1 of 1
10
per page

Inventory (Optional)
Completed

Discover
Not Started

Publish
Not Started

Inventory (Optional)
Files Inventoried 31,592

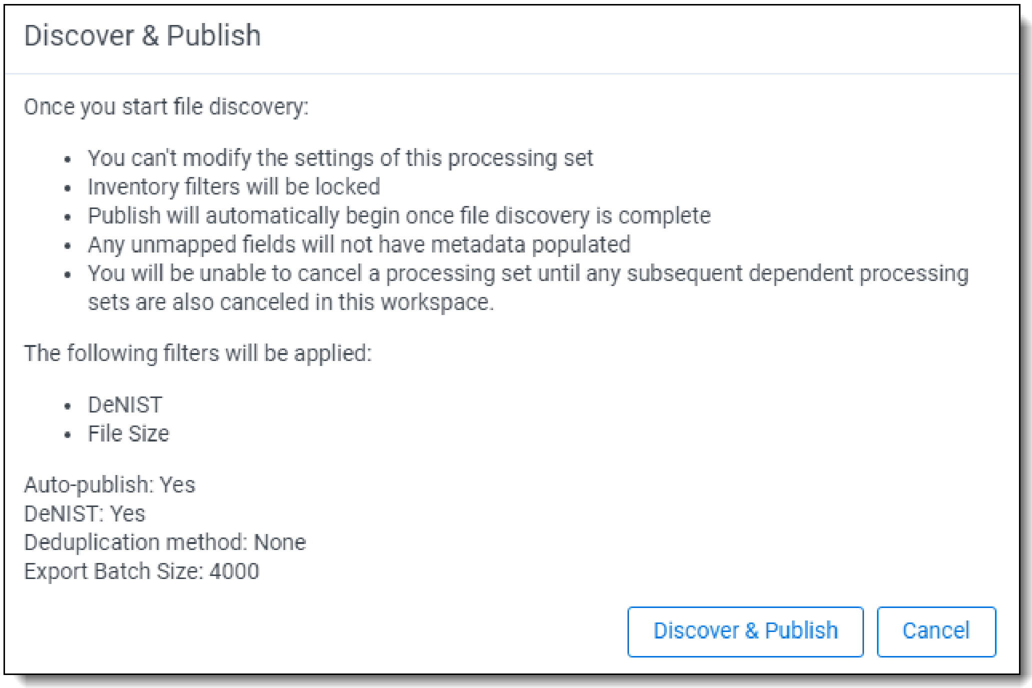
Discover
Not Started

Publish
Not Started

Errors
Unresolvable Job Errors 0
Available to Retry Job Errors 0
Available to Retry File Errors 0

	Status	Percent complete	Source path	Custodian	Document numberi...	Time zone	OCI
☐	Inventory files complete	100.00	\\files.t002.ctus01000	Howard, Johnson	REL	(UTC) Coordinated Universal Time	Eng

A confirmation message appears with the discovery settings and filters applied. Click **Discover** 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 & Publish**.



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
- File Size

Auto-publish: Yes
DeNIST: Yes
Deduplication method: None
Export Batch Size: 4000

Consider the following when discovering files:

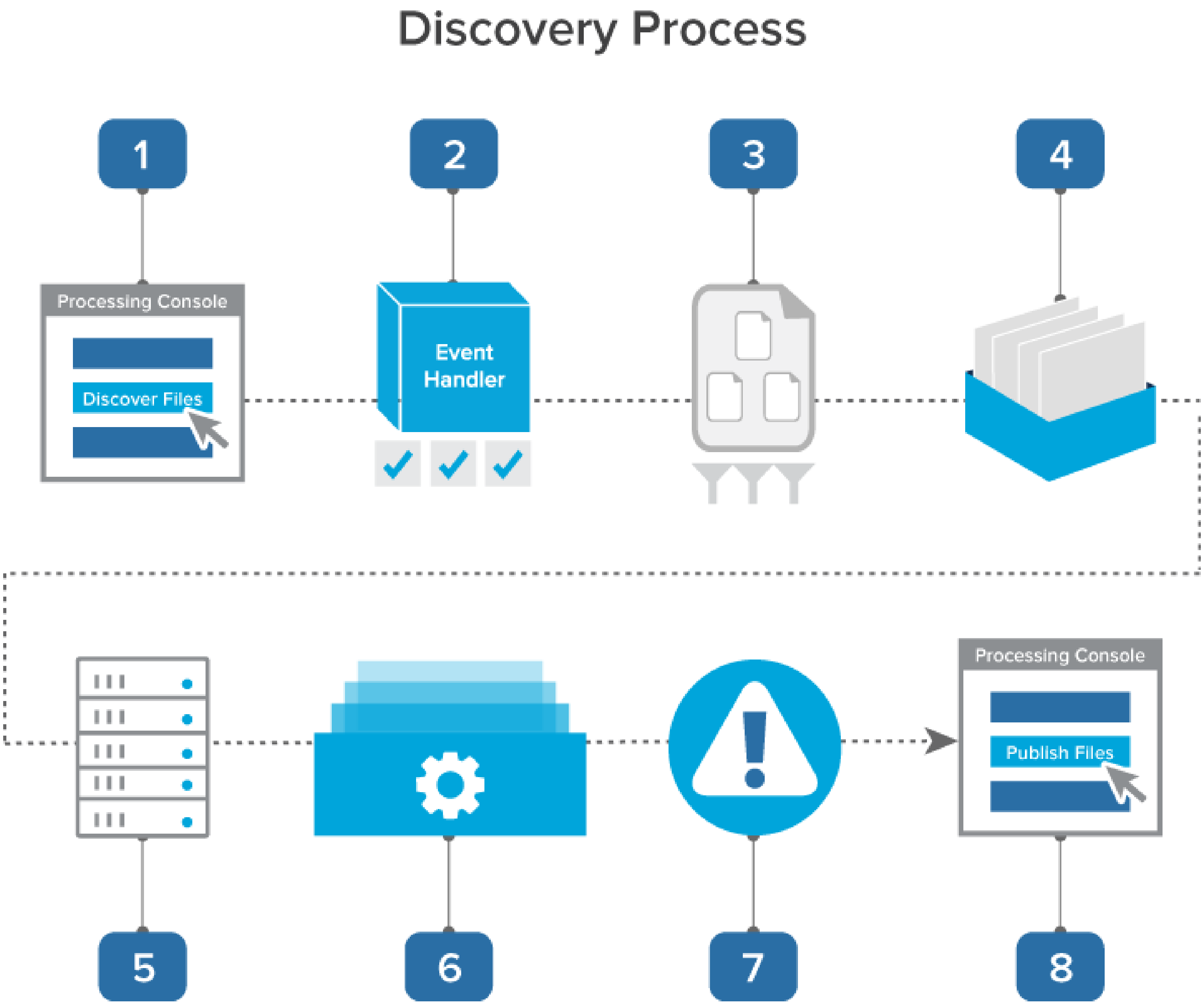
- Relativity does not re-extract text for a re-discovered file unless an extraction error occurred. This means that if you discover the same file twice and you change any settings on the profile, or select a different profile, between the two discovery jobs, Relativity will not re-extract the text from 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.
- If your discovery job becomes stuck for an inordinate length of time, do not disable the worker associated with that processing job, as that worker may also be performing other processing jobs in the environment.
- 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 41.

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 on page 184 for details.

16.1.1 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.

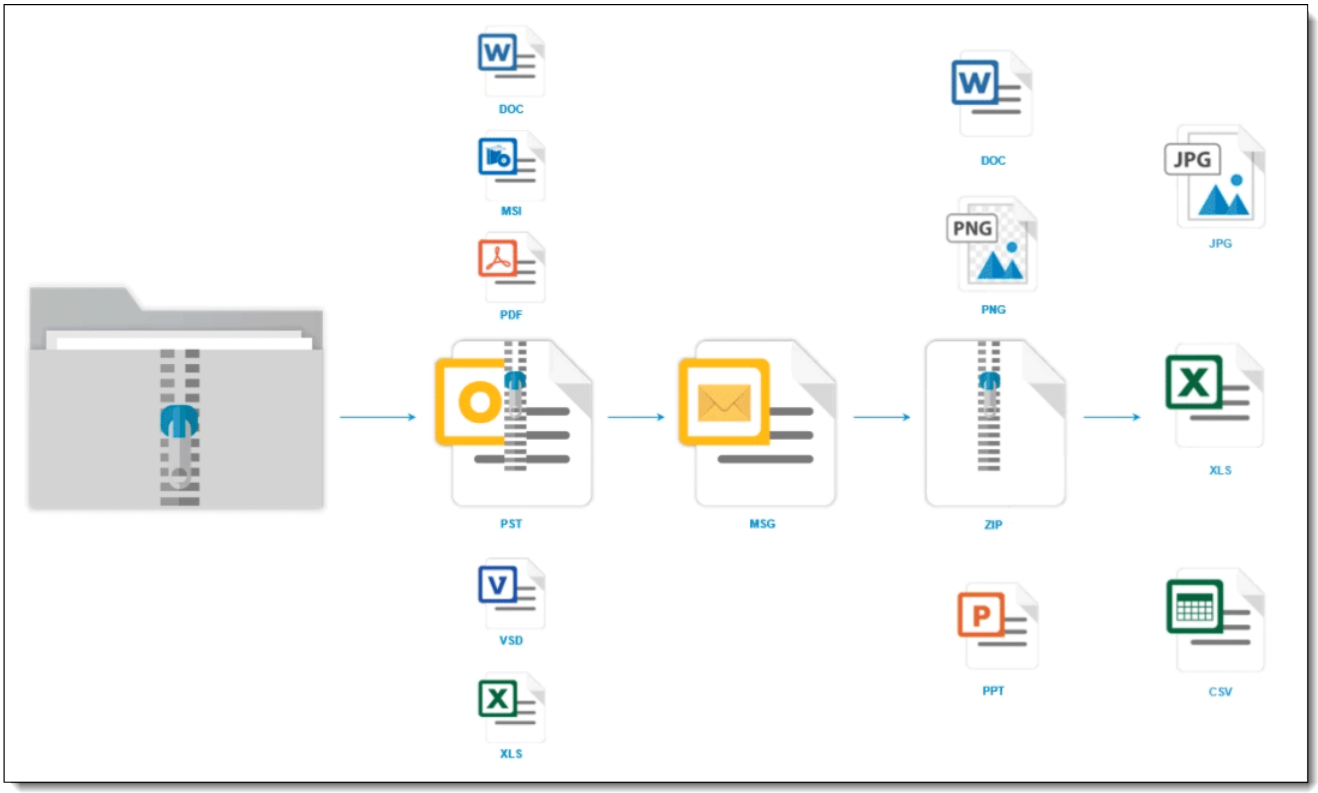


1. You click Discover Files on the processing set console.
2. A console event handler copies all settings from the processing profile to the data sources on the processing set and then checks to make sure that the set is valid and ready to proceed.
3. The event handler inserts all data sources into the 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.

16.1.2 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.



16.2 Special considerations - OCR and text extraction

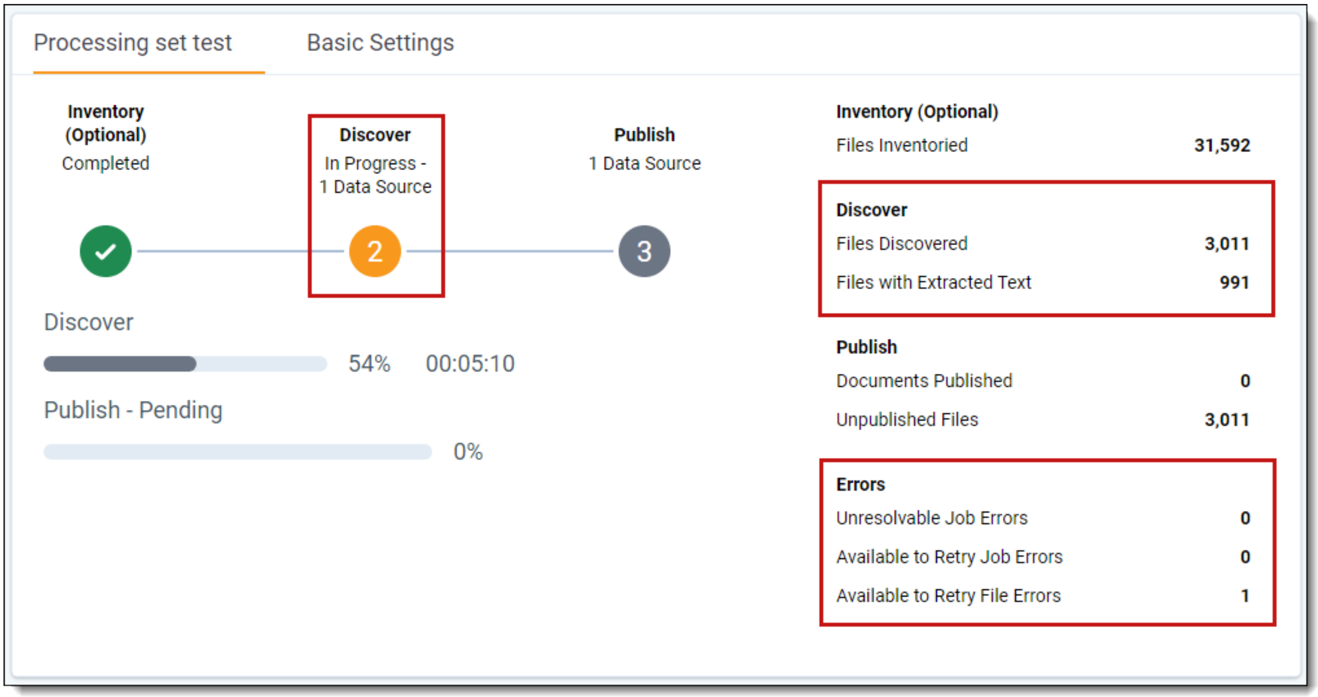
Consider the following regarding OCR and text extraction during discovery:

- During discovery, 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.
- Relativity populates the Extracted Text field when performing OCR during discovery. Relativity doesn’t overwrite metadata fields during OCR.
- 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. For each page of a document containing both native text and images, Relativity stores extracted text and OCR text separately.
- 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.
- Relativity extracts OCR to Unicode.
- 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)

16.3 Monitoring discovery status

You can monitor the progress of the discovery job through the information provided on the Processing set layout screen.

The Processing Set Layout view during Discovery:



Processing set test | Basic Settings

Inventory (Optional)
Completed

Discover
In Progress - 1 Data Source

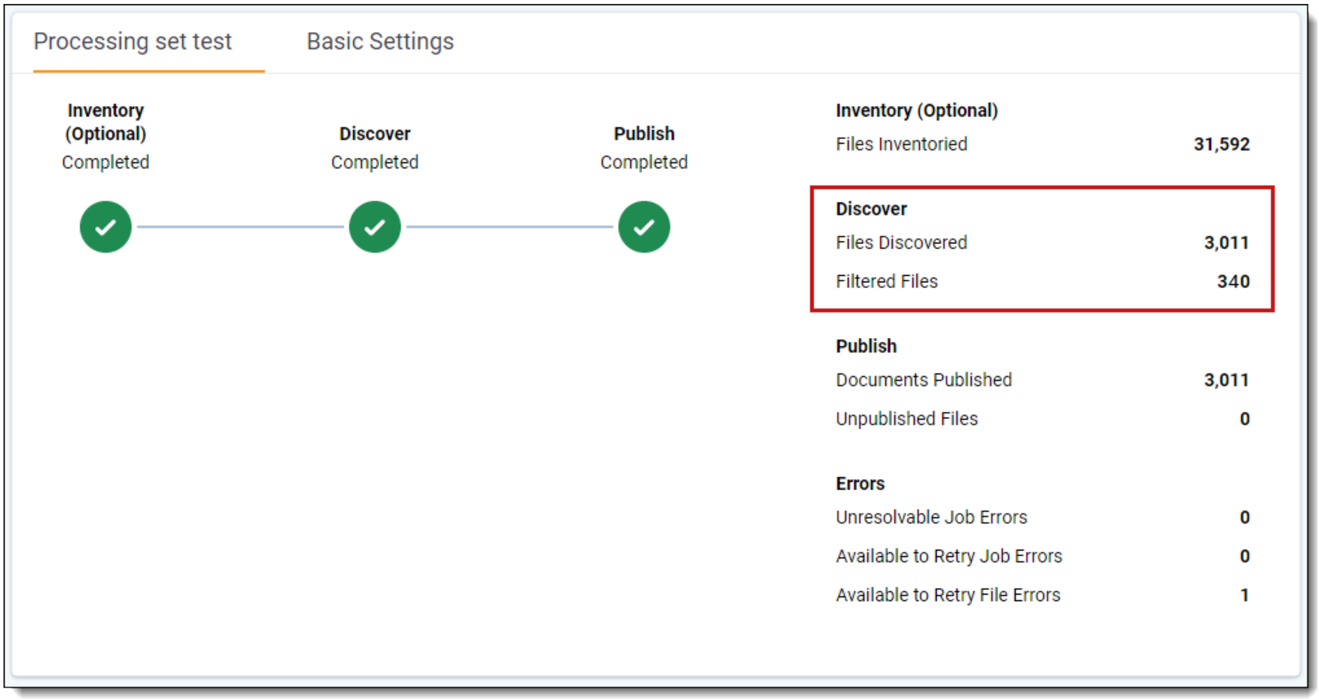
Publish
1 Data Source

Discover
54% 00:05:10

Publish - Pending
0%

Inventory (Optional)	
Files Inventoried	**31,592**
Discover	
Files Discovered	**3,011**
Files with Extracted Text	**991**
Publish	
Documents Published	**0**
Unpublished Files	**3,011**
Errors	
Unresolvable Job Errors	**0**
Available to Retry Job Errors	**0**
Available to Retry File Errors	**1**

The Processing Set Layout view after Discovery:



Processing set test | Basic Settings

Inventory (Optional)
Completed

Discover
Completed

Publish
Completed

Inventory (Optional)	
Files Inventoried	**31,592**
Discover	
Files Discovered	**3,011**
Filtered Files	**340**
Publish	
Documents Published	**3,011**
Unpublished Files	**0**
Errors	
Unresolvable Job Errors	**0**
Available to Retry Job Errors	**0**
Available to Retry File Errors	**1**

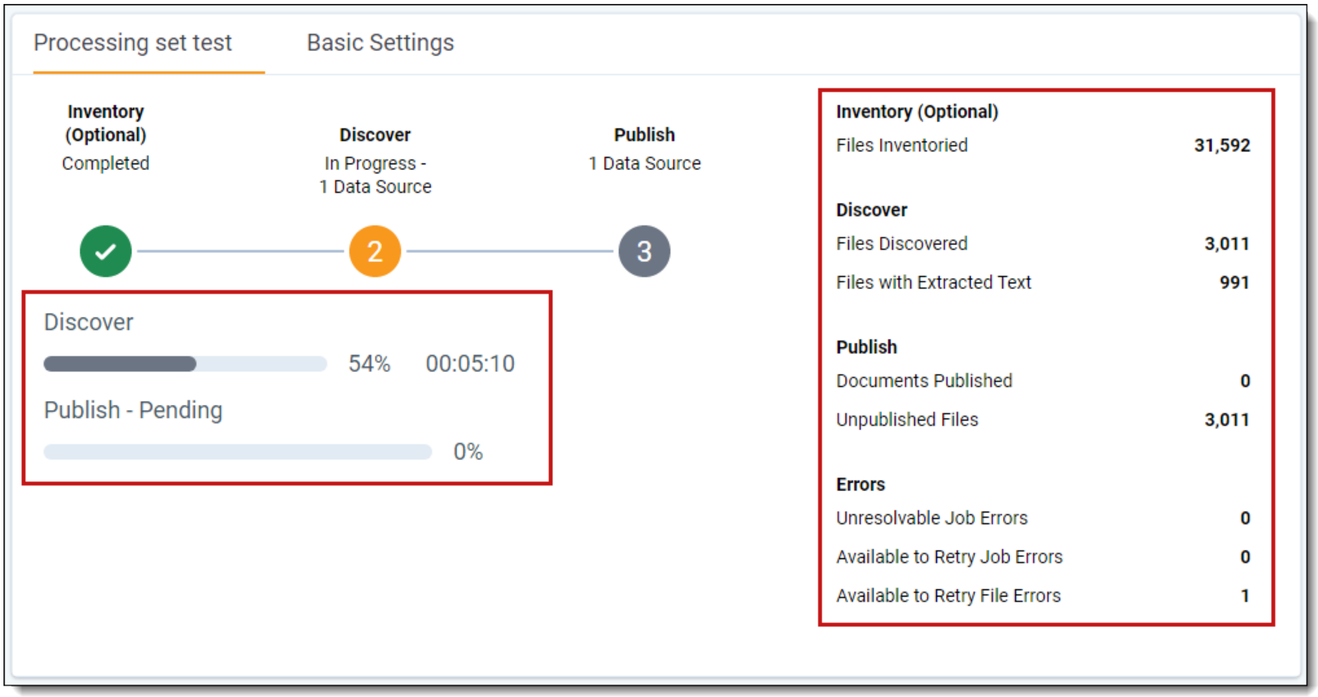
Throughout the processing job, you can monitor the following:

- **# of Data Sources**—the number of data sources currently in the processing queue.
- **Inventory (Optional)** > **Files Inventoried**—the number of files across all data sources submitted that the processing engine inventoried.

- **Discover** > **Files Discovered**—the number of files across all data sources submitted that the processing engine discovered.
- **Discover** > **Files with Extracted Text**—the number of files across all data sources submitted that have extracted text. This value only displays while the discovery jobs are still in progress. If the value is 0, text extraction has not started yet.
- **Discover** > **Filtered Files**—the number of files you excluded from discovery by applying any of the available filters from the Inventory tab. For example, if you applied a Date Range filter that excluded 10 files from your data after you inventoried it, the value displayed is 10. If you applied no filters in the Inventory tab, the value displayed is 0. This value does not include files excluded via the DeNIST setting on the processing profile associated with this set. This value only displays when the job is complete and takes the place of the Files with Extracted Text field.
- **Errors**—the number of errors that have occurred across all data sources submitted, which fall into the following categories:
    - **Unresolvable Job Errors**—errors that you cannot retry.
    - **Available to Retry Job Errors**—job errors that are available for retry.
    - **Available to Retry File Errors**—file errors that you have submitted for retry and are currently in the processing queue.

Note: Overall progress is calculated based on the number of data sources and percentage complete for each source.

If you enabled the auto-publish set option on the profile used by this set, you can monitor the progress for both discovery and publish.



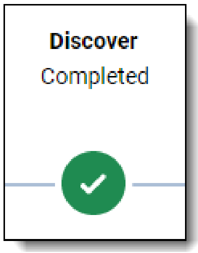
Discover 54% 00:05:10

Publish - Pending 0%

Inventory (Optional)	
Files Inventoried	31,592
Discover	
Files Discovered	3,011
Files with Extracted Text	991
Publish	
Documents Published	0
Unpublished Files	3,011
Errors	
Unresolvable Job Errors	0
Available to Retry Job Errors	0
Available to Retry File Errors	1

See Processing error overview for details.

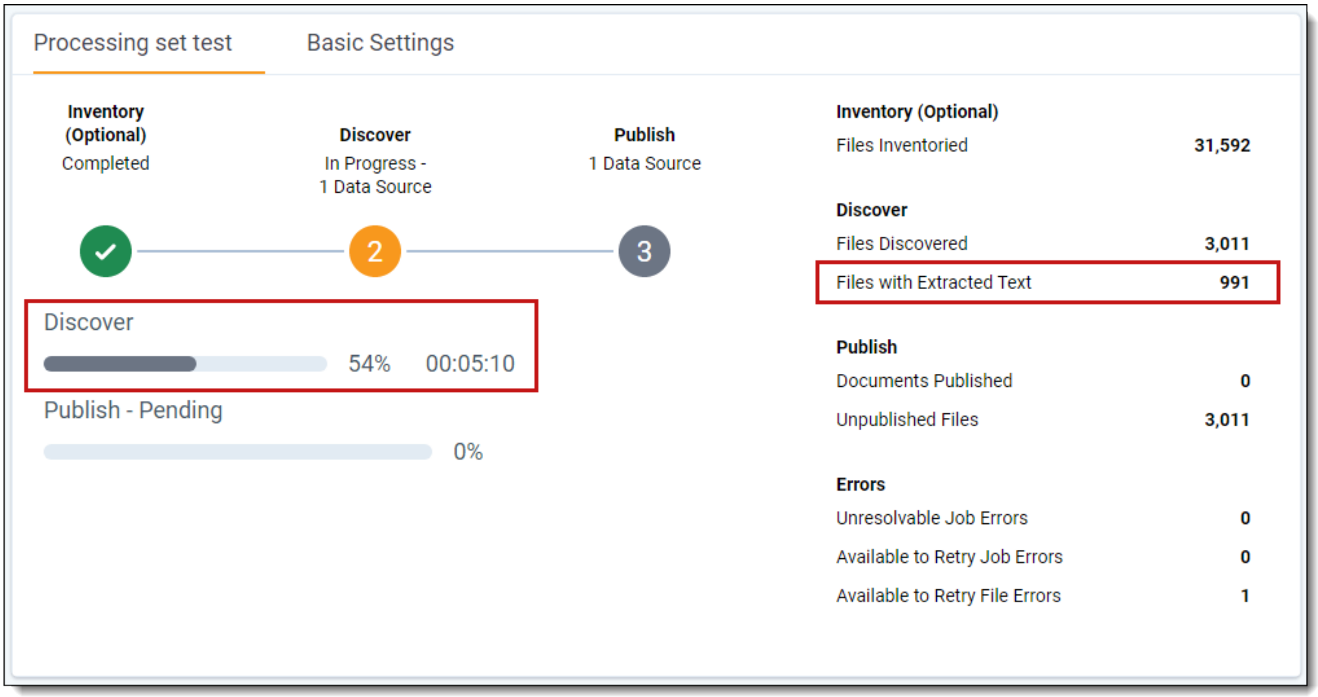
Once discovery is complete, the status section displays a check mark, indicating that you can move on to publishing your files. .



Discover
Completed

16.4 Viewing text extraction progress in processing sets

This feature is located below the **Discover** > **Files Discovered** count and shows the progress of text extraction by displaying an incrementing count of the number of files containing text within the processing sets. The extracted text files count remains at zero until the status bar passes 25%. The final count appears when the Discover process completes.



Processing set test
Basic Settings
Inventory (Optional)
Completed
Discover
In Progress - 1 Data Source
Publish
1 Data Source
2
3

Discover 54% 00:05:10

Publish - Pending 0%

Inventory (Optional)

Files Inventoried	31,592

Discover

Files Discovered	3,011
Files with Extracted Text	991

Publish

Documents Published	0
Unpublished Files	3,011

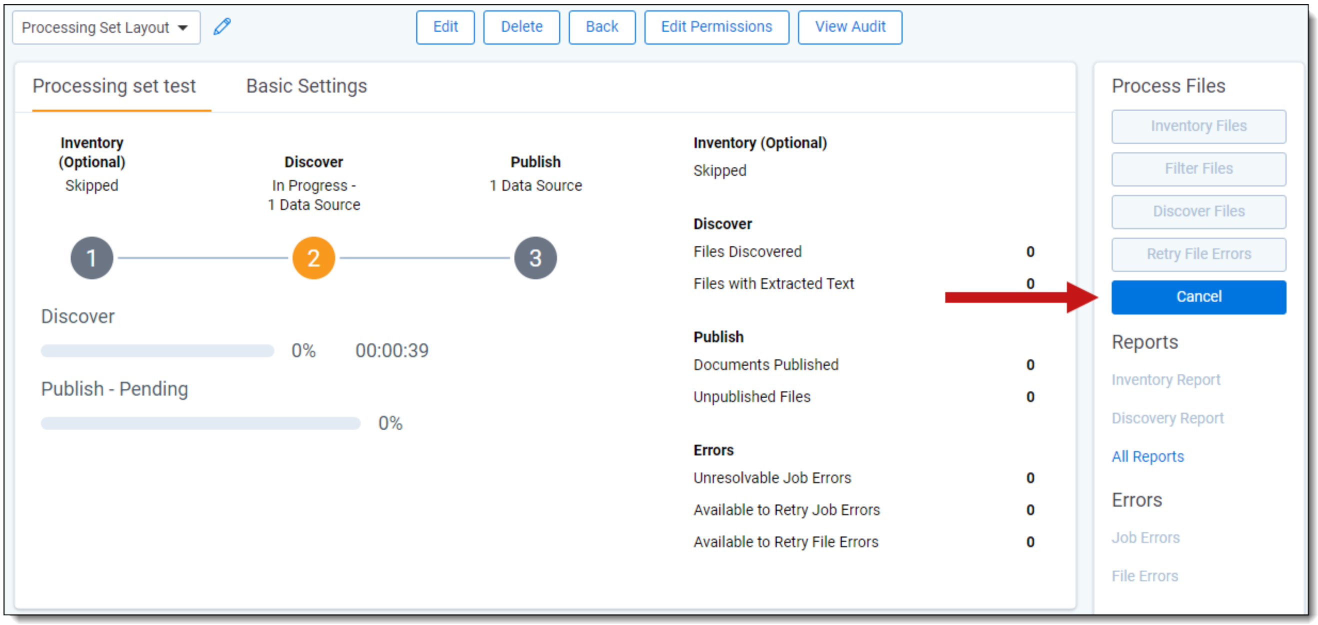
Errors

Unresolvable Job Errors	0
Available to Retry Job Errors	0
Available to Retry File Errors	1

16.5 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**.


Processing Set Layout
Edit
Delete
Back
Edit Permissions
View Audit
Processing set test
Basic Settings
Process Files
Inventory Files
Filter Files
Discover Files
Retry File Errors
Cancel
Reports
Inventory Report
Discovery Report
All Reports
Errors
Job Errors
File Errors

Inventory (Optional)
Skipped

Discover
In Progress -
1 Data Source

Publish
1 Data Source

Discover 0% 00:00:39

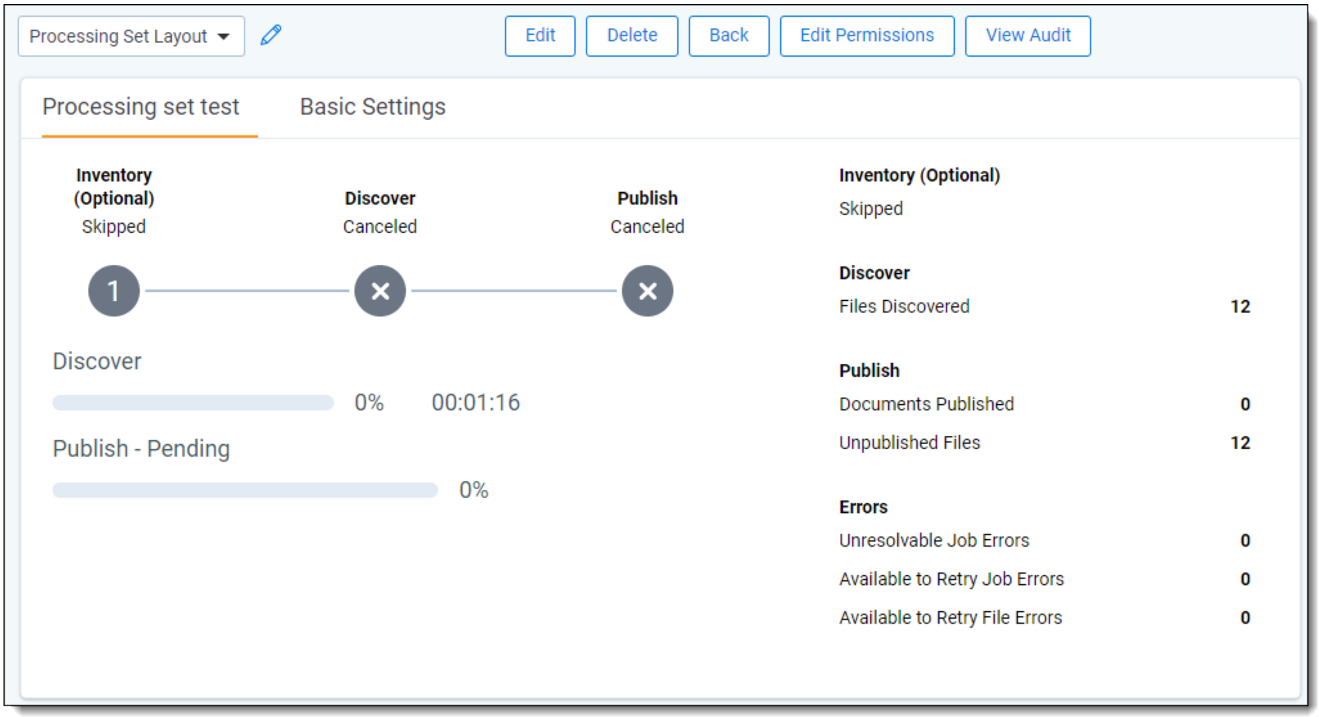
Publish - Pending 0%

Inventory (Optional)	
Skipped	
Discover	
Files Discovered	0
Files with Extracted Text	0
Publish	
Documents Published	0
Unpublished Files	0
Errors	
Unresolvable Job Errors	0
Available to Retry Job Errors	0
Available to Retry File Errors	0

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 Layout
Edit
Delete
Back
Edit Permissions
View Audit
Processing set test
Basic Settings

Inventory (Optional)
Skipped

Discover
Canceled

Publish
Canceled

1

Discover
0% 00:01:16

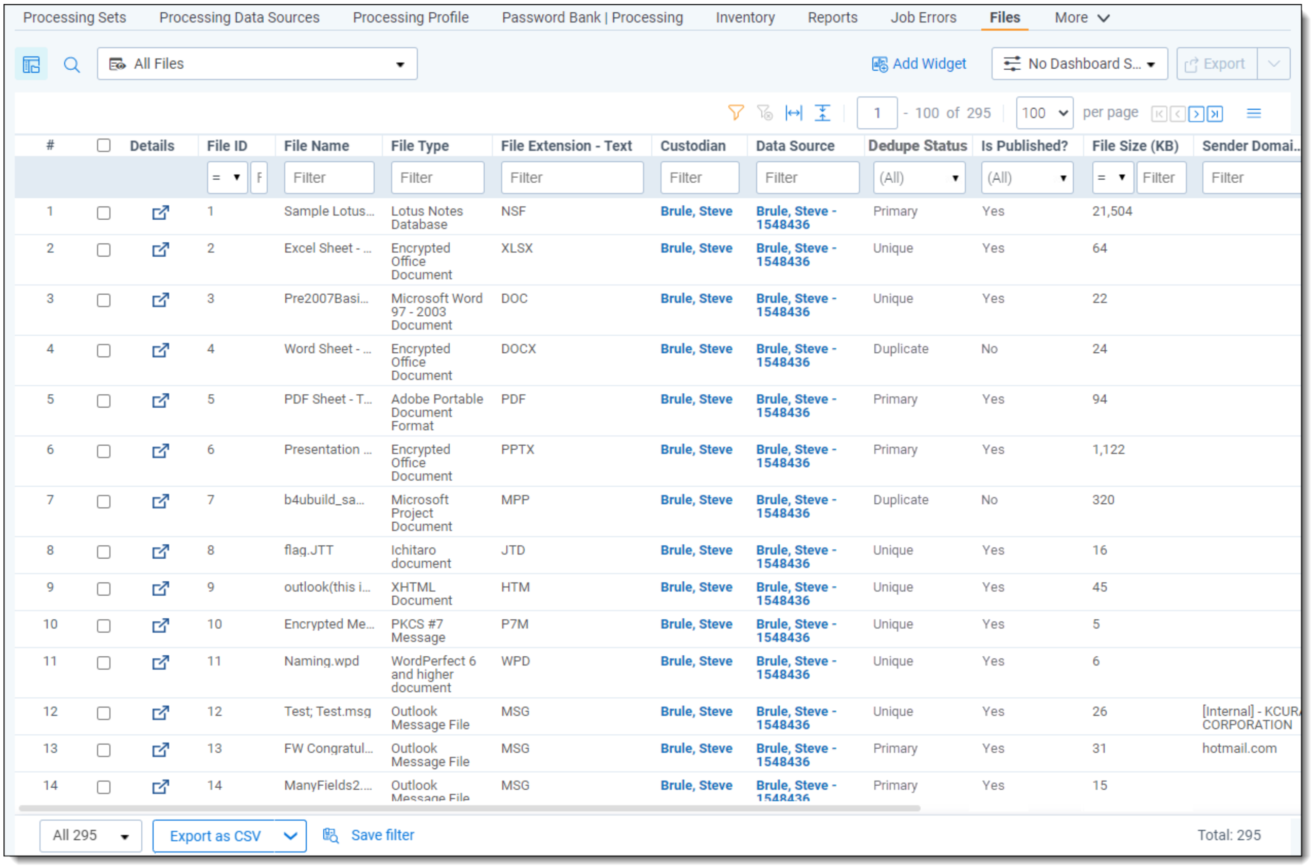
Publish - Pending
0%

Inventory (Optional)
Skipped

Discover	
Files Discovered	12
Publish	
Documents Published	0
Unpublished Files	12
Errors	
Unresolvable Job Errors	0
Available to Retry Job Errors	0
Available to Retry File Errors	0

17 Files tab

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.



#	☐	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	Filter	(All) ▾	(All) ▾	= ▾ Filter	Filter
1	☐	↗	1	Sample Lotus...	Lotus Notes Database	NSF	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	21,504	
2	☐	↗	2	Excel Sheet - ...	Encrypted Office Document	XLSX	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	64	
3	☐	↗	3	Pre2007Basi...	Microsoft Word 97 - 2003 Document	DOC	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	22	
4	☐	↗	4	Word Sheet - ...	Encrypted Office Document	DOCX	Brule, Steve	Brule, Steve - 1548436	Duplicate	No	24	
5	☐	↗	5	PDF Sheet - T...	Adobe Portable Document Format	PDF	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	94	
6	☐	↗	6	Presentation ...	Encrypted Office Document	PPTX	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	1,122	
7	☐	↗	7	b4ubuild_sa...	Microsoft Project Document	MPP	Brule, Steve	Brule, Steve - 1548436	Duplicate	No	320	
8	☐	↗	8	flag.JTT	Ichitaro document	JTD	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	16	
9	☐	↗	9	outlook(this i...	XHTML Document	HTM	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	45	
10	☐	↗	10	Encrypted Me...	PKCS #7 Message	P7M	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	5	
11	☐	↗	11	Naming.wpd	WordPerfect 6 and higher document	WPD	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	6	
12	☐	↗	12	Test; Test.msg	Outlook Message File	MSG	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	26	[Internal] - KCUR CORPORATION
13	☐	↗	13	FW Congratul...	Outlook Message File	MSG	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	31	hotmail.com
14	☐	↗	14	ManyFields2....	Outlook Message File	MSG	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	15	

17.1 Views on the Files tab

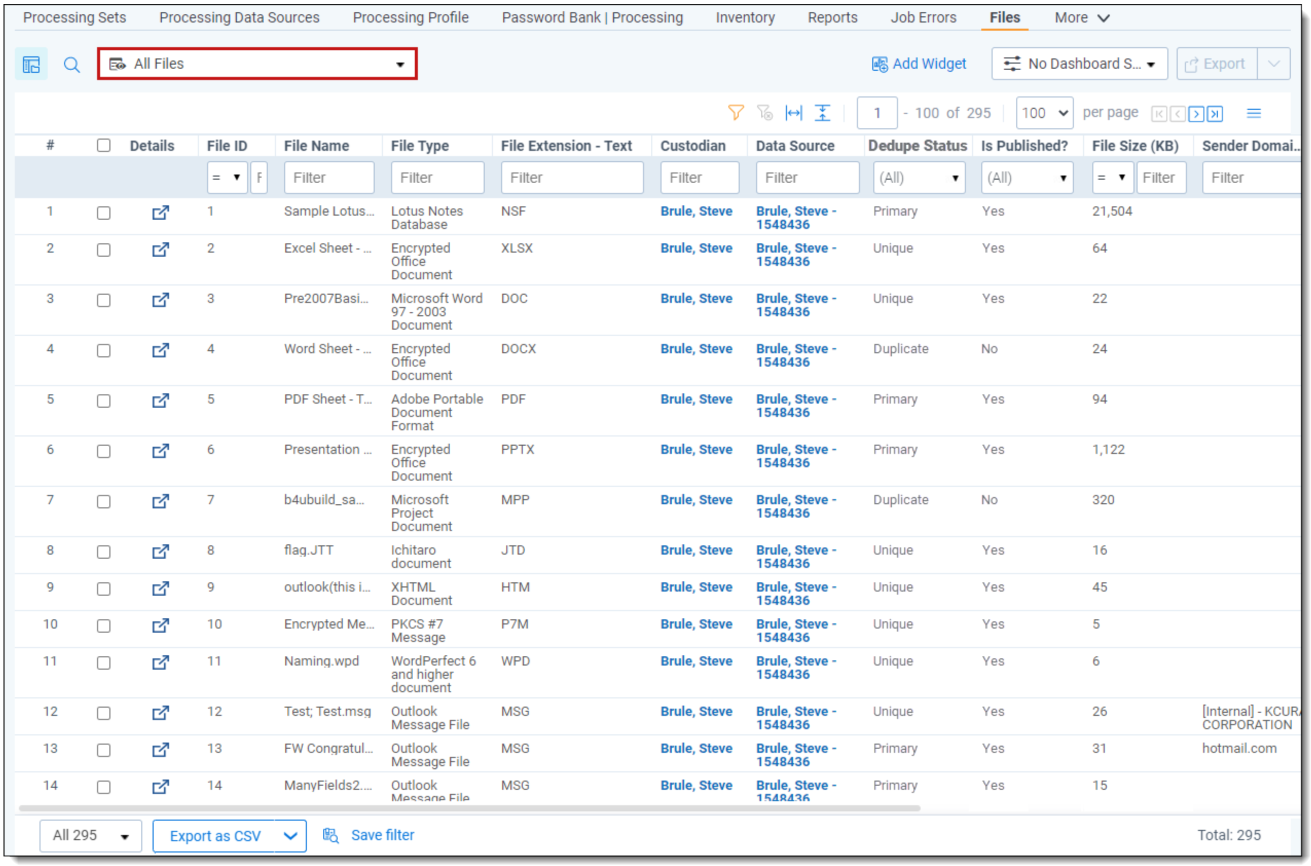
The Files tab contains the following views:

- **All Files**—lists all the files in your workspace.
- **Current Errored Files**—lists the documents that yielded errors in your workspace that currently have an Error Status value of Not Resolved.
- **All Errored Files**—lists the documents that yielded errors in your workspace, including those with a current Error Status value of Resolved and Unresolved.
- **Deleted Documents**—lists the documents you deleted from your workspace.
- Container files—lists the container files in your workspace.

Note: You can export any file list as a CSV file, which includes the total set of filtered results.

17.1.1 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.



#	☐	Details	File ID	File Name	File Type	File Extension - Text	Custodian	Data Source	Dedupe Status	Is Published?	File Size (KB)	Sender Domai...
			=	Filter	Filter	Filter	Filter	Filter	(All)	(All)	= Filter	Filter
1	☐		1	Sample Lotus...	Lotus Notes Database	NSF	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	21,504	
2	☐		2	Excel Sheet - ...	Encrypted Office Document	XLSX	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	64	
3	☐		3	Pre2007Basi...	Microsoft Word 97 - 2003 Document	DOC	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	22	
4	☐		4	Word Sheet - ...	Encrypted Office Document	DOCX	Brule, Steve	Brule, Steve - 1548436	Duplicate	No	24	
5	☐		5	PDF Sheet - T...	Adobe Portable Document Format	PDF	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	94	
6	☐		6	Presentation ...	Encrypted Office Document	PPTX	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	1,122	
7	☐		7	b4ubuild_sa...	Microsoft Project Document	MPP	Brule, Steve	Brule, Steve - 1548436	Duplicate	No	320	
8	☐		8	flag.JTT	Ichitaro document	JTD	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	16	
9	☐		9	outlook(this i...	XHTML Document	HTM	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	45	
10	☐		10	Encrypted Me...	PKCS #7 Message	P7M	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	5	
11	☐		11	Naming.wpd	WordPerfect 6 and higher document	WPD	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	6	
12	☐		12	Test; Test.msg	Outlook Message File	MSG	Brule, Steve	Brule, Steve - 1548436	Unique	Yes	26	[Internal] - KCUR CORPORATION
13	☐		13	FW Congratul...	Outlook Message File	MSG	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	31	hotmail.com
14	☐		14	ManyFields2....	Outlook Message File	MSG	Brule, Steve	Brule, Steve - 1548436	Primary	Yes	15	

Total: 295

This view contains the following fields:

- **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

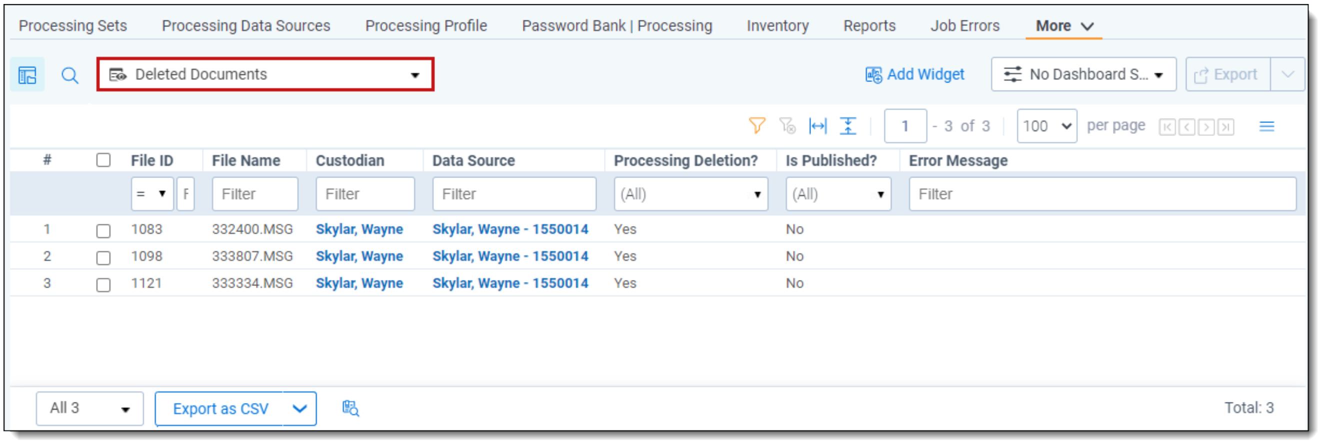
status from their parent.
Dedupe Status field values include:

  - **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.

- **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.

17.1.2 Deleted Documents view

The Deleted Documents view contains the files you deleted from the Documents tab after the files were published.



Deleted Documents

#	File ID	File Name	Custodian	Data Source	Processing Deletion?	Is Published?	Error Message
1	1083	332400.MSG	Skylar, Wayne	Skylar, Wayne - 1550014	Yes	No	
2	1098	333807.MSG	Skylar, Wayne	Skylar, Wayne - 1550014	Yes	No	
3	1121	333334.MSG	Skylar, Wayne	Skylar, Wayne - 1550014	Yes	No	

Total: 3

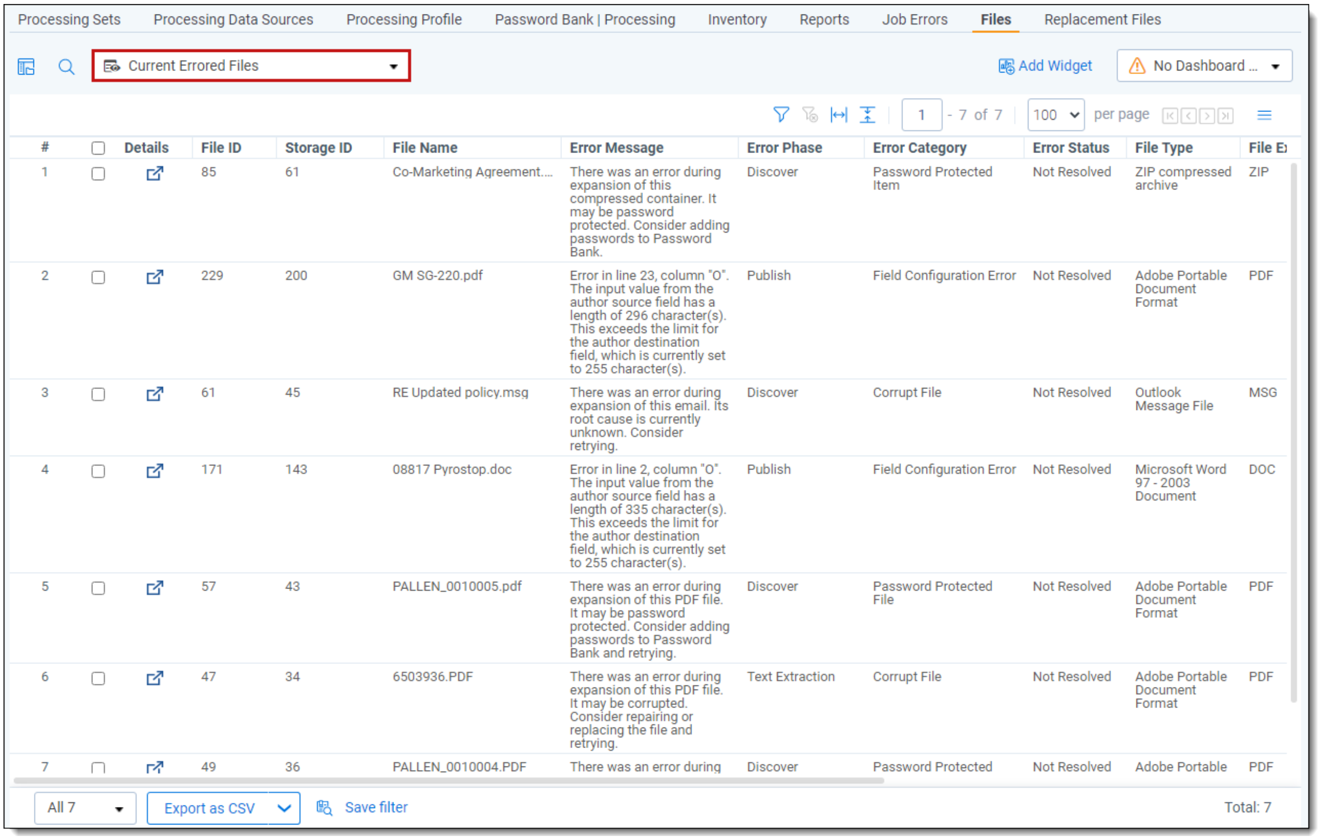
This view contains the following fields:

- **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

17.1.3 Current Errored Files view

The Current Errored Files view contains all the documents that yielded errors in your workspace that currently have an Error Status value of Not Resolved. By default, this view does not contain files with an Error Status of Resolved, as those can be found in the All Errored Files view.



#	☐	Details	File ID	Storage ID	File Name	Error Message	Error Phase	Error Category	Error Status	File Type	File E
1	☐		85	61	Co-Marketing Agreement....	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Item	Not Resolved	ZIP compressed archive	ZIP
2	☐		229	200	GM SG-220.pdf	Error in line 23, column "O". The input value from the author source field has a length of 296 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error	Not Resolved	Adobe Portable Document Format	PDF
3	☐		61	45	RE Updated policy.msg	There was an error during expansion of this email. Its root cause is currently unknown. Consider retrying.	Discover	Corrupt File	Not Resolved	Outlook Message File	MSG
4	☐		171	143	08817 Pyrostop.doc	Error in line 2, column "O". The input value from the author source field has a length of 335 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error	Not Resolved	Microsoft Word 97 - 2003 Document	DOC
5	☐		57	43	PALLEN_0010005.pdf	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved	Adobe Portable Document Format	PDF
6	☐		47	34	6503936.PDF	There was an error during expansion of this PDF file. It may be corrupted. Consider repairing or replacing the file and retrying.	Text Extraction	Corrupt File	Not Resolved	Adobe Portable Document Format	PDF
7	☐		49	36	PALLEN_0010004.PDF	There was an error during	Discover	Password Protected	Not Resolved	Adobe Portable	PDF

- **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.

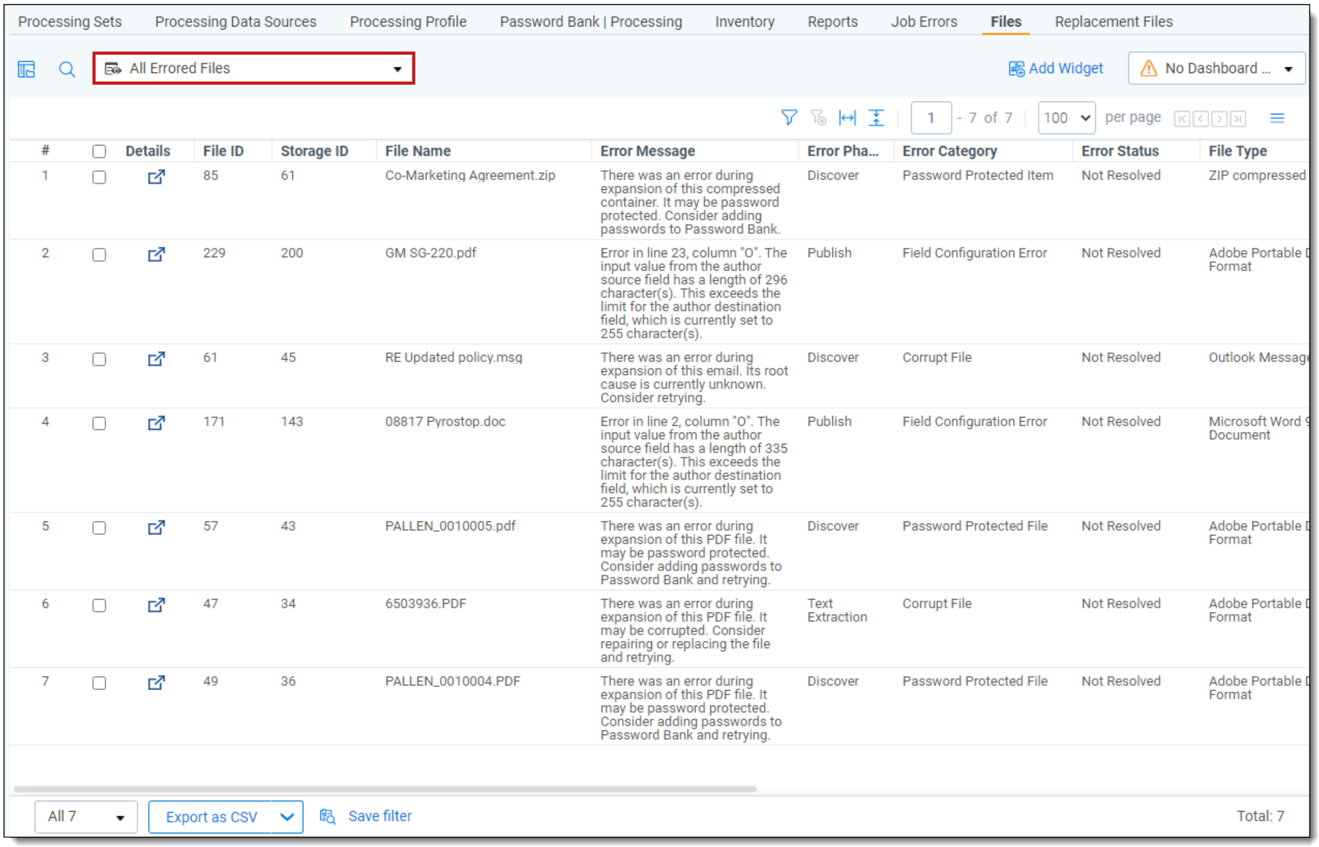
> **Note:** Differentiating between the File ID and the Storage ID. Every file is assigned a unique Storage ID. Every instance of a file is assigned a unique File ID. For example, you may have a Word document with a unique Storage ID. If that document is attached to two different emails, then each instance 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.

- **Error Message**—the message that details the error, cause, and suggested resolution of the error. 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 errors associated with it, this field will display a value of the Publish error.
    - Delete
    - Publish
    - Discover
    - Text Extraction
- **Error Phase**—the phase of processing in which the error 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 errors associated with it, this field will display a value of Publish.
    - Delete
    - Publish
    - Discover
    - Text Extraction
- **Error Category**—provides insight into the nature of the errors that have occurred on your processed files. For details, see Error categories and their probability of retry success.
- **Error Status**—the current status of the error. The Current Errored Files view only displays files with an Error 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.

17.1.4 All Errored Files view

The All Errored Files view contains all the documents that yielded errors in your workspace, including those with a current Error Status value of Resolved and Unresolved. These files are sorted by descending file size starting with the largest containers and ending with the smallest loose files.



#	☐	Details	File ID	Storage ID	File Name	Error Message	Error Pha...	Error Category	Error Status	File Type
1	☐	↗	85	61	Co-Marketing Agreement.zip	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Item	Not Resolved	ZIP compressed
2	☐	↗	229	200	GM SG-220.pdf	Error in line 23, column "O". The input value from the author source field has a length of 296 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error	Not Resolved	Adobe Portable Format
3	☐	↗	61	45	RE Updated policy.msg	There was an error during expansion of this email. Its root cause is currently unknown. Consider retrying.	Discover	Corrupt File	Not Resolved	Outlook Messag
4	☐	↗	171	143	08817 Pyrostop.doc	Error in line 2, column "O". The input value from the author source field has a length of 335 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error	Not Resolved	Microsoft Word Document
5	☐	↗	57	43	PALLEN_0010005.pdf	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved	Adobe Portable Format
6	☐	↗	47	34	6503936.PDF	There was an error during expansion of this PDF file. It may be corrupted. Consider repairing or replacing the file and retrying.	Text Extraction	Corrupt File	Not Resolved	Adobe Portable Format
7	☐	↗	49	36	PALLEN_0010004.PDF	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved	Adobe Portable Format

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.

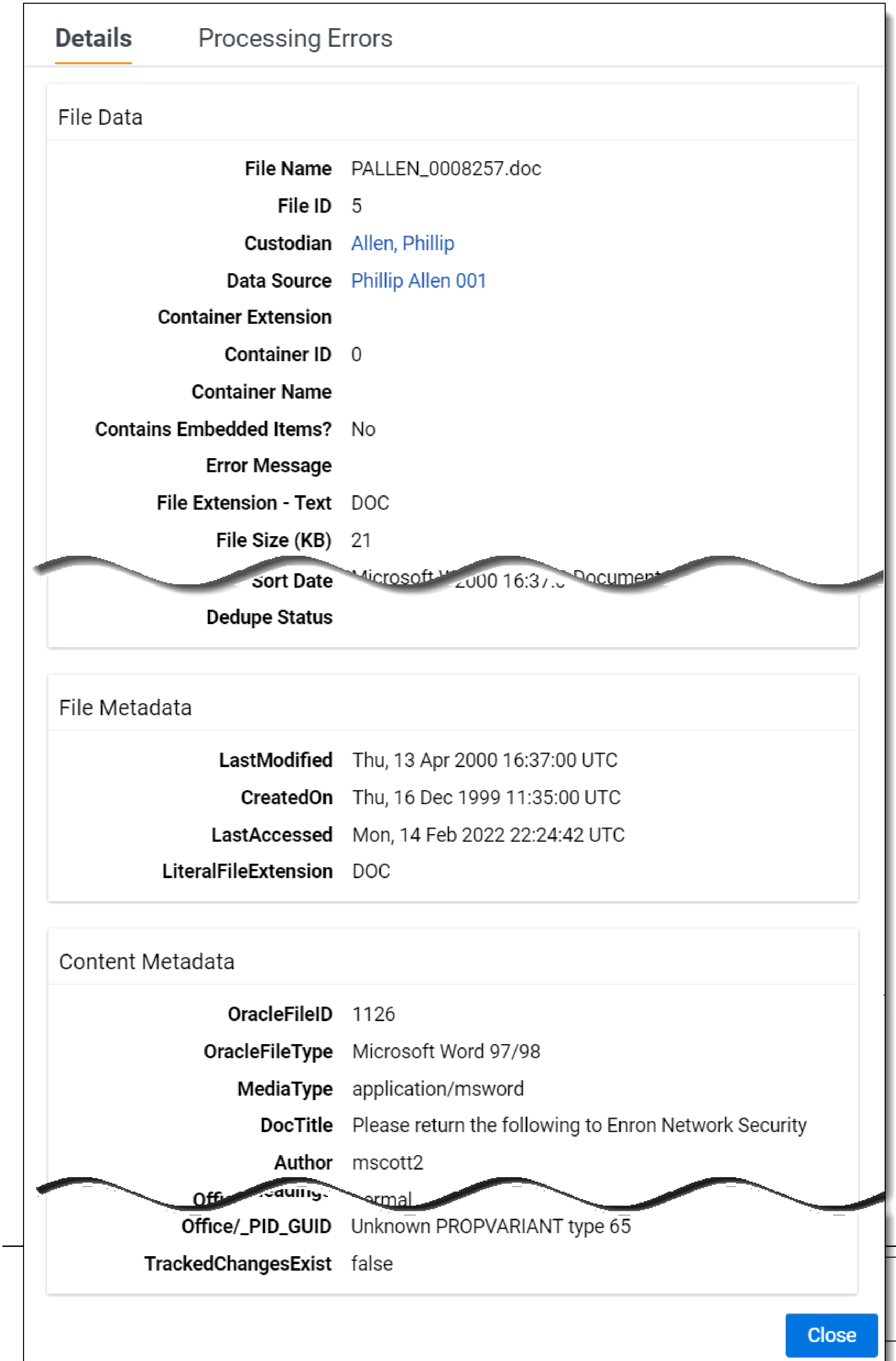
  Note: Differentiating between the File ID and the Storage ID. Every file is assigned a unique Storage ID. Every instance of a file is assigned a unique File ID. For example, you may have a Word document with a unique Storage ID. If that document is attached to two different emails, then each instance 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.
- **Error Message**—the message that details the error, cause, and suggested resolution of the error. 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 errors associated with it, this field will display a value of Publish.
  - Delete
  - Publish
  - Discover
  - Text Extraction

- **Error Phase**—the phase of processing in which the error 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 errors associated with it, this field will display a value of Publish.
  - Delete
  - Publish
  - Discover
  - Text Extraction
- **Error Status**—the current status of the error. This field displays any of the following values, depending on the current state of the file:
  - Ignored
  - Resolved
  - Resolving
  - 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.

17.2 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	PALLEN_0008257.doc
File ID	5
Custodian	Allen, Phillip
Data Source	Phillip Allen 001
Container Extension	
Container ID	0
Container Name	
Contains Embedded Items?	No
Error Message	
File Extension - Text	DOC
File Size (KB)	21
Sort Date	Microsoft W 2000 16:37 Document
Dedupe Status	

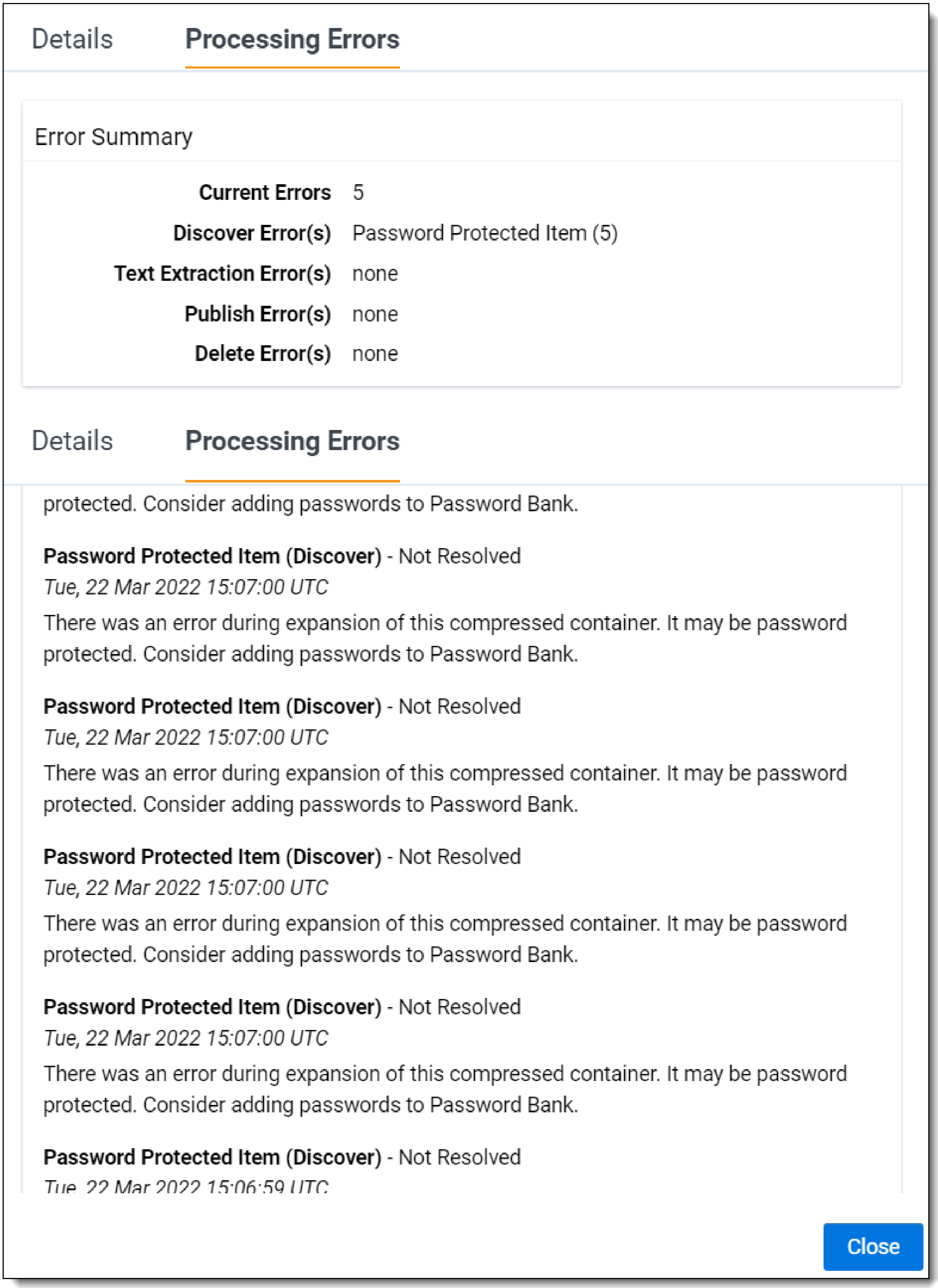
File Metadata

LastModified	Thu, 13 Apr 2000 16:37:00 UTC
CreatedOn	Thu, 16 Dec 1999 11:35:00 UTC
LastAccessed	Mon, 14 Feb 2022 22:24:42 UTC
LiteralFileExtension	DOC

Content Metadata

OracleFileID	1126
OracleFileType	Microsoft Word 97/98
MediaType	application/msword
DocTitle	Please return the following to Enron Network Security
Author	mscott2
Off eading:	ormal
Office/_PID_GUID	Unknown PROPVARIANT type 65
TrackedChangesExist	false

Click the **Processing Errors** tab to view a summary of errors, and the error history.



Error Summary

Current Errors	5
Discover Error(s)	Password Protected Item (5)
Text Extraction Error(s)	none
Publish Error(s)	none
Delete Error(s)	none

protected. Consider adding passwords to Password Bank.

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 protected. Consider adding passwords to Password Bank.

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 protected. Consider adding passwords to Password Bank.

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 protected. Consider adding passwords to Password Bank.

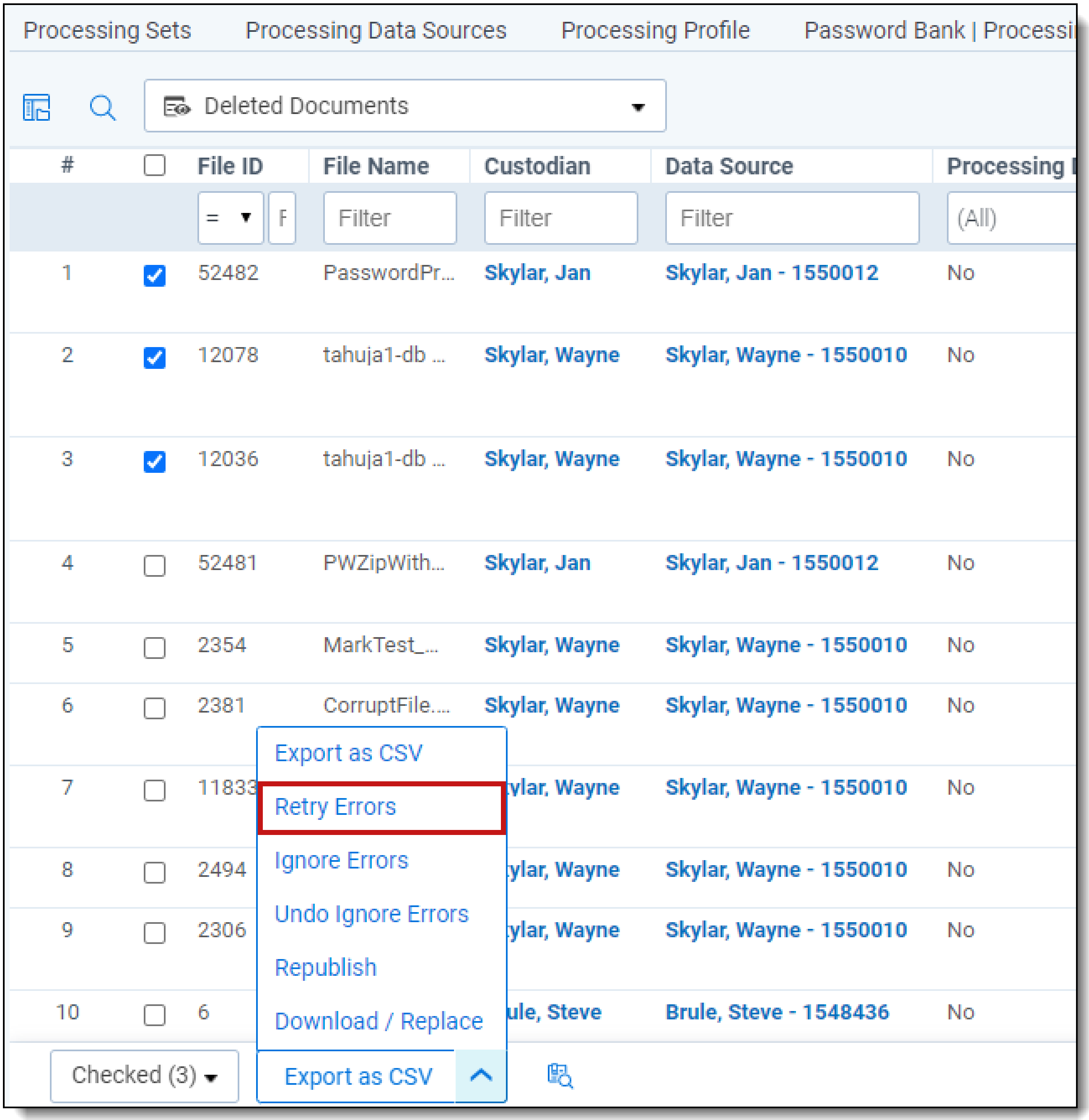
Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:07:00 UTC
There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.

Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:06:59 UTC

- The Error History section represents all errors that have ever occurred on a file. This acts as a timeline of the record’s errors, showing when they occurred, what they were about, and if any are still active. This includes errors resulting from retries of previous errors and contains category, phase, date/time, and message information. All times are kept in UTC format.
- The Error Summary section displays a count of all active errors along with their associated category and phase. This is especially important when investigating errors 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.

17.3 Retrying delete errors

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 Error Message to see the errors that occurred during deletion. Errors can be retried using the **Retry Errors** 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 Sets
Processing Data Sources
Processing Profile
Password Bank | Processi
Deleted Documents
Export as CSV
Retry Errors
Ignore Errors
Undo Ignore Errors
Republish
Download / Replace
Checked (3)
Export as CSV

#	☐	File ID	File Name	Custodian	Data Source	Processing I
		= ▾ F	Filter	Filter	Filter	(All)
1	☑	52482	PasswordPr...	Skylar, Jan	Skylar, Jan - 1550012	No
2	☑	12078	tahuja1-db ...	Skylar, Wayne	Skylar, Wayne - 1550010	No
3	☑	12036	tahuja1-db ...	Skylar, Wayne	Skylar, Wayne - 1550010	No
4	☐	52481	PWZipWith...	Skylar, Jan	Skylar, Jan - 1550012	No
5	☐	2354	MarkTest_...	Skylar, Wayne	Skylar, Wayne - 1550010	No
6	☐	2381	CorruptFile....	Skylar, Wayne	Skylar, Wayne - 1550010	No
7	☐	11833		ylar, Wayne	Skylar, Wayne - 1550010	No
8	☐	2494		ylar, Wayne	Skylar, Wayne - 1550010	No
9	☐	2306		ylar, Wayne	Skylar, Wayne - 1550010	No
10	☐	6		ule, Steve	Brule, Steve - 1548436	No

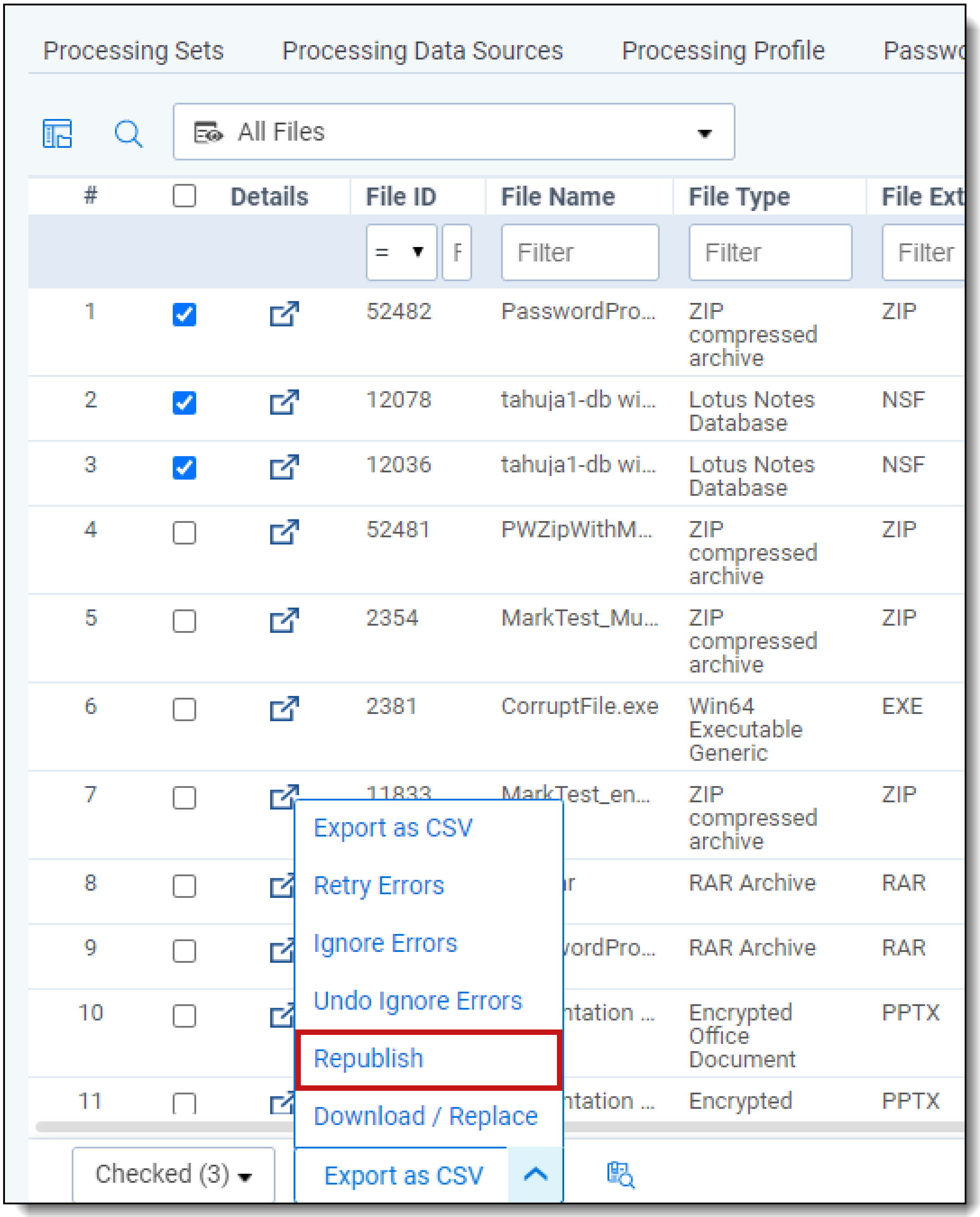
17.4 Republishing files from the Files tab

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 page 221 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 record of all discovered files and filter to published files via the Is Published? field published. You can republish files at the file level using the **Republish** mass operation.

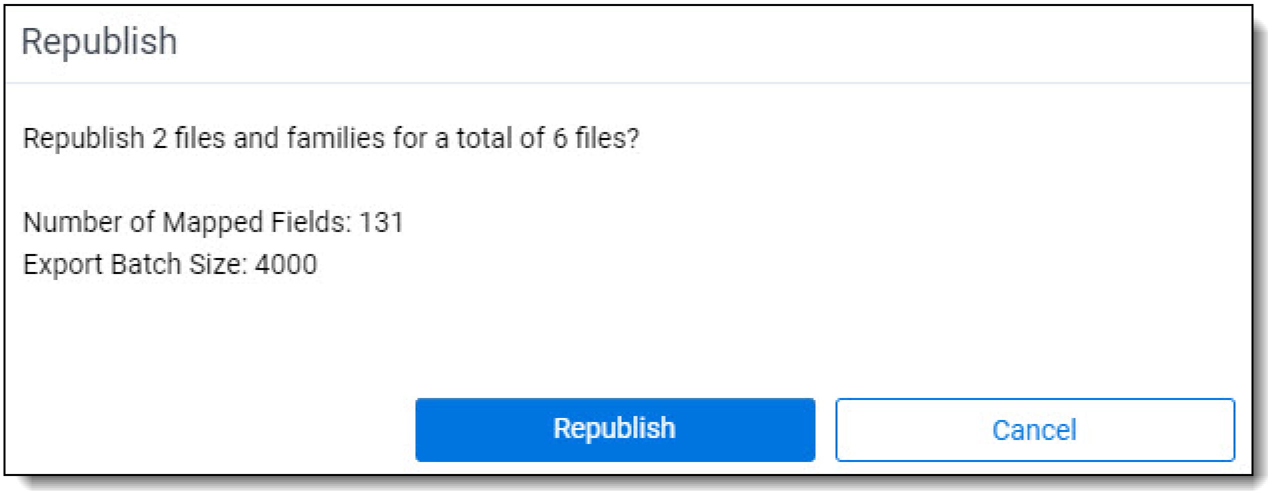

Processing Sets
Processing Data Sources
Processing Profile
Passwo
All Files
Export as CSV
Retry Errors
Ignore Errors
Undo Ignore Errors
Republish
Download / Replace
Checked (3)
Export as CSV

#	☐	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	☐	↗	11833	MarkTest_en...	ZIP compressed archive	ZIP
8	☐			r	RAR Archive	RAR
9	☐			ordPro...	RAR Archive	RAR
10	☐			ntation ...	Encrypted Office Document	PPTX
11	☐			ntation ...	Encrypted	PPTX

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 2 files and families for a total of 6 files?

Number of Mapped Fields: 131
Export Batch Size: 4000

If you have selected eligible files, click **Republish** to proceed or **Cancel** to return to the All Files view.

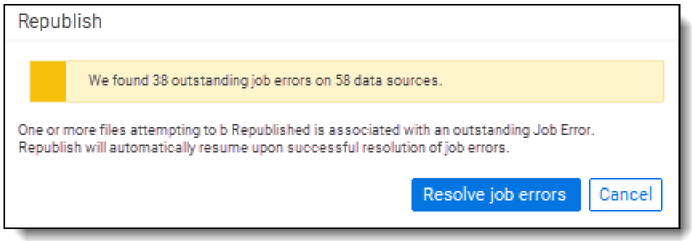
Note: Once you republish, you cannot cancel this job.

17.4.1 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.

   We found 38 outstanding job errors on 58 data sources.

   One or more files attempting to b Republished is associated with an outstanding Job Error.
   Republish will automatically resume upon successful resolution of job errors.

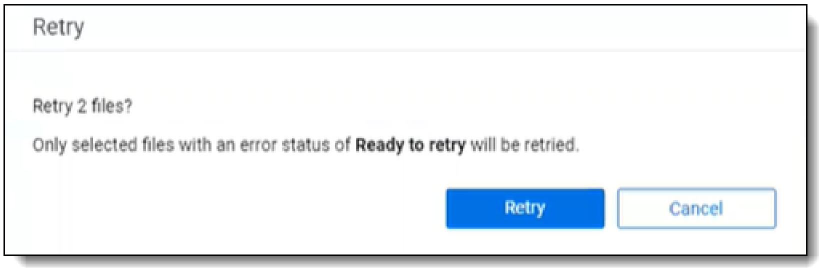
17.4.2 Common use cases for using the Republish mass operation

The following are common situations in which you would opt to use the Republish mass operation:

- 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.

17.5 Retrying errors

Navigate to the All Files view to see a record of all discovered files, including files with errors found during Discovery via the **Discover Error** field. The files that have ready to retry errors can be retried at file level using the **Retry Errors** mass operations option. When you click **Retry Errors**, you see a confirmation modal confirming the number of files you are about to submit for retry.

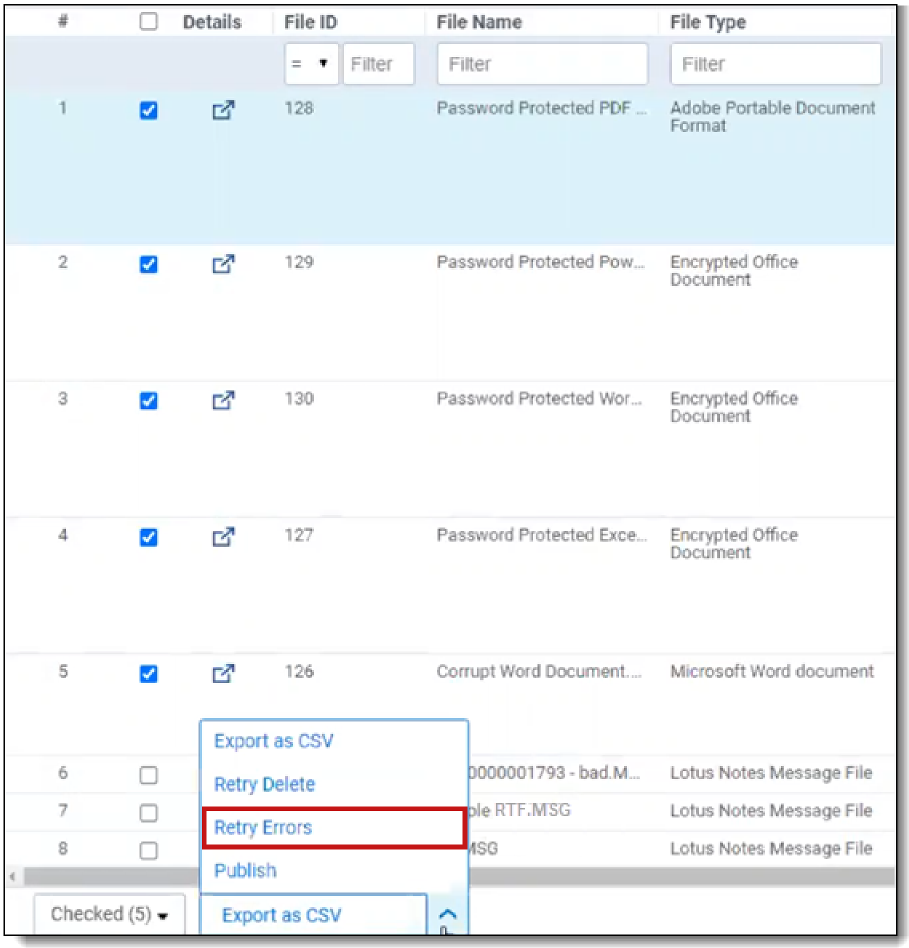


Retry

Retry 2 files?

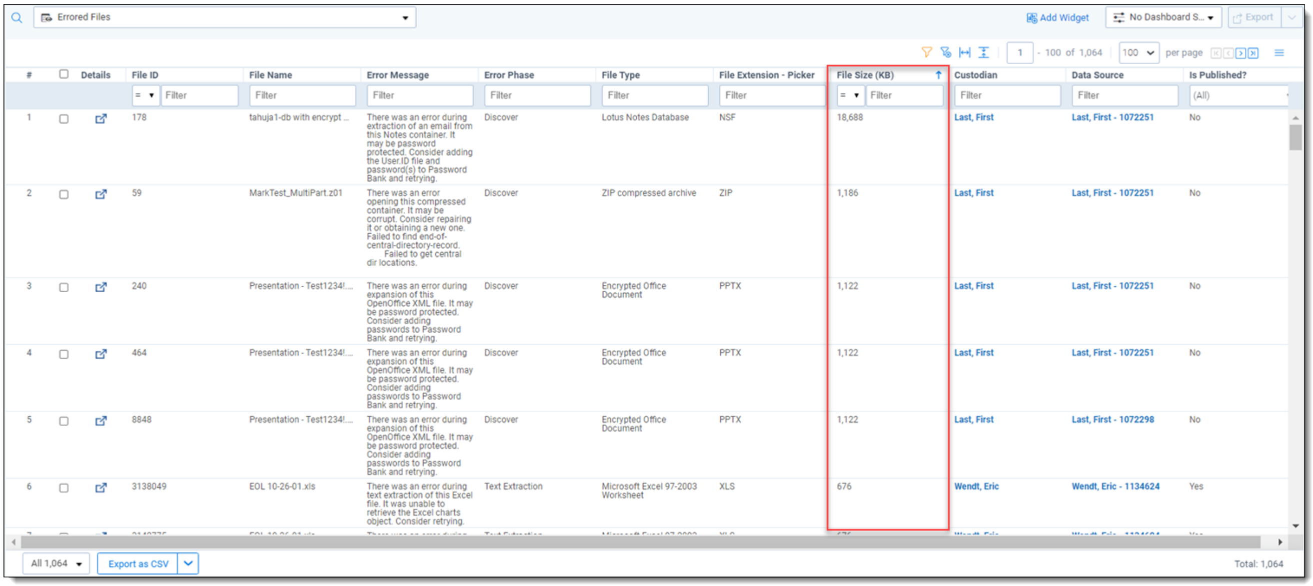
Only selected files with an error status of **Ready to retry** will be retried.

Click **Retry** on the retry errors confirmation modal.



#	☐	Details	File ID	File Name	File Type
1	☑		128	Password Protected PDF ...	Adobe Portable Document Format
2	☑		129	Password Protected Pow...	Encrypted Office Document
3	☑		130	Password Protected Wor...	Encrypted Office Document
4	☑		127	Password Protected Exce...	Encrypted Office Document
5	☑		126	Corrupt Word Document....	Microsoft Word document
6	☐			0000001793 - bad.M...	Lotus Notes Message File
7	☐			ple RTF.MSG	Lotus Notes Message File
8	☐			MSG	Lotus Notes Message File

Navigate to the **All Errored Files** view to see a record of all files with errors.

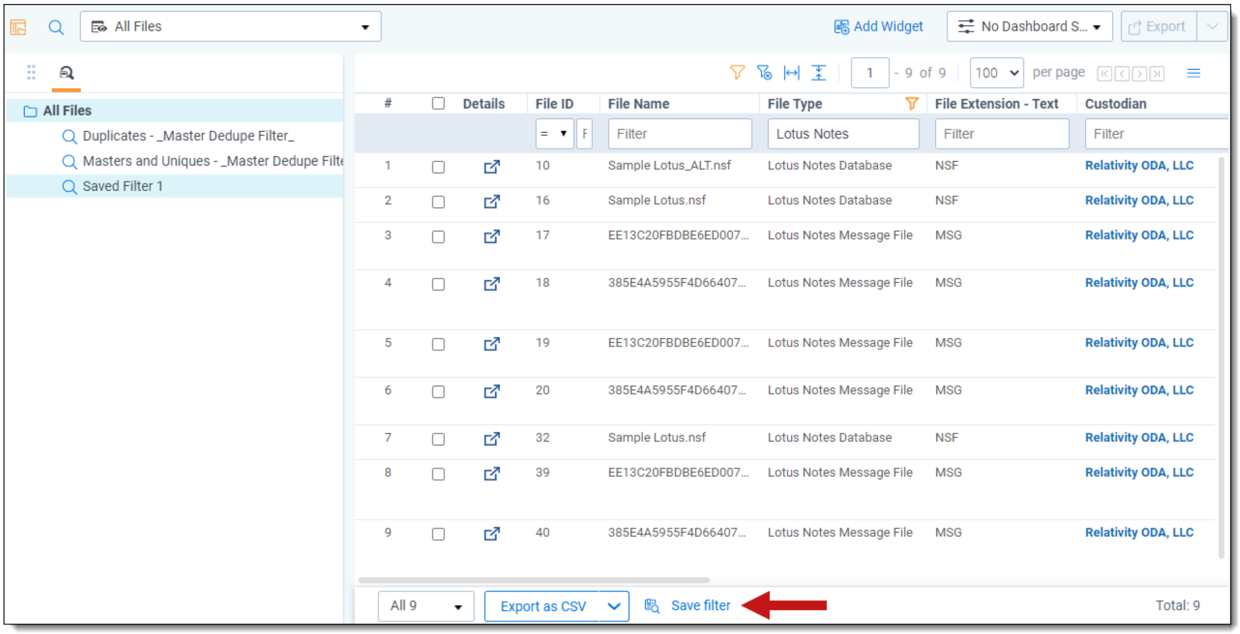


#	☐	Details	File ID	File Name	Error Message	Error Phase	File Type	File Extension - Picker	File Size (KB)	Custodian	Data Source	Is Published?
1	☐		178	tahuja1-db with encrypt ...	There was an error during extraction of an email from this Notes container. It may be password protected. Consider adding the User.ID file and password(s) to Password Bank and retrying.	Discover	Lotus Notes Database	NSF	18,688	Last, First	Last, First - 1072251	No
2	☐		59	MarkTest_MultiPart.z01	There was an error opening this compressed container. It may be corrupt. Consider repairing it or obtaining a new one. Failed to find end-of-central-directory-record. Failed to get central dir locations.	Discover	ZIP compressed archive	ZIP	1,186	Last, First	Last, First - 1072251	No
3	☐		240	Presentation - Test1234!....	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	Encrypted Office Document	PPTX	1,122	Last, First	Last, First - 1072251	No
4	☐		464	Presentation - Test1234!....	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	Encrypted Office Document	PPTX	1,122	Last, First	Last, First - 1072251	No
5	☐		8848	Presentation - Test1234!....	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	Encrypted Office Document	PPTX	1,122	Last, First	Last, First - 1072298	No
6	☐		3138049	EOL 10-26-01.xls	There was an error during text extraction of this Excel file. It was unable to retrieve the Excel charts object. Consider retrying.	Text Extraction	Microsoft Excel 97-2003 Worksheet	XLS	676	Wendt, Eric	Wendt, Eric - 1134624	Yes

17.6 Saved filters

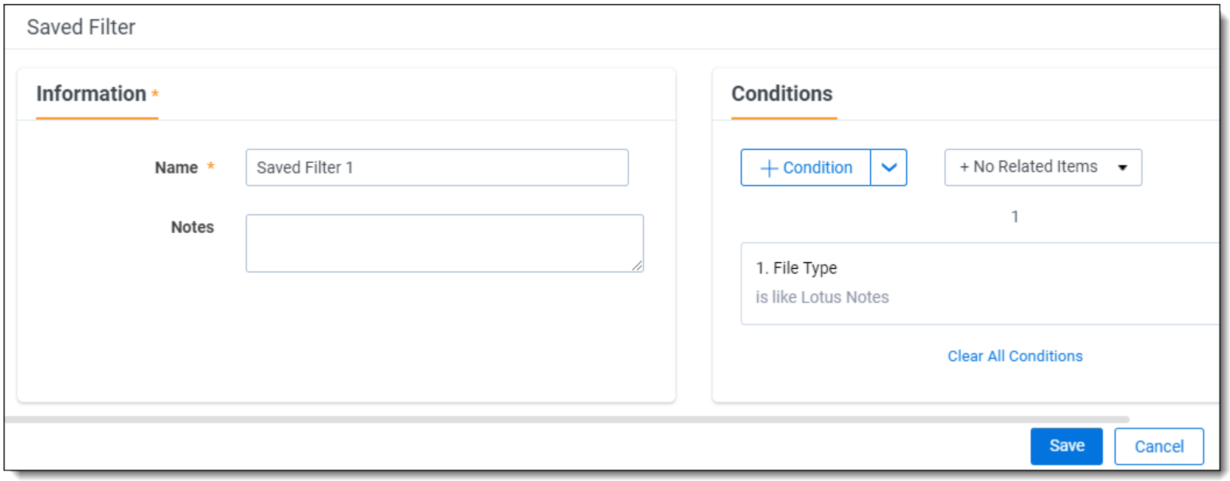
You can save any filters you set on any of the views in the Files tab and reuse them in future workflows. To do this:

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
All Files
Duplicates - _Master Dedupe Filter_
Masters and Uniques - _Master Dedupe Filte
Saved Filter 1
All 9
Export as CSV
Save filter
Total: 9

#	Details	File ID	File Name	File Type	File Extension - Text	Custodian
		=	Filter	Lotus Notes	Filter	Filter
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

2. In the Saved Filter modal, complete the Name, Notes (optional), and Conditions fields, and click **Save**.


Saved Filter
Information *
Conditions
+ Condition
+ No Related Items
Clear All Conditions
Save
Cancel

Name *: Saved Filter 1

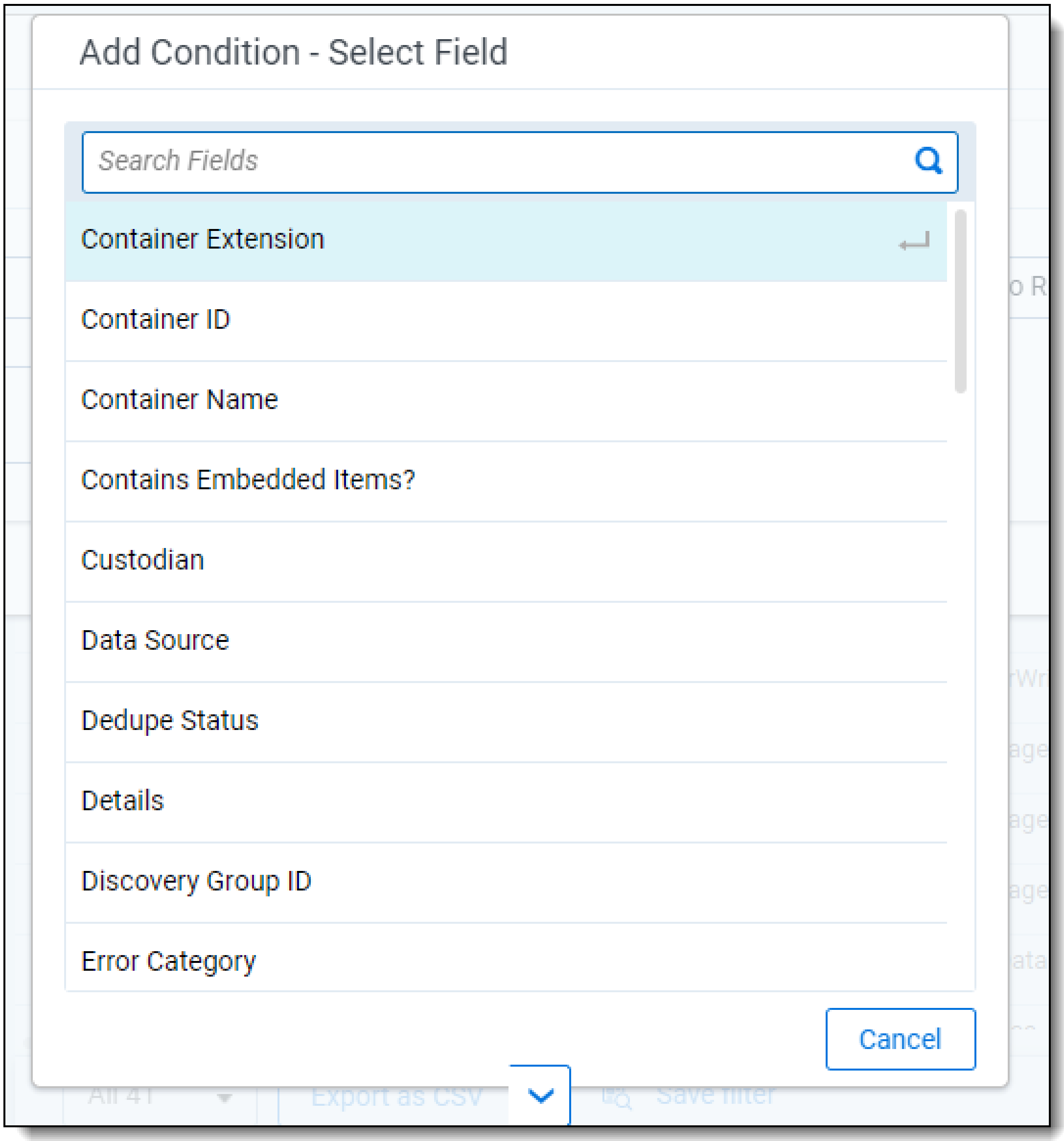
Notes:

1

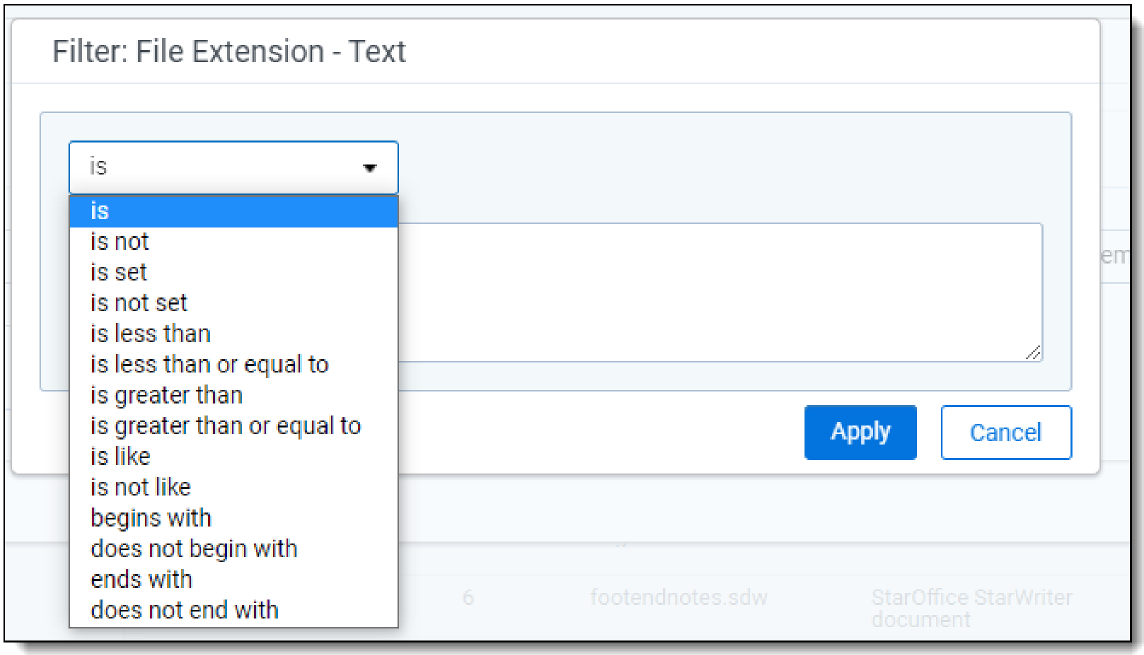
1. File Type
is like Lotus Notes

- **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

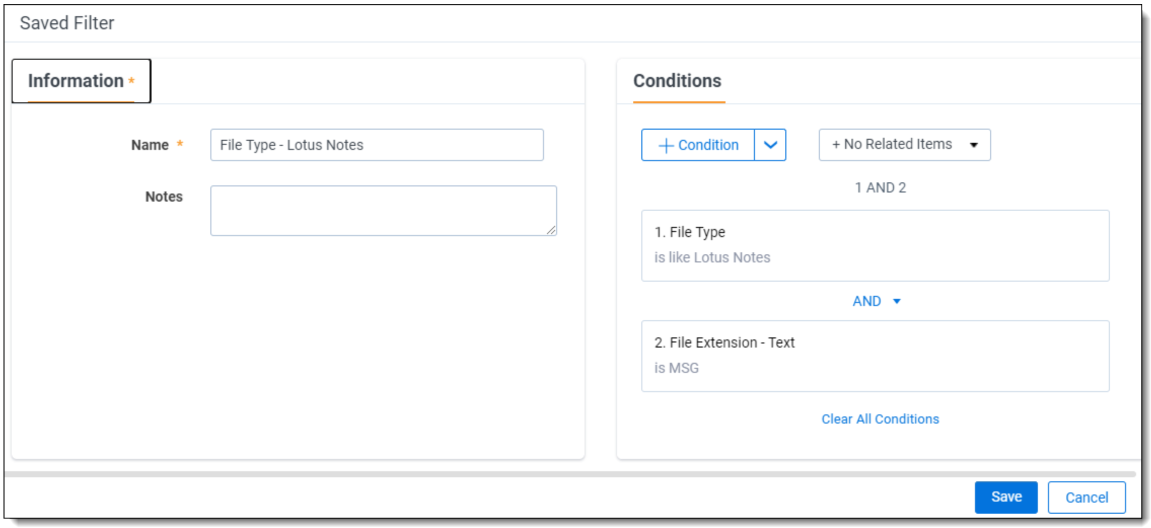
apply conditions.


Add Condition - Select Field
Search Fields
Container Extension
Container ID
Container Name
Contains Embedded Items?
Custodian
Data Source
Dedupe Status
Details
Discovery Group ID
Error Category
Cancel
All 41
Export as CSV
Save filter

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**.


Filter: File Extension - Text
is
is
is not
is set
is not set
is less than
is less than or equal to
is greater than
is greater than or equal to
is like
is not like
begins with
does not begin with
ends with
does not end with
Apply
Cancel
6
footendnotes.sdw
StarOffice StarWriter document

The new field(s) and condition(s) are visible in the Saved Filter modal. Click **Save** to save the filter set.


Saved Filter
Information *
Conditions
+ Condition
+ No Related Items
AND
Clear All Conditions
Save
Cancel

Name * File Type - Lotus Notes

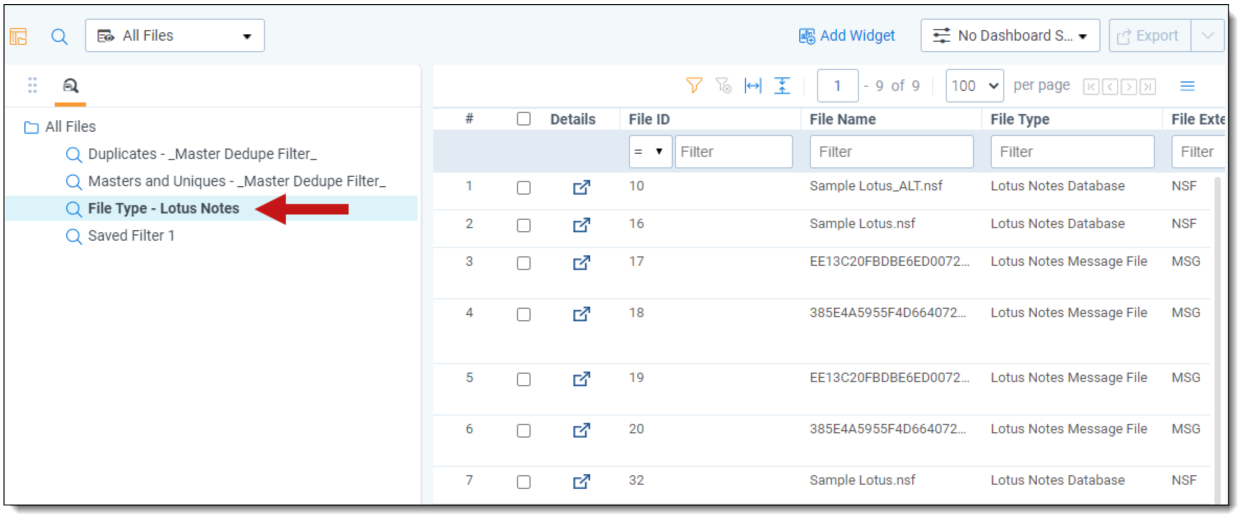
Notes

1 AND 2

1. File Type
is like Lotus Notes

2. File Extension - Text
is MSG

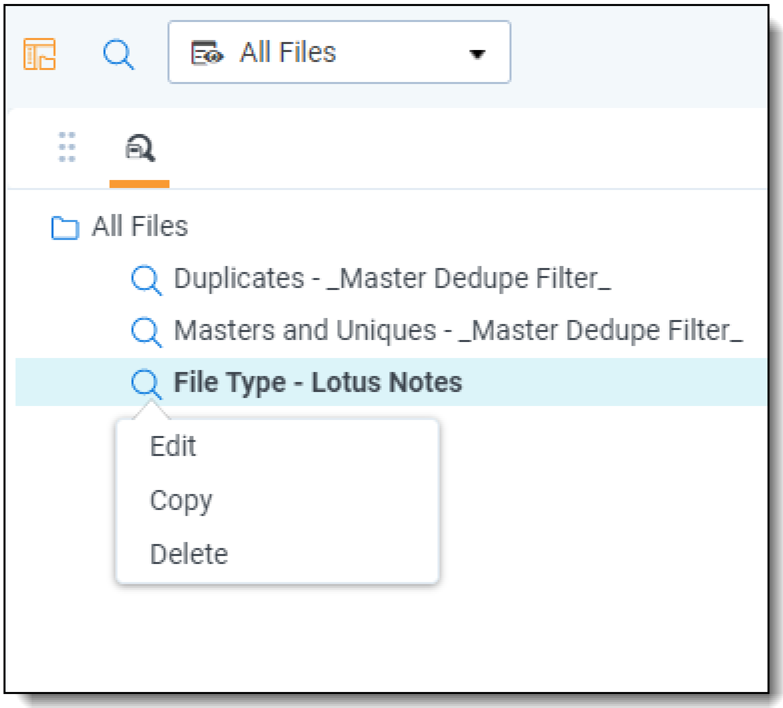
3. The saved filter is added to the list of Saved Searches.


All Files
Add Widget
No Dashboard S...
Export
1 - 9 of 9
100
per page
All Files
Duplicates - _Master Dedupe Filter_
Masters and Uniques - _Master Dedupe Filter_
File Type - Lotus Notes
Saved Filter 1

#	☐	Details	File ID	File Name	File Type	File Exte
1	☐		10	Sample Lotus_ALT.nsf	Lotus Notes Database	NSF
2	☐		16	Sample Lotus.nsf	Lotus Notes Database	NSF
3	☐		17	EE13C20FBDBE6ED0072...	Lotus Notes Message File	MSG
4	☐		18	385E4A5955F4D664072...	Lotus Notes Message File	MSG
5	☐		19	EE13C20FBDBE6ED0072...	Lotus Notes Message File	MSG
6	☐		20	385E4A5955F4D664072...	Lotus Notes Message File	MSG
7	☐		32	Sample Lotus.nsf	Lotus Notes Database	NSF

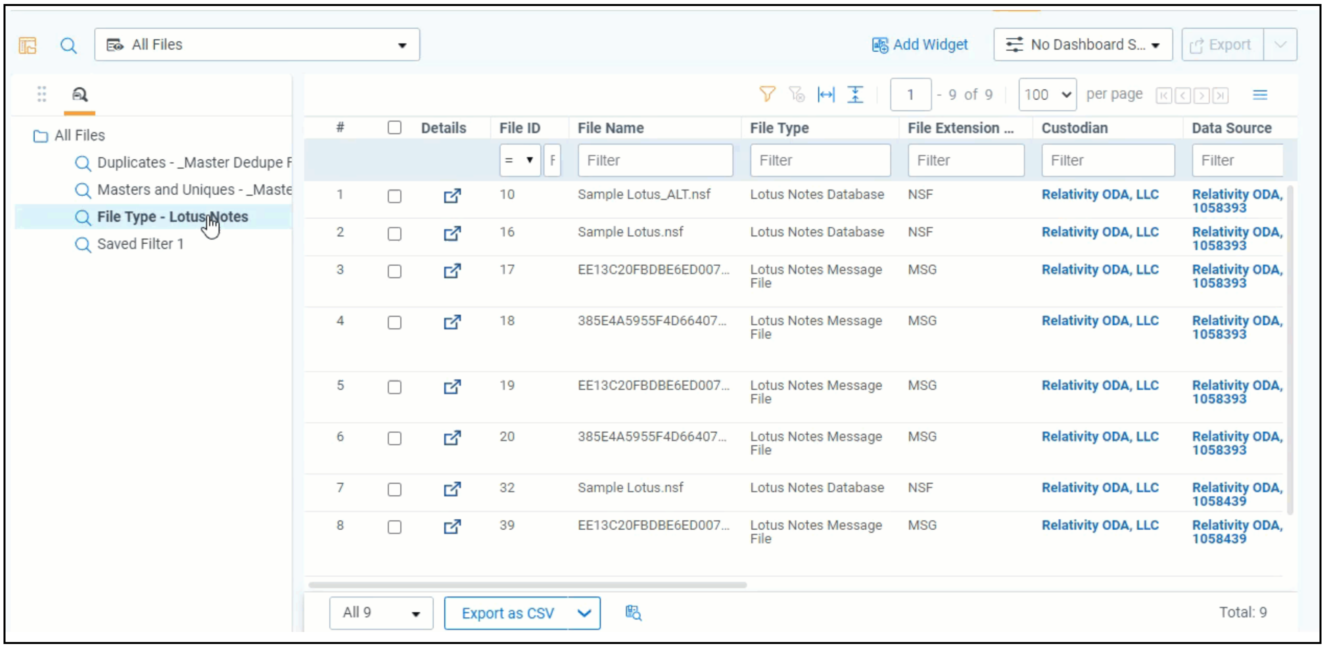
17.6.1 Right-click options

You can right-click any saved filter and choose to Edit, Copy, or Delete it.


All Files
All Files
Duplicates - _Master Dedupe Filter_
Masters and Uniques - _Master Dedupe Filter_
File Type - Lotus Notes
Edit
Copy
Delete

17.6.1.1 Edit

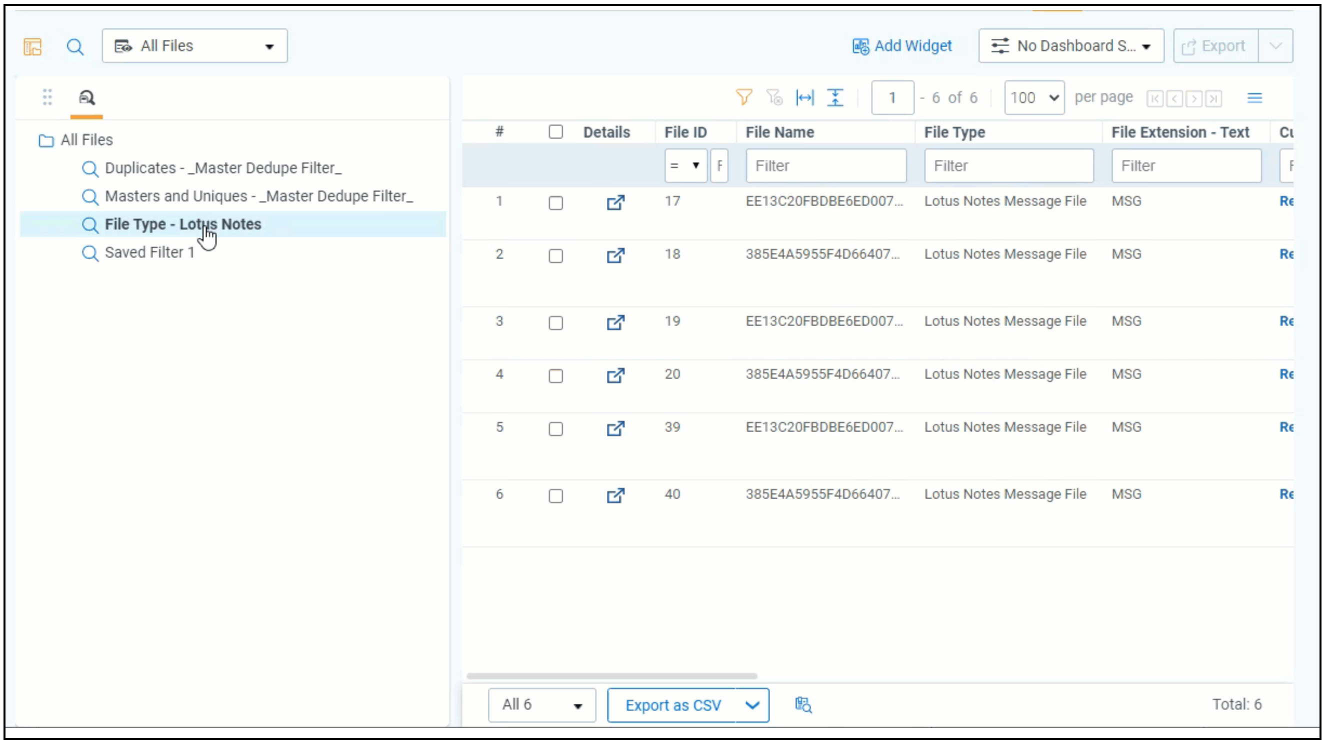
Clicking **Edit** opens the Saved Filter modal. You can add, remove, or modify fields and conditions, then click Save.


All Files
Add Widget
No Dashboard S...
Export
1 - 9 of 9
100
per page
All Files
Duplicates - _Master Dedupe F
Masters and Uniques - _Maste
File Type - Lotus Notes
Saved Filter 1
All 9
Export as CSV
Total: 9

#	☐	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	EE13C20FBDBE6ED007...	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	EE13C20FBDBE6ED007...	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	EE13C20FBDBE6ED007...	Lotus Notes Message File	MSG	Relativity ODA, LLC	Relativity ODA, 1058439

17.6.1.2 Copy

To make a copy of an existing filter set, right-click and select **Copy**.

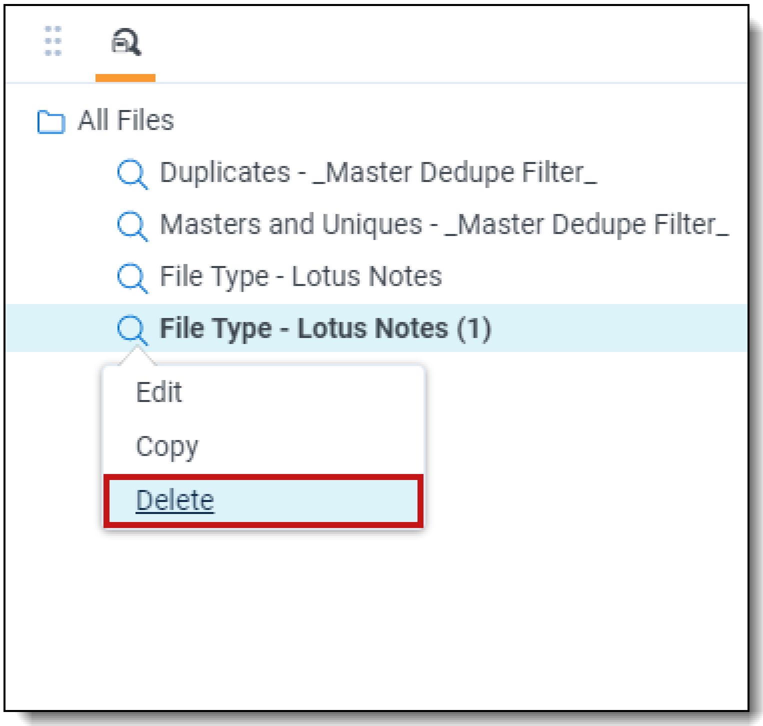

All Files
Add Widget
No Dashboard S...
Export
1 - 6 of 6
100
per page
All Files
Duplicates - _Master Dedupe Filter_
Masters and Uniques - _Master Dedupe Filter_
File Type - Lotus Notes
Saved Filter 1
All 6
Export as CSV
Total: 6

#	☐	Details	File ID	File Name	File Type	File Extension - Text	Cu
1	☐		17	EE13C20FBDBE6ED007...	Lotus Notes Message File	MSG	Re
2	☐		18	385E4A5955F4D66407...	Lotus Notes Message File	MSG	Re
3	☐		19	EE13C20FBDBE6ED007...	Lotus Notes Message File	MSG	Re
4	☐		20	385E4A5955F4D66407...	Lotus Notes Message File	MSG	Re
5	☐		39	EE13C20FBDBE6ED007...	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.

17.6.1.3 Delete

To remove a filter set from the saved filter browser entirely, right-click and select **Delete** .


All Files
Duplicates - _Master Dedupe Filter_
Masters and Uniques - _Master Dedupe Filter_
File Type - Lotus Notes
File Type - Lotus Notes (1)
Edit
Copy
Delete

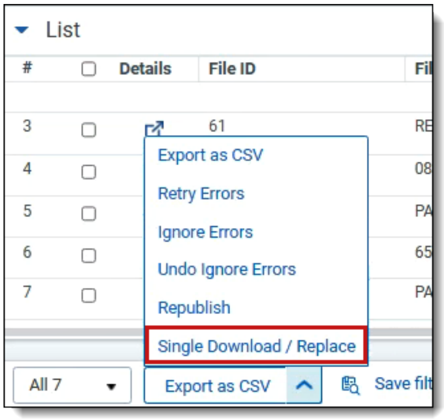
17.7 Single Download / Replace

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:

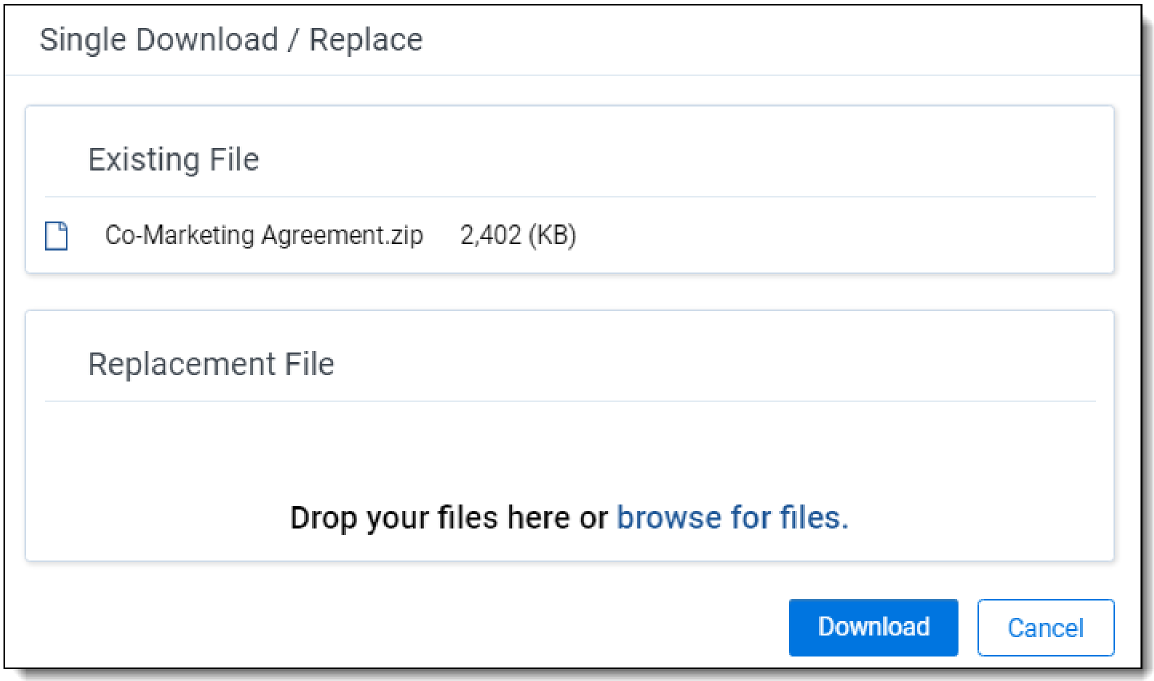
1. To take action on a specific file, select it, and then select the **Single Download / Replace** mass action.

   > **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.


List
#
Details
File ID
Fil
3
61
RE
4
08
5
PA
6
65
7
PA
Export as CSV
Retry Errors
Ignore Errors
Undo Ignore Errors
Republish
Single Download / Replace
All 7
Export as CSV
Save filt

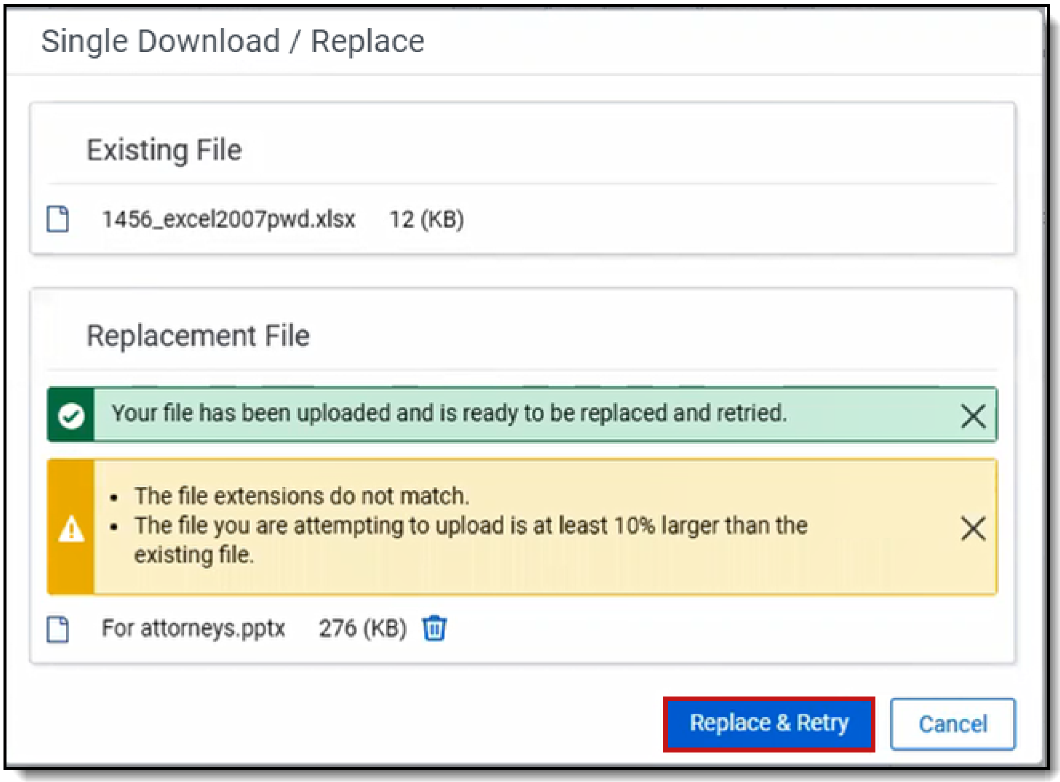
The Single Download / Replace dialog box opens allowing you to browse for or drop a replacement file.

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
For attorneys.pptx 276 (KB)
Replace & Retry
Cancel

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.

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.

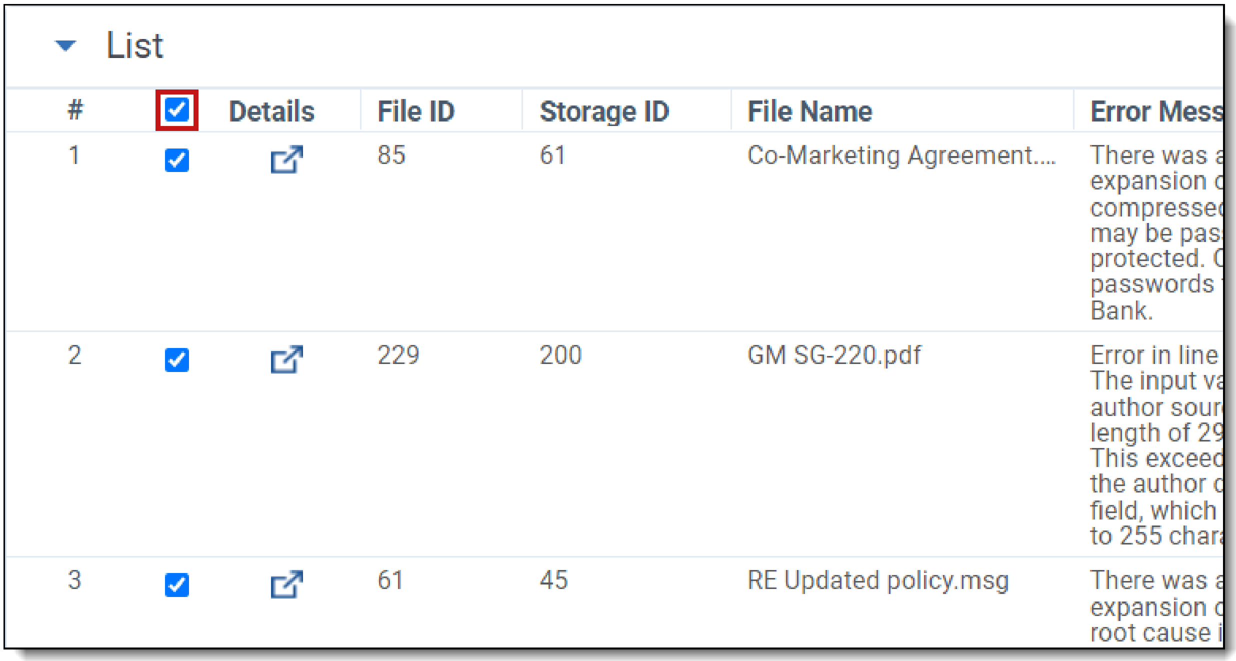
17.8 Mass download errored files

Use the Download mass operation to download errored files to your local drive. You can only use the Download mass operation on files with an error status of Not Resolved.

To mass download errored files:

1. Use the search bar to navigate to the **Files** page.
2. Select **Current Errored Files** 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.

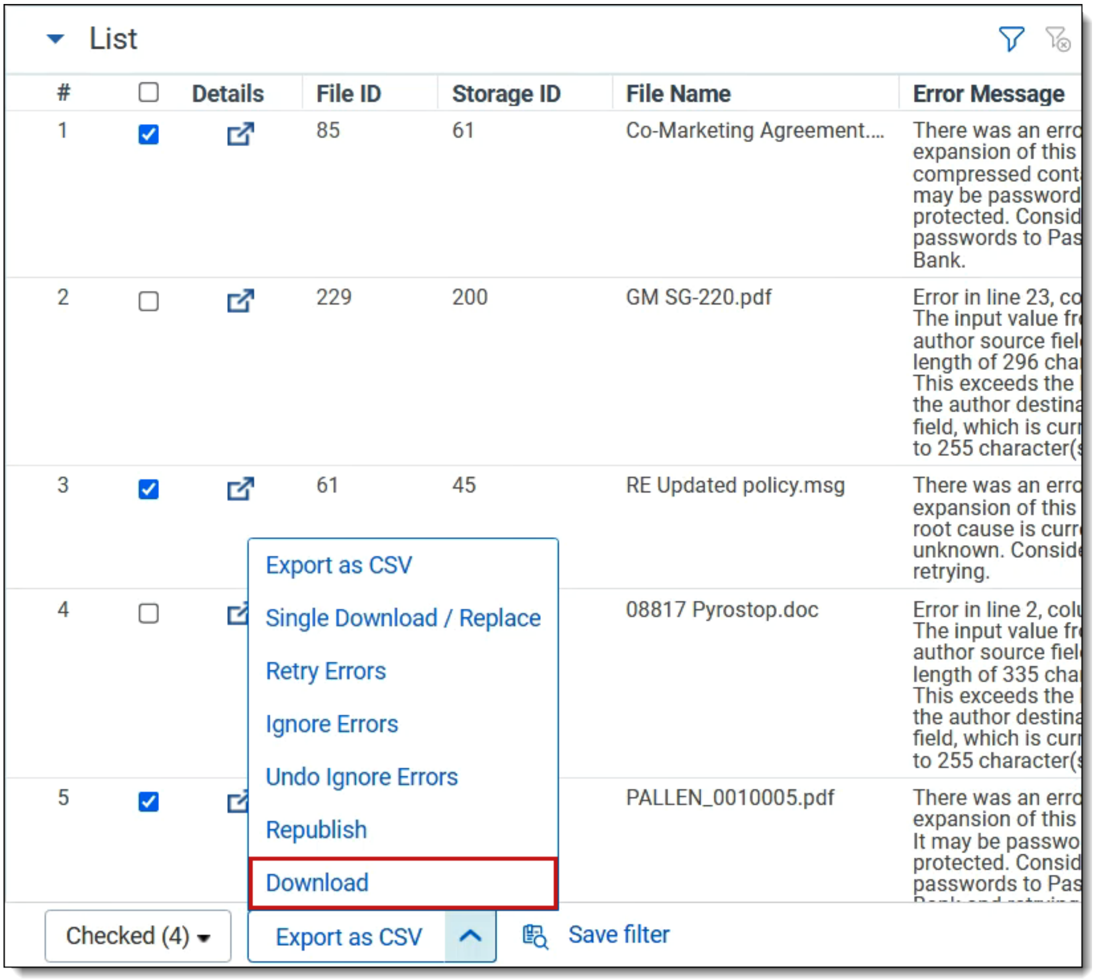
Note: Errored files must be in the Discover phase for mass download. You will see an error message if you attempt to download files in other phases, such as Publish or Text Extraction.



List

#	☑	Details	File ID	Storage ID	File Name	Error Mess
1	☑		85	61	Co-Marketing Agreement....	There was a expansion c 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 This exceed the author c field, which to 255 chara
3	☑		61	45	RE Updated policy.msg	There was a expansion c root cause i

4. Select **Download** from the mass operation drop-down menu.



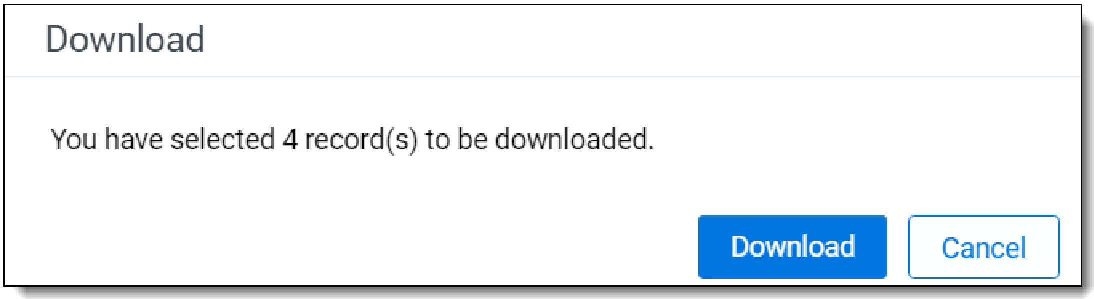
List

#	☐	Details	File ID	Storage ID	File Name	Error Message
1	☑		85	61	Co-Marketing Agreement....	There was an erro expansion of this compressed cont may be password protected. Consid passwords to Pas Bank.
2	☐		229	200	GM SG-220.pdf	Error in line 23, co The input value fr author source fiel length of 296 cha This exceeds the the author destina field, which is curr to 255 character(
3	☑		61	45	RE Updated policy.msg	There was an erro expansion of this root cause is curr unknown. Conside retrying.
4	☐				08817 Pyrostop.doc	Error in line 2, colu The input value fr author source fiel length of 335 cha This exceeds the the author destina field, which is curr to 255 character(
5	☑				PALLEN_0010005.pdf	There was an erro expansion of this It may be passwo protected. Consid passwords to Pas

Export as CSV
Single Download / Replace
Retry Errors
Ignore Errors
Undo Ignore Errors
Republish
Download
Checked (4)
Export as CSV
Save filter

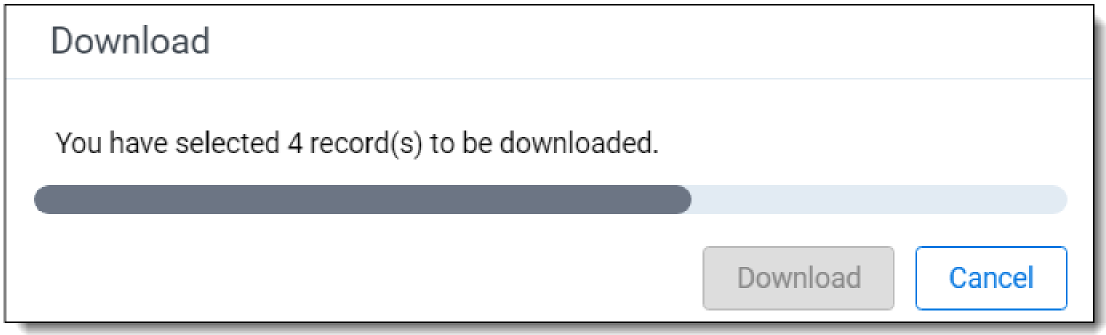
A confirmation message appears that displays the number of files downloading.

Note: Relativity downloads unique files from each selection. The number of files selected may differ from the number of files downloaded if duplicates exist.



You have selected 4 record(s) to be downloaded.

5. Click **Download**.
6. A status message appears that displays the overall progress.



You have selected 4 record(s) to be downloaded.

When complete, files are packaged in a zip container and saved to your local drive.

18 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:

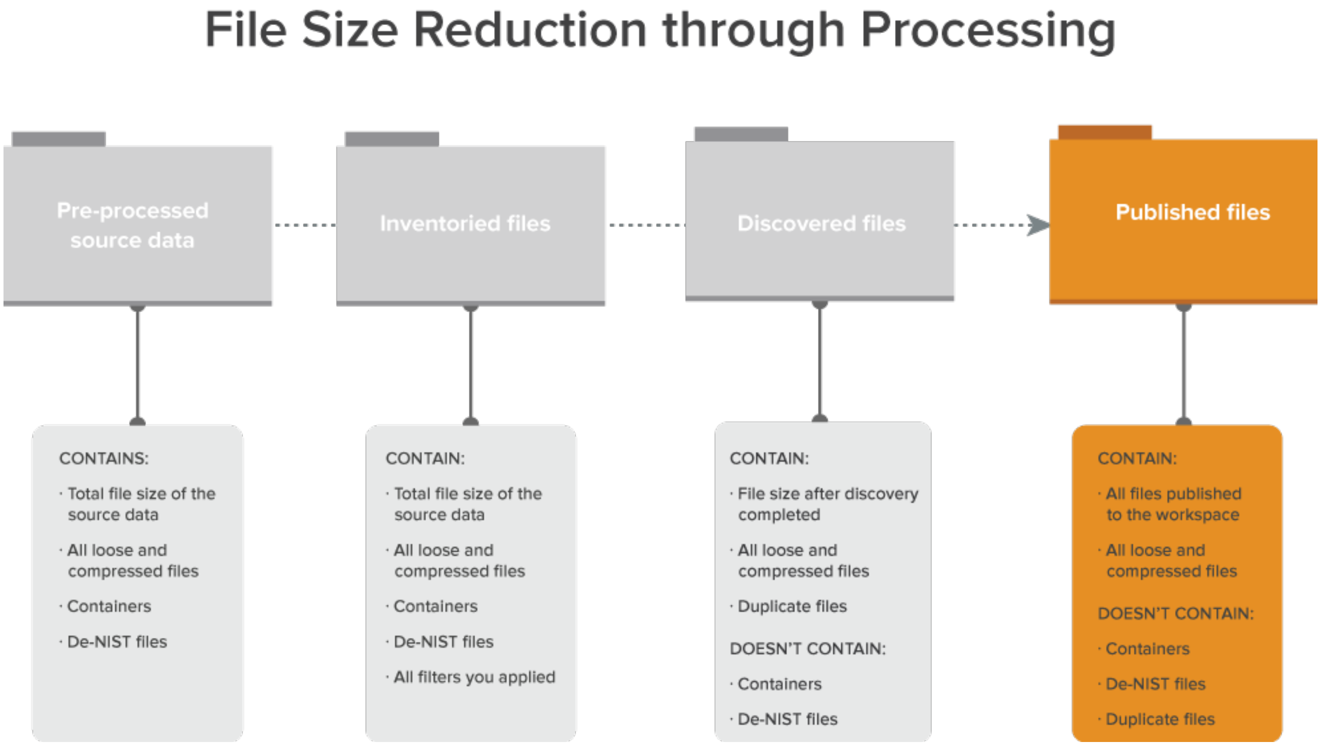
- Applies all the settings you specified on the profile to the documents you bring into the workspace.
- Determines which is the primary document and primary custodian and which are the duplicates.
- Populates the All Custodians, Other Sources, and other fields with data.

Note: For details on deleting files after publishing, see Post-publish delete.

Use the following guidelines when publishing files to a workspace:

- 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'd use to reduce the file size of a data set through processing. This workflow assumes that you’re applying some method of de-NIST and deduplication.


File Size Reduction through Processing
Pre-processed source data
Inventoried files
Discovered files
Published files
CONTAINS:
· Total file size of the source data
· All loose and compressed files
· Containers
· De-NIST files
CONTAIN:
· Total file size of the source data
· All loose and compressed files
· Containers
· De-NIST files
· All filters you applied
CONTAIN:
· File size after discovery completed
· All loose and compressed files
· Duplicate files
DOESN'T CONTAIN:
· Containers
· De-NIST files
CONTAIN:
· All files published to the workspace
· All loose and compressed files
DOESN'T CONTAIN:
· Containers
· De-NIST files
· Duplicate files

Note: Your environment has been enabled to dynamically scale your Invariant worker servers dependent on load. Sustained activity is automatically detected by the system, and Relativity will add workers to handle this work. Once the work is done, they will automatically scale back down. This feature is continually being improved to be smarter about when we add workers and how many we add.

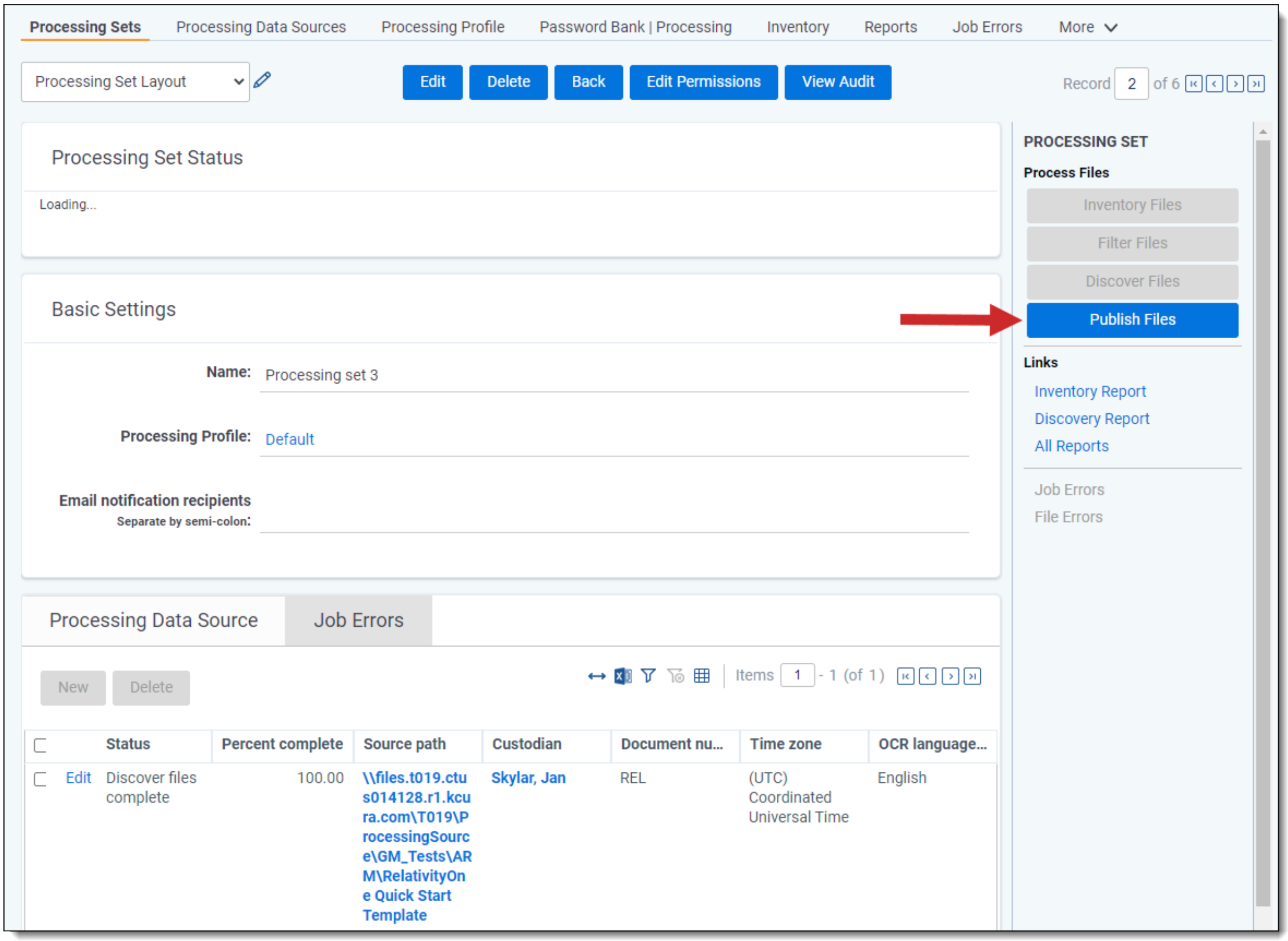
The following is a typical workflow that incorporates publish:

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.

18.1 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, and Primary Date/Time. The null value is excluded and not represented in the filtered list.


Processing Sets
Processing Data Sources
Processing Profile
Password Bank | Processing
Inventory
Reports
Job Errors
More
Processing Set Layout
Edit
Delete
Back
Edit Permissions
View Audit
Record 2 of 6
PROCESSING SET
Process Files
Inventory Files
Filter Files
Discover Files
Publish Files
Links
Inventory Report
Discovery Report
All Reports
Job Errors
File Errors
Processing Set Status
Basic Settings
Processing Data Source
Job Errors
New
Delete
Items 1 - 1 (of 1)

Loading...

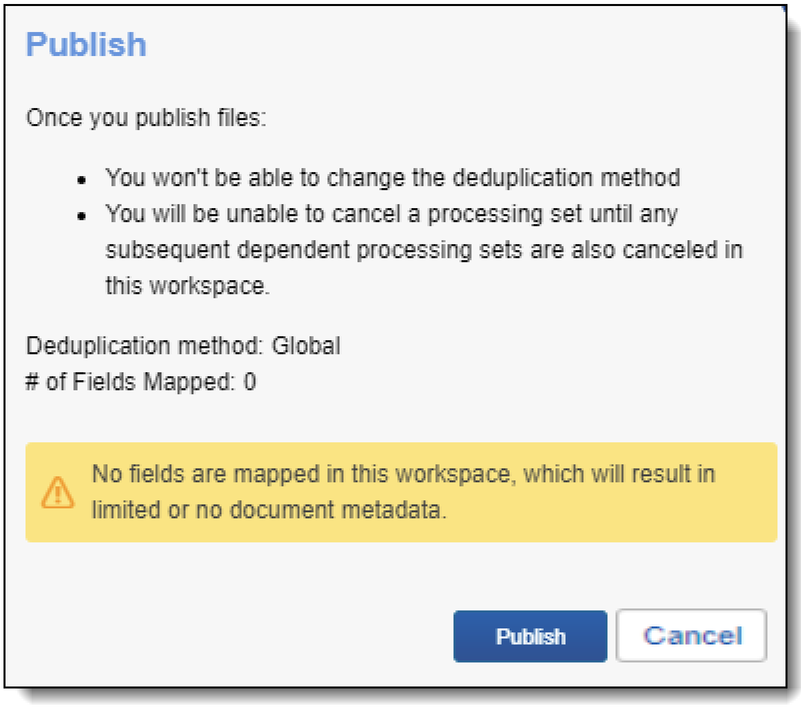
Name: Processing set 3

Processing Profile: Default

Email notification recipients Separate by semi-colon:

☐	Status	Percent complete	Source path	Custodian	Document nu...	Time zone	OCR language...
☐ Edit	Discover files complete	100.00	\\files.t019.ctus014128.r1.kcura.com\T019\ProcessingSource\GM_Tests\ARM\RelativityOne Quick Start Template	Skylar, Jan	REL	(UTC) Coordinated Universal Time	English

When you click Publish Files, you're presented with a confirmation message containing information about the job you're about to submit. If you haven't mapped any fields in the workspace, the message will reflect this. Click **Publish** to proceed or **Cancel** to return to the processing set layout.


Publish

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.

Deduplication method: Global
# of Fields Mapped: 0

⚠ No fields are mapped in this workspace, which will result in limited or no document metadata.

Consider the following when publishing files:

- During publish, Relativity assigns control numbers to documents from the top of the directory (source location) down. Duplicates do not receive unique control numbers.
- 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.
- Never disable a worker while it's completing a publish job.
- 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've arranged for auto-publish 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 aren't required to keep your source files in the location from which they were processed after you've 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 in the worker manager server, 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 will read "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 Support for guidance on what values to specify for these settings before performing an upgrade.
        - ProcessingMaxPublishJobCountPerRelativitySQLServer - 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 will result 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 can't allocate more jobs per workspace than what is allowed per SQL server.
        - ProcessingMaxPublishSubJobCountPerWorkspace - 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 will result 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 can't 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:

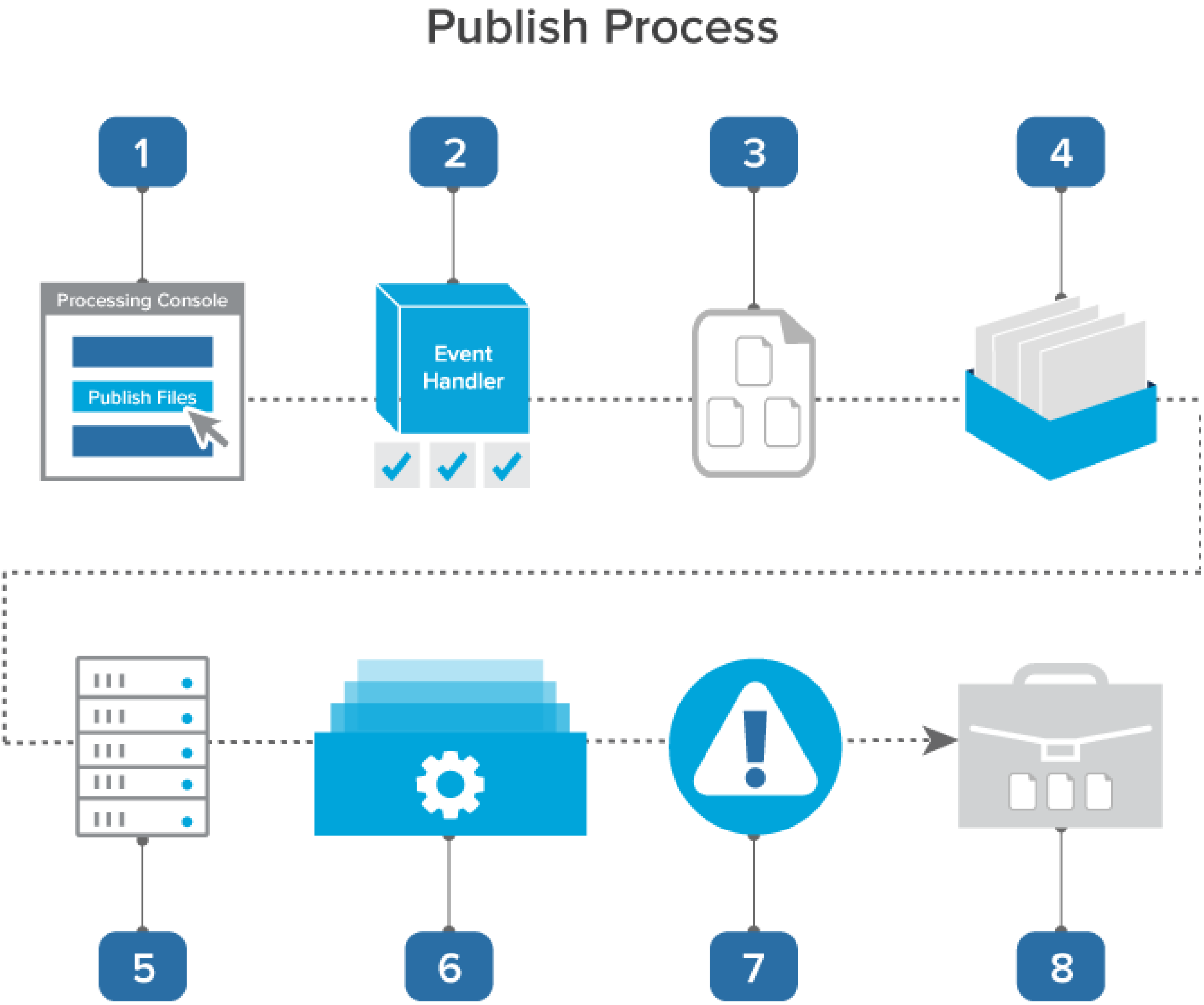
Environment setup	ProcessingMaxPublishSubJobCountPerWorkspace	ProcessingMaxPublishJobCountPerRelativitySQLServer
Tier 1 (see the System Requirements Guide for details)	3	7
Tier 2 (see the System Requirements Guide for details)	6	12
RelativityOne baseline	3	7

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.

18.1.1 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.



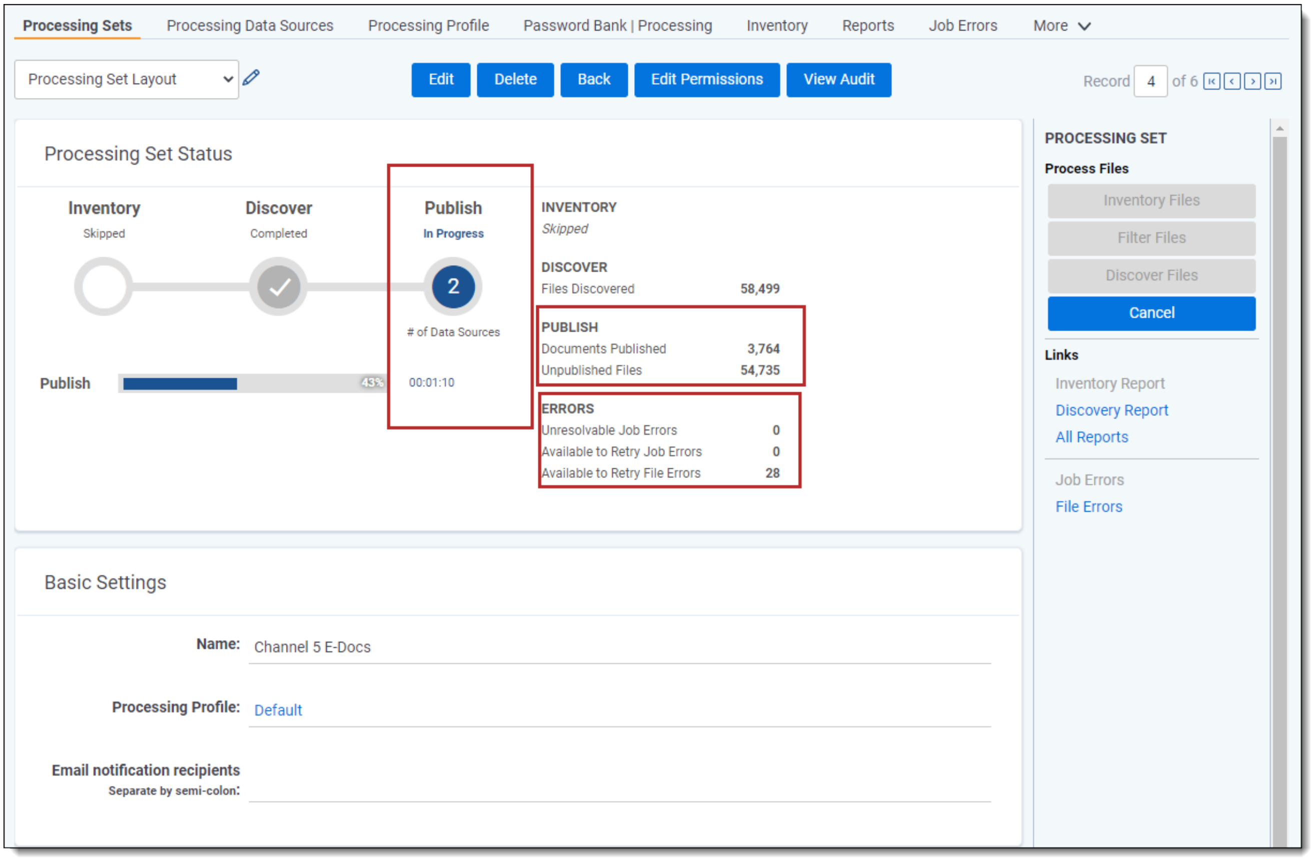
1. You click Publish Files on the processing set console. If you've arranged for auto-publish after discovery, publish will begin automatically and you aren't required to start it manually.
2. A console event handler checks to make sure that the set is valid and ready to proceed.
3. The event handler 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 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 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.
6. 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 performed by the worker manager server, in that it operates on batches of data for a specific amount of time before completing each transactional job and moving on.

7. 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.
8. 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.

18.2 Monitoring publish status

You can monitor the progress of the publish job through the information provided in the Processing Set Status display on the set layout.



Processing Sets
Processing Data Sources
Processing Profile
Password Bank | Processing
Inventory
Reports
Job Errors
More
Processing Set Layout
Edit
Delete
Back
Edit Permissions
View Audit
Record 4 of 6
Processing Set Status
Inventory
Skipped
Discover
Completed
Publish
In Progress
2
# of Data Sources
Publish
43%
00:01:10
PROCESSING SET
Process Files
Inventory Files
Filter Files
Discover Files
Cancel
Links
Inventory Report
Discovery Report
All Reports
Job Errors
File Errors
Basic Settings

INVENTORY
Skipped

DISCOVER
Files Discovered 58,499

PUBLISH
Documents Published 3,764
Unpublished Files 54,735

ERRORS
Unresolvable Job Errors 0
Available to Retry Job Errors 0
Available to Retry File Errors 28

Name: Channel 5 E-Docs
Processing Profile: Default
Email notification recipients
Separate by semi-colon:

Through this display, you can monitor the following:

- **# of Data Sources** - the number of data sources currently in the processing queue.
- **Publish | Documents Published** - the number of files across all data sources submitted that have been published to the workspace.

- **Publish | Unpublished Files** - the number of files across all data sources submitted that have yet to be published to the workspace.
- **Errors** - the number of errors that have occurred across all data sources submitted, which fall into the following categories:
    - **Unresolvable** - errors that you can't retry.
    - **Available to Retry** - errors that are available for retry.
    - **In Queue** - errors that you have submitted for retry and are currently in the processing queue.

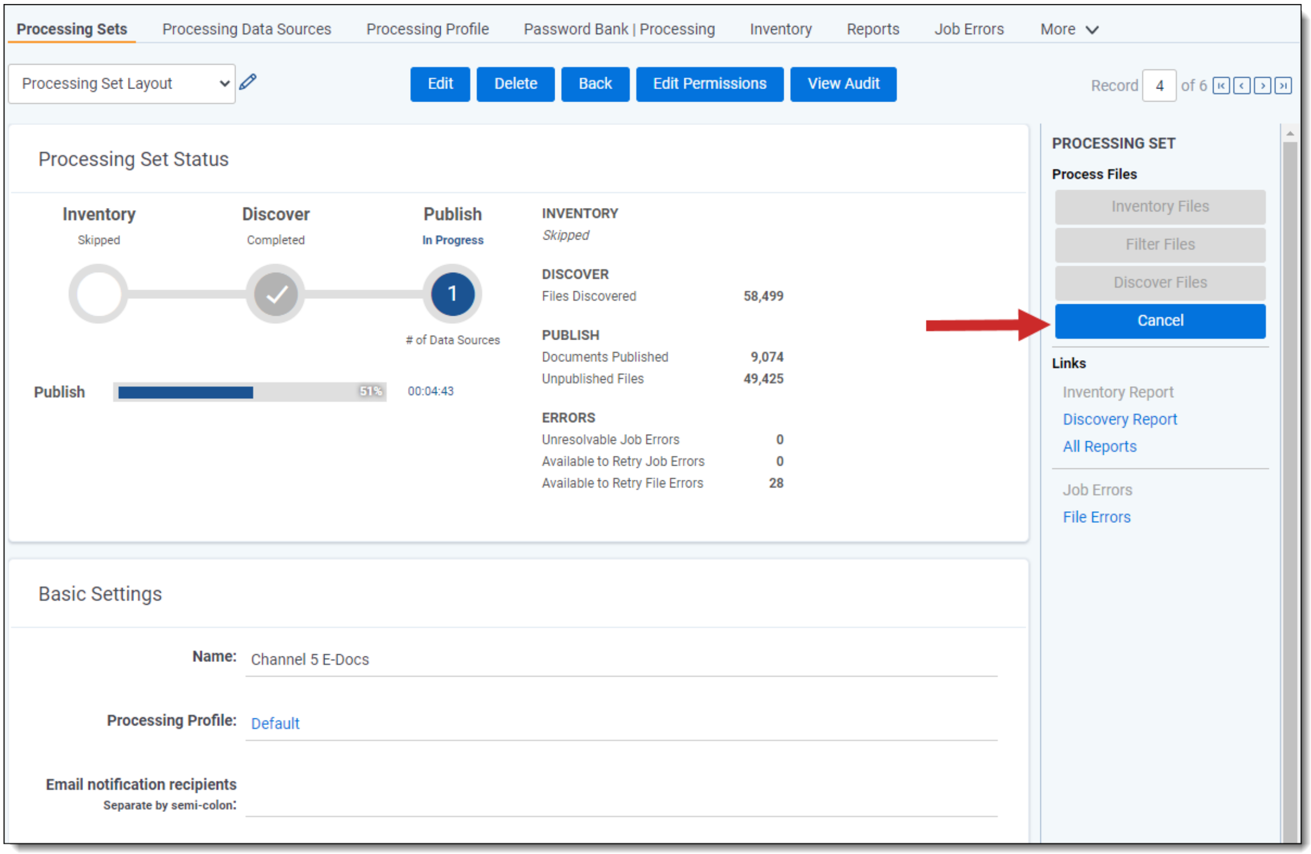
See Processing error overview for details.

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.

18.3 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
Inventory
Reports
Job Errors
More
Processing Set Layout
Edit
Delete
Back
Edit Permissions
View Audit
Record 4 of 6
Processing Set Status
Inventory
Skipped
Discover
Completed
Publish
In Progress
1
# of Data Sources
Publish
51%
00:04:43
PROCESSING SET
Process Files
Inventory Files
Filter Files
Discover Files
Cancel
Links
Inventory Report
Discovery Report
All Reports
Job Errors
File Errors
Basic Settings

INVENTORY
Skipped

DISCOVER

Files Discovered	58,499

PUBLISH

Documents Published	9,074
Unpublished Files	49,425

ERRORS

Unresolvable Job Errors	0
Available to Retry Job Errors	0
Available to Retry File Errors	28

Name: Channel 5 E-Docs

Processing Profile: Default

Email notification recipients
Separate by semi-colon:

Consider the following about canceling publish:

- You can't 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 can't be retried.
- Once the agent picks up the cancel publish job, you can't delete or edit those data sources.

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.

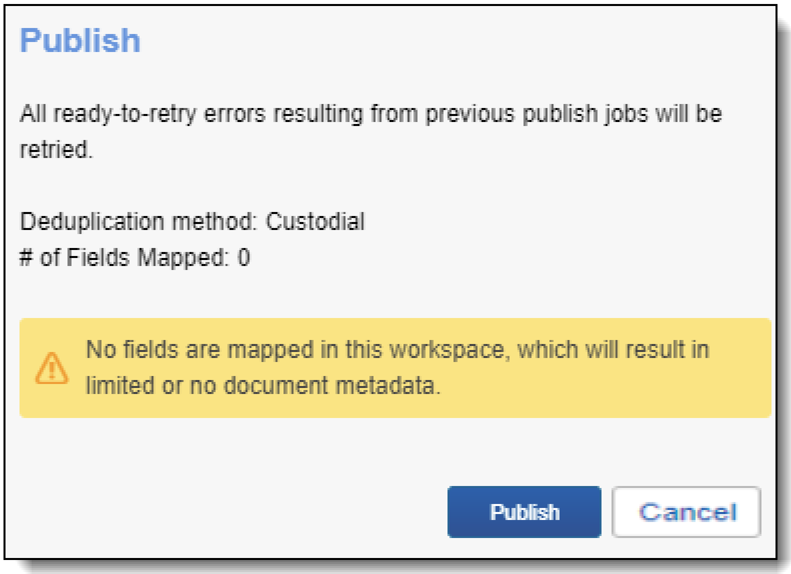
18.4 Republishing a processing set

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 from the Files tab on page 198 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 doesn't re-copy the settings from the profile to the data sources that you are publishing.

When you click Publish Files again, you're presented with a confirmation message containing information about the job you're about to submit. If you haven't mapped any fields in the workspace, the message will reflect this. Click **Publish** to proceed or **Cancel** to return to the processing set layout.


Publish

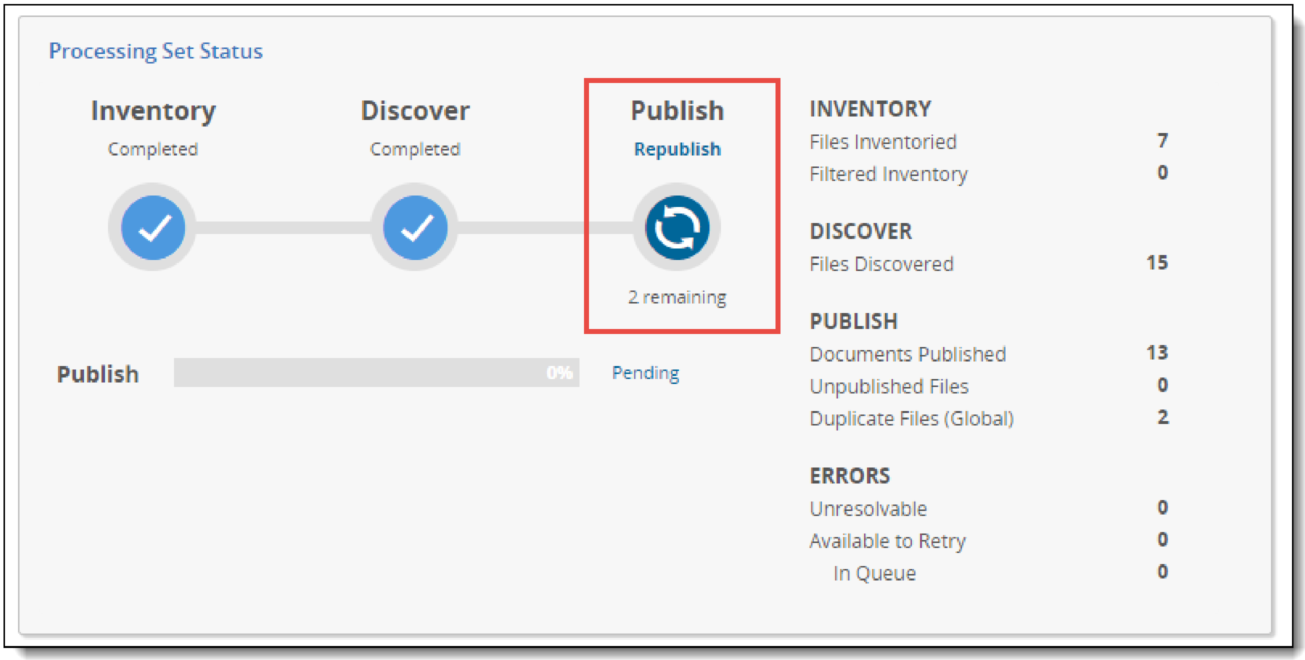
All ready-to-retry errors resulting from previous publish jobs will be retried.

Deduplication method: Custodial
# of Fields Mapped: 0

⚠ No fields are mapped in this workspace, which will result in limited or no document metadata.

Publish
Cancel

The status section is updated to display the in-progress republish job.


Processing Set Status
Inventory
Completed
Discover
Completed
Publish
Republish
2 remaining
Publish
0%
Pending

INVENTORY	
Files Inventoried	7
Filtered Inventory	0
DISCOVER	
Files Discovered	15
PUBLISH	
Documents Published	13
Unpublished Files	0
Duplicate Files (Global)	2
ERRORS	
Unresolvable	0
Available to Retry	0
In Queue	0

Consider the following when republishing files:

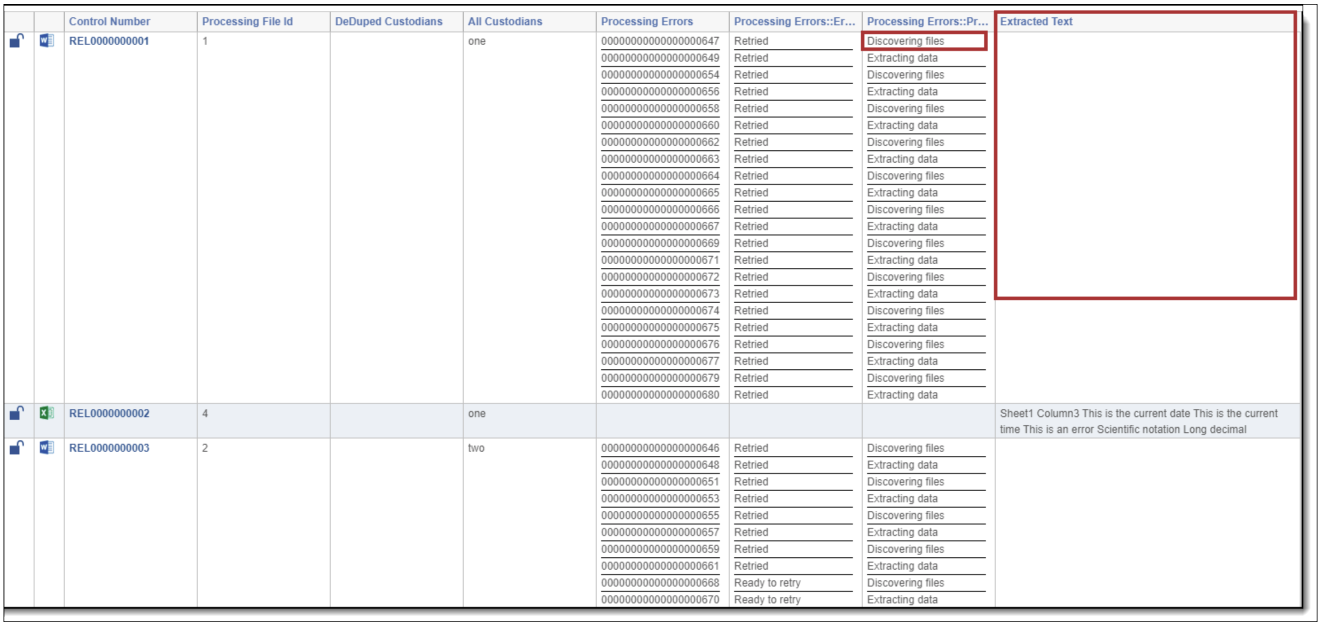
- All ready-to-retry errors resulting from this publish job are retried when you republish.
- Deduplication is respected on republish.
- When you resolve errors and republish the documents that contained those errors, Relativity performs an overlay, meaning that there's only one file for the republished document in the Documents tab.

- When you republish data, Relativity only updates field mappings for files that previously returned errors.
- 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 will be disallowed.
- Changes made to numbering type in a processing profile will not be respected after initial publishing. Data Source information cannot be changed after initial publishing.

18.5 Retrying errors after publish

You have the option of retrying errors generated during file discovery. When you discover corrupt or password-protected documents, these error files are still published into a Relativity workspace with their file metadata. This is important to remember if you have Auto-publish enabled. However, for documents with these types of errors, neither the document metadata nor the extracted text is available in the workspace.

Note: File metadata is derived from the file's operating system (e.g., File Extension) whereas document metadata is contained in the document itself (e.g., Is Embedded).



		Control Number	Processing File Id	DeDuped Custodians	All Custodians	Processing Errors	Processing Errors::Er...	Processing Errors::Pr...	Extracted Text
		REL0000000001	1		one	0000000000000000647 0000000000000000649 0000000000000000654 0000000000000000656 0000000000000000658 0000000000000000660 0000000000000000662 0000000000000000663 0000000000000000664 0000000000000000665 0000000000000000666 0000000000000000667 0000000000000000669 0000000000000000671 0000000000000000672 0000000000000000673 0000000000000000674 0000000000000000675 0000000000000000676 0000000000000000677 0000000000000000679 0000000000000000680	Retried Retried	Discovering files Extracting data Discovering files Extracting data Discovering files Extracting data Discovering files Extracting data Discovering files Extracting data Discovering files Extracting data Discovering files Extracting data Discovering files Extracting data Discovering files Extracting data Discovering files Extracting data Discovering files Extracting data	
		REL0000000002	4		one				Sheet1 Column3 This is the current date This is the current time This is an error Scientific notation Long decimal
		REL0000000003	2		two	0000000000000000646 0000000000000000648 0000000000000000651 0000000000000000653 0000000000000000655 0000000000000000657 0000000000000000659 0000000000000000661 0000000000000000668 0000000000000000670	Retried Retried Retried Retried Retried Retried Retried Retried Ready to retry Ready to retry	Discovering files Extracting data Discovering files Extracting data Discovering files Extracting data Discovering files Extracting data Discovering files Extracting data	

For resolvable issues such as password-protected files, you can retry these errors even after you publish the files into a workspace. If you provide a password via the password bank and successfully retry the file, then its document metadata and extracted text are made available in the workspace after the documents are republished.

19 Post-publish delete

Note: Some functionality described in this topic has been released via a phased rollout, meaning that certain customers receive it before others. For that reason, you may be seeing documentation for features that you do not yet have in your Relativity instance. For more information on phased rollouts see the Phased Rollout FAQ on the Community.

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.

19.1 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.

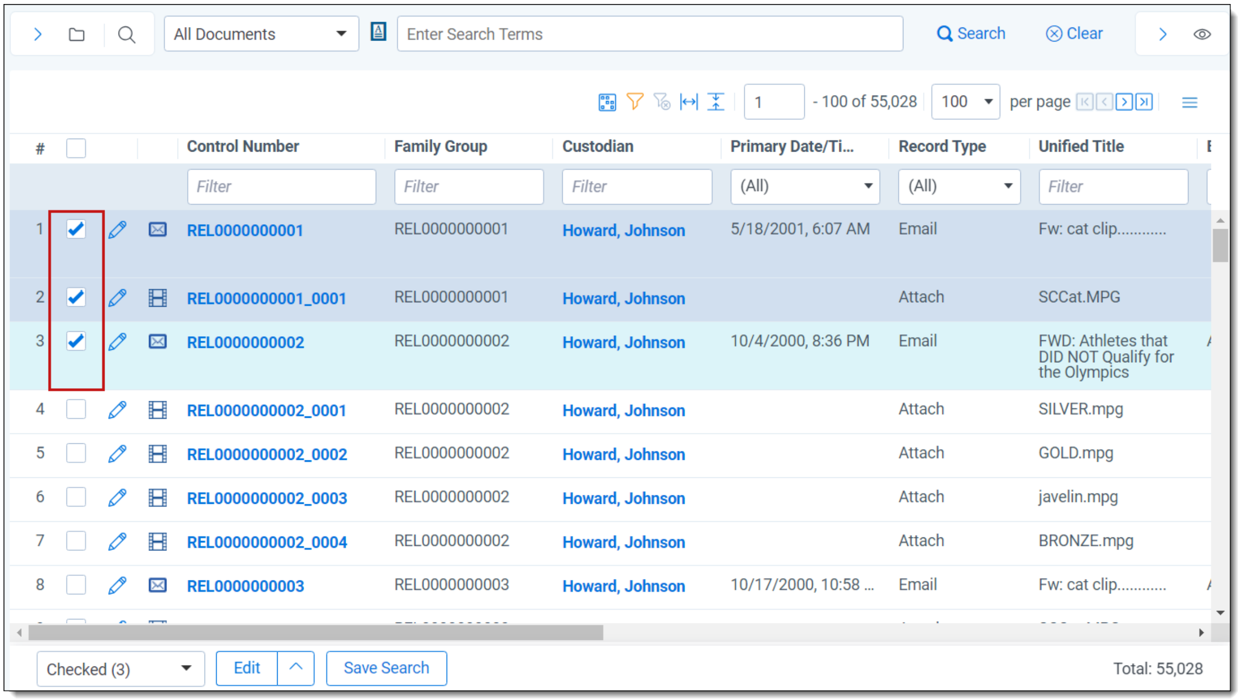
Since Relativity removes all duplicates within the same data source, you do not need to work with any automatically promoted files.

For information on reporting what was deleted, see Master Document Replacement Summary.

19.2 Running post-publish delete

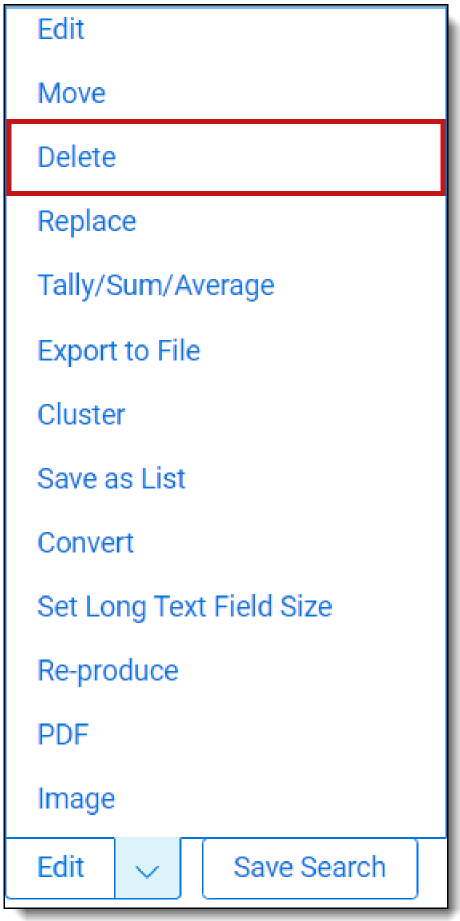
To mass delete documents after discovery and publish, perform the following:

1. Navigate to the Documents tab and select the documents you want to delete.


All Documents
Enter Search Terms
Search
Clear
1
- 100 of 55,028
100
per page
Checked (3)
Edit
Save Search
Total: 55,028

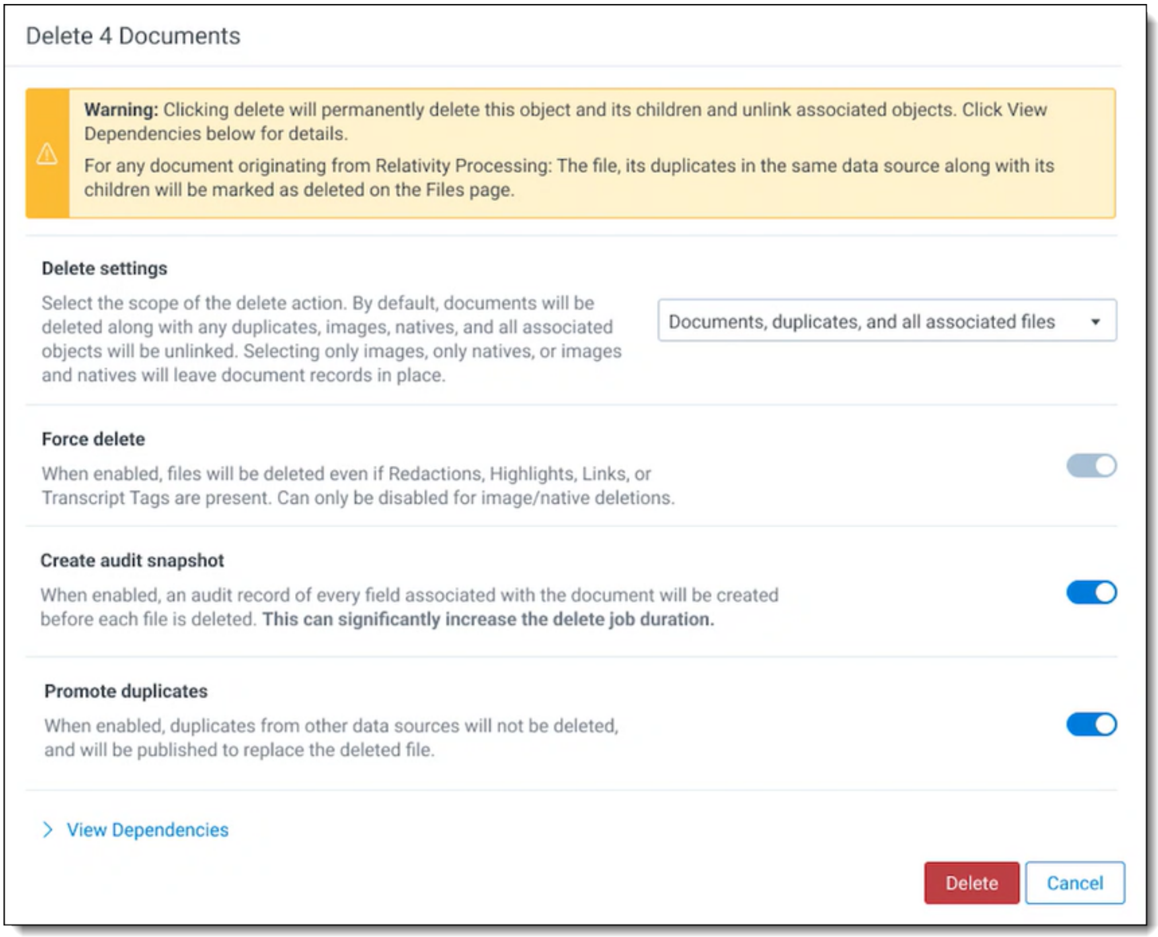
#	☐	Control Number	Family Group	Custodian	Primary Date/Ti...	Record Type	Unified Title
1	☑	REL0000000001	REL0000000001	Howard, Johnson	5/18/2001, 6:07 AM	Email	Fw: cat clip...........
2	☑	REL0000000001_0001	REL0000000001	Howard, Johnson		Attach	SCCat.MPG
3	☑	REL0000000002	REL0000000002	Howard, Johnson	10/4/2000, 8:36 PM	Email	FWD: Athletes that DID NOT Qualify for the Olympics
4	☐	REL0000000002_0001	REL0000000002	Howard, Johnson		Attach	SILVER.mpg
5	☐	REL0000000002_0002	REL0000000002	Howard, Johnson		Attach	GOLD.mpg
6	☐	REL0000000002_0003	REL0000000002	Howard, Johnson		Attach	javelin.mpg
7	☐	REL0000000002_0004	REL0000000002	Howard, Johnson		Attach	BRONZE.mpg
8	☐	REL0000000003	REL0000000003	Howard, Johnson	10/17/2000, 10:58 ...	Email	Fw: cat clip...........

2. Select the **Delete** mass operation.


Edit
Move
Delete
Replace
Tally/Sum/Average
Export to File
Cluster
Save as List
Convert
Set Long Text Field Size
Re-produce
PDF
Image
Edit
Save Search

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 images**—deletes only images and leaves parent documents in place.
     - **Only natives**—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 4 Documents

Warning: Clicking delete will permanently delete this object and its children and unlink 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.

Documents, duplicates, and all associated files

Force delete

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.

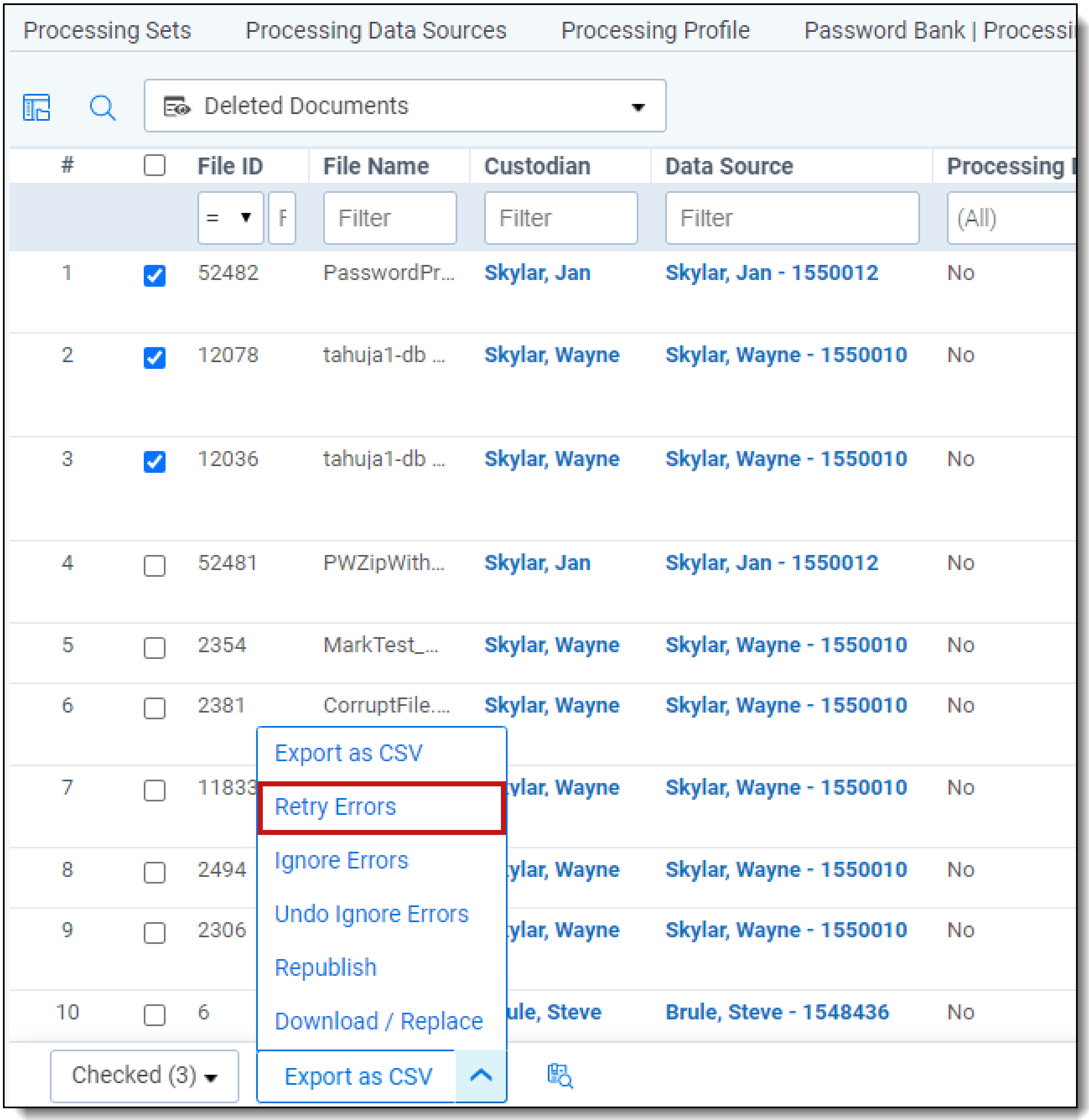
19.3 Interacting with deleted documents

To view deleted documents, navigate to the **Files** tab and select the **Deleted Documents** view.

To view errored documents, navigate to the **Files** tab and select either the **Current Errored Files** or **All Errored Files** view. See Processing error overview for more details on processing errors.

19.4 Retrying delete errors

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 Error Message to see the errors that occurred during deletion. Some errors can be retried using the **Retry Errors** mass actions option. Once deleted, documents are excluded from further processing operations (such as deduplication, retry, and republish.)


Processing Sets
Processing Data Sources
Processing Profile
Password Bank | Processi
Deleted Documents
Export as CSV
Retry Errors
Ignore Errors
Undo Ignore Errors
Republish
Download / Replace
Checked (3)
Export as CSV

#	☐	File ID	File Name	Custodian	Data Source	Processing I
		= ▾ F	Filter	Filter	Filter	(All)
1	☑	52482	PasswordPr...	Skylar, Jan	Skylar, Jan - 1550012	No
2	☑	12078	tahuja1-db ...	Skylar, Wayne	Skylar, Wayne - 1550010	No
3	☑	12036	tahuja1-db ...	Skylar, Wayne	Skylar, Wayne - 1550010	No
4	☐	52481	PWZipWith...	Skylar, Jan	Skylar, Jan - 1550012	No
5	☐	2354	MarkTest_...	Skylar, Wayne	Skylar, Wayne - 1550010	No
6	☐	2381	CorruptFile....	Skylar, Wayne	Skylar, Wayne - 1550010	No
7	☐	11833		ylar, Wayne	Skylar, Wayne - 1550010	No
8	☐	2494		ylar, Wayne	Skylar, Wayne - 1550010	No
9	☐	2306		ylar, Wayne	Skylar, Wayne - 1550010	No
10	☐	6		ule, Steve	Brule, Steve - 1548436	No

20 Error overview

You may encounter errors from time to time while running processing jobs. 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 errors you may encounter and where to find information on errors in Relativity.

For information on resolving processing errors, see Processing error resolution.

This page details the following:

- Types of processing errors
- Processing phases where errors may occur
- Processing error tabs
    - Files
        - File error fields
        - Viewing file error details
    - Job Errors
        - Job error fields
        - Viewing job error details
- Viewing errors from a processing set
- Required security permissions

20.1 Types of processing errors

There are two types of processing errors: file errors and job errors.

- **File errors**—affect a single file. Overall, the processing job continues, even if file errors exist. File errors are resolvable and have a status of **Not Resolved** in the Files tab. An example of a file error might include a password-protected file where Relativity cannot find the password.
- **Job errors**—affect the entire processing job and block publishing. You must resolve job errors before you can continue processing.

20.2 Processing phases where errors may occur

Errors can occur during any processing phase, such as Inventory, Discover, and Publish.

- **Inventory**—all Inventory errors are unresolvable, and you cannot retry them. To address inventory errors, rerun Inventory from the same processing set.
- **Discover**—the Discover phase covers file discovery and text extraction.
    - **Discover**—job errors that occur during discovery are unresolvable, and you cannot retry them. You should always address unresolvable discover job errors. Not addressing unresolvable errors may result in incomplete metadata in published documents. Address these errors by creating a new processing set (or sets) and rediscovering the data.

  - **Text-extraction**—job errors that are resolvable have an error status of **Ready to retry**. Retry the error from the same processing set.
- **Publish**—job errors during the publishing phase that are resolvable have an error status of **Ready to retry**. Retry the error from the same processing set.

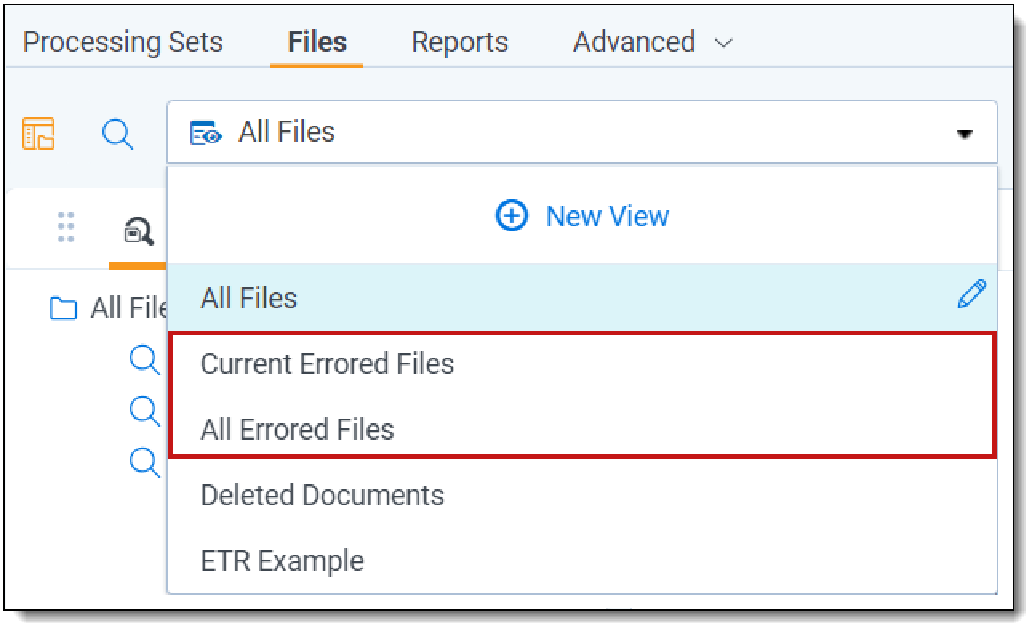
20.3 Processing error tabs

Use the Files tab and Job Errors tab to view errors that occurred during processing. Relativity notifies you of errors on the processing set page when the job completes.

20.3.1 Files tab

The **Files** tab displays a list of all documents within your workspace. The Files tab has two views related specifically to file errors:

- **Current Errored Files**—displays all outstanding errors from processing jobs. This is the primary location for workflows like error retry, ignore, and file replacement.
- **All Errored Files**—displays a historical report of errors from all processing sets, regardless of error resolution.



For more information on the Files tab, see the Files topic.

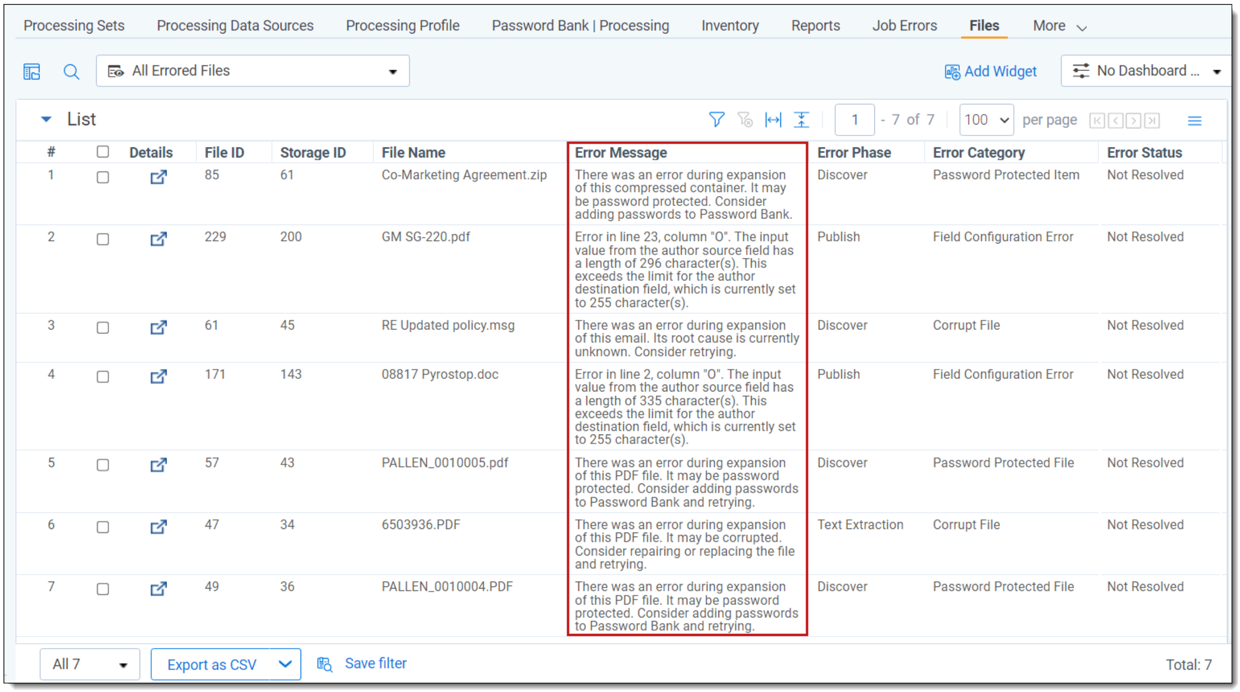
20.3.1.1 File error fields

The Current Errored Files and All Errored Files views display the following fields:

- **Details**—click this icon to open a dialog containing information on the file properties and error. See Viewing file error details for more detailed information.
- **File ID**—the unique identifier assigned to an instance of a file.
- **Storage ID**—the unique identifier assigned to a file.

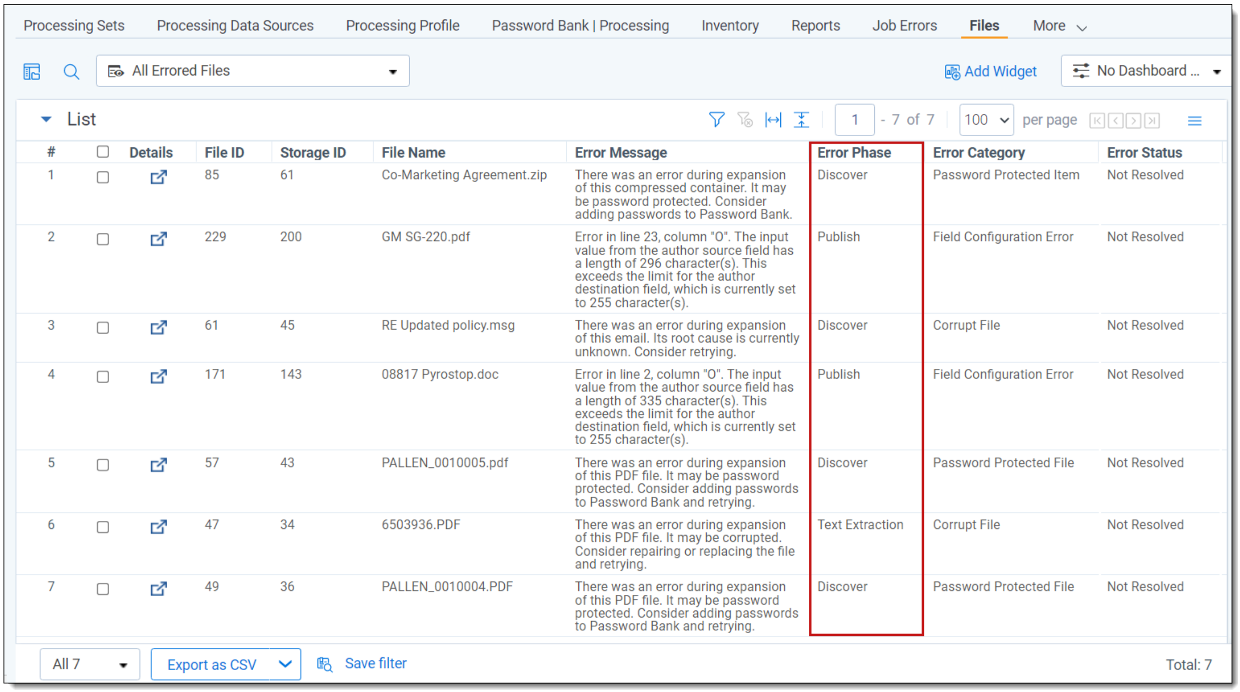
Note: Differentiating between the File ID and the Storage ID. Every file is assigned a unique Storage ID. Every instance of a file is assigned a unique File ID. For example, you may have a Word document with a unique Storage ID. If that document is attached to two different emails, then each instance 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.
- **Error Message**—displays a message about the error, along with possible resolution.



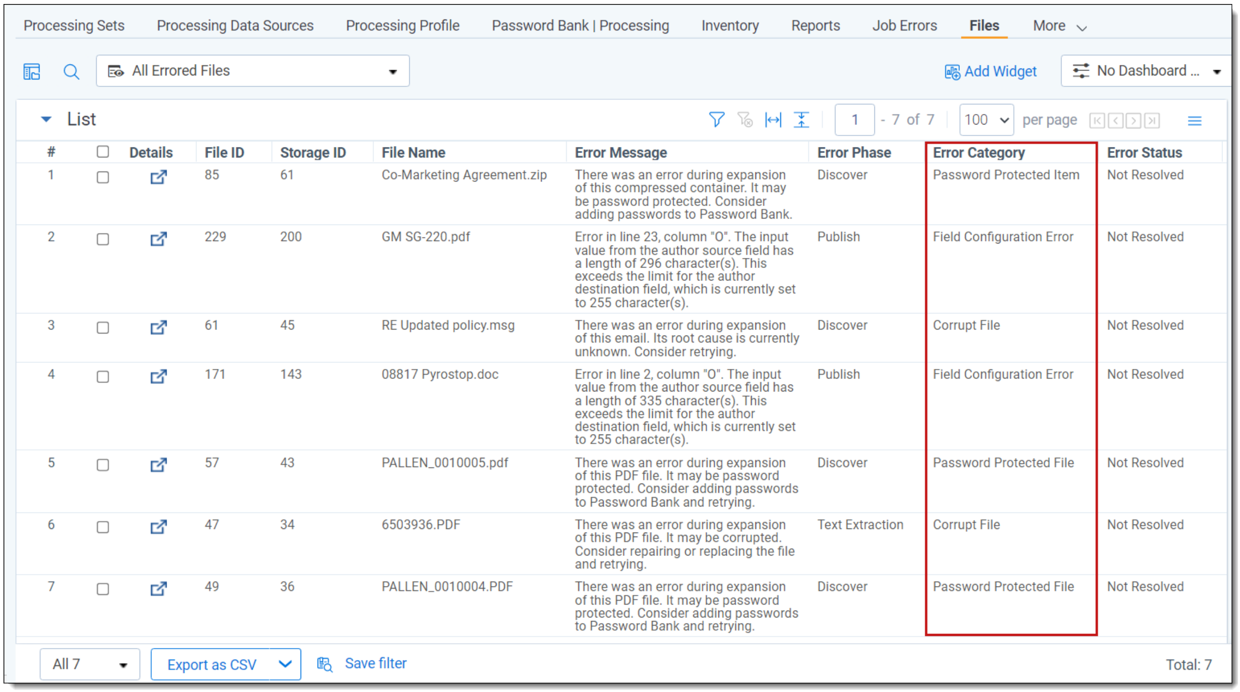
#	☐	Details	File ID	Storage ID	File Name	Error Message	Error Phase	Error Category	Error Status
1	☐		85	61	Co-Marketing Agreement.zip	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Item	Not Resolved
2	☐		229	200	GM SG-220.pdf	Error in line 23, column "0". The input value from the author source field has a length of 296 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error	Not Resolved
3	☐		61	45	RE Updated policy.msg	There was an error during expansion of this email. Its root cause is currently unknown. Consider retrying.	Discover	Corrupt File	Not Resolved
4	☐		171	143	08817 Pyrostop.doc	Error in line 2, column "0". The input value from the author source field has a length of 335 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error	Not Resolved
5	☐		57	43	PALLEN_0010005.pdf	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved
6	☐		47	34	6503936.PDF	There was an error during expansion of this PDF file. It may be corrupted. Consider repairing or replacing the file and retrying.	Text Extraction	Corrupt File	Not Resolved
7	☐		49	36	PALLEN_0010004.PDF	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved

- **Error Phase**—displays the processing phase in which the error occurred. Processing phases include Inventory, Discover, and Publish.


Processing Sets
Processing Data Sources
Processing Profile
Password Bank | Processing
Inventory
Reports
Job Errors
Files
More
All Errored Files
Add Widget
No Dashboard ...
List
1
- 7 of 7
100
per page
All 7
Export as CSV
Save filter
Total: 7

#	☐	Details	File ID	Storage ID	File Name	Error Message	Error Phase	Error Category	Error Status
1	☐		85	61	Co-Marketing Agreement.zip	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Item	Not Resolved
2	☐		229	200	GM SG-220.pdf	Error in line 23, column "O". The input value from the author source field has a length of 296 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error	Not Resolved
3	☐		61	45	RE Updated policy.msg	There was an error during expansion of this email. Its root cause is currently unknown. Consider retrying.	Discover	Corrupt File	Not Resolved
4	☐		171	143	08817 Pyrostop.doc	Error in line 2, column "O". The input value from the author source field has a length of 335 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error	Not Resolved
5	☐		57	43	PALLEN_0010005.pdf	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved
6	☐		47	34	6503936.PDF	There was an error during expansion of this PDF file. It may be corrupted. Consider repairing or replacing the file and retrying.	Text Extraction	Corrupt File	Not Resolved
7	☐		49	36	PALLEN_0010004.PDF	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved

- **Error Category**—displays the error category such as a password protected file where Relativity cannot find the password. See the table below the image for a complete list of error categories.


Processing Sets
Processing Data Sources
Processing Profile
Password Bank | Processing
Inventory
Reports
Job Errors
Files
More
All Errored Files
Add Widget
No Dashboard ...
List
1
- 7 of 7
100
per page
All 7
Export as CSV
Save filter
Total: 7

#	☐	Details	File ID	Storage ID	File Name	Error Message	Error Phase	Error Category	Error Status
1	☐		85	61	Co-Marketing Agreement.zip	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Item	Not Resolved
2	☐		229	200	GM SG-220.pdf	Error in line 23, column "O". The input value from the author source field has a length of 296 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error	Not Resolved
3	☐		61	45	RE Updated policy.msg	There was an error during expansion of this email. Its root cause is currently unknown. Consider retrying.	Discover	Corrupt File	Not Resolved
4	☐		171	143	08817 Pyrostop.doc	Error in line 2, column "O". The input value from the author source field has a length of 335 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error	Not Resolved
5	☐		57	43	PALLEN_0010005.pdf	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved
6	☐		47	34	6503936.PDF	There was an error during expansion of this PDF file. It may be corrupted. Consider repairing or replacing the file and retrying.	Text Extraction	Corrupt File	Not Resolved
7	☐		49	36	PALLEN_0010004.PDF	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved

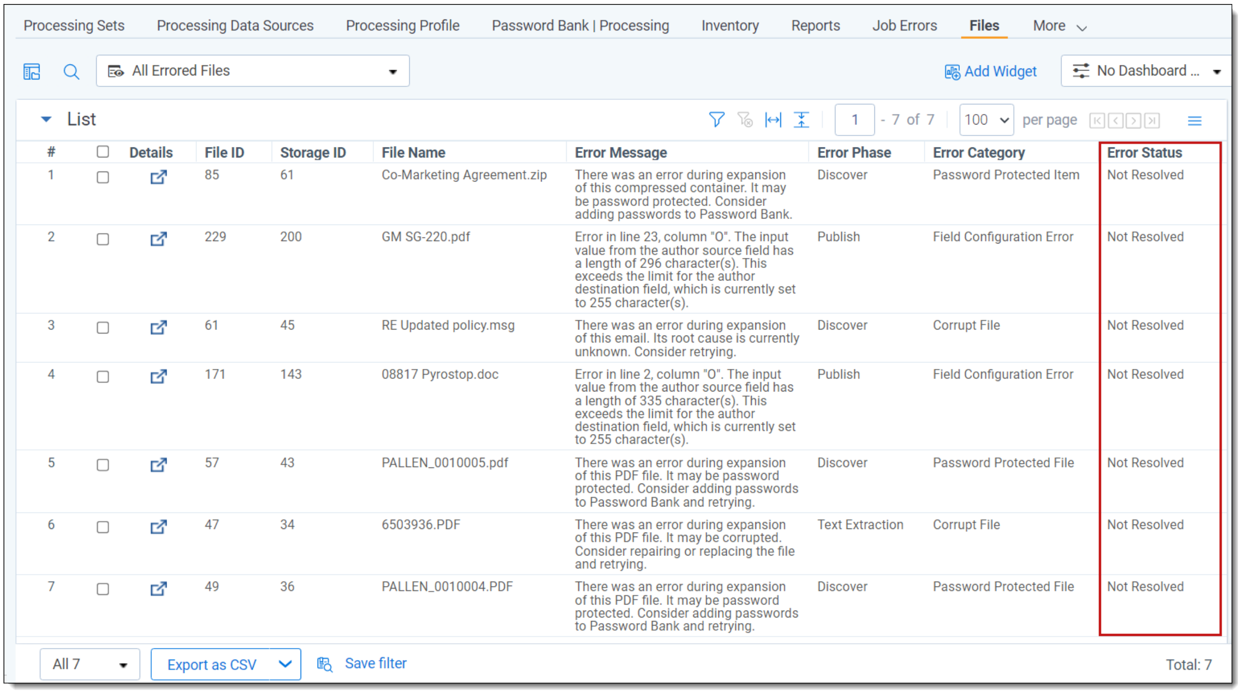
The following table provides a listing of values for the **Error Category** field, along with a description of the issues associated with the error.

Category name	Description
Corrupt Container	These errors are exclusive to container files that have encountered

Category name	Description
	corruption when attempting to open and locate files within the container itself. When containers have these errors associated to them, you will not see any extracted loose files. The errors are either ignored or downloaded offline for you to investigate whether the corruption is fixable, then subsequently replaced and retried.
Corrupt File	These errors are exclusive to non-container files that have found elements of corruption during Processing. The errors are either ignored or downloaded offline for you to investigate whether the corruption is fixable, then subsequently replaced and retried.
Could Not Identify	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.
Environmental	Environmental errors are a result of issues in the Relativity Processing environment. Retry and resolve these errors when encountered.
File Read / Write Error	These errors are a subset of Environmental issues specifically caused by file system issues. Retry and resolve these errors when encountered.
Missing Attachment	Relativity could not extract an attachment from a document or email from its file.
Missing File Metadata	A file is missing a piece of metadata.
Missing Extracted Text	These errors 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 errors where possible.
Partially Corrupted Container	These errors are exclusive to container files that have encountered corruption during extraction of specific records. When containers have these errors associated to them, you may see some files extracted, but not all. The errors are either ignored or downloaded offline for you to investigate whether the corruption is fixable, then subsequently replaced and retried.
Password Protected Container	These errors 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 error. For more information, see Password Bank. **Note:** When investigating publish errors, if you see five password protected errors 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.
Password Protected File	These errors 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 error. For more information, see Password Bank.

Category name	Description
Relativity Field Configuration	These errors represent issues with Field Mapping during publish jobs. They are usually associated to a specific setting like length or an Associative Object Type. Resolve the field settings errors according to the error message.
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.

- **Error Status**—provides information on where the file is in error remediation.



All Errored Files

1 - 7 of 7 | 100 per page

#	☐	Details	File ID	Storage ID	File Name	Error Message	Error Phase	Error Category	Error Status
1	☐		85	61	Co-Marketing Agreement.zip	There was an error during expansion of this compressed container. It may be password protected. Consider adding passwords to Password Bank.	Discover	Password Protected Item	Not Resolved
2	☐		229	200	GM SG-220.pdf	Error in line 23, column "O". The input value from the author source field has a length of 296 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error	Not Resolved
3	☐		61	45	RE Updated policy.msg	There was an error during expansion of this email. Its root cause is currently unknown. Consider retrying.	Discover	Corrupt File	Not Resolved
4	☐		171	143	08817 Pyrostop.doc	Error in line 2, column "O". The input value from the author source field has a length of 335 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error	Not Resolved
5	☐		57	43	PALLEN_0010005.pdf	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved
6	☐		47	34	6503936.PDF	There was an error during expansion of this PDF file. It may be corrupted. Consider repairing or replacing the file and retrying.	Text Extraction	Corrupt File	Not Resolved
7	☐		49	36	PALLEN_0010004.PDF	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved

All 7

Total: 7

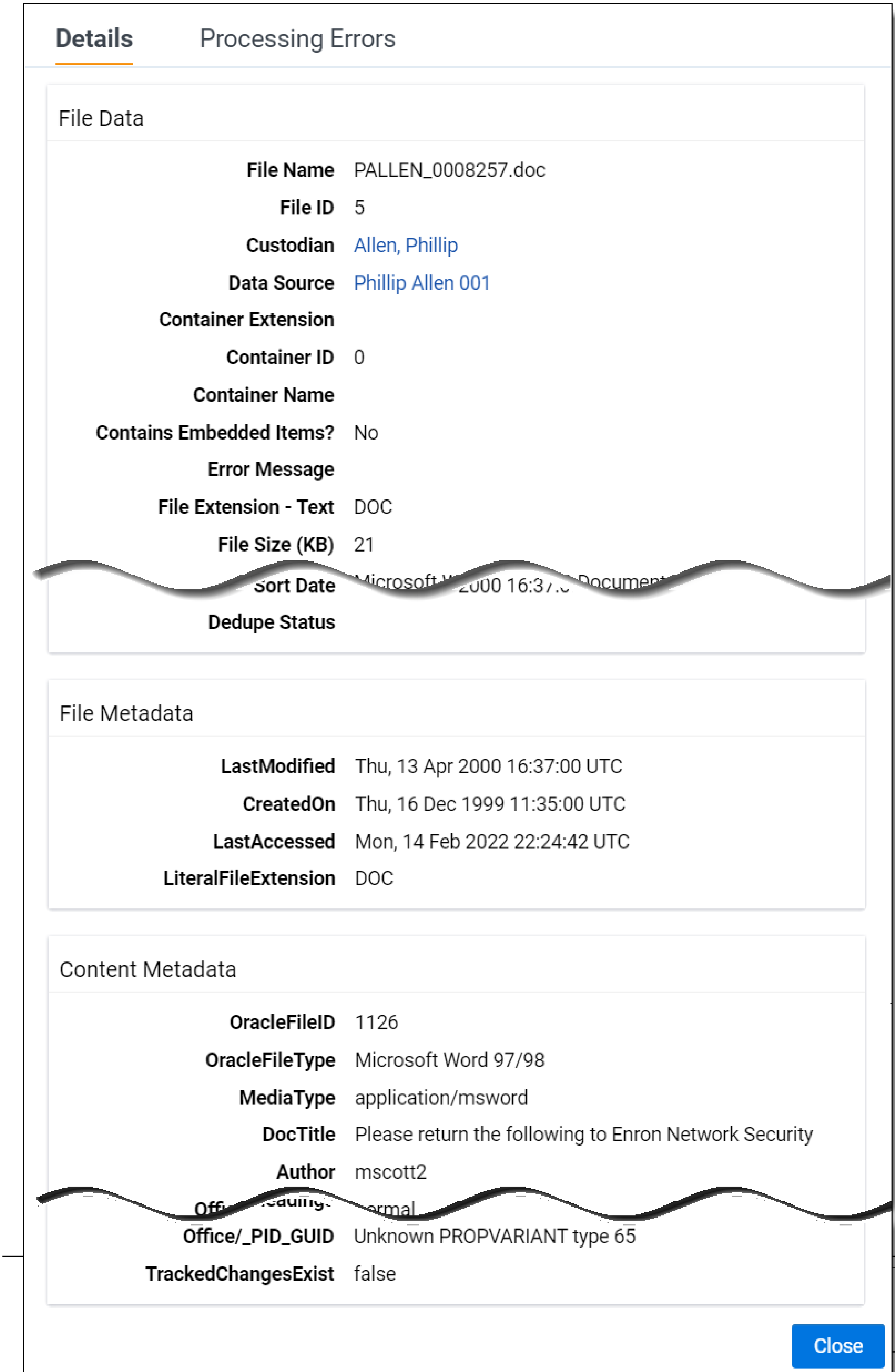
The error status value helps in determining if you need to take action on a file, or to see if the file had previous errors. When a file has all its errors resolved, the **Error Message**, **Error Category**, and **Error Phase** fields no longer display any content. However, the **Error Status** field displays Resolved to indicate that the file had issues that are now resolved. Error status values include:

- **Not Resolved**—the error is still outstanding.
- **Resolving**—the error is in the process of being retried.
- **Resolved**—error resolved.
- **Ignored**—error ignored.

Note: In some cases, the **Error Status** field on the **Files** tab may not be in sync with the **Error Status** field on the **Documents** tab. The mismatch occurs when you resolve an error but do not republish the file. For example, after discovery, you notice a file error due to a missing password. The error status fields on the Files and Documents tabs show the same error status. At some point, you fix the password and retry the error. The error status on the Files tab now shows resolved, while the Documents tab still shows an error. Republishing syncs the fields once again.

20.3.1.2 Viewing file error 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	PALLEN_0008257.doc
File ID	5
Custodian	Allen, Phillip
Data Source	Phillip Allen 001
Container Extension	
Container ID	0
Container Name	
Contains Embedded Items?	No
Error Message	
File Extension - Text	DOC
File Size (KB)	21
Sort Date	[illegible] 2000 16:37 [illegible]
Dedupe Status	

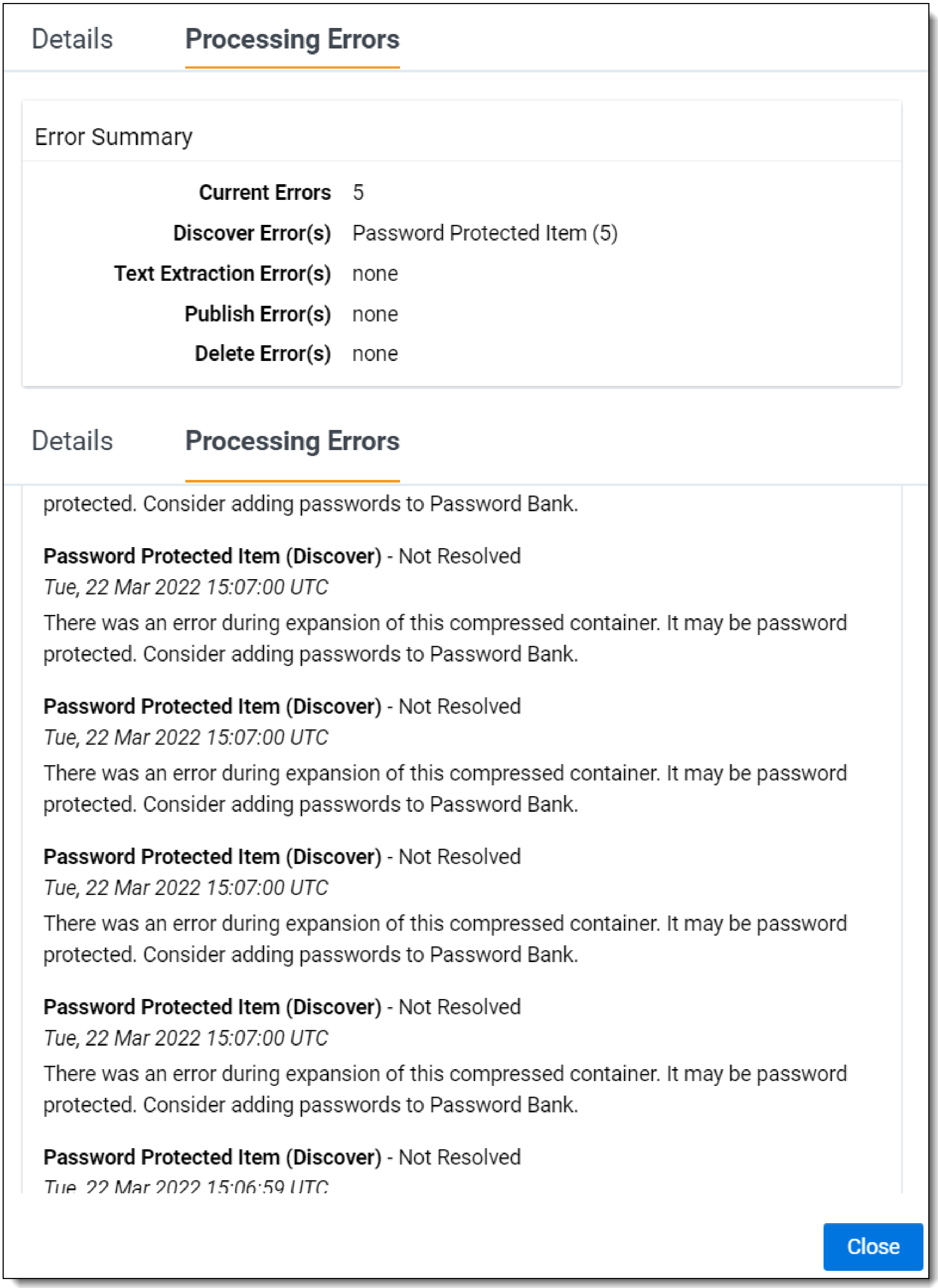
File Metadata

LastModified	Thu, 13 Apr 2000 16:37:00 UTC
CreatedOn	Thu, 16 Dec 1999 11:35:00 UTC
LastAccessed	Mon, 14 Feb 2022 22:24:42 UTC
LiteralFileExtension	DOC

Content Metadata

OracleFileID	1126
OracleFileType	Microsoft Word 97/98
MediaType	application/msword
DocTitle	Please return the following to Enron Network Security
Author	mscott2
[illegible]	[illegible]ormal
Office/_PID_GUID	Unknown PROPVARIANT type 65
TrackedChangesExist	false

Click the **Processing Errors** tab to view a summary of errors, and the error history.



Error Summary

Current Errors 5

Discover Error(s) Password Protected Item (5)

Text Extraction Error(s) none

Publish Error(s) none

Delete Error(s) none

protected. Consider adding passwords to Password Bank.

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 protected. Consider adding passwords to Password Bank.

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 protected. Consider adding passwords to Password Bank.

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 protected. Consider adding passwords to Password Bank.

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 protected. Consider adding passwords to Password Bank.

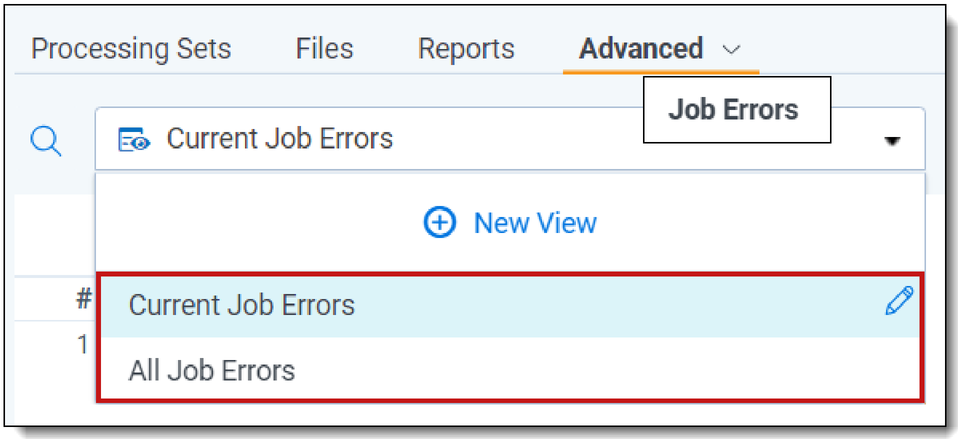
Password Protected Item (Discover) - Not Resolved
Tue, 22 Mar 2022 15:06:59 UTC

- The Error History section represents all errors that have ever occurred on a file. This acts as a timeline of the record’s errors, showing when they occurred, what they were about, and if any are still active. This includes errors resulting from retries of previous errors and contains category, phase, date/time, and message information. All times are kept in UTC format.
- The Error Summary section displays a count of all active errors along with their associated category and phase. This is especially important when investigating errors 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.

20.3.2 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:

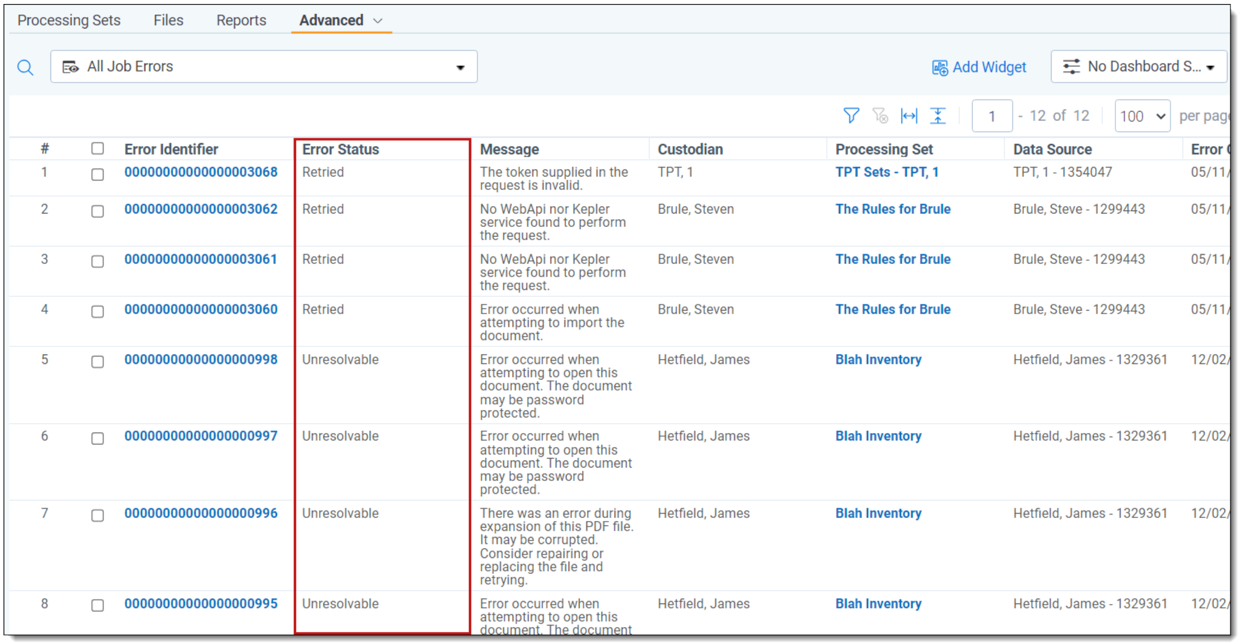
- **Current Job Errors**—displays all unresolved job errors.
- **All Job Errors**—displays job errors that occurred throughout the lifecycle of the matter.



20.3.2.1 Job error fields

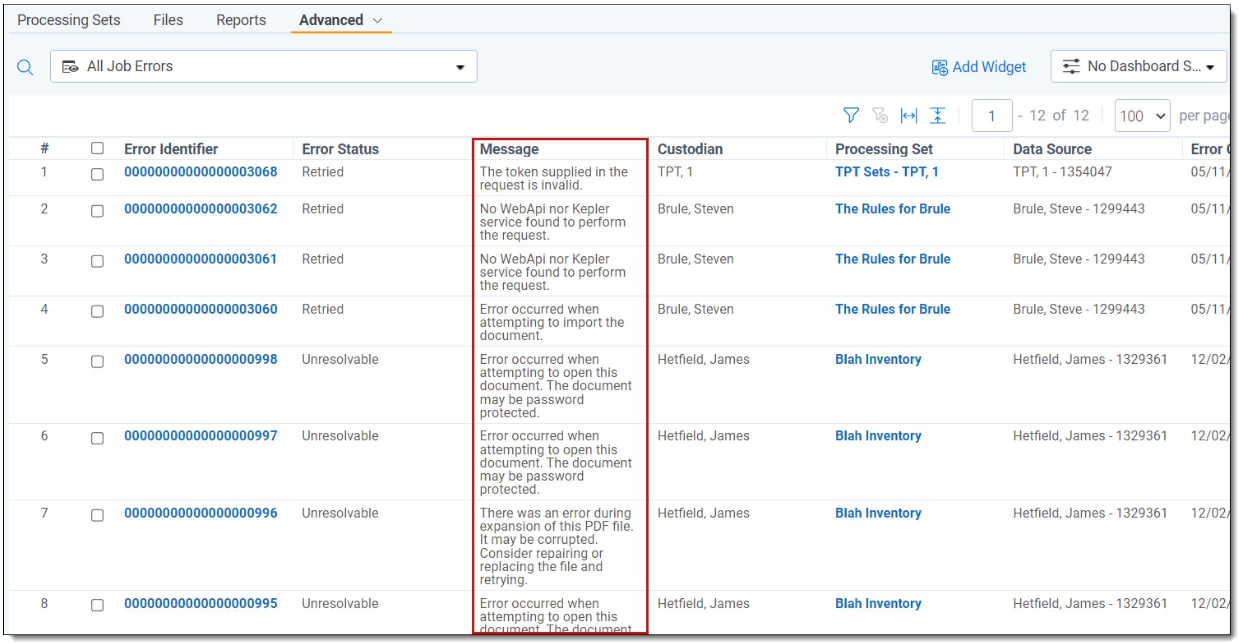
The Current Job Errors and All Job Errors views display the following fields:

- **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 errors, the console is disabled because you cannot take any actions on that error from inside Relativity.
- **Error status**—provides the status of the error.



#	☐	Error Identifier	Error Status	Message	Custodian	Processing Set	Data Source	Error
1	☐	0000000000000003068	Retried	The token supplied in the request is invalid.	TPT, 1	TPT Sets - TPT, 1	TPT, 1 - 1354047	05/11
2	☐	0000000000000003062	Retried	No WebApi nor Kepler service found to perform the request.	Brule, Steven	The Rules for Brule	Brule, Steve - 1299443	05/11
3	☐	0000000000000003061	Retried	No WebApi nor Kepler service found to perform the request.	Brule, Steven	The Rules for Brule	Brule, Steve - 1299443	05/11
4	☐	0000000000000003060	Retried	Error occurred when attempting to import the document.	Brule, Steven	The Rules for Brule	Brule, Steve - 1299443	05/11
5	☐	0000000000000000998	Unresolvable	Error occurred when attempting to open this document. The document may be password protected.	Hetfield, James	Blah Inventory	Hetfield, James - 1329361	12/02
6	☐	0000000000000000997	Unresolvable	Error occurred when attempting to open this document. The document may be password protected.	Hetfield, James	Blah Inventory	Hetfield, James - 1329361	12/02
7	☐	0000000000000000996	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	Blah Inventory	Hetfield, James - 1329361	12/02
8	☐	0000000000000000995	Unresolvable	Error occurred when attempting to open this document. The document	Hetfield, James	Blah Inventory	Hetfield, James - 1329361	12/02

- 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.



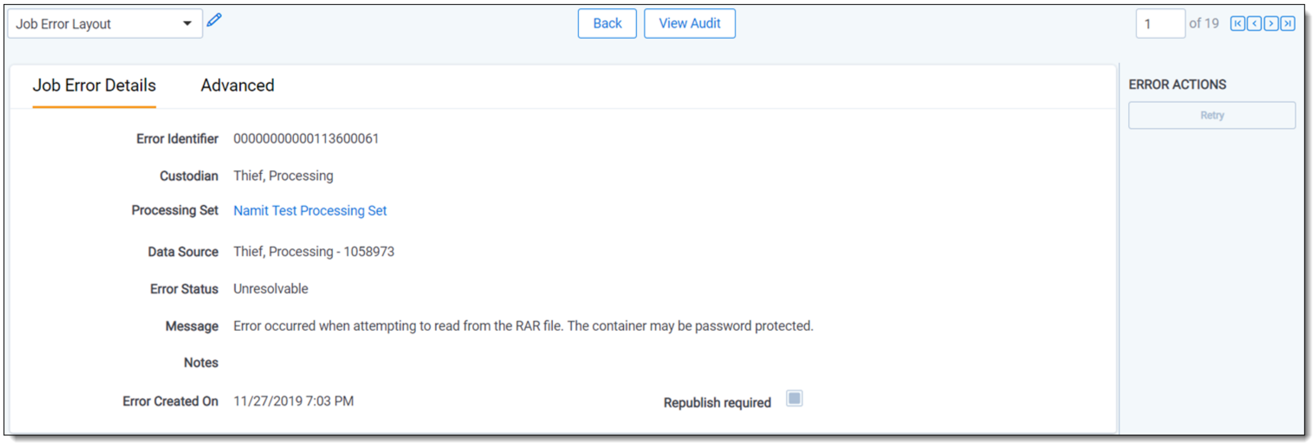
#	☐	Error Identifier	Error Status	Message	Custodian	Processing Set	Data Source	Error
1	☐	0000000000000003068	Retried	The token supplied in the request is invalid.	TPT, 1	TPT Sets - TPT, 1	TPT, 1 - 1354047	05/11
2	☐	0000000000000003062	Retried	No WebApi nor Kepler service found to perform the request.	Brule, Steven	The Rules for Brule	Brule, Steve - 1299443	05/11
3	☐	0000000000000003061	Retried	No WebApi nor Kepler service found to perform the request.	Brule, Steven	The Rules for Brule	Brule, Steve - 1299443	05/11
4	☐	0000000000000003060	Retried	Error occurred when attempting to import the document.	Brule, Steven	The Rules for Brule	Brule, Steve - 1299443	05/11
5	☐	0000000000000000998	Unresolvable	Error occurred when attempting to open this document. The document may be password protected.	Hetfield, James	Blah Inventory	Hetfield, James - 1329361	12/02
6	☐	0000000000000000997	Unresolvable	Error occurred when attempting to open this document. The document may be password protected.	Hetfield, James	Blah Inventory	Hetfield, James - 1329361	12/02
7	☐	0000000000000000996	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	Blah Inventory	Hetfield, James - 1329361	12/02
8	☐	0000000000000000995	Unresolvable	Error occurred when attempting to open this document. The document	Hetfield, James	Blah Inventory	Hetfield, James - 1329361	12/02

- **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.

20.3.2.2 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 errors, since you cannot retry or ignore job errors as you can with document errors.



Error Identifier 00000000000113600061

Custodian Thief, Processing

Processing Set Namit Test Processing Set

Data Source Thief, Processing - 1058973

Error Status Unresolvable

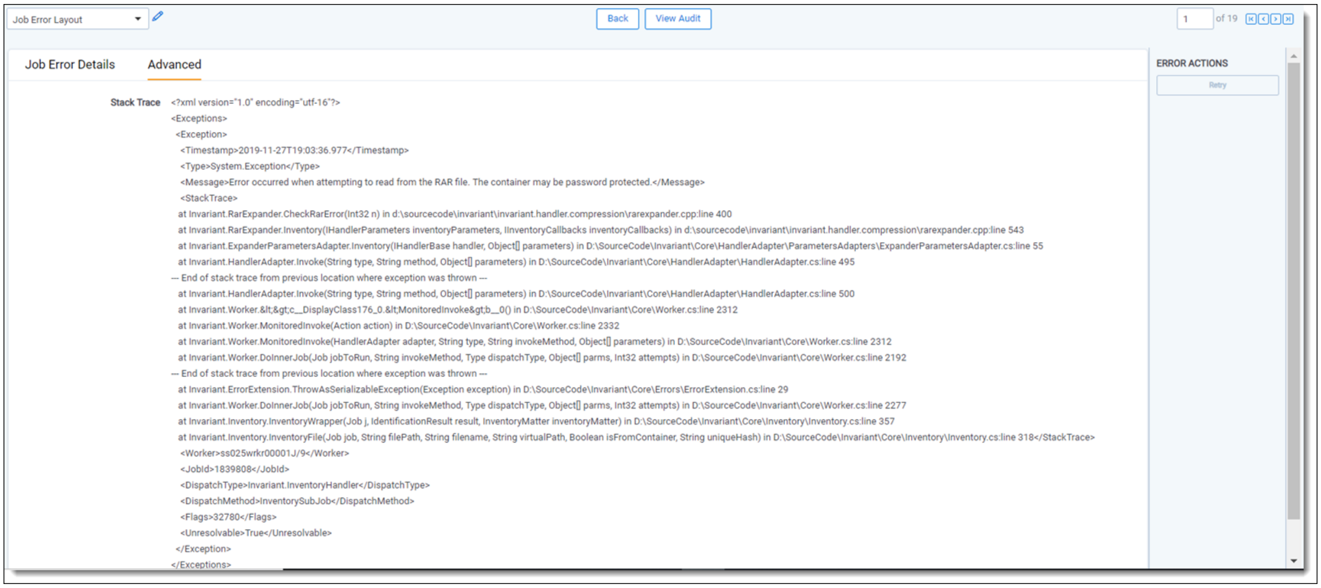
Message Error occurred when attempting to read from the RAR file. The container may be password protected.

Notes

Error Created On 11/27/2019 7:03 PM

Republish required ☐

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.

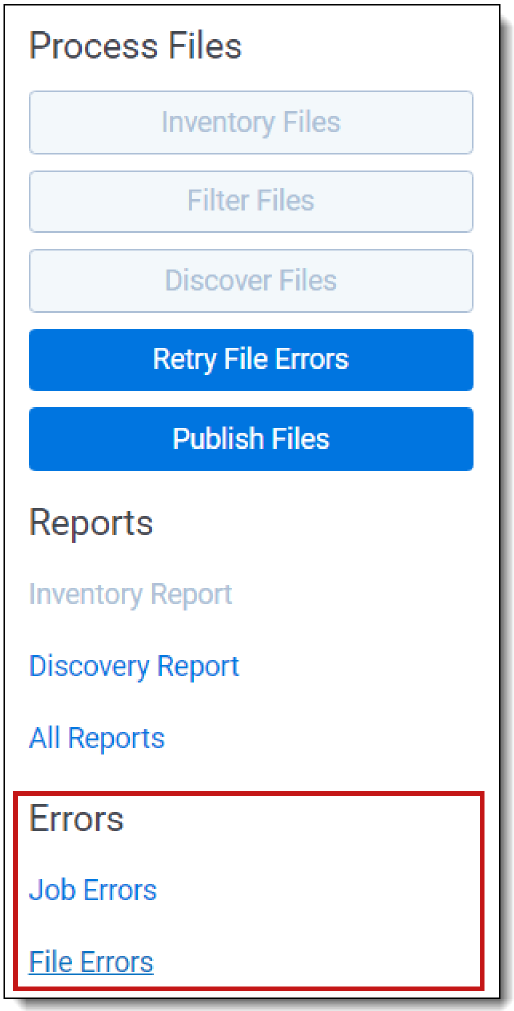


Stack Trace

```
<?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.&lt;&gt;c__DisplayClass176_0.&lt;MonitoredInvoke&gt;b__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>
```

20.4 Viewing errors from a processing set

From within your processing set, use the links in the **Process Files** console to view job errors and file errors. Click the **File Errors** link to open the **Files** tab > **Current Errored Files** view. Click the **Job Errors** link to open the **Job Errors** > **Current Job Errors** view.


Process Files
Inventory Files
Filter Files
Discover Files
Retry File Errors
Publish Files
Reports
Inventory Report
Discovery Report
All Reports
Errors
Job Errors
File Errors

20.5 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 ■ Files

21 Error resolution

Two distinct types of errors occur during processing: exceptions and system errors. Exceptions are 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. System errors occur when the processing engine cannot complete the job. For example, a system 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 system errors may be more complex and require support intervention.

For information on where to find processing error information, see Processing error overview.

This page details the following:

- Types of processing errors
- Processing phases where errors may occur
- File errors
    - File error actions
    - Resolving file errors
        - Retrying an error from the processing set
            - Error categories and their probability of retry success
        - Retry file errors mass action
        - Replacing a corrupted file
        - Resolving a password-protected file error
- Job errors
    - Job error workflow
    - Resolving job errors

21.1 Types of processing errors

There are two types of processing errors: file errors and job errors.

- **File errors**—affect a single file. Overall, the processing job continues, even if file errors exist. File errors are resolvable and have a status of **Not Resolved** in the Files tab. An example of a file error might include a password-protected file where Relativity cannot find the password.
- **Job errors**—affect the entire processing job and block publishing. You must resolve job errors before you can continue processing.

21.2 Processing phases where errors may occur

Errors can occur during any processing phase, such as Inventory, Discover, and Publish.

- **Inventory**—all Inventory errors are unresolvable, and you cannot retry them. To address inventory errors, rerun Inventory from the same processing set.

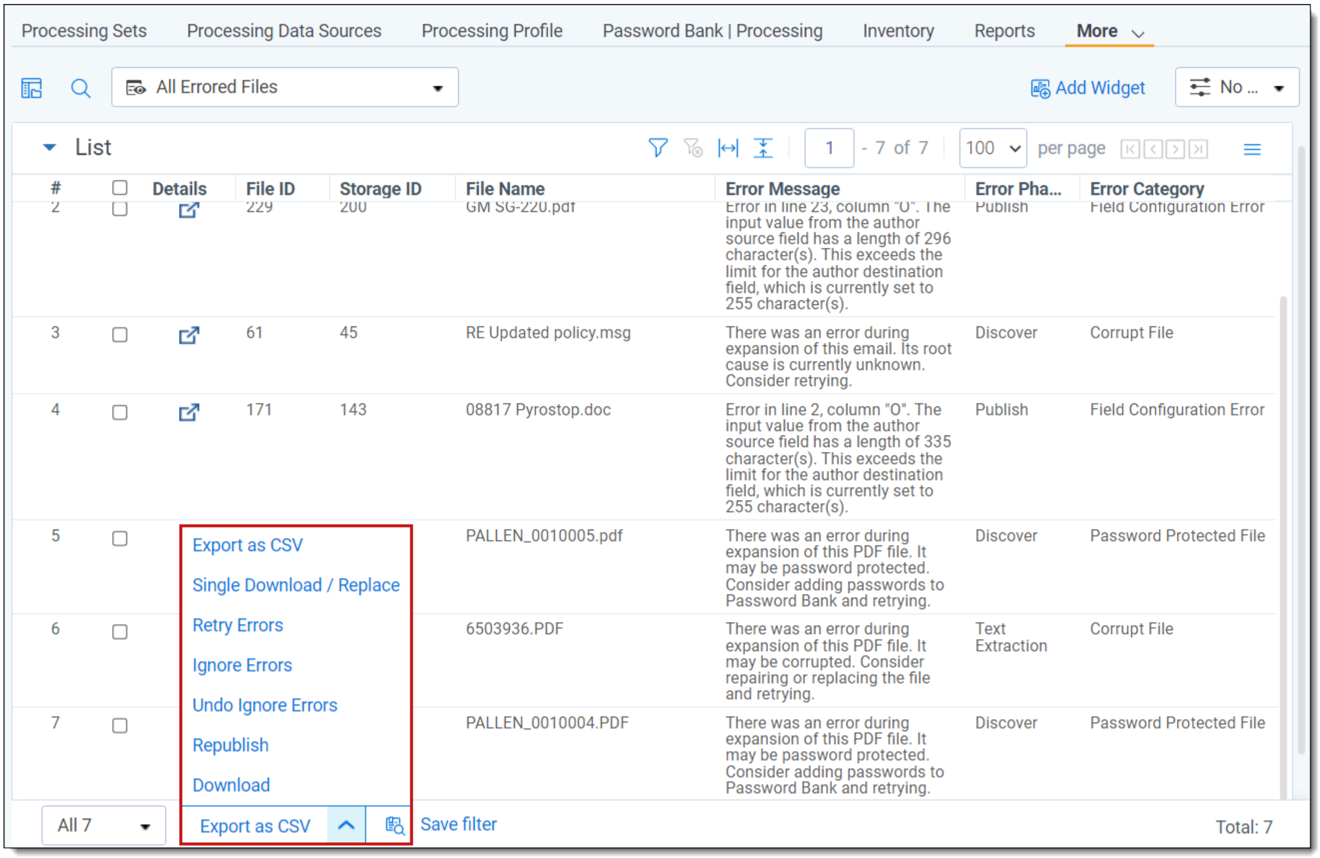
- **Discover**—the Discover phase covers file discovery and text extraction.
    - **Discover**—job errors that occur during discovery are unresolvable, and you cannot retry them. You should always address unresolvable discover job errors. Not addressing unresolvable errors may result in incomplete metadata in published documents. Address these errors by creating a new processing set (or sets) and rediscovering the data.
    - **Text-extraction**—job errors that are resolvable have an error status of **Ready to retry**. Retry the error from the same processing set.
- **Publish**—job errors during the publishing phase that are resolvable have an error status of **Ready to retry**. Retry the error from the same processing set.

21.3 File errors

After completing the discovery or publish phases, you notice the processing job encountered file errors. You want to investigate and resolve the errors to maximize the data published to your workspace.

21.3.1 File error actions

You can take action on file errors 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
More
All Errored Files
Add Widget
No ...
List
1 - 7 of 7
100
per page
Export as CSV
Single Download / Replace
Retry Errors
Ignore Errors
Undo Ignore Errors
Republish
Download
All 7
Export as CSV
Save filter
Total: 7

#	☐	Details	File ID	Storage ID	File Name	Error Message	Error Pha...	Error Category
2	☐		229	200	GM SG-220.pdf	Error in line 23, column "O". The input value from the author source field has a length of 296 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error
3	☐		61	45	RE Updated policy.msg	There was an error during expansion of this email. Its root cause is currently unknown. Consider retrying.	Discover	Corrupt File
4	☐		171	143	08817 Pyrostop.doc	Error in line 2, column "O". The input value from the author source field has a length of 335 character(s). This exceeds the limit for the author destination field, which is currently set to 255 character(s).	Publish	Field Configuration Error
5	☐				PALLEN_0010005.pdf	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File
6	☐				6503936.PDF	There was an error during expansion of this PDF file. It may be corrupted. Consider repairing or replacing the file and retrying.	Text Extraction	Corrupt File
7	☐				PALLEN_0010004.PDF	There was an error during expansion of this PDF file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File

The following mass actions are available:

Mass operation	Description
Export as CSV	This exports the list of processing errors as a CSV file.
Republish	Gives you the option of republishing errored files after you resolve the erorrs. If the republish is successful, the files are available in the Documents list and removed from the Current Errored Files view. For details on how to republish files from the **Files** tab, see Republishing files.
Retry Errors	This action provides the ability to resolve issues occurring during discover and publish. You can see these issues on the Current Errored Files and All Errored Files views within the **Files** tab. For details on how to retry errors from the Deleted Documents view in the **Files** tab, see Retrying delete errors. ■ You must have edit permissions on the Discovered Files object to be able to retry file errors. ■ Note the following regarding retrying errors: ○ Auto-publish is not enabled when you retry errors. If you resolve any discover or text extraction errors, you must manually publish them into your workspace by navigating back to the processing set and clicking the **Retry** button. ○ The discovery retry of errors 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 errors as these errors represent data not in review. ○ If an error occurs on a loose file during discovery, Relativity still attempts to publish it. For example, if a Password Protected error 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 errors for a set when you retry any error 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 errors with an Error Status of **Not Resolved** for retry.
Ignore Errors	This provides the ability to set a file's Error Status to Ignored, which removes it from the Current Errored Files view. The record is visible in the All Files and All Errored Files views.
Undo Ignore Errors	Sets a file's **Error Status** field back to its original value after you pre-

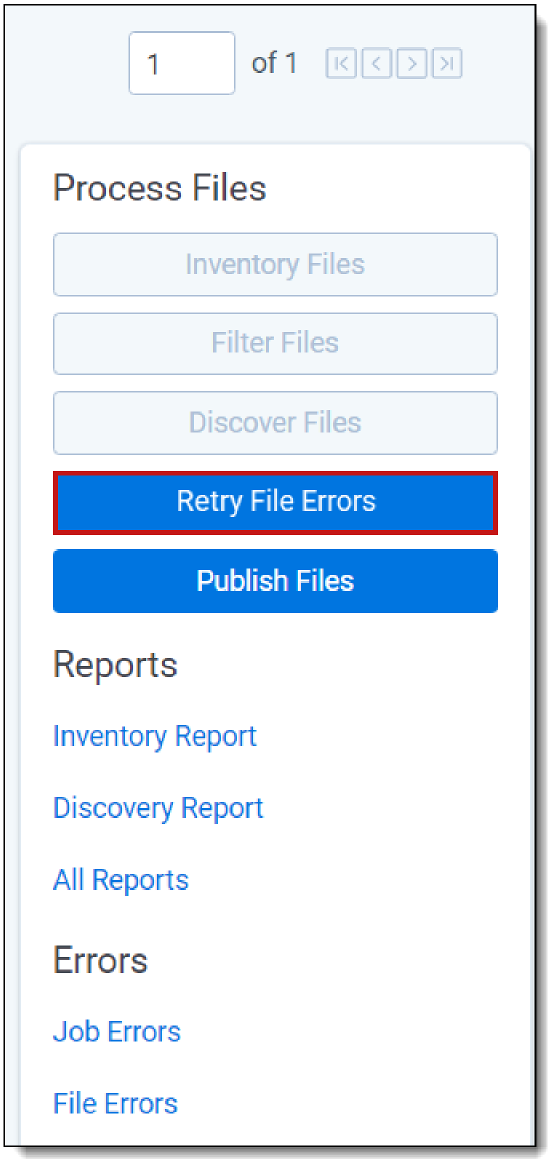
Mass operation	Description
	viously ignored it.
Single Download / Replace	Use the Single Download / Replace mass operation to download a file to your local machine for investigation. After fixing the error, you can use the same mass operation to replace the original file with a newer version. Note the 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 error 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 errors once completed. ■ After uploading a new document, and you select Replace & Retry, the native file updates before you republish. ■ The retry action for job errors only retries errors in a Ready to Retry state. For more information on the Single Download / Replace mass operation, see Single Download / Replace on the **Files tab** page.
Download	Use the Download mass operation to download multiple errored 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 error status of Not Resolved. ■ When complete, files are packaged in a zip container and saved to your local drive. ■ You must use the Single File / Replace mass operation to replace a file. ■ There is a 2.14 GB limitation on the zip container file size. For more information on the Download mass operation, see Mass download errored files on the **Files tab** page.

21.4 Resolving file errors

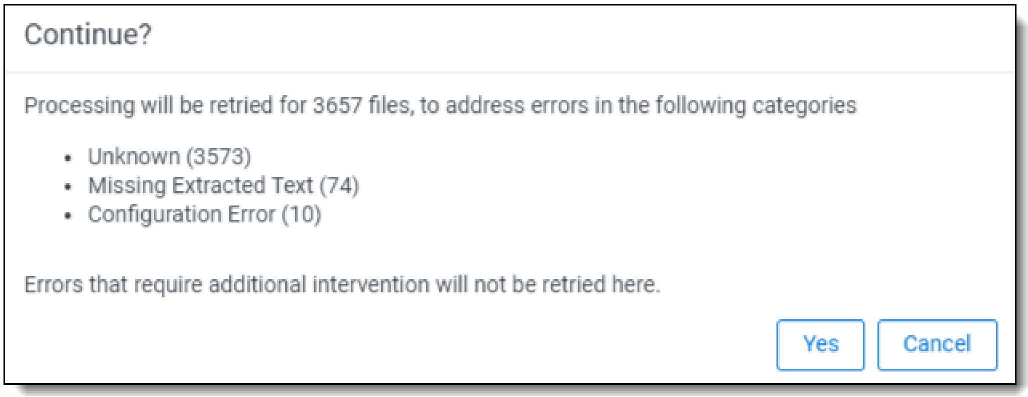
The following sections describe actions you may take to resolve file errors.

21.4.1 Retrying an error from the processing set

You can retry file errors from the processing set by clicking the **Retry File Errors** button located in the **Process Files** console.


1
of 1
Process Files
Inventory Files
Filter Files
Discover Files
Retry File Errors
Publish Files
Reports
Inventory Report
Discovery Report
All Reports
Errors
Job Errors
File Errors

A confirmation message appears with information about the number of errors you are about to retry, along with the associated error 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.

21.4.1.1 Container file errors

File errors 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 error, even though you see a single file error on the container. For container file errors, extract the files and then create a new processing set with the extracted files.

21.4.1.2 Error categories and their probability of retry success

Relativity retries file errors that have a high chance of success. The error category associated with the file determines the probability of success. Error 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 errors that Relativity retries.

Error Category	Included in Retry Button
Corrupt Container	No
Corrupt File	No
Could Not Identify	No
Environmental	Yes
File Read/Write Issue	Yes
Missing Attachment	Yes
Missing child items due to password protection	No
Missing Extracted Text	Yes
Missing File Metadata	Yes
Partially Corrupted Container	No
Password Protected Container	No
Password Protected File	No
Relativity Field Configuration	No
Unsupported	No

21.4.2 Retry file errors mass action

Starting from your processing set, perform the following steps:

1. From the **Process Files** console, click **File Errors** to open the **Current Errored Files** view. The list is automatically filters for your processing set.

2. You can filter the errored files that are the most important to resolve. Some common filters are:
    - **Error Category**—group issues of a similar type.
    - **Error 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 Errors** mass action to begin the process. Alternatively, you can retry all errored files without filtering.

You can now track the progress of the error retry through the processing set's progress bar or by navigating to the Processing and Imaging Queue and viewing active jobs.

21.4.3 Replacing a corrupted file

Sometimes, files reach processing in a corrupted state. Here is a workflow to replace corrupted files with non-corrupted versions so you can get the most out of your data. This works on encrypted documents as well.

Starting from your processing set, perform the following steps:

1. From the **Process Files** console, click **File Errors** to open the **Current Errored Files** view. The list is automatically filters for your processing set.
2. Locate the file you need to replace. Common techniques are:
    - Filter the Error Category field for either Corrupted File or Corrupted Container.
    - Filter by specific file names.
    - Filter for specific error messages.
3. Select the appropriate checkbox for the file you want to replace.
4. Select the **Single Download / Replace** option in the mass action drop-down menu.
5. From here, two options are available:
    - To download a corrupted file: confirm the modal displays the correct container and click the **Download** button.
    - To replace a corrupted file: drag the replacement file into the modal or select **browse for files** to locate the container.
6. Once you add the replacement file to the modal, it automatically uploads to Relativity. A quick verification process lets you know if there are any issues or if there are any significant differences between the original and replaced files.
7. Select the **Replace & Retry** button to complete the replacement and retry any Discovery related errors.

Note: When replacing a file, the metadata associated with the new file overwrites any metadata associated with the original file. For example, if the original file's author is Steve Bruhle and the replacement file's author is Dave Crews, the metadata in Relativity displays Dave Crews.

For more information on replacement considerations, see Single Download / Replace.

21.4.4 Resolving a password-protected file error

Some files are password-protected. If Relativity cannot find the password during the Discover phase, you will see an error with the category, Password Protected Container. The resolution for this error 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 error from either the processing set Process Files console, or from the Files tab mass actions drop-down menu.

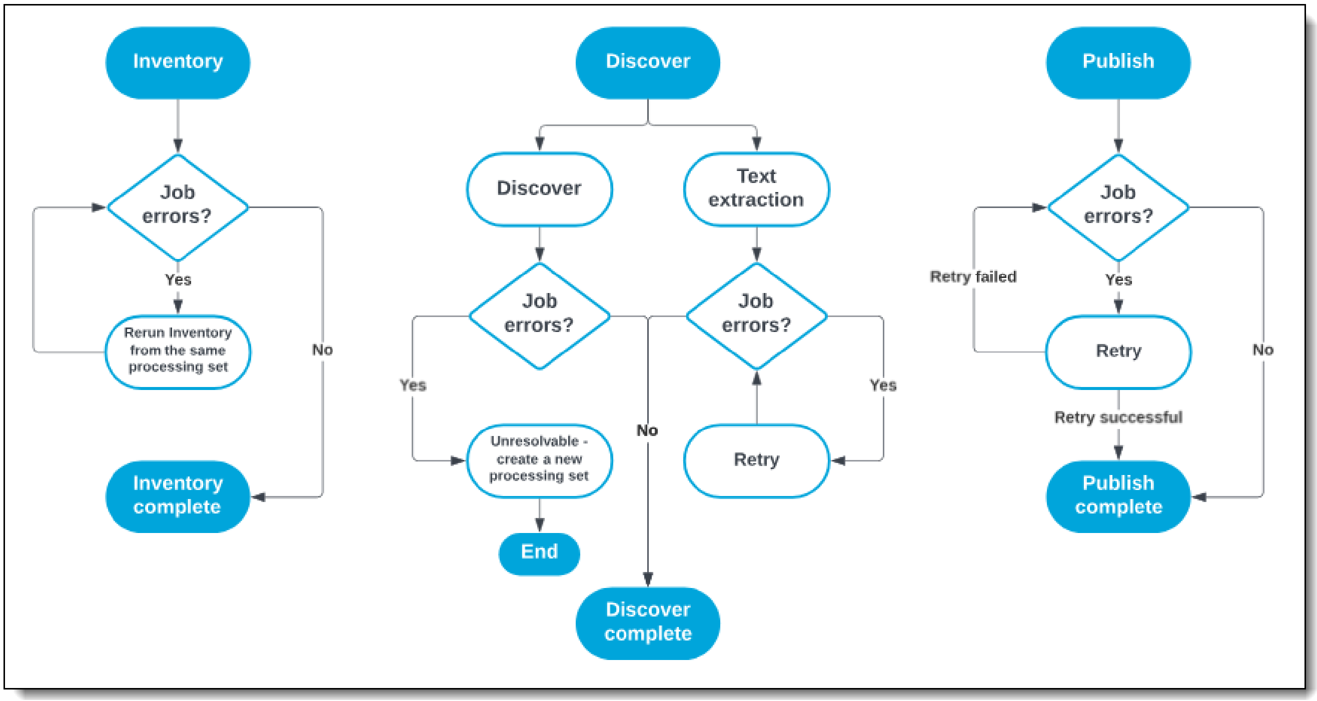
For completed details on storing passwords, see Password Bank.

21.5 Job errors

Navigate to the **Job Errors** tab to view a list of job errors. The **Current Job Errors** view displays all outstanding job errors. The **All Job Errors** view displays all job errors across time, regardless of their resolution. Review the **Error Status** and **Message** fields for an overview of the error and the potential for remediation.

21.5.1 Job error workflow

The following diagram shows the job error workflows for Inventory, Discover, and Publish jobs.


Inventory
Job errors?
Yes
Rerun Inventory from the same processing set
No
Inventory complete
Discover
Discover
Text extraction
Job errors?
Job errors?
Yes
Unresolvable - create a new processing set
End
No
Yes
Retry
Discover complete
Publish
Job errors?
Retry failed
Yes
Retry
No
Retry successful
Publish complete

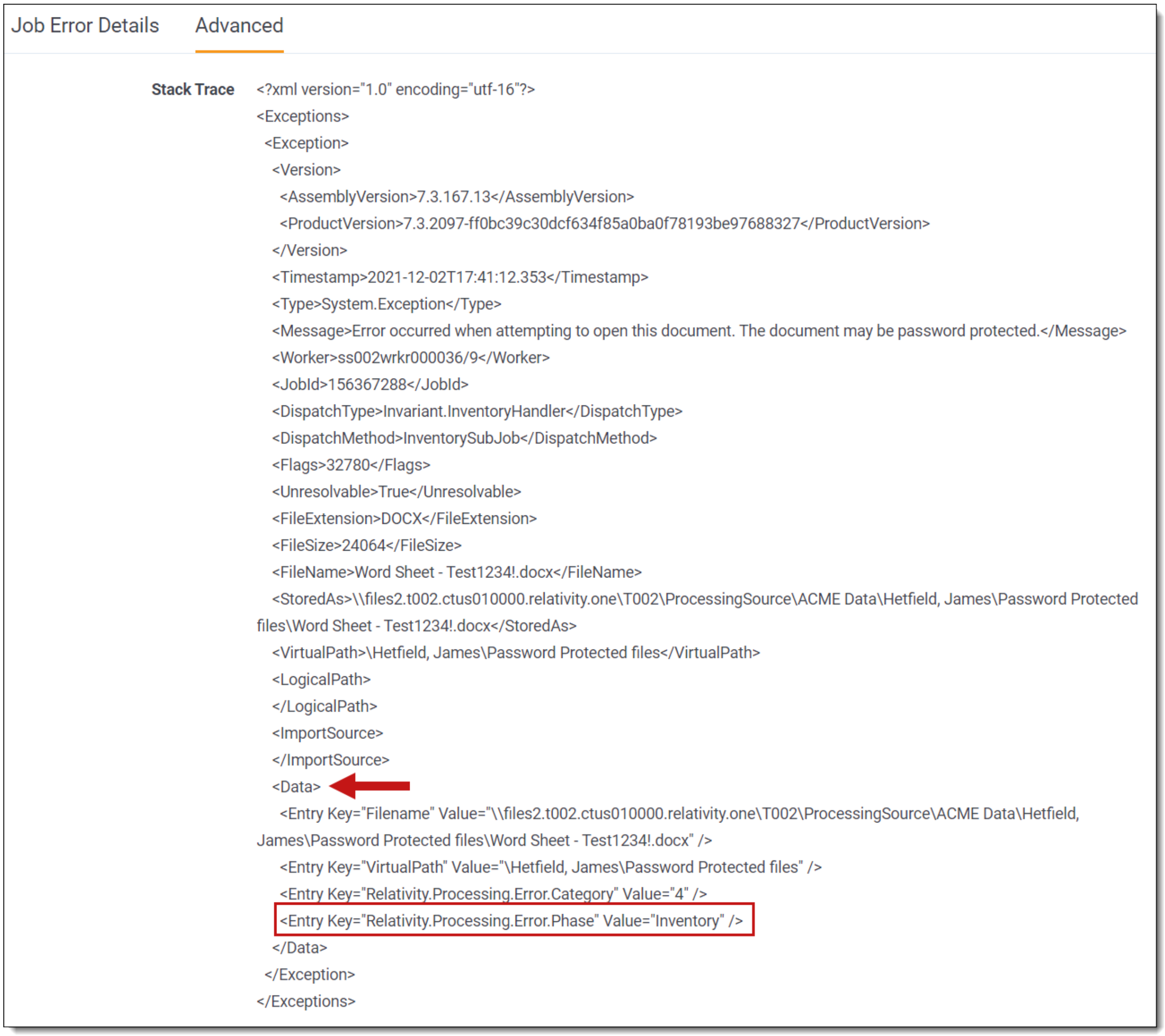
21.5.2 Resolving job errors

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.



Stack Trace

```
<?xml version="1.0" encoding="utf-16"?>
<Exceptions>
 <Exception>
  <Version>
   <AssemblyVersion>7.3.167.13</AssemblyVersion>
   <ProductVersion>7.3.2097-ff0bc39c30dcf634f85a0ba0f78193be97688327</ProductVersion>
  </Version>
  <Timestamp>2021-12-02T17:41:12.353</Timestamp>
  <Type>System.Exception</Type>
  <Message>Error occurred when attempting to open this document. The document may be password protected.</Message>
  <Worker>ss002wrkr000036/9</Worker>
  <JobId>156367288</JobId>
  <DispatchType>Invariant.InventoryHandler</DispatchType>
  <DispatchMethod>InventorySubJob</DispatchMethod>
  <Flags>32780</Flags>
  <Unresolvable>True</Unresolvable>
  <FileExtension>DOCX</FileExtension>
  <FileSize>24064</FileSize>
  <FileName>Word Sheet - Test1234!.docx</FileName>
  <StoredAs>\\files2.t002.ctus010000.relativity.one\T002\ProcessingSource\ACME Data\Hetfield, James\Password Protected
files\Word Sheet - Test1234!.docx</StoredAs>
  <VirtualPath>\Hetfield, James\Password Protected files</VirtualPath>
  <LogicalPath>
  </LogicalPath>
  <ImportSource>
  </ImportSource>
  <Data>
   <Entry Key="Filename" Value="\\files2.t002.ctus010000.relativity.one\T002\ProcessingSource\ACME Data\Hetfield,
James\Password Protected files\Word Sheet - Test1234!.docx" />
   <Entry Key="VirtualPath" Value="\Hetfield, James\Password Protected files" />
   <Entry Key="Relativity.Processing.Error.Category" Value="4" />
   <Entry Key="Relativity.Processing.Error.Phase" Value="Inventory" />
  </Data>
 </Exception>
</Exceptions>
```

21.5.2.1 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.

21.5.2.2 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.

21.5.2.3 Publish

Job errors during the publishing phase that are resolvable have an error status of **Ready to retry**. Retry the error from the same processing set.

For detailed information on the Publish phase and errors, see Publishing files.

22 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.

Using processing reports

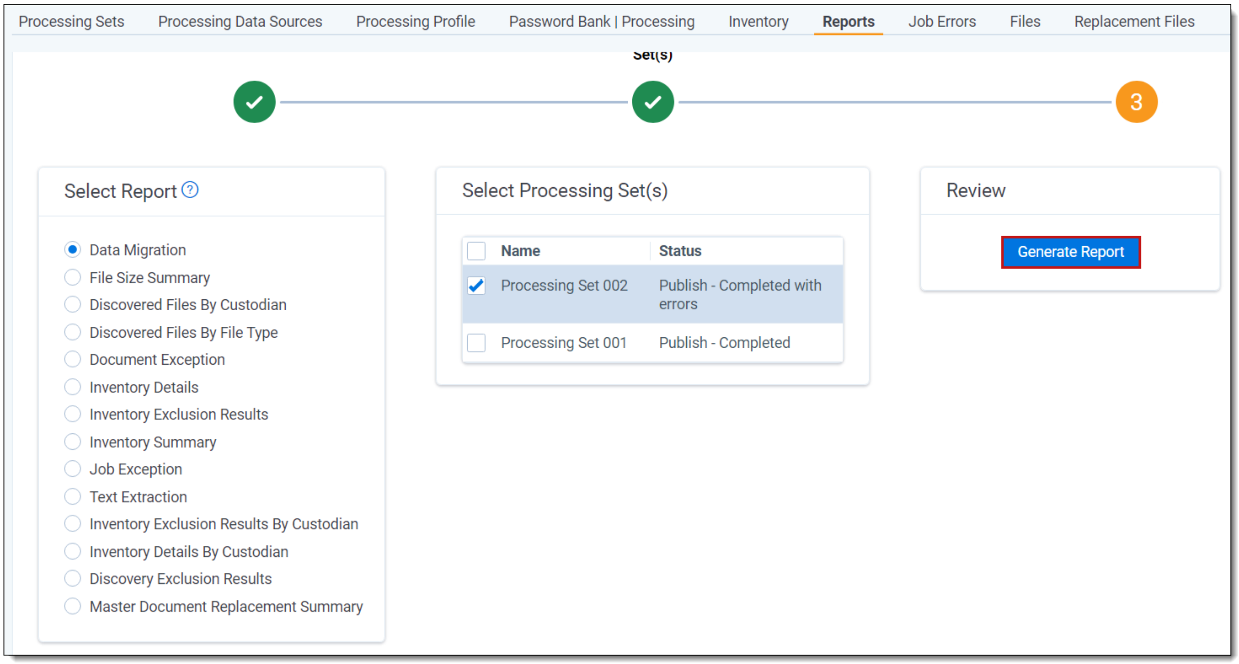
Imagine you're a litigation support specialist, and someone in your firm needs to see a hard copy of a report that shows them how many files have been discovered in their processing workspace per custodian. They need this report quickly because they're afraid that certain custodians were accidentally associated with the wrong data sources and processing sets.

To produce this, go to the Reports tab under Processing and run the Discovered Files by Custodian report for the processing set(s) that your firm suspects are wrong.

22.1 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**.



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

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.

22.2 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.

22.2.1 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 files that may have been denisted.
- **Excluded Files**—each custodian's total number of excluded files.
- **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.

22.2.2 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.

22.2.3 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.

22.3 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.

22.3.1 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.

22.3.2 Replacements Master Documents

This table lists all files deleted during discovery 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.

22.3.3 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.

22.4 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.

22.4.1 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.

22.4.2 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.

22.4.3 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.

22.5 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.

22.5.1 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.

22.5.2 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.

22.5.3 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.

22.5.4 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.

22.5.5 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.

22.5.6 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.

22.6 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.

22.6.1 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.

22.6.2 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.

22.6.3 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.

22.6.4 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.

22.6.5 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.

22.6.6 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.

22.7 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.

22.7.1 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.

22.7.2 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.

22.7.3 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.

22.8 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.

22.8.1 Pre-Processed File Size

This table lists the pre-processed file size for all loose and compressed file sizes in the source location.

22.8.2 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

22.8.3 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

22.9 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 Reports on page 254 for more information. You can run this report on inventoried processing sets.

22.9.1 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.

22.9.2 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.

22.9.3 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.

22.9.4 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.

22.9.5 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.

22.10 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.

22.10.1 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:

22.10.2 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.

22.10.3 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.

22.10.4 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.

22.10.5 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.

22.11 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 Reports on page 254 for more information. You can run this report on inventoried processing sets.

22.11.1 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.

22.11.2 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.

22.11.3 Location | Excluded File Count

This table lists all file locations that were filtered out and the number of files per location that were excluded.

22.11.4 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.

22.11.5 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.

22.12 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:

22.12.1 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.

22.12.2 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.

22.12.3 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.

22.12.4 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.

22.13 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 Reports on page 254 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.

22.13.1 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.

22.13.2 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.

22.13.3 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.

22.13.4 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.

22.14 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.

22.14.1 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.

22.14.2 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.

22.15 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.

22.15.1 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.

22.15.2 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.

22.15.3 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.

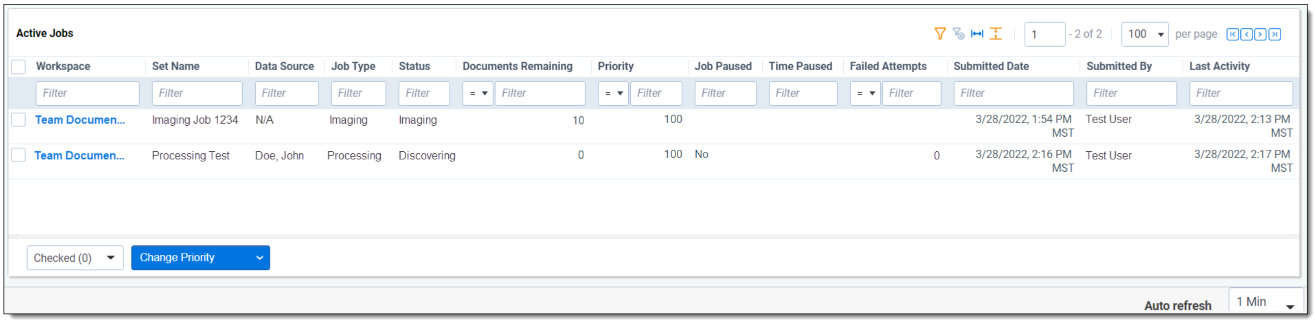
22.15.4 Processing Sets

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.

23 Processing administration

Processing administration consists of managing active jobs and viewing processing history. You can manage active jobs via the Queue Management > Processing and Imaging Queue tabs. You will find the Queue Management tab on the Relativity home page. Access Processing History from your workspace menu options. You can also use the search bar to jump directly to either page.



☐	Workspace	Set Name	Data Source	Job Type	Status	Documents Remaining	Priority	Job Paused	Time Paused	Failed Attempts	Submitted Date	Submitted By	Last Activity
	Filter	Filter	Filter	Filter	Filter	= Filter	= Filter	Filter	Filter	= Filter	Filter	Filter	Filter
☐	Team Documen...	Imaging Job 1234	N/A	Imaging	Imaging	10	100				3/28/2022, 1:54 PM MST	Test User	3/28/2022, 2:13 PM MST
☐	Team Documen...	Processing Test	Doe, John	Processing	Discovering	0	100	No		0	3/28/2022, 2:16 PM MST	Test User	3/28/2022, 2:17 PM MST

23.1 Security considerations for processing administration

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.

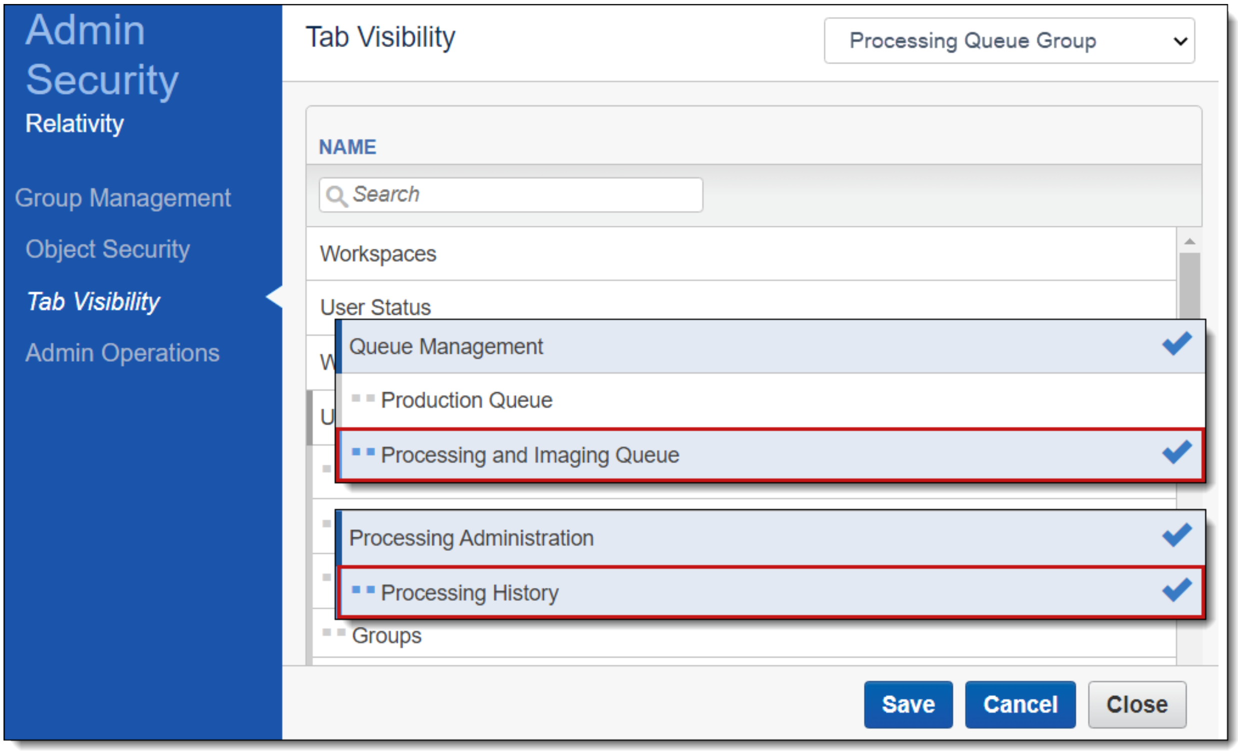
Note: To change the priority of a job where **Customer lockbox** is enabled, you must be in a permission group having access to the workspace where the job originated, along with the System Administrator's group.

Note: Only System Administrators can modify processing jobs on the Processing and Imaging Queue tab. Other users can see the Processing and Imaging Queue tab with instance level permissions, but see an error when attempting to modify processing jobs.

Groups do not have access to the Queue Management or Processing Administration tab or sub-tabs by default. To grant them access, perform the following steps:

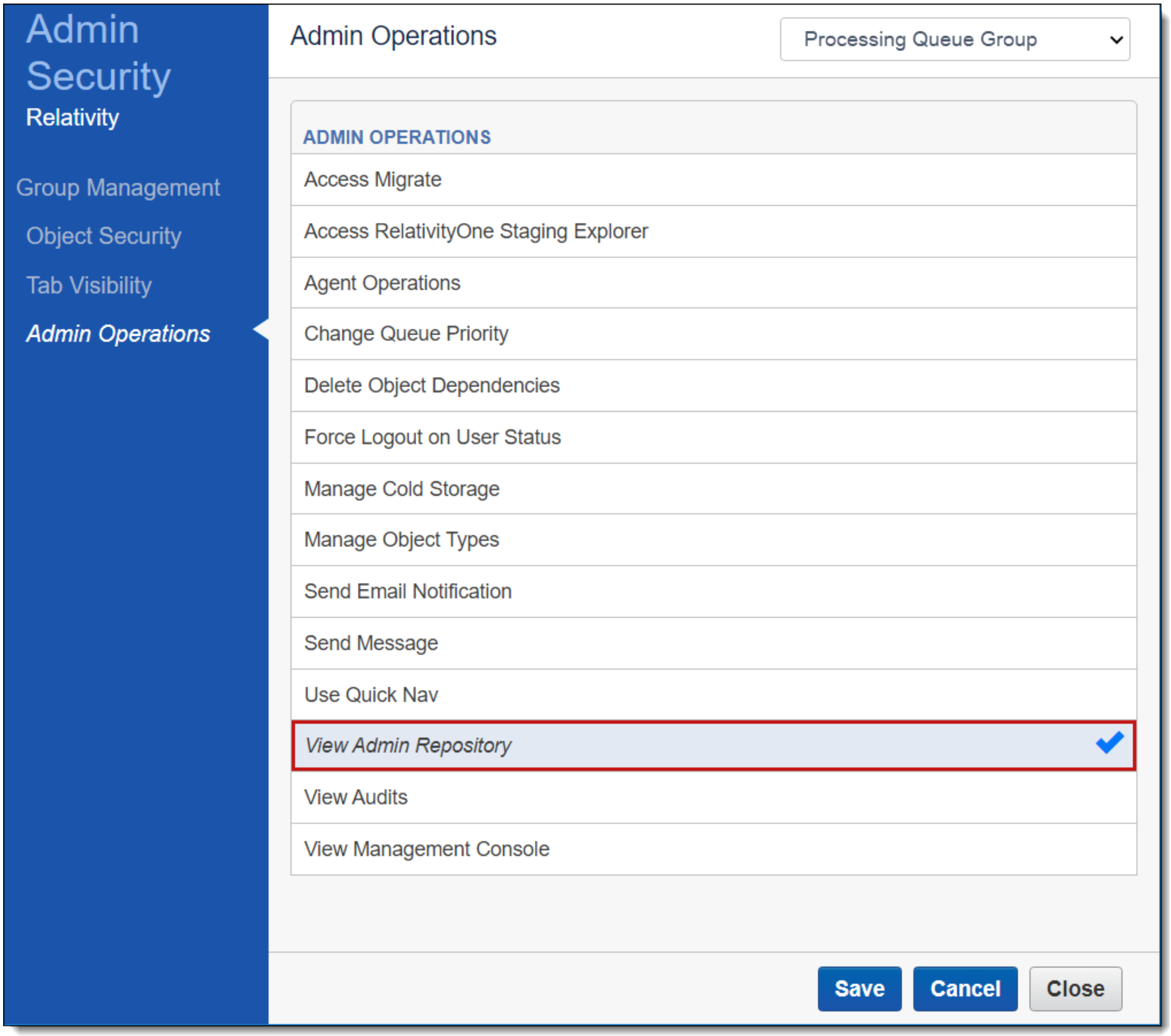
1. From the Relativity home page, navigate to the **Instance Details** tab. (Use the search bar to jump directly to the page.)
2. In the Security section, click **Manage Permissions**.
3. In the Admin Security window, select **Tab Visibility**.
4. From the drop-down list at the top right, select the group to whom you want to grant access.

5. Select **Processing and Imaging Queue** (under the **Queue Management** heading), and **Processing History** (under the Processing Administration heading.)


Admin Security
Relativity
Group Management
Object Security
Tab Visibility
Admin Operations
Tab Visibility
Processing Queue Group
NAME
Search
Workspaces
User Status
Queue Management ✓
Production Queue
Processing and Imaging Queue ✓
Processing Administration ✓
Processing History ✓
Groups
Save
Cancel
Close

6. Click **Save**.

You must also have the **View Admin Repository** permission set in the **Admin Operations** console.

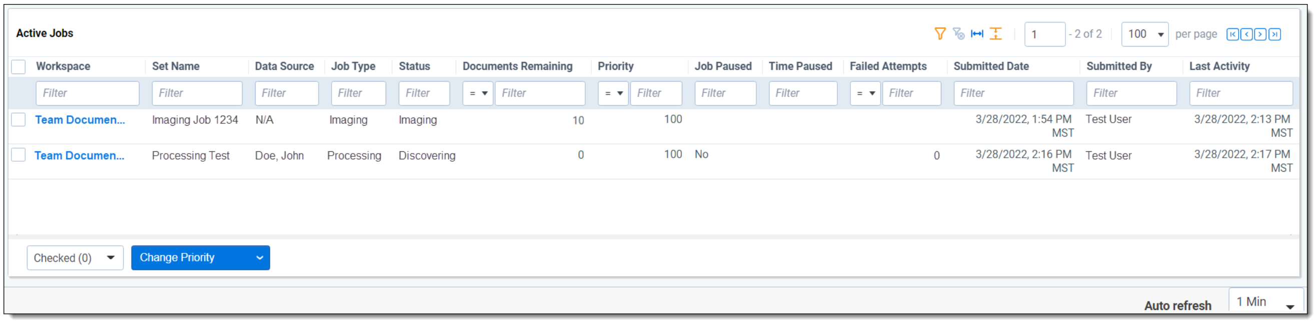

Admin Security
Relativity
Group Management
Object Security
Tab Visibility
Admin Operations
Admin Operations
Processing Queue Group
ADMIN OPERATIONS
Access Migrate
Access RelativityOne Staging Explorer
Agent Operations
Change Queue Priority
Delete Object Dependencies
Force Logout on User Status
Manage Cold Storage
Manage Object Types
Send Email Notification
Send Message
Use Quick Nav
View Admin Repository ✓
View Audits
View Management Console
Save
Cancel
Close

23.2 Monitoring active jobs

To see all active processing and imaging jobs in the environment, view the Active Jobs view in the Processing and Imaging Queue tab. If no jobs are visible in this view, it means there are no jobs currently running in the environment.

- Jobs that are running in workspaces to which you don't have permissions display the placeholder text "Item Secured" in the Active Jobs view. Actual job details are not visible. To permit visibility, see Workspace Security.
- The Workspaces tree on the left only contains workspaces in which an active job is currently running.

The following columns appear on Active Jobs view:



☐	Workspace	Set Name	Data Source	Job Type	Status	Documents Remaining	Priority	Job Paused	Time Paused	Failed Attempts	Submitted Date	Submitted By	Last Activity
☐	Team Documen...	Imaging Job 1234	N/A	Imaging	Imaging	10	100				3/28/2022, 1:54 PM MST	Test User	3/28/2022, 2:13 PM MST
☐	Team Documen...	Processing Test	Doe, John	Processing	Discovering	0	100	No		0	3/28/2022, 2:16 PM MST	Test User	3/28/2022, 2:17 PM MST

- **Workspace** – the workspace in which the job was created. Click the name of a workspace to navigate to the main tab in that workspace.
- **Set Name** – 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. The worker manager server handles processing and imaging jobs.

  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
  - Imaging
  - Initializing
  - Retrieving/Retrying Errors
  - Submitting Job
- **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.

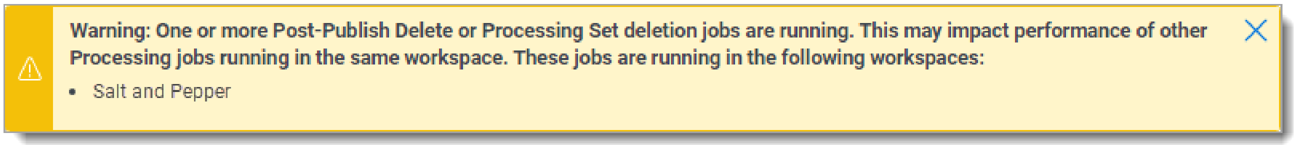
    - Discovery, publishing, and imaging jobs are multi-threaded and can run in parallel, depending on the number of agents available.
    - Job types have the following priorities set by default:
        - Imaging/TIFF-on-the-fly jobs have a priority of 1 by default and will always run first.
        - Publishing jobs have a priority of 90 and will always run after any imaging on the fly jobs and before all other jobs.
        - Inventory, Discovery, Mass Imaging/Imaging Set 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 Support.
- **Job 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.
- **Paused Time** - 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.
- **Last Activity** - the date and time at which a job last communicated to the worker.

At the bottom of the screen, the active jobs mass operations buttons appear.

23.2.0.1 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.

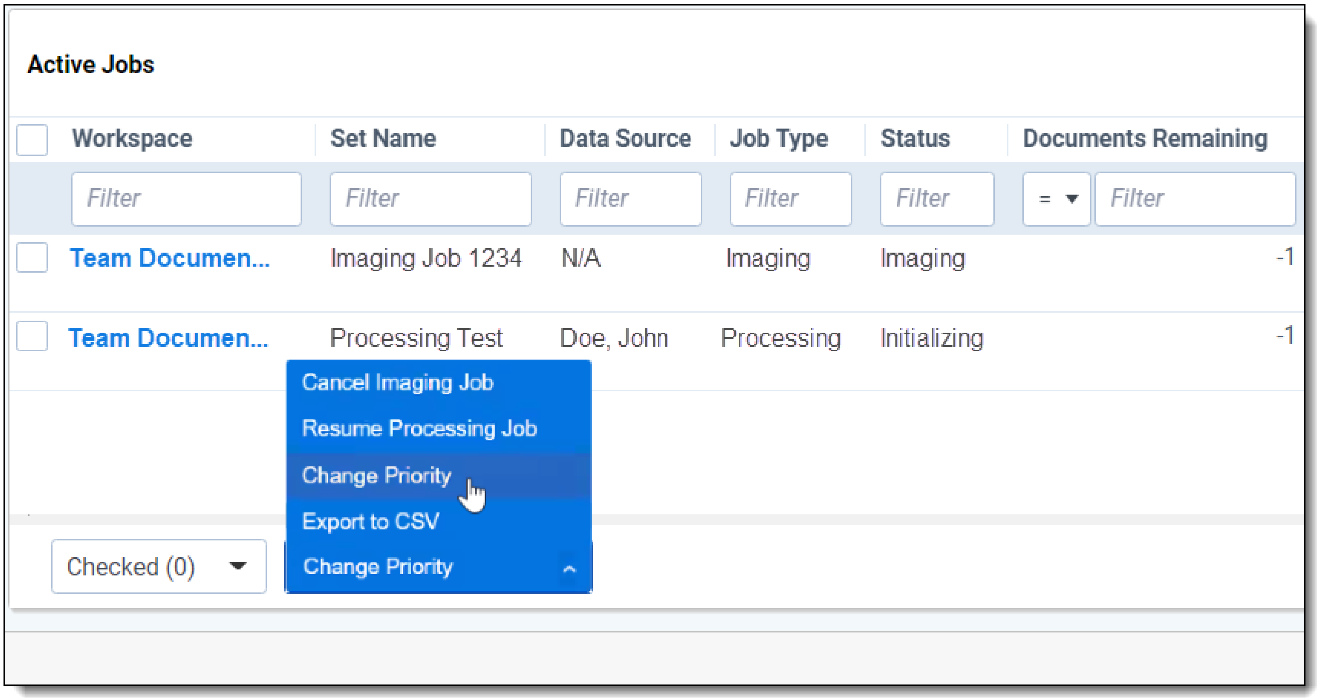


⚠ Warning: One or more Post-Publish Delete or Processing Set deletion jobs are running. This may impact performance of other Processing jobs running in the same workspace. These jobs are running in the following workspaces:

- Salt and Pepper

23.2.1 Active jobs mass operations

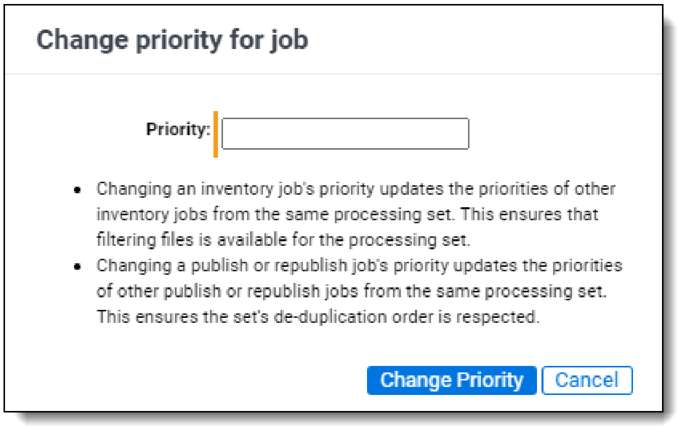
A number of mass operations are available on the Active Jobs view.



Active Jobs

☐	Workspace	Set Name	Data Source	Job Type	Status	Documents Remaining
	Filter	Filter	Filter	Filter	Filter	= Filter
☐	Team Documen...	Imaging Job 1234	N/A	Imaging	Imaging	-1
☐	Team Documen...	Processing Test	Doe, John	Processing	Initializing	-1

- **Cancel Imaging Job** - cancel an imaging job. If you have processing jobs selected when you click **Cancel Imaging Job**, the processing jobs are skipped over and are allowed to proceed. When you cancel an imaging job, it signals to the workers to finish their current batch of work, which may take a few minutes.
- **Resume Processing 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.

- 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.

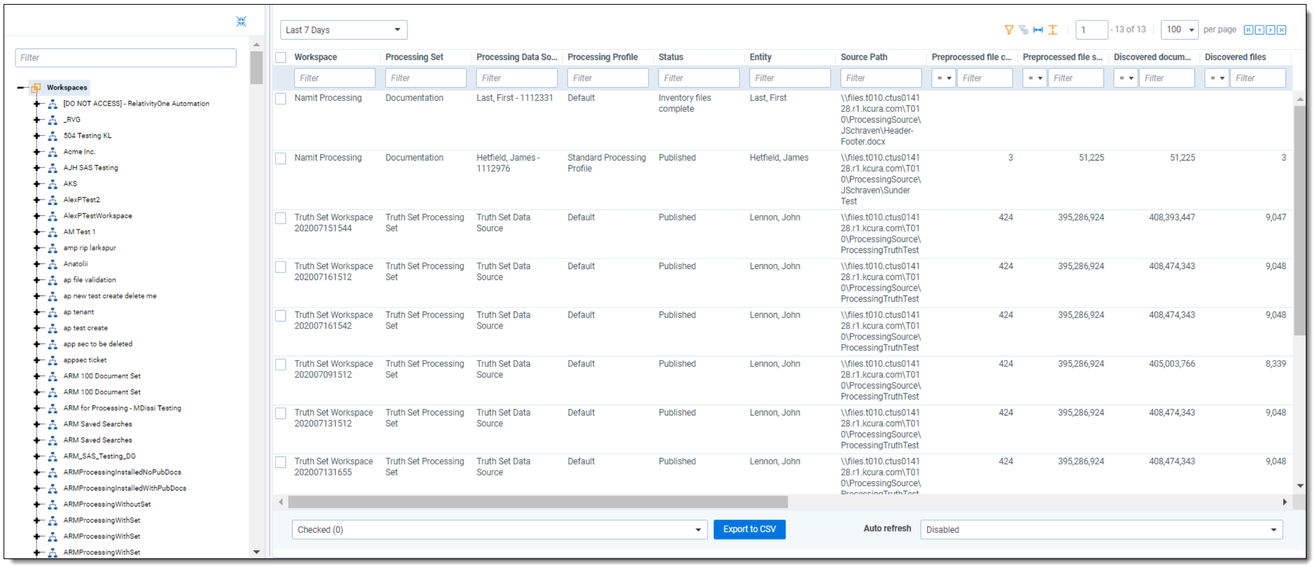
23.3 Using the Processing History sub-tab

To view the details of all processing actions taken on all data sources in the environment, navigate to the **Processing History** sub-tab.

In the Workspaces tree on the left, you'll see all workspaces in the environment that have at least one published document in them. You can expand the tree and click on processing sets and data sources to filter on them.

If you don't have permissions to a workspace, you'll see an "Item Restricted" message for that workspace.

The Processing History view provides the following fields:

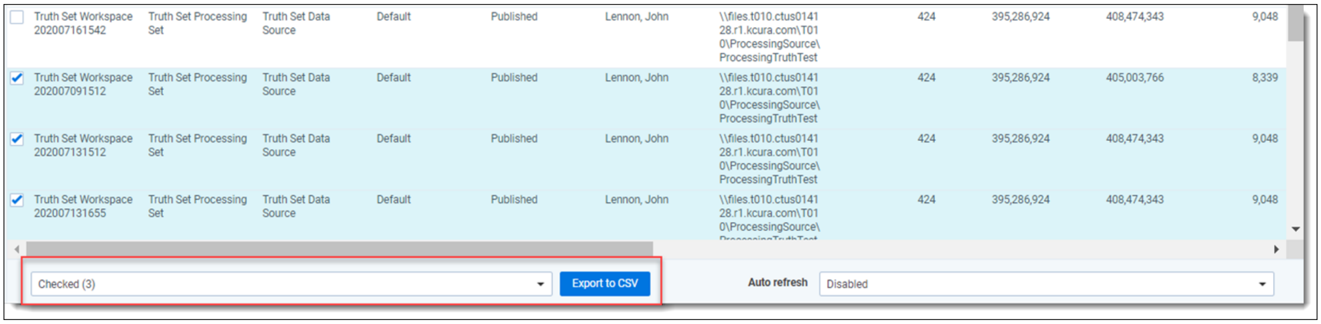

Filter
Workspaces
[DO NOT ACCESS] - RelativityOne Automation
_RVG
504 Testing KL
Acme Inc.
AJH SAS Testing
AKS
AlexPTest2
AlexPTestWorkspace
AM Test 1
amp rip larkspur
Anatolii
ap file validation
ap new test create delete me
ap tenant
ap test create
app sec to be deleted
appsec ticket
ARM 100 Document Set
ARM 100 Document Set
ARM for Processing - MDissi Testing
ARM Saved Searches
ARM Saved Searches
ARM_SAS_Testing_DG
ARMProcessingInstalledNoPubDocs
ARMProcessingInstalledWithPubDocs
ARMProcessingWithoutSet
ARMProcessingWithSet
ARMProcessingWithSet
ARMProcessingWithSet
Last 7 Days
1
- 13 of 13
100
per page
Checked (0)
Export to CSV
Auto refresh
Disabled

☐	Workspace	Processing Set	Processing Data So...	Processing Profile	Status	Entity	Source Path	Preprocessed file c...	Preprocessed file s...	Discovered docum...	Discovered files
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	= Filter	= Filter	= Filter	= Filter
☐	Namit Processing	Documentation	Last, First - 1112331	Default	Inventory files complete	Last, First	\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\JSchraven\Header-Footer.docx				
☐	Namit Processing	Documentation	Hetfield, James - 1112976	Standard Processing Profile	Published	Hetfield, James	\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\JSchraven\Sunder Test	3	51,225	51,225	3
☐	Truth Set Workspace 202007151544	Truth Set Processing Set	Truth Set Data Source	Default	Published	Lennon, John	\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\ProcessingTruthTest	424	395,286,924	408,393,447	9,047
☐	Truth Set Workspace 202007161512	Truth Set Processing Set	Truth Set Data Source	Default	Published	Lennon, John	\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\ProcessingTruthTest	424	395,286,924	408,474,343	9,048
☐	Truth Set Workspace 202007161542	Truth Set Processing Set	Truth Set Data Source	Default	Published	Lennon, John	\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\ProcessingTruthTest	424	395,286,924	408,474,343	9,048
☐	Truth Set Workspace 202007091512	Truth Set Processing Set	Truth Set Data Source	Default	Published	Lennon, John	\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\ProcessingTruthTest	424	395,286,924	405,003,766	8,339
☐	Truth Set Workspace 202007131512	Truth Set Processing Set	Truth Set Data Source	Default	Published	Lennon, John	\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\ProcessingTruthTest	424	395,286,924	408,474,343	9,048
☐	Truth Set Workspace 202007131655	Truth Set Processing Set	Truth Set Data Source	Default	Published	Lennon, John	\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\	424	395,286,924	408,474,343	9,048

- **Workspace** - the name of the workspace in which the processing job was run.
- **Processing Set** - the name of the processing set that was run.
- **Processing Data Source** - the name and artifact ID of the data source attached to the processing set.
- **Processing Profile** - the profile associated with the processing set.
- **Status** - the current status of the processing job.
- **Entity** - the entity associated with the data source.
- **Source Path** - the location of the data that was processed, as specified on the data source.
- **Preprocessed file count** - the count of all native files before extraction/decompression, as they exist in storage.
- **Preprocessed file size** - the sum of all the native file sizes, in bytes, before extraction/decompression, as they exist in storage.
- **Discovered document size** - the sum of all native file sizes discovered, in bytes, that aren’t classified as containers as they exist in storage.
- **Discovered files** - the number of files from the data source that were discovered.
- **Nisted file count** - the count of all files denisted out during discovery, if deNIST was enabled on the processing profile.
- **Nisted file size** - the sum of all the file sizes, in bytes, denisted out during discovery, if deNIST was enabled on the processing profile.
- **Published documents size** - the sum of published native file sizes, in bytes, associated to the user, processing set and workspace.
- **Published documents** - the count of published native files associated to the user, processing set and workspace.

- **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.
- **Last publish time submitted** - the date and time at which publish was last started on the processing set.
- **Discover time submitted** - the date and time at which discovery was last started on the processing set.
- **Last activity** - the date and time at which any action was taken on the processing set.

You have the option of exporting any available processing history data to a CSV file through the Export to CSV mass operation at the bottom of the view.



☐	Truth Set Workspace 202007161542	Truth Set Processing Set	Truth Set Data Source	Default	Published	Lennon, John	\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\ProcessingTruthTest	424	395,286,924	408,474,343	9,048
☑	Truth Set Workspace 202007091512	Truth Set Processing Set	Truth Set Data Source	Default	Published	Lennon, John	\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\ProcessingTruthTest	424	395,286,924	405,003,766	8,339
☑	Truth Set Workspace 202007131512	Truth Set Processing Set	Truth Set Data Source	Default	Published	Lennon, John	\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\ProcessingTruthTest	424	395,286,924	408,474,343	9,048
☑	Truth Set Workspace 202007131655	Truth Set Processing Set	Truth Set Data Source	Default	Published	Lennon, John	\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\	424	395,286,924	408,474,343	9,048

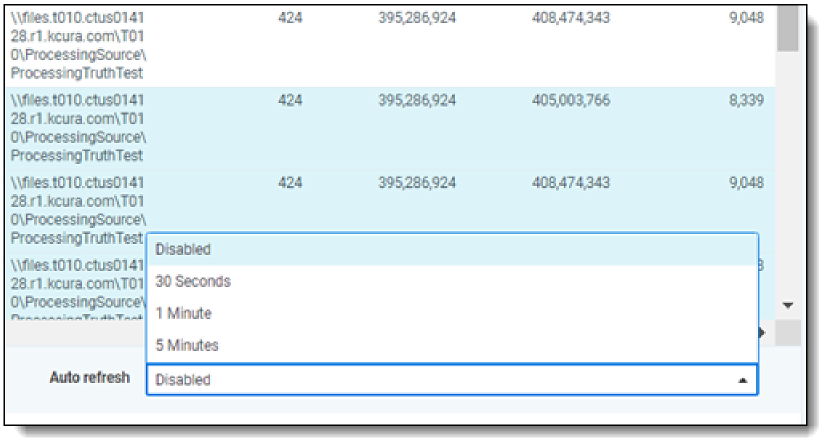
Checked (3)

Auto refresh: Disabled

23.3.1 Auto refresh options for processing history

The processing history tab receives processing information when loaded and update every time the page refreshes.

To configure the rate at which the view automatically refresh, select a value from the **Auto refresh** drop-down at the bottom right of the view.



\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\ProcessingTruthTest	424	395,286,924	408,474,343	9,048
\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\ProcessingTruthTest	424	395,286,924	405,003,766	8,339
\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\ProcessingTruthTest	424	395,286,924	408,474,343	9,048
\\files.t010.ctus014128.r1.kcura.com\T010\ProcessingSource\				

Auto refresh: Disabled

- **Disabled** - prevents the automatic refresh of the view and makes it so that processing history information only updates when you manually refresh the page. This option is useful at times of heavy processing usage, in that it offers you more control over the refresh rate and prevents the information from constantly changing often while you monitor the work being performed. We've set this as the default because if your environment contains many workspaces and data sources, it could take a long time to load all of the data, which you may not want to update on an auto-refresh interval.

- **30 seconds** - arranges for the processing history view to automatically refresh every thirty seconds.
- **1 minute** - arranges for the processing history view to automatically refresh every one minute.
- **5 minutes** - arranges for the processing history view to automatically refresh every five minutes.

24 Frequently asked questions

If you have a question about processing, consult the following FAQs:

Can file names be appended after discovery?
There's currently not a way to append a file name after discovery but prior to publishing.

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.

Does Relativity allow the use of a custom NIST?
There's no official support for a custom NIST list.

Does Relativity process password protected PST or OST files?
Passwords on PST and OST files are bypassed automatically by Relativity.

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.

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 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.

How does processing handle regional setting changes?
When a user submits a ticket to change the worker regional setting (worker time zone/culture), it is best to have no processing jobs running. If a job is running and the regional setting changes, it may affect how imaging and text extraction interpret things such as dates, currency formats, and deduplication on the jobs that are in progress.

How does processing handle time zones?
Discovery is performed on all natives in UTC. Processing uses the timezone as defined in the processing set settings to convert metadata dates and times into the selected timezone. For Daylight Savings, there is a table called dbo.TimeZone in the Invariant database that is used to account for Daylight Savings Time on a year-by-year basis. This way, we always use the accurate DST rule for the given year.

For example, a change to how we observe DST went into effect in 1996, and we have this stored. The TimeZone table also keeps track of all of the half-hour time zones, i.e. parts of India.

Can the date and time format be modified?
Yes. The date and time format reflect the region where Relativity is deployed and will change depending on the location. Contact your Customer Success Manager for questions regarding this and other instance settings.

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.

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.

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.

©2024. Relativity ODA LLC. All rights reserved. Relativity® is a registered trademark of Relativity ODA LLC.