Processing User Guide

February 29, 2024 | Server 2022

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

Table of Contents

1 Processing **10**

1.1 Application version considerations 10

1.2 Basic processing workflow 11

1.3 Logging for processing 11

2 Installing and configuring Processing **13**

2.1 Upgrade considerations for Relativity Juniper 13

2.2 Installation process 13

2.3 Throttle settings for distributed publish 14

2.4 License considerations 15

2.5 Importing the Processing application 16

2.6 Worker manager server 16

2.6.1 Designating a worker for processing 16

2.7 Entering the ProcessingWebAPiPath instance setting 17

2.8 Adding the worker manager server to a resource pool 19

2.9 Configuring processing agents 20

2.10 Creating a choice for the processing source location 21

2.11 Logging for processing 22

2.12 Security permissions 23

3 Processing to Data Grid **24**

3.1 Enabling processing to Data Grid 24

4 Supported file types for processing **27**

4.1 Supported file types 27

4.1.1 Excel file considerations 30

4.1.2 Relativity Short Message Format mapping considerations 30

4.1.3 Multi-part forensic file considerations 31

4.1.4 Native text extraction and OCR 31

4.1.5 Support for password-protected Roshal Archive files 31

4.1.6 Outlook message item (.msg) to MIME encapsulation (.mht) conversion considerations 32

4.1.7 Email image extraction support 33

4.1.8 Microsoft Office child extraction support 33

4.2 Notable unsupported file types 35
4.3 Supported container file types 38
4.3.1 Lotus Notes considerations 39
4.3.2 Multi-part container considerations 41
4.3.3 Calendar file, vCard file considerations 41
4.4 Container file types supported for the password bank 44
4.4.1 Non-container file types supported for Password Bank in Inventory 45
5 Password bank 46
5.1 Password bank in processing workflow 46
5.2 Password Bank in imaging workflow 47
5.3 Creating or deleting a Password Bank entry 48
5.3.1 Fields 48
5.3.2 Example password 51
5.4 Validations, errors, and exceptions 51
5.5 Viewing audits 52
6 Mapping processing fields 53
6.1 Mapping fields 53
6.1.1 Processing system field considerations 54
6.1.2 Field mapping validations 55
6.2 System-mapped processing fields 55
6.3 Optional processing fields 56
6.4 Email Store Name details 79
6.5 Virtual path details 80
6.6 Processing folder path details 81
6.7 Email folder path details 82
6.8 Source path details 82
6.9 Message ID considerations 83
6.10 Comments considerations 83
6.11 Deduped custodian and path considerations 83
7 Processing profiles 85
7.1 Creating or editing a processing profile 85
7.1.1 Fields 86

7.1.2 dtSearch special considerations 108
7.1.3 Text extraction method considerations 109
8 Deduplication considerations 112
8.1 Global deduplication 113
8.2 Custodial deduplication 114
8.3 No deduplication 114
8.4 Global deduplication with attachments 115
8.5 Global deduplication with document-level errors 116
8.6 Technical notes for deduplication 117
8.6.1 Calculating MD5/SHA1/SHA256 hashes 117
8.6.2 Calculating deduplication hashes for emails 118
8.6.3 Calculating the Relativity deduplication hash 119
9 Quick-create set(s) 121
9.1 Required security permissions 121
9.2 Using quick-create set(s) 123
9.2.1 Validations and errors 130
10 Processing sets 132
10.1 Processing sets default view 133
10.2 Creating a processing set 135
10.3 Processing Set Fields 135
10.4 Adding a data source 137
10.5 Data Source Fields 139
10.5.1 Order considerations 144
10.5.2 Edit considerations for data sources 144
10.5.3 Processing Data Source view 145
10.5.4 Job Errors View 147
10.6 Processing Data Sources tab 148
10.7 Deleting a processing set 152
10.8 Avoiding data loss across sets 155
10.9 Copying natives during processing 155
11 Inventory 159
11.1 Running inventory 160

11.1.1 Inventory process 162
11.1.2 Monitoring inventory status 163
11.1.3 Canceling inventory 164
11.2 Filtering files 165
11.2.1 Applying a Date range filter 168
11.2.2 Applying a File Size filter 171
11.2.3 Applying a deNIST filter 172
11.2.4 Applying a Location filter 172
11.2.5 Applying a File Type filter 173
11.2.6 Applying a Sender Domain filter 174
11.3 Removing filters 178
11.4 Inventory progress 178
11.5 Discovering files from Inventory 180
11.6 Inventory errors 180
11.6.1 Inventory error scenarios 181
11.7 Re-inventory 181
12 Discovering files 184
12.1 Running file discovery 185
12.1.1 Discovery process 189
12.1.2 Container extraction 191
12.2 Special considerations - OCR and text extraction 191
12.3 Monitoring discovery status 192
12.4 Viewing text extraction progress in processing sets 194
12.5 Canceling discovery 195
12.5.1 Canceling discovery 195
13 Files tab 198
13.1 Views on the Files tab 198
13.1.1 All Files view 199
13.1.2 Deleted Documents view 200
13.1.3 Current Errored Files view 201
13.1.4 All Errored Files view 202
13.2 Details modal 204

13.3 Retrying delete errors 207
13.4 Republishing files from the Files tab 207
13.4.1 Common use cases for using the Republish mass operation 209
13.5 Saved filters 210
13.5.1 Right-click options 213
13.6 Download / Replace 215
14 Publishing files 219
14.1 Running file publish 220
14.1.1 Publish process 224
14.2 Monitoring publish status 226
14.3 Canceling publishing 227
14.4 Republishing a processing set 228
14.5 Retrying errors after publish 230
15 Post-publish delete 231
15.1 Post-publish delete overview 231
15.2 Publishing a new master document 231
15.3 Deleting documents within a family 234
15.4 Retrying delete errors 237
16 Processing error workflow 239
16.1 Required security permissions 239
16.2 Processing errors tabs 239
16.2.1 Files tab 239
16.2.2 Job Errors tab 240
16.2.3 Job Error layout 242
16.3 Useful error field information 244
16.3.1 Combined error fields 244
16.3.2 Error status information 245
16.3.3 Error Category list 245
16.3.4 Details modal 247
16.3.5 Pivotable error fields 250
16.4 File error actions 250
16.4.1 Processing Set error retry 250

16.5 Files tab error actions 252
16.6 Common workflows 254
16.6.1 Identifying and resolving errors 254
16.6.2 Replacing a corrupted file 255
17 Reports 256
17.1 Generating a processing report 256
17.2 Data Migration 257
17.2.1 Excluded Files 257
17.2.2 Summary Statistics: Data Migration 258
17.2.3 Processing Sets 258
17.3 Master Document Replacement Summary 258
17.3.1 Deleted Master Documents 258
17.3.2 Replacements Master Documents 259
17.3.3 Processing Sets 259
17.4 Discovery File Exclusion 259
17.4.1 Discover Filter Settings 259
17.4.2 File Type | File Size | Excluded File Count 259
17.4.3 Processing Sets 259
17.5 Discovered Files by Custodian 260
17.5.1 Discovered Files by Custodian 260
17.5.2 File Types Discovered - Processable 260
17.5.3 File Types Discovered - Processable(By Custodian) 260
17.5.4 File Types Discovered - Unprocessable 260
17.5.5 File Types Discovered - Unprocessable (by Custodian) 260
17.5.6 Processing Sets 261
17.6 Discovered Files by File Type 261
17.6.1 Discovered Files by Custodian 261
17.6.2 File Types Discovered - Processable 261
17.6.3 File Types Discovered - Processable (By File Type) 261
17.6.4 File Types Discovered - Unprocessable 261
17.6.5 File Types Discovered - Unprocessable (By File Type) 262
17.6.6 Processing Sets 262

17.7 Document Exception 262
17.7.1 Document Level Errors - Discovery 262
17.7.2 Document Level Errors - Publishing 262
17.7.3 Processing Sets 263
17.8 File Size Summary 263
17.8.1 Pre-Processed File Size 263
17.8.2 Processed File Size 263
17.8.3 Published File Size 263
17.9 Inventory Details 263
17.9.1 Inventory Filter Settings 264
17.9.2 Excluded by File Type Filter | Excluded File Count 264
17.9.3 Excluded by Location Filter | Excluded File Count 264
17.9.4 Excluded by Sender Domain Filter | Excluded File Count 264
17.9.5 Processing Sets 264
17.10 Inventory Details by Custodian 264
17.10.1 Inventory Filter Settings 264
17.10.2 Custodian | Excluded by File Type Filter | Excluded File Count 265
17.10.3 Custodian | Excluded by File Location Filter | Excluded File Count 265
17.10.4 Custodian | Excluded by Sender Domain | Excluded File Count 265
17.10.5 Processing Sets 265
17.11 Inventory Exclusion Results 265
17.11.1 Inventory Filter Settings 265
17.11.2 File Type | Excluded File Count 266
17.11.3 Location | Excluded File Count 266
17.11.4 Sender Domain | Excluded File Count 266
17.11.5 Processing Sets 266
17.12 Inventory Exclusion Results by Custodian 266
17.12.1 Custodian | Excluded by File Type Filter | Excluded File Count 266
17.12.2 Custodian | Excluded by File Location Filter | Excluded File Count 266
17.12.3 Custodian | Excluded by Sender Domain | Excluded File Count 266
17.12.4 Processing Sets 267
17.13 Inventory Summary 267

17.13.1 Initial Inventory Results 267
17.13.2 Filtering Summary 267
17.13.3 Final Inventory Results 268
17.13.4 Processing Sets 268
17.14 Job Exception 268
17.14.1 Job Level Errors 268
17.14.2 Processing Sets 268
17.15 Text Extraction 268
17.15.1 Text Extraction by Custodian 269
17.15.2 Text Extraction by File Type 269
17.15.3 Breakdown by Error Message 269
17.15.4 Processing Sets 269
18 Processing Administration 271
18.1 Security considerations for processing administration 271
18.2 Monitoring active jobs 273
18.2.1 Active jobs mass operations 275
18.3 Checking worker and thread status 276
18.3.1 Worker mass operations 280
18.3.2 Auto refresh options 281
18.3.3 Thread data visibility 282
18.3.4 Errors 282
18.4 Using the Processing History sub-tab 283
18.4.1 Auto refresh options for processing history 285
19 Managing processing jobs in the queue 286
20 Processing FAQs 289

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.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

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: Processing does not perform language identification. For information on how to perform language identification using Analytics, see the Language identification section of the Analytics Guide.

To gain control over more complex processing jobs on a granular level, you can use the Relativity Processing Console desktop application. For more information, see the Relativity Processing Console Guide.

1.1 Application version considerations

All the content in this section and its related pages correspond to the latest version of the Processing application, which is updated on a monthly basis with the release of each patch of Relativity Server 2022.

If the processing components in your environment do not match the descriptions in this content exactly, it may be because you are using an older version of the Processing application. To get the newest version of the Processing application, upgrade to the latest product update of Relativity Server 2022.

For a list of changes made to processing per monthly product update, see the Relativity Server 2022 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.

1.2 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. Processing sets on page 132
   - 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 on page 159
   - 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. Discovering files on page 184and Publishing files on page 219
   - Post publish—view, ignore, or retry any errors that occurred during any phase of the processing job. If needed, republish the files.

1.3 Logging for processing

The logging framework enables you to efficiently gather runtime diagnostic information. You can use logging for troubleshooting application problems when you need a very granular level of detail, for example, when working with a Relativity Support representative.

Relativity system components that can log messages are identified based on the system-subsystem-application designation. When troubleshooting, use the system-subsystem-application matrix to configure logging to target a specific Relativity component, such as the Processing application.

Note: It is recommended that you not set your logging to verbose when publishing documents to a workspace, as doing so can cause your worker to run out of resources, such as CPU, RAM, disk usage, or others, which then causes your publish job to cease entirely. If you need to use verbose logging in order to collect detailed logs, do so for short periods of time only, under 5 minutes, and have a developer on hand to troubleshoot if any issues occur.

For more information, see the Relativity Logging guide.

2 Installing and configuring Processing

This topic provides information on installing and configuring processing so that you can run it in your Relativity environment.

You must have the following in order to use processing:

- A processing license. For steps on obtaining a Processing license, see the Licensing Guide.
- The worker manager server installed and configured. For more information, see the Upgrade Guide.
- The worker manager server attached to the resource pool in which the workspace resides.
- A token-authentication processing Web API path specified for the ProcessingWebApiPath entry in the Instance Settings table.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

2.1 Upgrade considerations for Relativity Juniper

Starting in Server 2021, Invariant requires worker server installation on each worker to function properly. This ensures workers have the appropriate Microsoft Visual C++ 2019 redistributables necessary.

When installing the RPC, the username must be EDDSDBO for Relativity Servercustomers.

2.2 Installation process

The following steps make up the comprehensive procedure for installing processing:

1. Review the most current system processing requirements to ensure that you have enough resources to run processing. For more information, see the System Requirements guide.
2. Review the worker manager server pre-installation steps to ensure that your environment is updated with the required security and other configuration settings. For more information, see the Worker Manager Server Installation guide.
3. Run the Invariant installer for a fresh install or an upgrade. For more information, see the Worker Manager Server Installation guide.
4. Add/configure the worker manager server in the Servers tab. For more information, see the Servers chapter of the Admin Guide.
5. Add the worker manager server to a resource pool. For more information, see Adding the worker manager server to a resource pool.

6. Enter the ProcessingWebAPIPath instance setting. For more information, see Entering the ProcessingWebAPIiPath instance setting.
7. Import the processing application. For more information, see Importing the Processing application.
8. Configure the processing agents (if they weren't automatically created when you imported the Processing application). For more information, see Configuring processing agents.
9. Create a choice for the processing source location. For more information, see Creating a choice for the processing source location.

2.3 Throttle settings for distributed publish

Note the following recommended throttle settings for distributed publish:

- The following instance settings have been added to facilitate the work of distributed publish.
  - ProcessingMaxPublishSubJobCountPerRelativitySQLServer - the maximum number of publish jobs per Relativity SQL server that may be worked on in parallel.
    - This puts an absolute limit on the number of publish jobs that occur in parallel for a given SQL server, independent of how many workspaces may be publishing simultaneously. This means that it overrides the limit set by ProcessingMaxPublishSubJobCountPerWorkspace.
    - The default value is 21. Leaving this setting at its default value will result in increased throughput; however, we recommend contacting Support before you upgrade for guidance on what value will be most beneficial to you based on your environment setup.
    - 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.
  - ProcessingMaxPublishSubJobCountPerWorkspace- the maximum number of publish jobs per workspace that may be worked on in parallel.
    - You can't allocate more jobs per workspace than what is allowed per SQL server. This means that if this value is set to be higher than the value for the MaxPublishJobCountPerRelativitySQLServer instance setting, then Relativity only allows the maximum of jobs per SQL server. For example, if you have a workspace limit of 4 and a server limit of 8 and all of your workspaces are on the same SQL server, you will have at most 8 publish sub jobs running concurrently.
    - The default value is 7. Leaving this setting at its default value will result in increased throughput; however, we recommend contacting Support before you upgrade for guidance on what value will be most beneficial to you based on your environment setup.

      Note: The default value of this setting was changed from 3 to 5 in Relativity 9.6.202.10.

- 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.

- The **ProcessingExportMaxThreads** instance setting has been deprecated in accordance with the addition of the ProcessingMaxPublishSubJobCountPerWorkspace and ProcessingMaxPublishSubJobCountPerRelativitySQLServer instance settings, which facilitate the work of distributed publish.

The following table provides the recommended values for each instance setting per environment setup:

Environment setup	ProcessingMaxPublishSubJobCountPerWorkspace	ProcessingMaxPublishSubJobCountPerRelativitySQLServer
Tier 1 (see the System Requirements Guide for details)	5	7
Tier 2 (see the System Requirements Guide for details)	6	12
RelativityOne baseline	5	7

2.4 License considerations

You are unable to process data in Relativity if any of the following conditions are true:

- You don't have a processing license associated with your environment.
- The processing license associated with your environment is invalid.
- The processing license associated with your environment is expired.
- The worker manager server associated with the resource pool is not included in the processing license.

Contact your system admin if any of these occur. See the Admin Guide for more information on Licensing.

Note: You can add processing capacity to your environment by adding hardware and additional licenses. For more information, contact your system admin.

2.5 Importing the Processing application

To install processing in your Relativity environment, import the Processing application from the application library. To do this, you must have the appropriate system admin rights.

You must have obtained a processing license before you can import the Processing application. For steps on obtaining a Processing license, see the Licensing Guide.

To import the Processing application:

1. Navigate to the **Relativity Applications** tab.
2. Click **New Relativity Application**.
3. Select **Select from Application Library**.
4. Click ... on the **Choose from Application Library** field.
5. Select **Processing** and click **OK**.
6. Click **Import**.

2.6 Worker manager server

The worker manager server uses workers to perform imaging, and all phases of processing, including inventory, discovery, and publish. You can configure the default queue priorities for your entire environment on the Worker Manager Server layout. If you are not licensed for processing, then the worker manager server only handles save as PDF and imaging.

To enable processing in your workspace, you must add a worker manager server to your Relativity environment through the Servers tab available from Home. For information on how to do this, see the Servers section of the Admin Guide.

Note: Don't restart a worker manager server if there are currently processing jobs running on it, as you'll need to recreate those jobs and re-run them once the server has completed restarting.

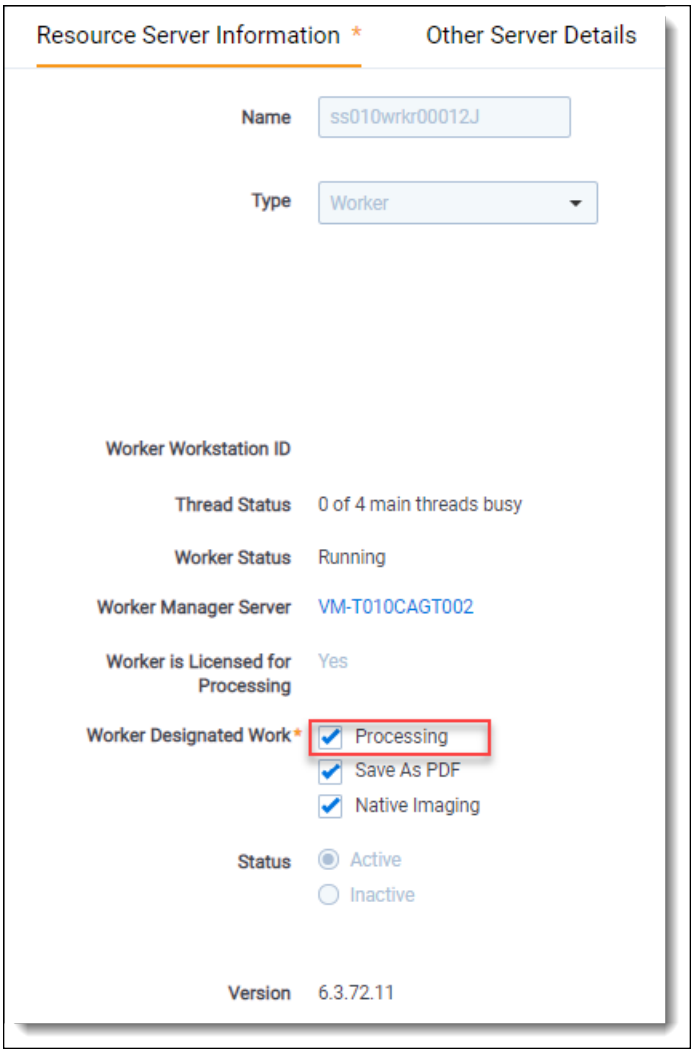
2.6.1 Designating a worker for processing

In order to process files, you need to designate at least one worker for processing.

To designate a worker for processing, perform the following steps:

1. Navigate to the **Servers** sub-tab.
2. From the list of servers, select the worker(s) on your worker manager server that you need to perform processing jobs.
3. Click **Edit** on the worker layout and navigate to the **Worker Designated Work** field.

4. Check the box next to the **Processing** choice.



5. Click **Save**.

2.7 Entering the ProcessingWebAPiPath instance setting

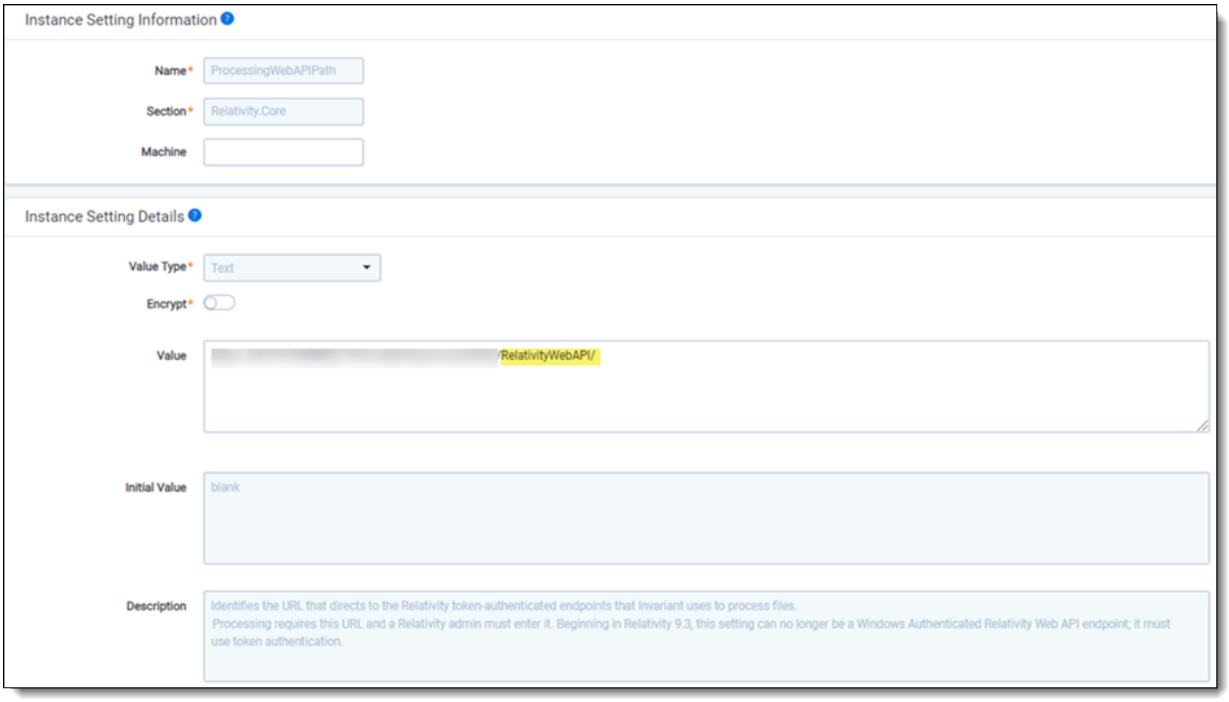
You must enable token authentication on your web server for certain Relativity features, such as the worker manager server, which requires this authentication type for processing.

You must also edit the ProcessingWebAPIPath Instance Setting. This setting identifies the URL that directs to the Relativity token-authenticated endpoints that Invariant uses to process and image files. Invariant requires this URL and a Relativity admin must enter it.

To do this, perform the following steps to comply with this change:

1. While in Home mode, navigate to the **Instance Settings** sub-tab.
2. In the default All Instance Settings view, enable filters and enter **ProcessingWebAPIPath** in the **Name** field.
3. Click the **ProcessingWebAPIPath** name and click **Edit** in the instance setting layout.

4. In the **Value** field change the existing ProcessingWebAPI URL to the RelativityWebAPI URL.



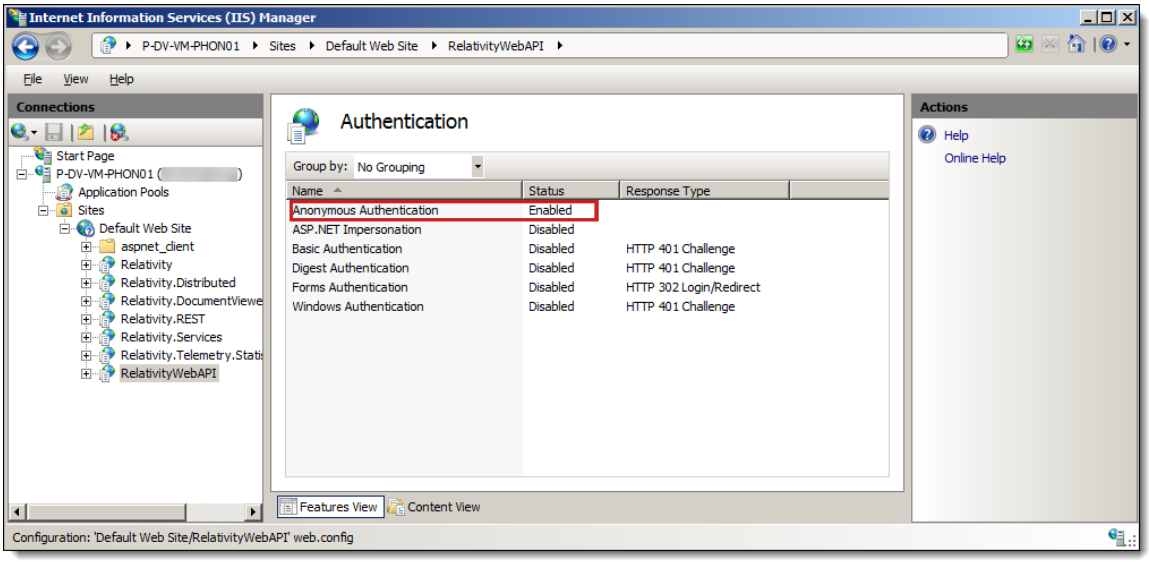
5. Click **Save**.

Depending on what Relativity version you're installing or upgrading, you may need to enable the RelativityWebAPI setting in IIS for Anonymous authentication in order to publish documents to a workspace.

To do this, perform the following steps:

1. Open IIS.
2. To enable anonymous authentication, complete the following steps:
   a. Click on the **RelativityWebAPI** site.
   b. In the Features view, click **Authentication**.

c. In the Authentication view, right-click on **Anonymous Authentication** and click Enable.



d. To update the web.config file, locate it in the following folder:
C:\Program Files\Relativity Corporation\Relativity\WebAPI

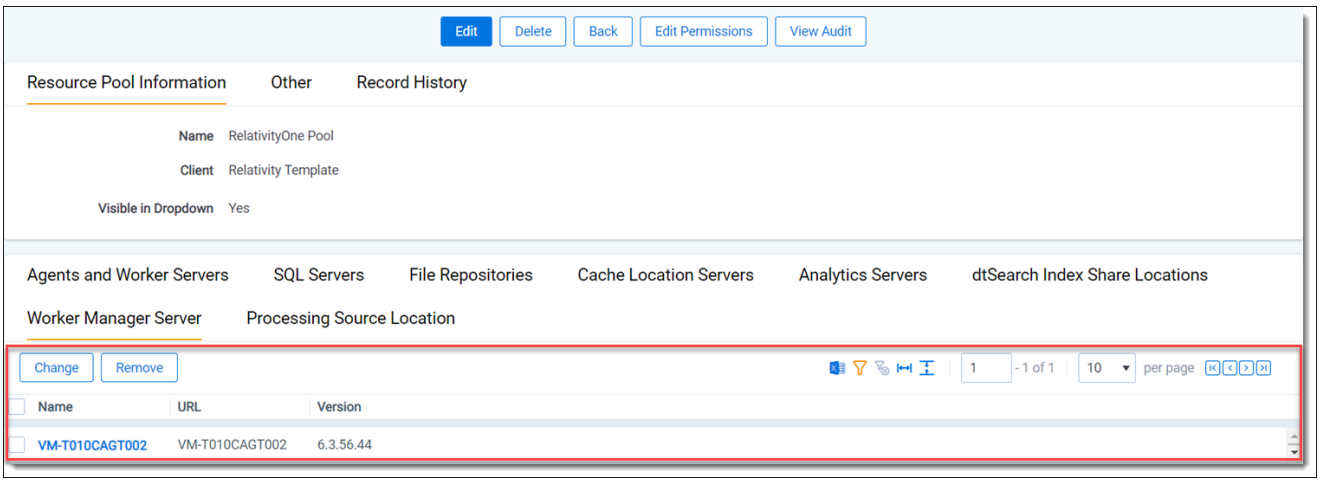
e. Open the file in an editor. Update the authentication mode and authorization sections as follows:

```
<system.web>
     <authentication mode="None" />
     <authorization><allow users="*" />
</authorization>
</system.web>
```

3. Restart IIS.

2.8 Adding the worker manager server to a resource pool

You must add the worker manager server to the resource pool associated with the workspace that is hosting processing. You can only have one worker manager server per resource pool.



Note: Don't change the worker manager server in a resource pool after you've processed data in a workspace that uses that resource pool. Changing the worker manager server after data has been processed causes unexpected results with retrying errors, deduplication, and document numbering. This is because a new server is not aware of what has happened in the workspace before it was added.

2.9 Configuring processing agents

The Processing application uses the following agents:

- **Server Manager** - retrieves version information from the worker manager server and updates the processing queue tab with this information.
- **Processing Set Manager** - manages the running of processing sets, retrieves errors encountered while sets are running, and picks up processing set deletion jobs and submits them to the worker manager server.

We recommend running two Processing Set Manager agents and adding more of them as needed.

To manually install processing agents, perform the following steps:

1. Navigate to the **Agents** tab.
2. Click **New Agent** and complete the following required fields:
   - **Agent Type** - click ... to display a list of agents. Filter for one of the processing agents, select the agent, and click **OK**.
   - **Number of agents** - enter the number of agents you want to add.
   - **Agent Server** - click ... to display a list of servers, then select a server and click **OK**. Select the regular agent server here, not the processing worker or processing queue manager.
   - **Run interval** - enter the interval, in seconds, at which the agent should check for available jobs.

- **Logging level of event details** - select **Log critical errors only** (recommended), **Log warnings and errors**, or **Log all messages**.
- **Enabled -** select **Yes.**

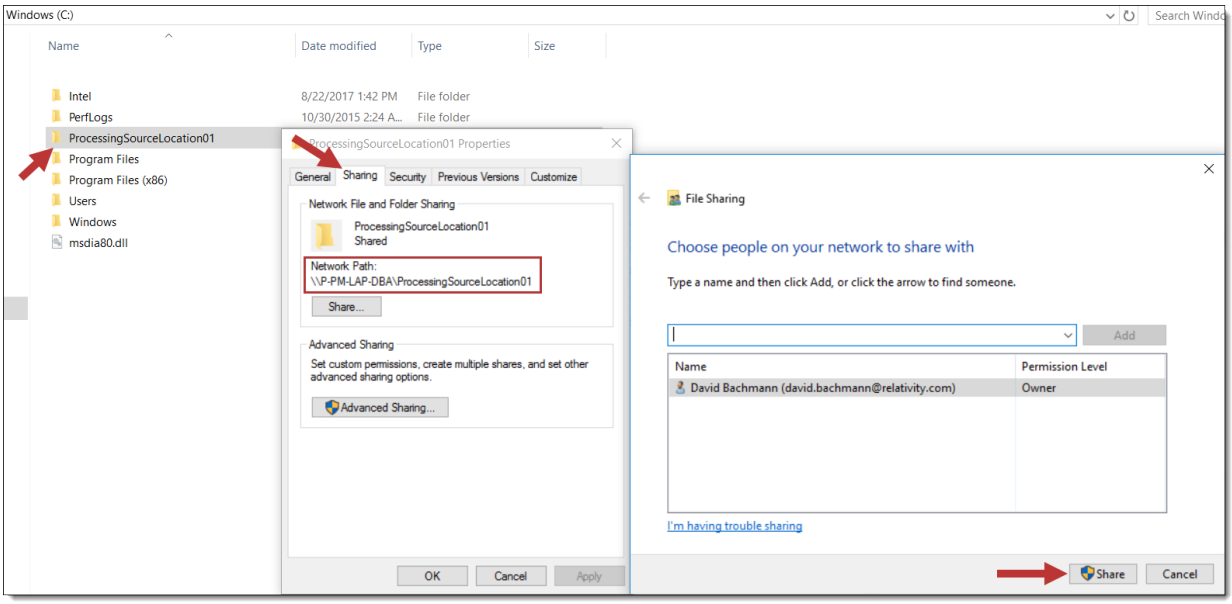
3. Click **Save**.

2.10 Creating a choice for the processing source location

After saving a processing set, you must select a value for the **Select source for files to process** field on the data sources you add to the set. To make a value available for this field, you must create a choice for the **Processing Source Location** field.

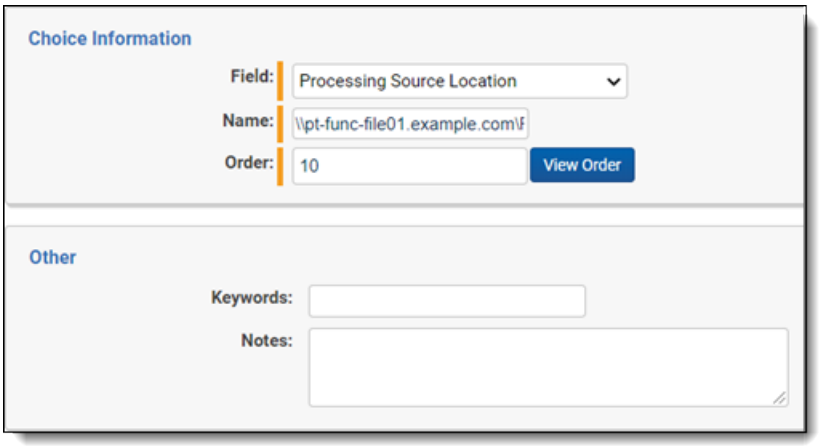
To create a choice for the **Processing Source Location** field:

1. In your File Explorer, locate the folder containing the files that you intend to publish, right-click on it and select **Properties**.
2. In the Properties window, select the **Sharing** tab and then click the **Share** button in the Network File and Folder Sharing section.
3. In the File Sharing window, add the appropriate user and click the **Share** button.
4. Return to the Sharing tab in the Properties window and copy the folder path displayed in the **Network Path** field. When you create the corresponding choice in Relativity, you'll use this path as the name of that choice.



5. Log in to Relativity and navigate to the **Choices** sub-tab.
6. Click **New Choice**.

7. Enter the following values for the following required fields:



   - **Field** - select **Processing Source Location**. The Processing Source Location field is automatically created for you.
   - **Name** - the name of the repository containing the files you want to process. Enter an absolute network path (UNC). For example, *\\pt-func-file01.example.com\FileShare\Custodian\MJones*.

     Note: The Relativity Service Account must have read access to the processing source location.

   - **Order** - the desired order of the choice.

8. Add the source location you just created to the resource pool:
   a. Navigate to the **Resource Pools** sub-tab.
   b. Select the pool to which you want to add the source location.
   c. Click **Add** on the Processing Source Locations tab.
   d. Select the source location choice you created and move it to the right column.
   e. Click **Apply**. The source location is now attached to the resource pool.

2.11 Logging for processing

The logging framework enables you to efficiently gather runtime diagnostic information. You can use logging for troubleshooting application problems when you need a very granular level of detail, for example, when working with a Relativity Support representative.

Relativity system components that can log messages are identified based on the system-subsystem-application designation. When troubleshooting, use the system-subsystem-application matrix to configure logging to target a specific Relativity component, such as the Processing application.

Note: It is recommended that you not set your logging to verbose when publishing documents to a workspace, as doing so can cause your worker to run out of resources, such as CPU, RAM, disk usage, or others, which then causes your publish job to cease entirely. If you need to use verbose logging in order to collect detailed logs, do so for short periods of time only, under 5 minutes, and have a developer on hand to troubleshoot if any issues occur.

For more information, see the Relativity Logging guide.

2.12 Security permissions

The following security permissions are the bare minimum required to publish files to a workspace with Relativity Processing.

Object Security	Tab Visibility
▪ **Processing set** - Add, Edit, View ▪ **Processing Data Source** - Add, View ▪ **Document** - Add, View	▪ Documents ▪ Processing (parent) ◦ Processing Sets (child)

If you want access to view, add, and edit other processing objects, such as profiles, errors, reports, and the password bank, you must configure these options in the Tab Visibility and Object Security windows in the Workspace Security console.

You're finished configuring processing in your Relativity environment. You can now move on to using the processing feature in your workspace through the following components and phases:

- Password bank
- Processing profiles
- Processing sets
- Inventory
- Discovering files
- Publishing files
- Processing error workflow
- Reports

3 Processing to Data Grid

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

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

In order to process data to Data Grid, you must first install all the required components, agents, and applications. For information on how to do this, see the Relativity Data Grid guide .

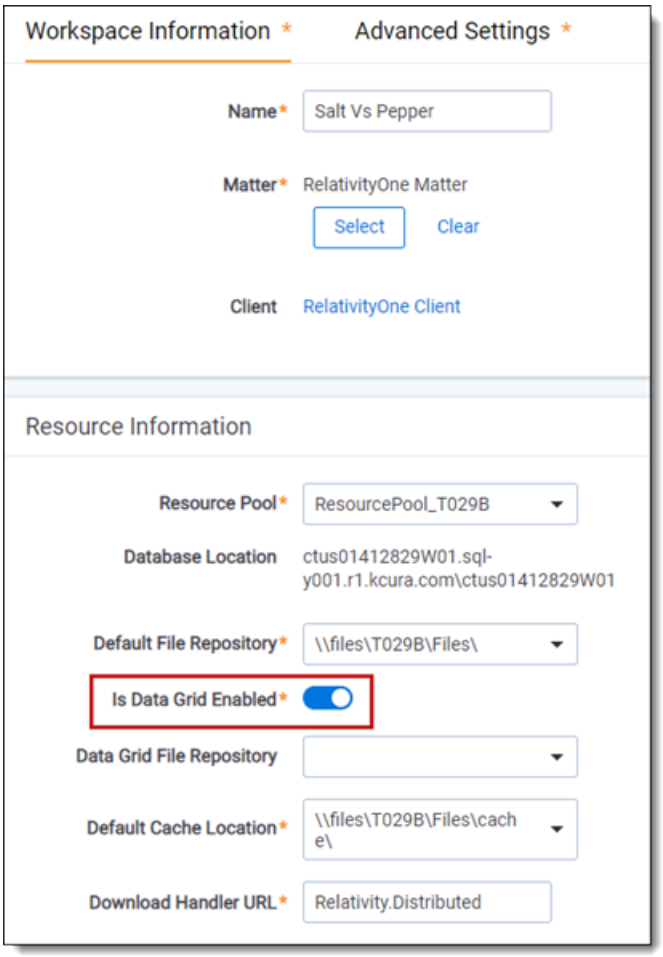
3.1 Enabling processing to Data Grid

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.



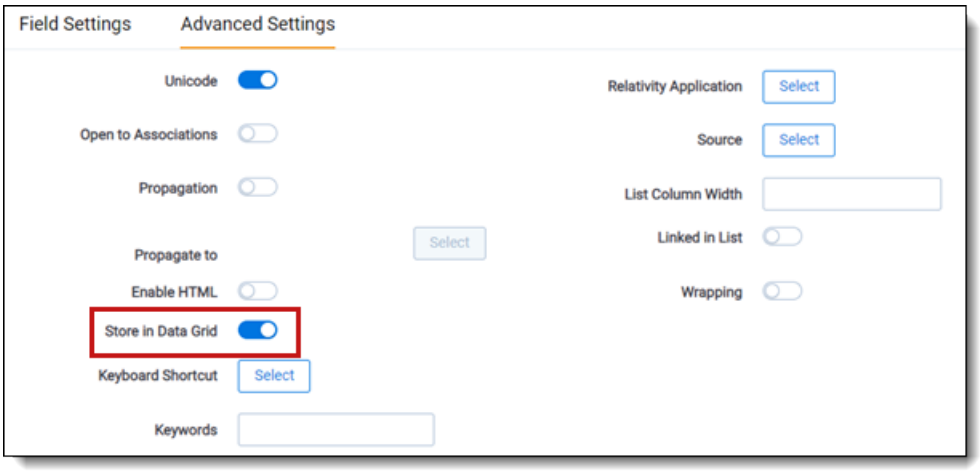
3. (Optional) Next to **Data Grid File Repository**, select the path for the physical location of the text files used by Data Grid. If no file repository is specified for this field, and Data Grid is enabled, Data Grid stores text in the default file repository.

   Note: If you run out of space in this repository, you can specify a new repository. Data Grid will continue to read from the old repository as well as the new repository.

4. Click **Save**.

To enable the extracted text field for Data Grid, perform the following steps:

1. Navigate to the **Fields** tab.
2. Locate the extracted text field and click the **Edit** link next to it.
3. Enable the **Store in Data Grid** field under the Advanced Settings tab.



Note: If you are storing extracted text in Data Grid, the **Include in Text Index** field is set to No because there is no SQL text index. If you want to search using dtSearch, you must follow best practice of creating a saved search of fields you want to index.

4. Click **Save**.

Note: Enabling extracted text fields for Data Grid works for new workspaces only. You can't enable Data Grid for fields that already have text in SQL. If you want to migrate fields from SQL to Data Grid, you must use the Data Grid Text Migration application.

Now that you've enabled processing to Data Grid, you can proceed to running a processing set the way you normally would.

Note: The processing engine and Data Grid do not 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.

4 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 what types are supported and unsupported, as well as any caveats involved with processing those file types.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the and/or the for details on all new functionality.

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.

4.1 Supported file types

The following file types and extensions are supported by Relativity 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 ▪ 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 ▪ For processing and imaging data sets containing CAD files, you can configure the timeout value in the AppSettings table. 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.

File type	Extensions
	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 file considerations.
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 versions	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 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 file considerations.
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 file considerations.

File type	Extensions
	Note: If you save a Powerpoint or Excel document in pre-2007 format, like .PPT or .XLS, 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.
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 not currently supported for PowerPoint 2021+.
	Note: If you save a Powerpoint or Excel document in pre-2007 format, like .PPT or .XLS, 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
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 Relativity Short Message Format metadata and mapping, see RSMF mapping considerations. ▪ For technical details, see the Relativity Short Message Files Guide.
Text files	Such as .txt or .csv.

File type	Extensions
	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.

4.1.1 Excel file considerations

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.

4.1.2 Relativity Short Message Format mapping considerations

Generally, Relativity maps all metadata based on electronic email (.eml) files, but the following RSMF-specific mappings are considered non-standard.

Note: All electronic email (.eml) header strings are case insensitive, which is not unique to RSMF files.

Electronic email (.eml) header	Relativity source field name
X-RSMF-BeginDate	Rsmf/BeginDate EmailSentOn CreatedOn InternalCreatedOn
X-RSMF-EndDate	Rsmf/EndDate LastModified
X-RSMF-EventCount	Rsmf/MessageCount
X-RSMF-Participants	Rsmf/Participants

Electronic email (.eml) header	Relativity source field name
X-RSMF-Version	Rsmf/Version
X-RSMF-Generator	Rsmf/Generator
X-RSMF-Application	Rsmf/Application
X-RSMF-Custodian	Rsmf/Custodian
X-RSMF-AttachmentCount	Rsmf/AttachmentCount
X-RSMF-EventCollectionID	Rsmf/EventCollectionID

If an RSMF file does not include a Sent Date, and the X-RSMF-BeginDate header exists, that header is mapped to the Sent Date field.

For more technical details on how Processing handles Relativity Short Message Format files, see the Relativity Short Message Files Guide.

4.1.3 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 will fail with an unrecoverable error.

This is because inventory looks at files where they reside in the processing source folder and does not copy them to the repository. If only the first file is selected, during discovery, that file only is copied to the repository, and the workers will 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

4.1.4 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.

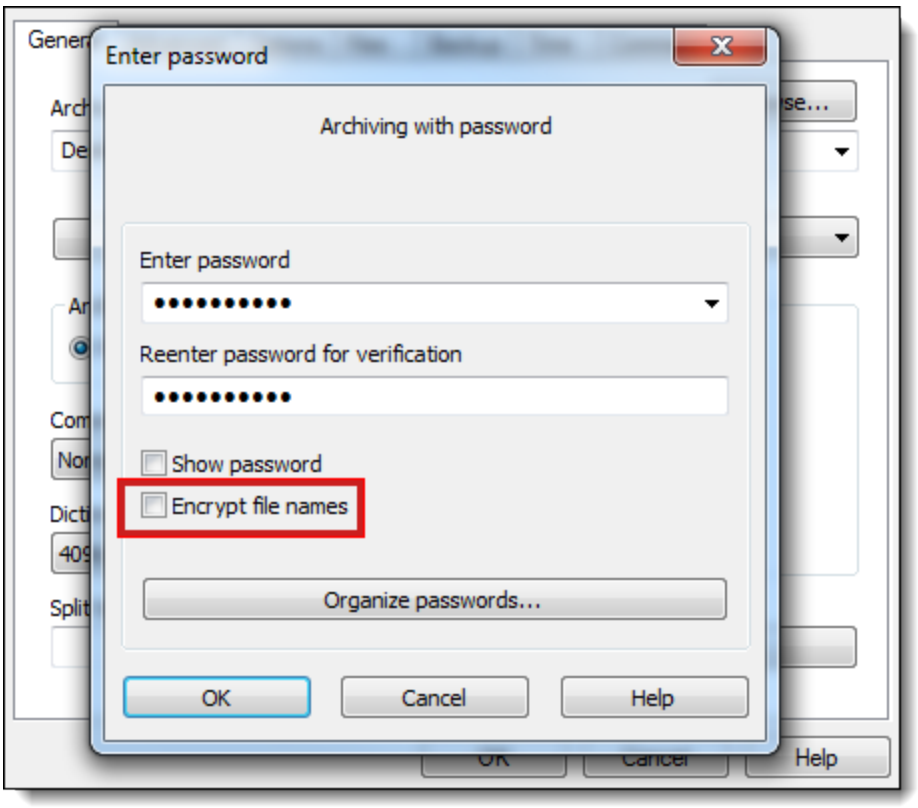
Accordingly, Relativity performs both native text extraction and OCR on the following file formats:

- 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.

4.1.5 Support for password-protected Roshal Archive files

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



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

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

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

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

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

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	Show Time As does not appear in the extracted text if the default

Category	Field/attribute	Outlook message item (.msg)	MIME encapsultation (.mht)
		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.	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.

4.1.7 Email image extraction support

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

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

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

4.1.8 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	√	√
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		

Office program	File extension	Embedded object extraction	Embedded image extraction
Publisher	.pub	✓	
PowerPoint	.pptx	✓	✓
PowerPoint	.pptm	✓	✓
PowerPoint	.ppsx	✓	✓
PowerPoint	.ppsm	✓	✓
PowerPoint	.potx	✓	✓
PowerPoint	.ppt	✓	✓
PowerPoint	.pps	✓	✓
PowerPoint	.pot	✓	✓
OneNote	.one	✓	

4.2 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 ▪ 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

Product category	Product name and version
Windows Word Processors	■ 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 ■ Adobe FrameMaker (.mif) Version 6.0 ■ JustSystems Ichitaro Versions 5.0, 6.0, 8.0, 13.0, 2004 ■ JustWrite Versions through 3.0 ■ Legacy Versions through 1.1 ■ Lotus AMI/AMI Professional Versions through 3.1 ■ Lotus Word Pro Millenium Versions 96 through Edition 9.6, text only ■ Novell Perfect Works Version 2.0 ■ Professional Write Plus Version 1.0 ■ Q&A Write Version 3.0 ■ WordStar Version 1.0
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

Product category	Product name and version
	■ 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

Note: For information on the limitations and exceptions to our supported file types, see Supported file types.

4.3 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.
EnCase	.e01, .l01, .lx01, .ex01
AccessData Logical Image	.ad1 Relativity supports processing both single and multi-part non-encrypted .ad1 files. For encrypted .ad1 files, only single part files are supported. For multi-part .ad1 files, you must decrypt the files prior to processing. See Multi-part container considerations for more information.
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 considerations 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 Off-line 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 Port-	.pdf

File type	Extensions
folio	
RAR	.rar You do not need to combine multi-part .rar files before processing them.
TAR (Tape Archive)	.tar Relativity does not handle multi-part .tar files.
ZIP	See Compressed files.

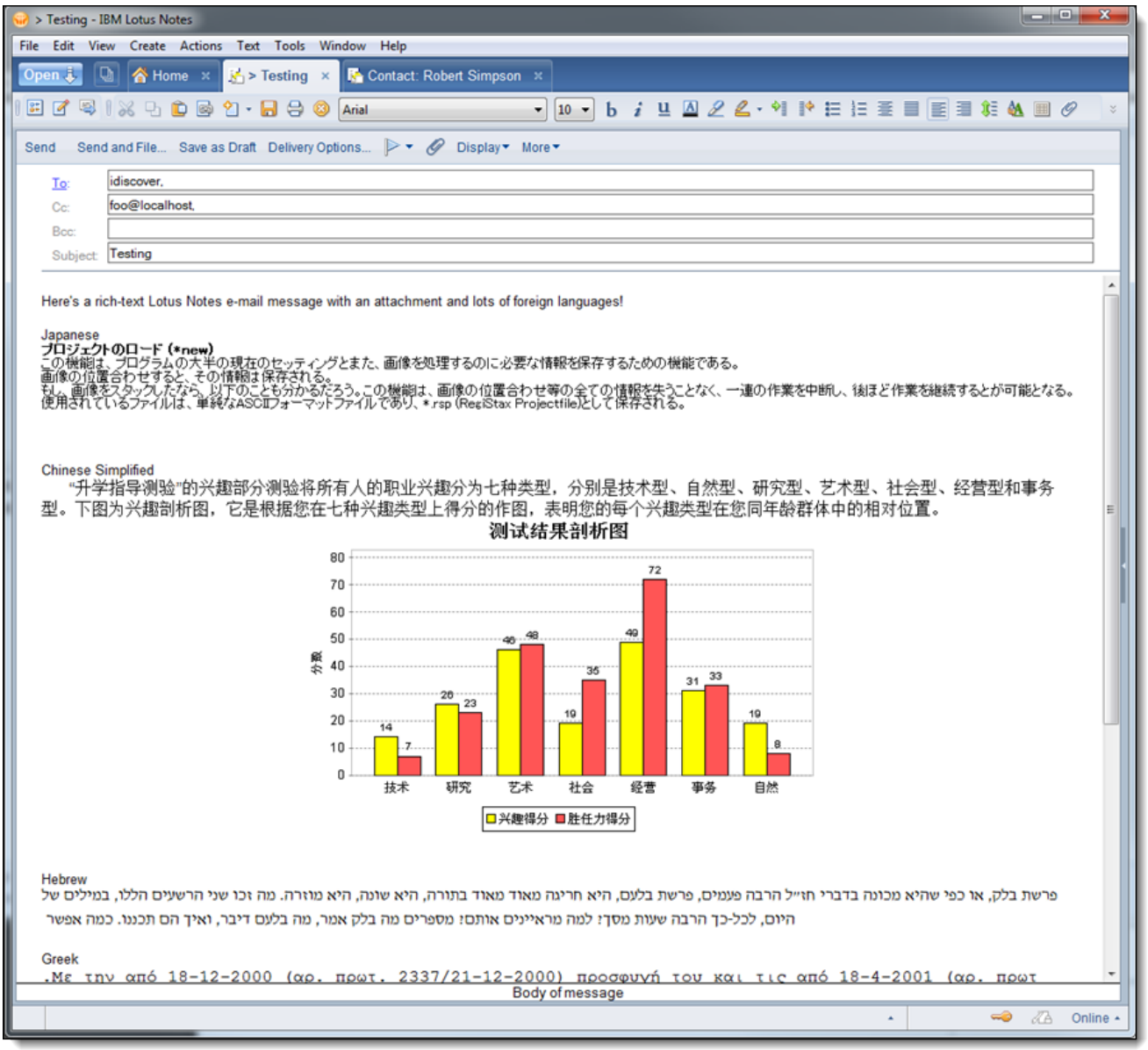
Note: Container files do not store encoding information. Because of this, you may see garbled characters in the file names of children files processed from a container file (such as a .zip file) if the originating locale differs from the processed locale. For example, if the originating container file's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

4.3.1 Lotus Notes considerations

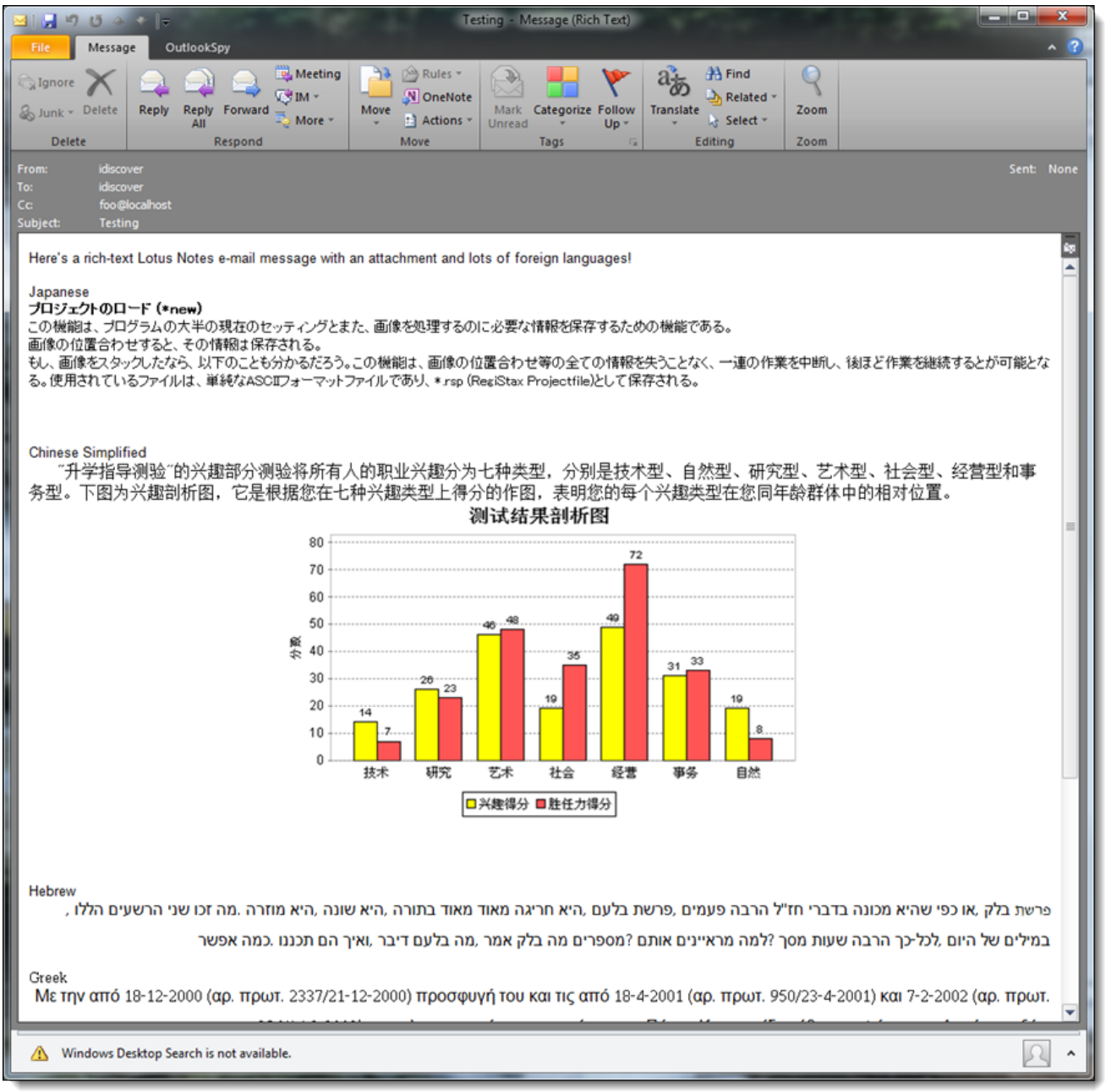
Note the following about how Processing handles note storage facility files:

- 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:



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



4.3.1.1 Lotus Notes supported versions

Lotus Notes are supported through Version 10.

4.3.2 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.

4.3.3 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]
EmailCCSmtp	CC (SMTP Address)
EmailConversation	Conversation
EmailFrom	[other metadata]
EmailImportance	Importance
EmailSenderName	Sender Name

Processing engine property name	Relativity property 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]
Email/EmailEmailAddress	[other metadata]
Email/EmailOriginalDisplayName	[other metadata]
Email/FileUnder	[other metadata]
Email/Generation	[other metadata]
Email/GivenName	[other metadata]

Processing engine property name	Relativity property name
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]

4.4 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

File type	Extensions
PST	.pst
RAR	.rar

4.4.1 Non-container file types supported for Password Bank in Inventory

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

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

* Except .drm files or custom encryption

5 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.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

Using the password bank

Imagine you're a project manager, and you've 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're 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're bringing the most relevant material into your workspace.

5.1 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'd 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, Save as PDF isn't able to connect with the password bank to grab passwords for encrypted documents. That connection is only available for native imaging and processing.

5.2 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.

To view and resolve password-protection errors:

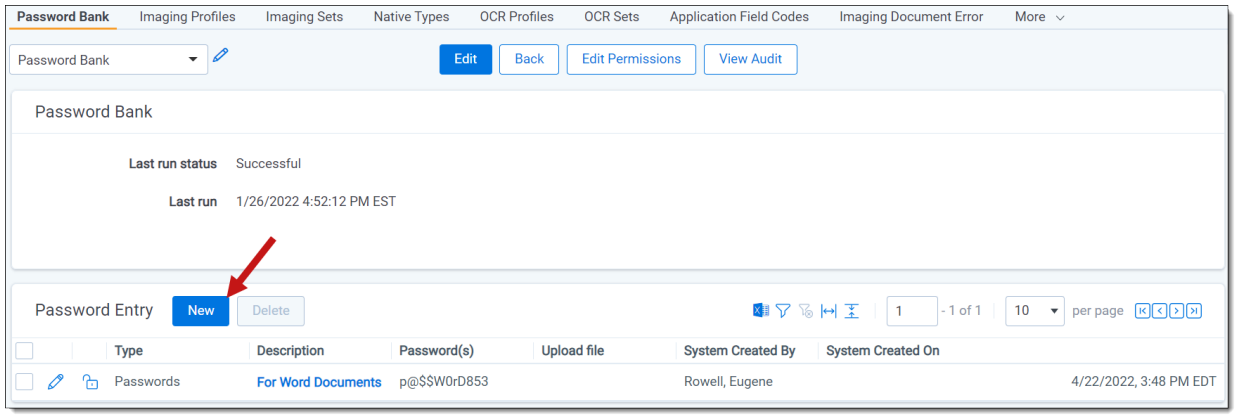
1. Click View Document Errors in the Imaging Set console after you run an imaging set.
2. Outside of Relativity, locate the passwords designated to unlock those files.
3. Return to Relativity, go to the Password Bank, and create entries for every password that corresponds with the errored files.
4. Click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

5.3 Creating or deleting a Password Bank entry

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

To create a new entry in the bank:

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



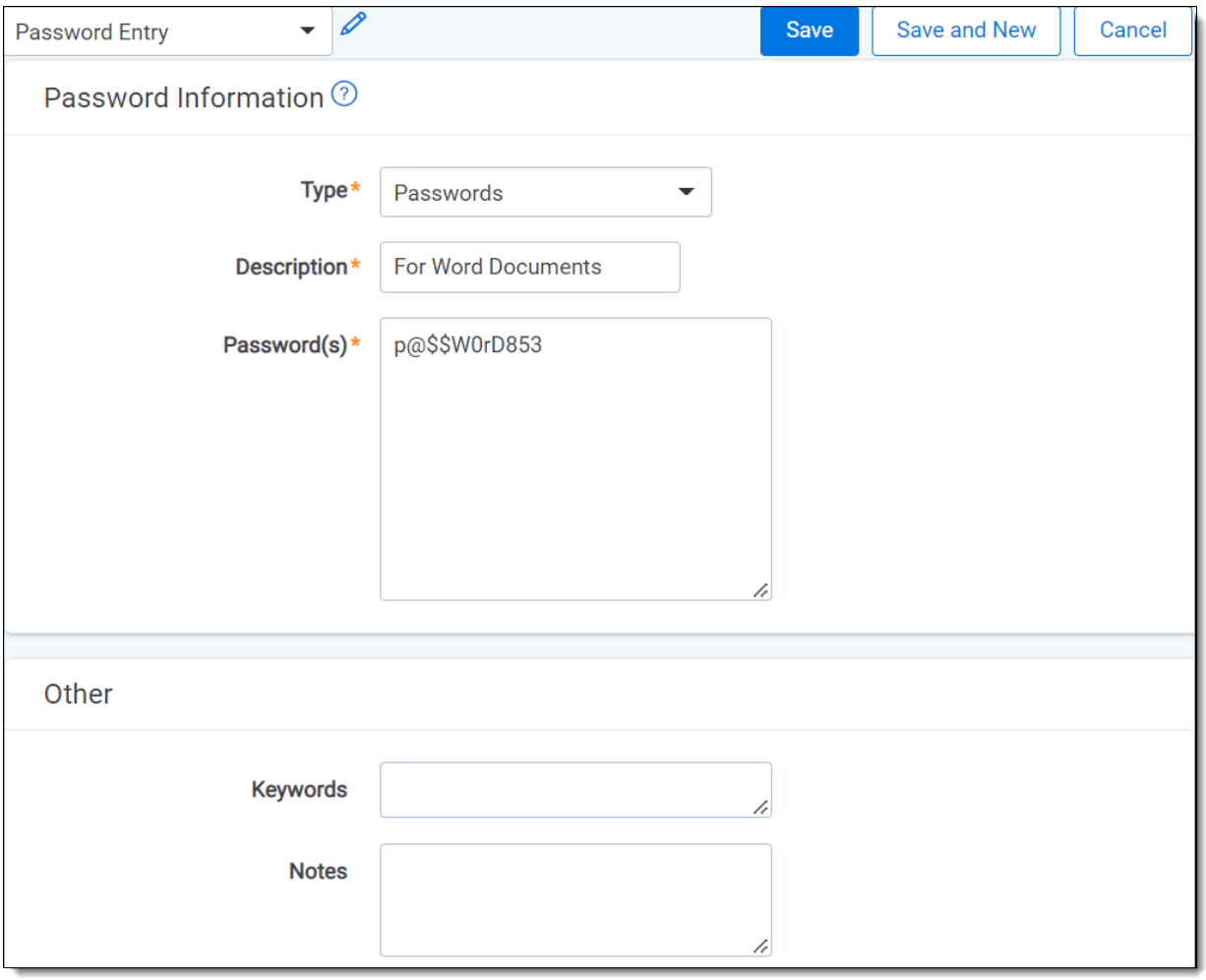
3. Complete the fields on the Password Entry Layout. See Fields below for more information.
4. Click **Save**. The entry appears among the others under the Password Entry object.

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 will continue to try that password.

5.3.1 Fields

The Password Bank layout contains the following fields:



- **Type** - the type of password entry you're 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're able to process EnCase Logical Evidence files, the password bank doesn't support password-protected Encase files.
    - When you select this type, you must enter at least one password in the Passwords field in order to save.
    - The password bank doesn't 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 doesn't display as being 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 processing is 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 doesn't display as being 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 doesn't display as being 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'll receive an error if you attempt 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 in order to save.
  - You can also add values here if you are uploading certificates that don't have passwords. See Example password on the next page.
  - Unicode passwords for zip files aren't supported.
  - Relativity bypasses passwords on .pst and .ost files automatically during file discovery; thus, passwords aren't required for these files to get discovered.

- **Upload file** - the accompanying file you're 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, like .PPT or .XLS, 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.

5.3.2 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.

5.4 Validations, errors, and exceptions

Note the following:

- Including a password that doesn't belong to a document in your data set doesn't 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 can't 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.

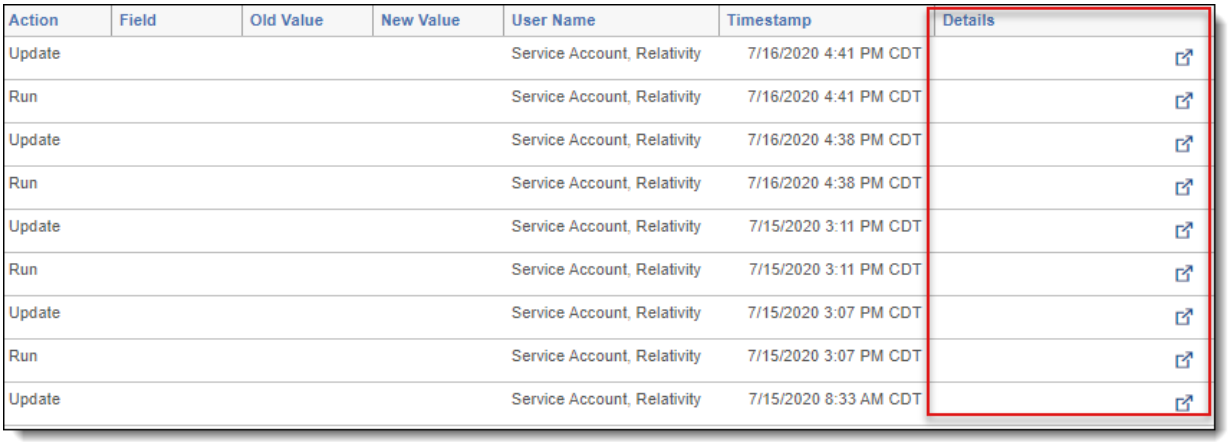
Note: When you supply a valid password to the password bank, the processing engine extracts metadata and extracted text from the document that the password unlocks. However, when you publish that document, its password security isn't removed, in which case it technically remains in an encrypted state even after it's published to the workspace. However, you can view the still-encrypted document in the viewer, because the viewer will recognize that a valid password has been supplied. If the Password Protected field indicates that a document has been 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.

5.5 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, as well as 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.

6 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.

This section provides information on all system-mapped fields in Relativity, as well as the optional metadata fields available to you to map to your data.

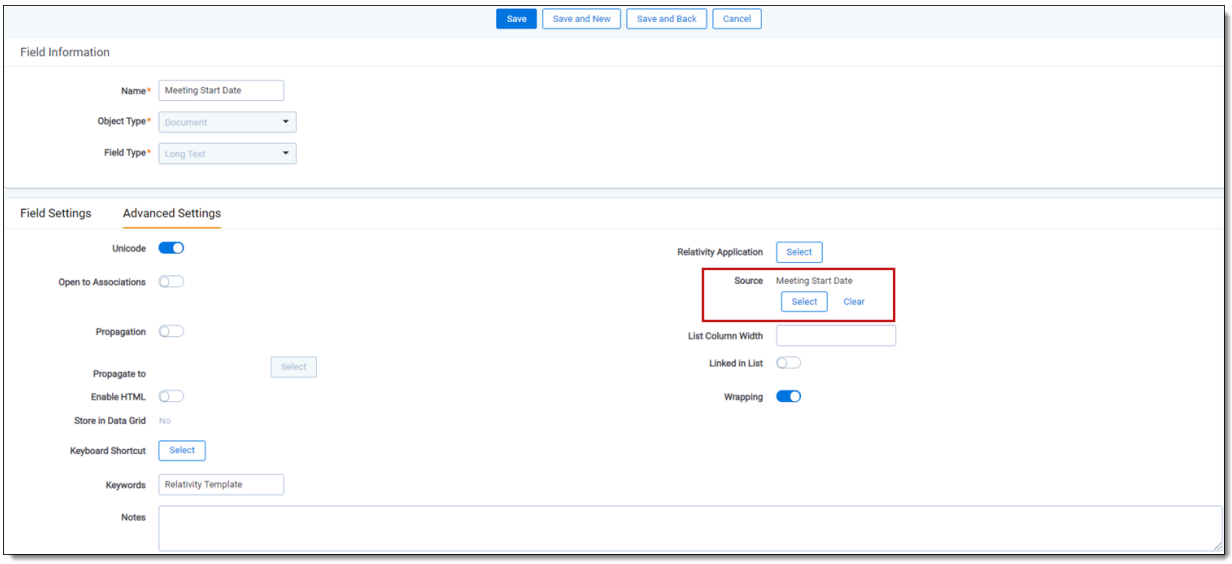
The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

6.1 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 on the Advanced Settings tab.

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



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.

6.1.1 Processing system field considerations

Note the following regarding processing system fields:

- Processing system fields are mapped by default and cannot be modified.
- Processing system fields aren't 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. This means 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.

6.1.2 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."

6.2 System-mapped processing fields

The following system-created metadata fields are always populated when data is processed.

Note: These fields are automatically mapped when you install or upgrade the Processing application from a version earlier than 9.4. They are not available for manually mapping through the Source field on the Field layout:

Processing Field Name	Field Type	Description
Container Extension	Fixed-Length Text	Document extension of the container file in which the document originated.
Container ID	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	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	Long Text	Complete text extracted from content of electronic files or OCR data field. This field holds the hidden comments of MS Office files.
Last Published On	Date	Date on which the document was last updated via re-publish.
Level	Whole Number	Numeric value indicating how deeply nested the document is within the family. The higher the number, the deeper the document is nested.
Originating Processing Set	Single Object	The processing set in which the document was processed.
Originating Processing Data Source	Single Object	A single object field that refers to the processing data source.
Processing File ID	Fixed-Length Text	Unique identifier of the document in the processing engine database.

Processing Field Name	Field Type	Description
Processing Folder Path	Long Text	The folder structure and path to the file from the original location, which is used to generate the Relativity folder browser for your documents. This field is populated every time you process documents. See Processing folder path details on page 81 for more information.
Relativity Attachment ID	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	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	Long Text	Folder structure and path to file from the original location identified during processing. See Virtual path details on page 80 for more information.

6.3 Optional processing fields

The following metadata fields can be, but are not required to be, mapped through the Field Catalog. The Field Catalog contains a list of all available fields to map regardless of discovered data. Select **Field Catalog** from the drop-down menu in the Source field modal if you are setting up Processing prior to Discovery and Publish. The All Fields selection contains a list of all available fields where data has been discovered. Select **All Fields** from the drop-down menu prior to Publish to display available fields with data to publish.

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.

Processing field/ source name	Field type	Description	Example value
All Custodians	Multi Object	All custodians, deduped and master, associated with a file. The All Custodians field is mapped to a document and is updated only when Global or Custodial deduplication 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 Path-s/Locations	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 deduplication 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 Folder-s\Discussion threads
Attachment Document IDs	Long Text	Attachment document IDs of all child items in family group, delimited by semicolon, only present on parent items.	KL0000000031.0001;KL0000000031.0002
Attachment List	Long Text	Attachment file names of all child items in a family group, delimited by semi-colon, only present on parent items.	EC PRC Meeting Agenda.doc;Map to The St.Re-gis.doc
Author	Fixed-Length Text (50)	Original composer of doc-ument or sender of email message. This field has a maximum length of 50 alpha-numeric char-acters.	Jane Doe
BCC	Long Te-	The names, when avail-	Capellas Michael D. [Michael.Capel-

Processing field/ source name	Field type	Description	Example value
	xt	able, and email addresses of the Blind Carbon Copy recipients of an email message.	las@COMPAQ.com]
BCC (SMTP Addres-s)	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 avail-able, and email addresses of the Carbon Copy recipients of an email message.	Capellas Michael D. [Michael.Capel-las@COMPAQ.com]
CC (SMTP Address)	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	Long Text	Attachment MD5 hash value of all child items in a family group, only present on parent items. Relativity cannot calculate this value if you have FIPS (Federal Information Processing Standards cryptography) enabled for the worker manager server.	BA8F37866F59F269AE1D62D962B887B6;5DE747-4 D13679D9388B75C95EE7780FE
Child SHA1 Hash Values	Long Text	Attachment SHA1 hash value of all child items in a family group, only present on parent items.	1989C1E539B5AE98182064862395487 2BEE3E48-3; 58D9E4B4A3068DA6E9BCDD969523288CF38F9F-B3
Child SHA256 Hash Values	Long Text	Attachment SHA256 hash value of all child items in a family group, only present on parent items.	7848EEFC40C40F868929600BF033617642E0D37-C2 F5FA444C7EF83350AE19883;628B6233DD6E0C8-9 F32D6EFF2885F26917F144B19F3678265BEBAC7 E9ACAAF5B
Comments	Long Text	Comments extracted from the metadata of the native file. For more information, see Com-	Oracle 8i ODBC QueryFix Applied

Processing field/ source name	Field type	Description	Example value
		ments considerations.	
Company	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	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	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	Fixed-Length Text (50)	The identifier of the last document in a family group. This field is also populated for documents with no family members. This field has a maximum length of 50 alpha-numeric characters.	KL0000000001.0002
Conversation	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	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	01C9D1FD002240FB633CEC894C1985845049B1886B67

Processing field/ source name	Field type	Description	Example value
		initial email.	
Conversation Index	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.	01C9D1FD002240FB633CEC894C1985845049B1886B67
Created Date	Long Text	The date on which a file was created.	12/24/2015
Created Date/Time	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's the only date created value available. If a document has both a filesystem date created value and a document metadata date created value, this field will display the document metadata date created value.	"12/24/2015 11:59 PM"
Created Time	Long Text	The time at which a file was created.	11:59 PM
DeDuped Count	Whole Number	The number of duplicate files related to a master 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 pop-	2

Processing field/ source name	Field type	Description	Example value
		ulated on the master document. You are not able to retroactively populate this field with custodian information.	
DeDuped Custodians	Multiple Object	The custodians associated with the deduped records of a file. The DeDuped Custodians file is mapped to a document and is present only when Global or Custodial Deduplication is enabled and duplicates are present. This is populated on the master 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	Long Text	The virtual paths of duplicates of a file. This is present only when Global or Custodial Deduplication is enabled and duplicates are present. Each path contains the associated custodian. This is populated on the master document. You are not able to retroactively populate this field with path information.	Lay, Kenneth|\Lay, Kenneth\kenneth_lay_000_1_2_1.pst \lay-k\Kenneth_Lay_Dec2000\Notes Folders\Notes inbox| Doe, John|\Doe, John\John_Doe_000_1_2_1.pst\Doe-J \John_Doe_Dec2000\Notes Folders\Discussion threads
Delivery Receipt Requested	Yes/No	Indicates whether a delivery receipt was requested for an email.	No
Discover Errors on Child Documents	Multiple Object	Identifier of the file that contains the parent document on which the error	

Processing field/ source name	Field type	Description	Example value
		occurred.	
Document Subject	Long Text	Subject of the document extracted from the properties of the native file.	RE: Our trip to Washington
Document Title	Long Text	The title of a non-email document. This is blank if there is no value available.	Manual of Standard Procedures
Email Categories	Long Text	Categories assigned to an email message.	Personal
Email Created Date/Time	Date	The date and time at which an email was created.	"12/24/2015 11:59 PM"
Email Entry ID	Long Text	The unique Identifier of an email in an mail store.	000000005B77B2A7467F56468D820375BC3DC58-2 44002000
Email Folder Path	Long Text	The folder path in which a custodian stored an email. See Email folder path details on page 82 for more information.	Inbox\New Business
Email Format	Single Choice	The indicator of whether an email is HTML, Rich Text, or Plain Text.	HTML
Email Has Attachments	Yes/No	The yes/no indicator of whether an email has children, attachments.	Yes
Email In Reply To ID	Long Text	The internal metadata value within an email for the reply-to ID.	<F9B1A278195DF640A4CC6EC973DFF0C85FBB-EDEB @Prod-EX-MB-01.company.corp>
Email Last Modified Date/Time	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	Single Choice	The indicator of the privacy level of an email.	Company Confidential
Email Sent Flag	Yes/No	The yes/no indicator of whether an email was sent, versus saved as a draft.	Yes

Processing field/ source name	Field type	Description	Example value
Email Store Name	Fixed-Length Text (255)	Any email, contact, appointment, or other data that is extracted from an email container, .pst, .ost, .nsf, .mbox, or any other, will have this field populated with the name of that email container. Any children of those extracted emails, contacts, and appointments will not have anything populated in this field. For more information on this field, see Email Store Name details on page 79. This field has a maximum length of 255 alpha-numeric characters.	kenneth_lay_000_1_1_1_1.pst
Email Unread	Yes/No	The yes/no indicator of whether an email was not read.	Yes
Error Category	Fixed-Length Text (50)	The category assigned by the system to a processing error.> This field was introduced in Server 2022.	Password Protected Container
Error Message	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. This field was introduced in Server 2022.	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	Fixed-Length Text (50)	The phase of processing in which the error occurred—discovery, text extraction, publish, file deletion. This field was introduced in Server 2022.	Discovery

Processing field/ source name	Field type	Description	Example value
Error Status	Fixed-Length Text (200)	The status of the error—undetermined, ready to retry, retried, submitted, unresolvable. This field was introduced in Server 2022.	Ready to retry.
Excel Hidden Columns	Yes/No	The yes/no indicator of whether an Excel file contains one or more hidden columns.	No
Excel Hidden Rows	Yes/No	The yes/no indicator of whether an Excel file contains one or more hidden rows.	Yes
Excel Hidden Worksheets	Yes/No	The yes/no indicator of whether an Excel file contains one or more hidden worksheets.	No
Excel Pivot Tables	Yes/No	The yes/no indicator of whether an Excel file contains pivot tables.	Yes
Family Group (formerly "Group Identifier")	Fixed-Length Text (40)	Group the file belongs to, used to identify the group if attachment fields are not used. This field has a maximum length of 40 alpha-numeric characters.	KL0000000002
File Extension	Fixed-Length Text (25)	The extension of the file, as assigned by the processing engine after it reads the header information from the original file. This may differ from the value for the Original File Extension field. If you publish processing sets without mapping the File Extension processing field, the Text Extraction report does not accurately report document counts by file type. This field has a	MSG

Processing field/ source name	Field type	Description	Example value
		maximum length of 25 alpha-numeric characters.	
File Name	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	Decimal	Generally a decimal number indicating the size in bytes of a file.	15896
File Type	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	Fixed-Length Text (255)	The name, when available, and email address of the sender of an email message. This field has a maximum length of 255 alpha-numeric characters.	Capellas Michael D. [Michael.Capellas@COMPAQ.com]
From (SMTP Address)	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	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.	Yes

Processing field/ source name	Field type	Description	Example value
		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.	
Has OCR Text	Yes/No	The yes/no indicator of whether the extracted text field contains OCR text.	Yes
Image Taken Date/Time	Date	The date and time at which an original image, such as a document scan or .jpg, was taken.	"12/24/2015 11:59 PM"
Importance	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	Yes/No	The yes/no indicator of whether a file is embedded in a Microsoft Office document.	No
Is Parent	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

Processing field/ source name	Field type	Description	Example value
Keywords	Long Text	The internal value entered for keywords associated with a Microsoft Office document.	Enron, Security Agreement
Last Accessed Date	Long Text	The date on which a loose file was last accessed.	12/24/2015
Last Accessed Date/Time	Date	The date and time at which the loose file was last accessed.	"12/24/2015 11:59 PM"
Last Accessed Time	Long Text	The time at which the loose file was last accessed.	11:59 PM
Last Modified Date	Long text	The date on which changes to a file were last saved.	12/24/2015
Last Modified Date/Time	Date	The date and time at which changes to a file were last saved.	"12/24/2015 11:59 PM"
Last Modified Time	Long Text	The time at which changes to a file were last saved.	11:59 PM
Last Printed Date	Long Text	The date on which a file was last printed.	12/24/2015
Last Printed Date/Time	Date	The date and time at which a file was last printed.	"12/24/2015 11:59 PM"
Last Printed Time	Long Text	The time at which a file was last printed.	11:59 PM
Last Saved By	Fixed-Length Text (255)	The internal value indicating the last user to save a document. This field has a maximum length of 255 alpha-numeric characters.	ymendez
Last Saved Date	Long Text	The date on which a file was last saved.	12/24/2015
Last Saved Date/Time	Date	The internal value entered for the date and time at which a document was last saved.	"12/24/2015 11:59 PM"

Processing field/ source name	Field type	Description	Example value
Last Saved Time	Long Text	The time at which a file was last saved.	11:59 PM
Lotus Notes Other Folders	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	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 field has a maximum length of 40 alpha-numeric characters.	21A74B494A1BFC2FE217CC274980E915
MS Office Document Manager	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	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	Single Choice	A standard identifier used on the Internet to indicate the type of data that a file contains.	application/msword
Meeting End Date	Long Text	The date on which a meeting item in Outlook or Lotus Notes ended.	12/24/2015

Processing field/ source name	Field type	Description	Example value
Meeting End Date/Time	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	Long Text	The time at which a meeting item in Outlook or Lotus Notes ended.	11:59 PM
Meeting Start Date	Long Text	The date on which a meeting item in Outlook or Lotus Notes started.	12/24/2015
Meeting Start Date/Time	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	Long Text	The time at which a meeting item in Outlook or Lotus Notes started.	11:59 PM
Message Class	Single Choice	The type of item from an email client—email, contact, calendar, and others.	IPM.Note
Message Header	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	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 83. This field has a maximum length of 255 alpha-numeric characters.	<PLSRGLMRNQWEDFYPJL5ZJFF41USDEIQHB @zlsvr22>

Processing field/ source name	Field type	Description	Example value
Message Type	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	Long text	The path to a copy of a file for loading into Relativity.	\\files2.T026.c-tus014128.r1.company.com\T026\Files\EDDS2544753\Processing\1218799\INV2544753\SOURCE\0\982.MSG
Number of Attachments	Whole Number	Number of files attached to a parent document.	2
Original Author Name	Fixed-Length Text (50)	The display name of the original author of an email. This field has a maximum length of 50 alpha-numeric characters.	Jane Doe
Original Email Author	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	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	Long Text	Metadata extracted during processing for additional fields beyond the list of processing fields available for	Excel/HasHiddenColumns=True;Office/Application=Microsoft Excel;InternalCreatedOn=7/25/1997 9:14:12 PM;Office/Security=2;Office/PROPID_23=528490;O-

Processing field/ source name	Field type	Description	Example value
		mapping. This includes TrackChanges, HiddenText, HasOCR, and dates of calendar items. Field names and their corresponding values are delimited by a semicolon.	ffice/ Scale=0;Office/ LinksDirty=0;Office/PROPID_19=0;Office /PROPID_22=0; Office/Parts=sum,ENRON; Office/Headings= Worksheets,2;Office/_PID_GUID=Unknown PROPVARIANT type 65; Excel/HasHiddenRows=True; LiteralFileExtension=XLS
Outlook Flag Status	Single Choice	The indicator of which flag, if any, an Outlook item has assigned to it—NoFlag, FlagMarked, or FlagComplete.	NoFlag
Password Protected	Single Choice	Indicates the documents that were password protected. It contains the value Decrypted if the password was identified, Encrypted if the password was not identified, or no value if the file was not password protected.	Encrypted
PowerPoint Hidden Slides	Yes/No	The yes/no indicator of whether a PowerPoint file contains hidden slides.	Yes
Primary Date/Time	Date	Date taken from Sent Date, Received Date, or Last Modified Date in the order of precedence.	"12/24/2015 11:59 PM"
Processing Errors	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	Yes/No	Indicates whether a read receipt was requested for an email.	Yes
Received Date	Long Text	The date on which an email message was received.	12/24/2015

Processing field/ source name	Field type	Description	Example value
Received Date/Time	Date	The date and time at which an email message was received.	"12/24/2015 11:59 PM"
Received Time	Long Text	The time at which an email message was received.	11:59 PM
Recipient Count	Whole Number	The total count of recipients in an email which includes the To, CC, and BCC fields.	1
Recipient Domains (BCC)	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 multi-object fields, see Relativity Objects. The Domains processing fields listed in this table eliminate the need to perform domain parsing using transform sets for the processed documents.	enron.com;bellatlantic.com
Recipient Domains (CC)	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 multi-object fields, see Relativity Objects. The Domains processing fields listed in this table eliminate the need to perform domain parsing using transform sets for the processed documents.	enron.com;bellatlantic.com
Recipient Domains (To)	Multiple Object	The domains of the 'To' recipients of an email. For information on domains and steps to create the Domains	enron.com;bellatlantic.com

Processing field/ source name	Field type	Description	Example value
		object and associative multi-object fields, see Relativity Objects.	
		The Domains processing fields listed in this table eliminate the need to perform domain parsing using transform sets for the processed documents.	
Recipient Name (To)	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
RSMF Begin Date	Date	The timestamp (ISO8601) of the earliest short message event within the file.	11:59 PM
RSMF End Date	Date	The timestamp (ISO8601) of the latest short message event within the file.	11:59 PM
RSMF Event Count	Whole Number	The number of short message events captured within the file.	5
SHA1 Hash	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.	D4C4EB398980E82B4B3064CC2005F04D04BBAA-E6
SHA256 Hash	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	4F8CA841731A4A6F78B919806335C963EE039F3-3 214A041F0B403F3D156938BC

Processing field/ source name	Field type	Description	Example value
		70 alpha-numeric characters.	
Sender Domain	Multiple Object	The domain of the sender of an email.	enron.com
Sender Name	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	Long text	The date on which an email message was sent.	12/24/2015
Sent Date/Time	Date	The date and time at which an email message was sent.	"12/24/2015 11:59 PM"
Sent Time	Long Text	The time at which an email message was sent.	11:59 PM
Sort Date/Time	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.	"12/24/2015 11:59 PM"

Processing field/ source name	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	Long Text	The folder structure and path to the file from the original location identified during processing. For emails, this displays the subject rather than the email's entry ID. This provides you with better context of the origin of the email. Previously, the Virtual Path field displayed the entry ID with the email file name, and if you followed this virtual path, it was difficult to tell by that entry ID where the email came from. See Source path details on page 82 for more information.	Reports\User**Sample.pst\Inbox\ Requested February report**
Speaker Notes	Yes/No	The yes/no indicator of whether a PowerPoint file has speaker notes associated with its slides.	Yes
Subject	Long Text	The subject of the email message.	Blackmore Report - August

Processing field/ source name	Field type	Description	Example value
Suspect File Extension	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	Single Choice	The method used to run text extraction.	Excel
Title	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	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)	Long Text	The full SMTP value for the recipient of an email message, for example, “bob@example.com.”	Michael.Capellas@COMPAQ.com
TrackChanges	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 looking only at	Yes

Processing field/ source name	Field type	Description	Example value
		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	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: ■ The Track Changes button is enabled in the	

Processing field/ source name	Field type	Description	Example value
		document. ■ There is actual Tracked Change content in the document.	
Unprocessable	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 imaging/OCRing that document type. For example, it's 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 asso-	No

Processing field/ source name	Field type	Description	Example value
		ciated with them because they never reach a point at which they can register a processing error.	

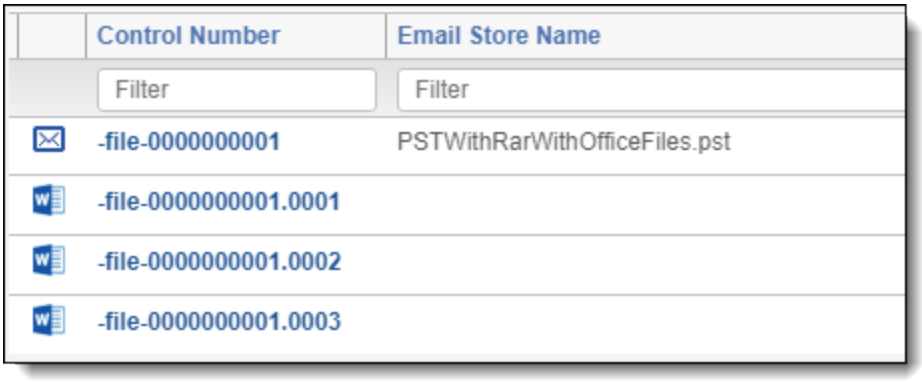
Note: Extracted Text Size in KB is also an available mappable field outside of the Field Catalog. This field was introduced in Relativity 9.4, and it indicates the size of the extracted text field in kilobytes. To map this field, you can edit the corresponding Relativity field, open the Field Catalog via the Source field, select the All Fields view, and select **Extracted Text Size in KB** as the Source value.

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.

6.4 Email Store Name details

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

- When an email comes from .pst, 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.



  - 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
-file-0000000001.0001		Kolhe, Aneeta - 1048232	Word.rar
-file-0000000001.0002		Kolhe, Aneeta - 1048232	Word.rar
-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 named PSTContainingEmbeddedPSTInFolders.pst.

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

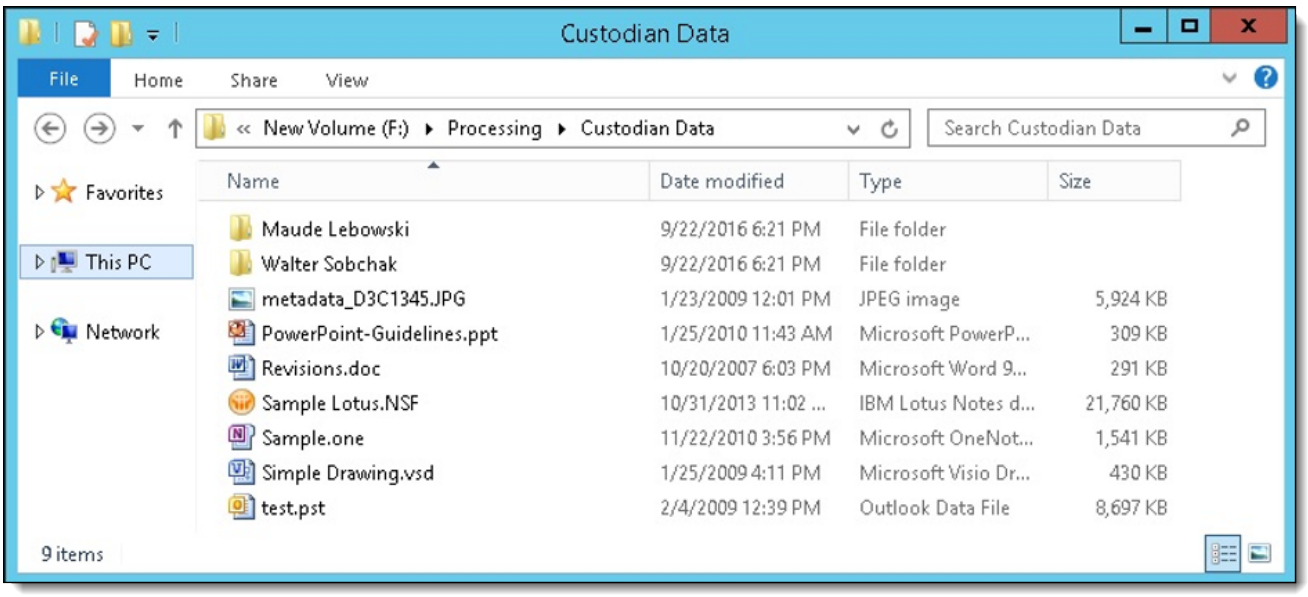
- For an email taken from a zip folder, the Email Store Name field is blank.

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

6.5 Virtual path details

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.



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.

6.6 Processing folder path details

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.

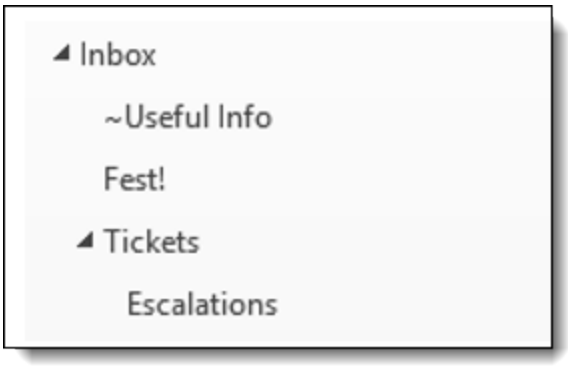
The following are examples of virtual paths and corresponding processing folder paths.

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

6.7 Email folder path details

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."



6.8 Source path details

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.

6.9 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, and more.
- 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 deduplication. For the purposes of deduplication, 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.

6.10 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 does not automatically mean that the document has inline comments.

6.11 Deduped custodian and path considerations

If you run deduplication 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 multi-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 deduplicated 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
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.

7 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.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

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.

7.1 Creating or editing a processing profile

To create or edit a processing profile:

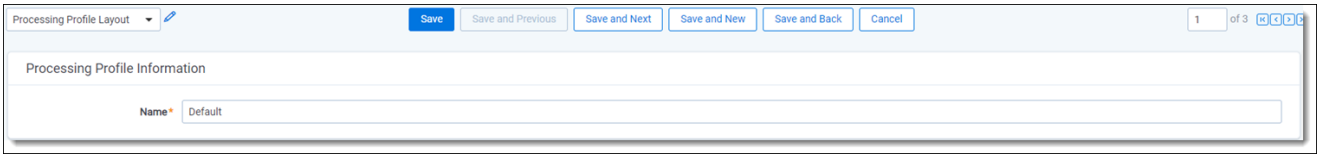
1. Go to the **Processing Profile** tab.
2. Click **New Processing Profile** or select any profile in the list.
3. Complete or modify the fields on the Processing Profile layout. See Processing profiles above.
4. Click **Save**. Once you save the processing profile, you can associate it with a processing set. For more information, see Processing sets on page 132.

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.

7.1.1 Fields

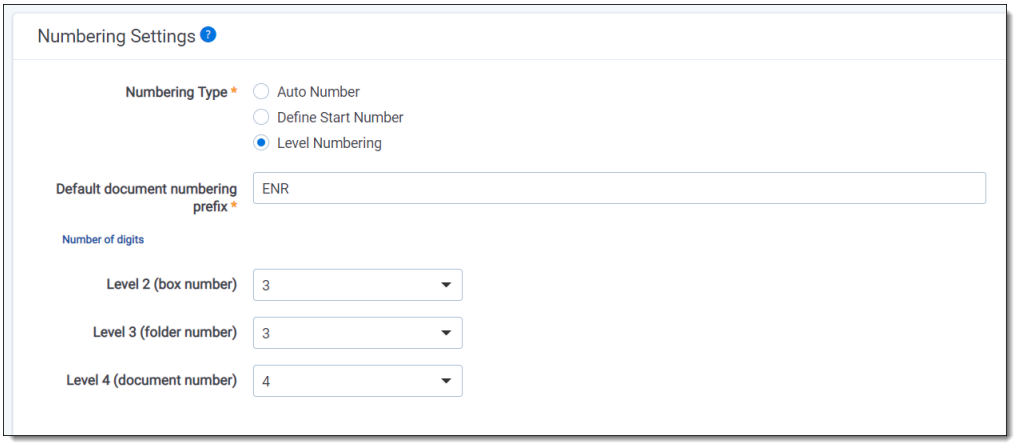
Note: Relativity doesn't re-extract text for a re-discovered file unless an extraction error occurred. This means that if you discover the same file twice and you change any settings on the profile, or select a different profile, between the two discovery jobs, Relativity will not re-extract the text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

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



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

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



- **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 105.
    - **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 on page 107 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.

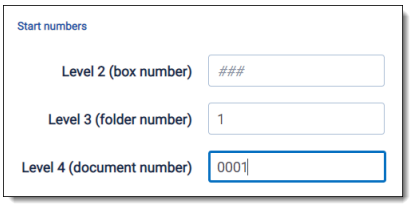
Level 2 (box number) - corresponds to the BBB level in the PPP.BBB.FFF.NNNN format. Default value is 3 digits.

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.



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

7.1.1.1 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
```

7.1.1.2 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 2 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 2 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.

7.1.1.3 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

7.1.1.4 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.

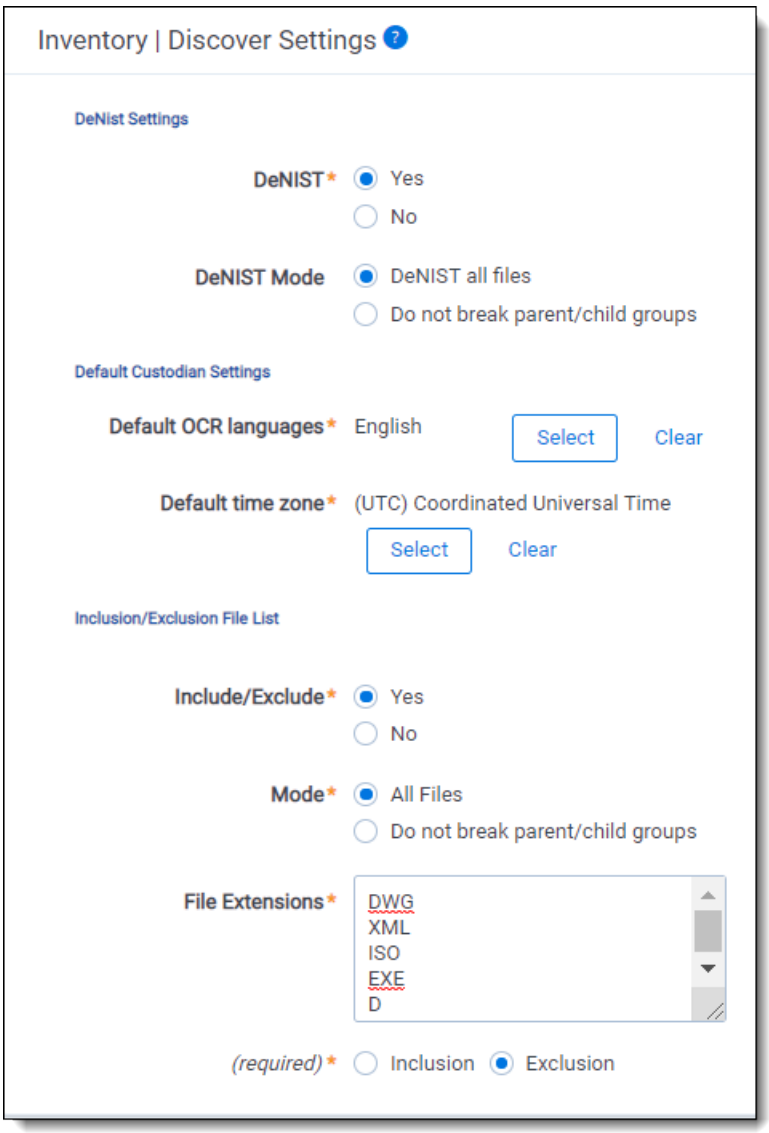
Initial Publish:

```
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
```

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



- **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's 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 data source on page 137.
- **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 data source on page 137.

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
  ```

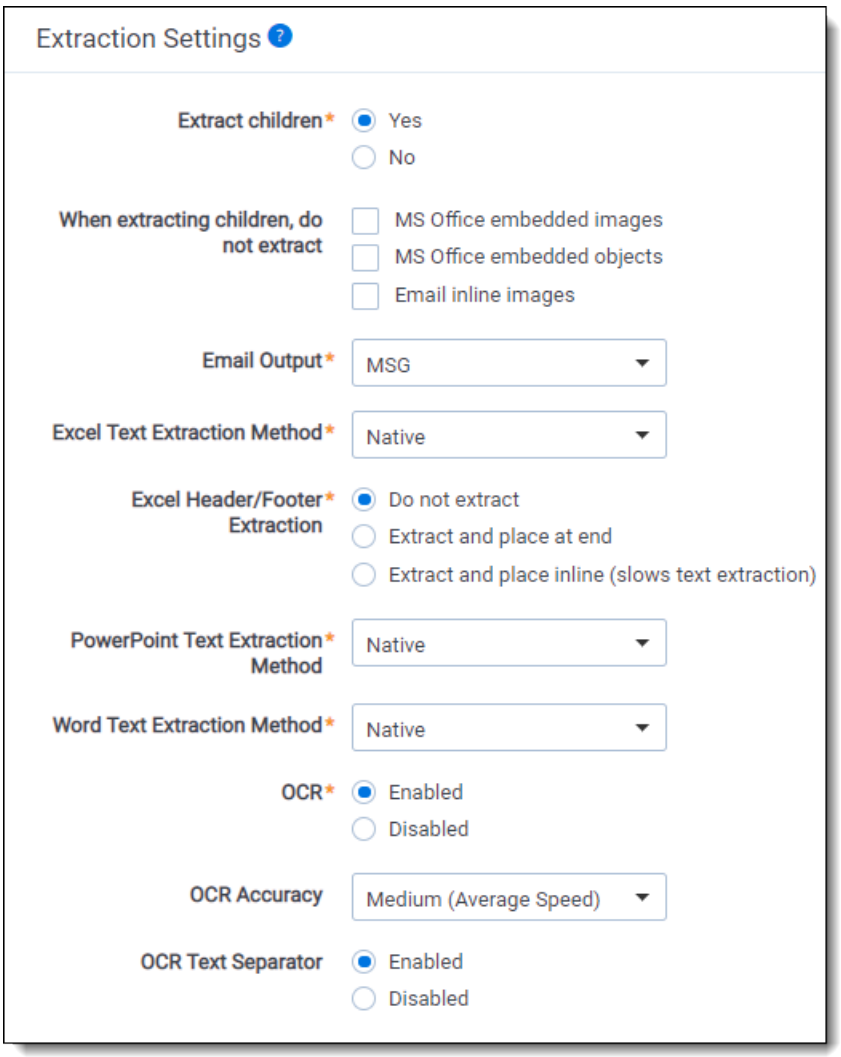
  will create a list of DWG, XML, ISO, EXE, D to exclude from Discovery.

  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.

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.



- **Extract children** - arranges for the removal of child items during discovery, including attachments, embedded objects and images and other non-parent files. The options are:
  - **Yes** - extracts all children files during discovery so that both children and parents are included in the processing job.
  - **No** - does not extract children, so that only parents are included in the processing job.

  Note: You don't 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** - exclude one or all of the following file types when extracting children. You can't make a selection here if you set the Extract children field to No.
  - **MS Office embedded images** - excludes images of various file types found inside Microsoft Office files—such as .jpg, .bmp, or .png in a Word file—from discovery so that embedded images aren't published separately in Relativity.

- **MS Office embedded objects** - excludes objects of various file types found inside Microsoft Office files—such as an Excel spreadsheet inside a Word file—from discovery so that the embedded objects aren't published separately in Relativity. MS Office embedded objects will not have text extracted and will not be searchable.

  Note: Relativity currently doesn't 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 on page 33.

- **Email inline images** - excludes images of various files types found inside emails—such as .jpg, .bmp, or .png in an email—from discovery so that inline images aren't published separately in Relativity.

  Note: For a detailed list of the kinds of attachments that Relativity treats as inline, or embedded, images during processing, see Email image extraction support on page 33.

- **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 on page 108.
- **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 on page 108.

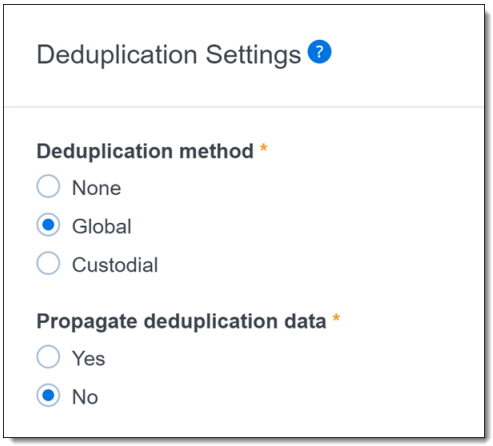
- **Word Text Extraction Method** - determines whether the processing engine uses Word, Relativity, or dtSearch to extract text from Word files during publish.
  - **Relativity** -tells Relativity to use Relativity's built-in engine to extract text from Word files. This setting is recommended for performance 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 on page 108.
- **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 includes 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.

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



- **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 on page 112:

  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.

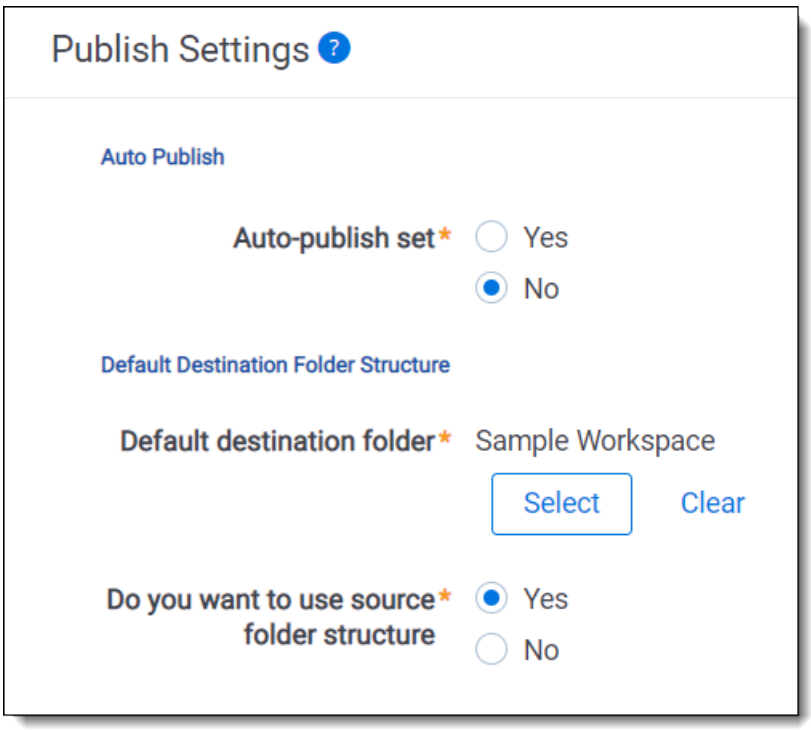
> **Note:** If you select Global, there should be no exact e-mail duplicates in the workspace after you publish. The only exception is a scenario in which two different e-mail systems are involved, and the e-mails are different enough that the processing engine can't exactly match them. In the rare case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

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



- **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.

7.1.1.5 Parent/child numbering type examples

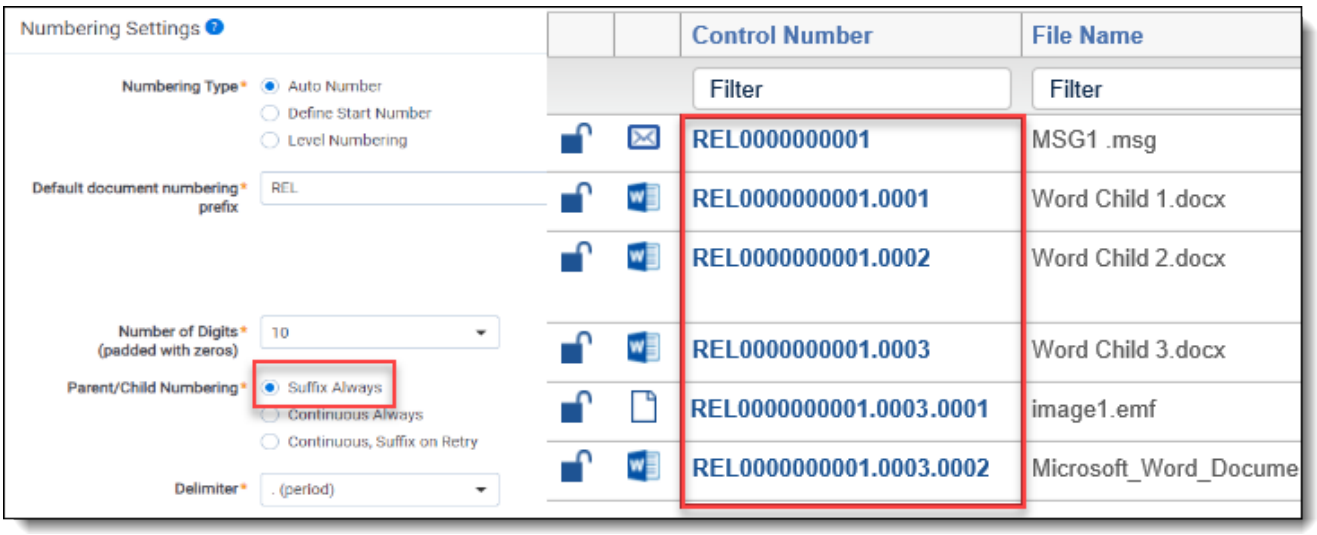
To better understand how each parent/child numbering option appears for published documents, consider the following scenario.

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

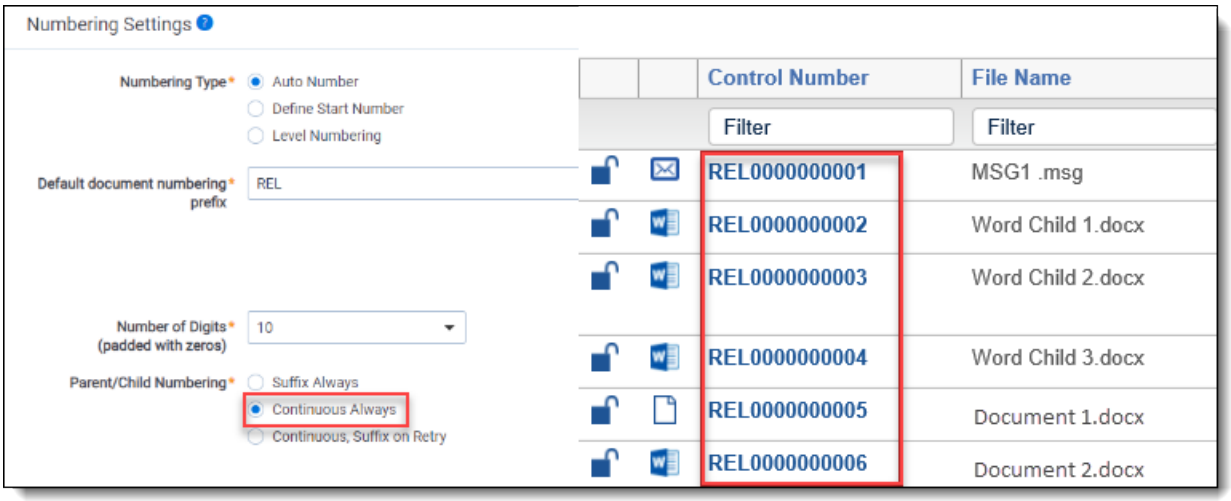
- MSG
  - Word Child 1
  - Word Child 2
  - Word Child 3 (password protected)
    - sub child 1
    - sub child 2

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:

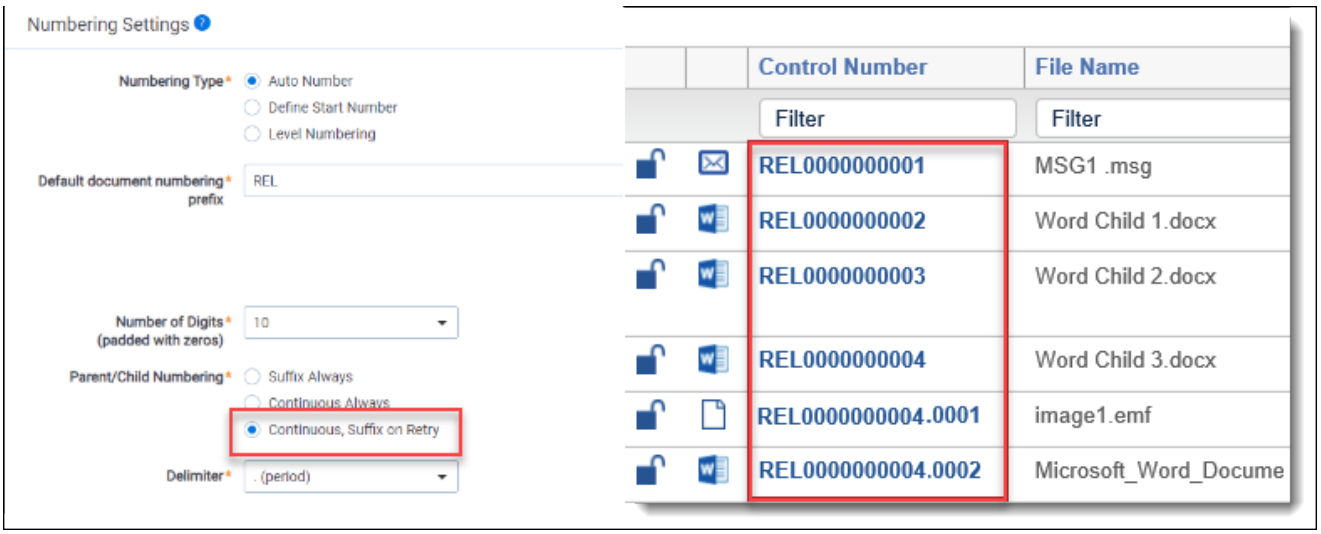


If you'd chosen Continuous Always for the Parent/Child Numbering field on the profile, the identifiers of the published documents would appear as the following:



- 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:



- 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.

7.1.1.6 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

7.1.1.7 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 master (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 master and non-error child documents remain without suffixes, based on the original discovery.

7.1.2 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, cleared 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.

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

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

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

8 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.

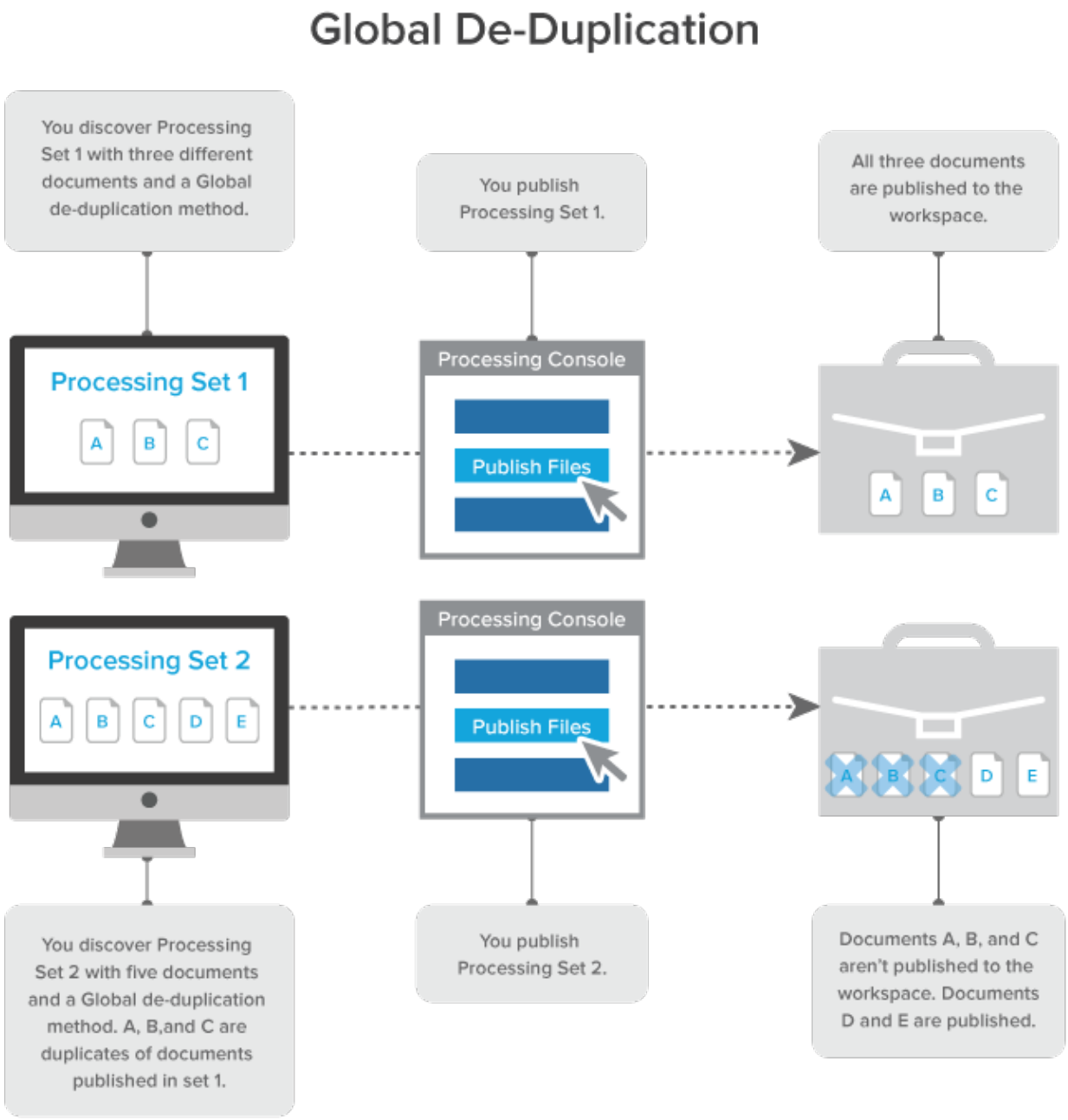
Note the following special considerations regarding deduplication:

- 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.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

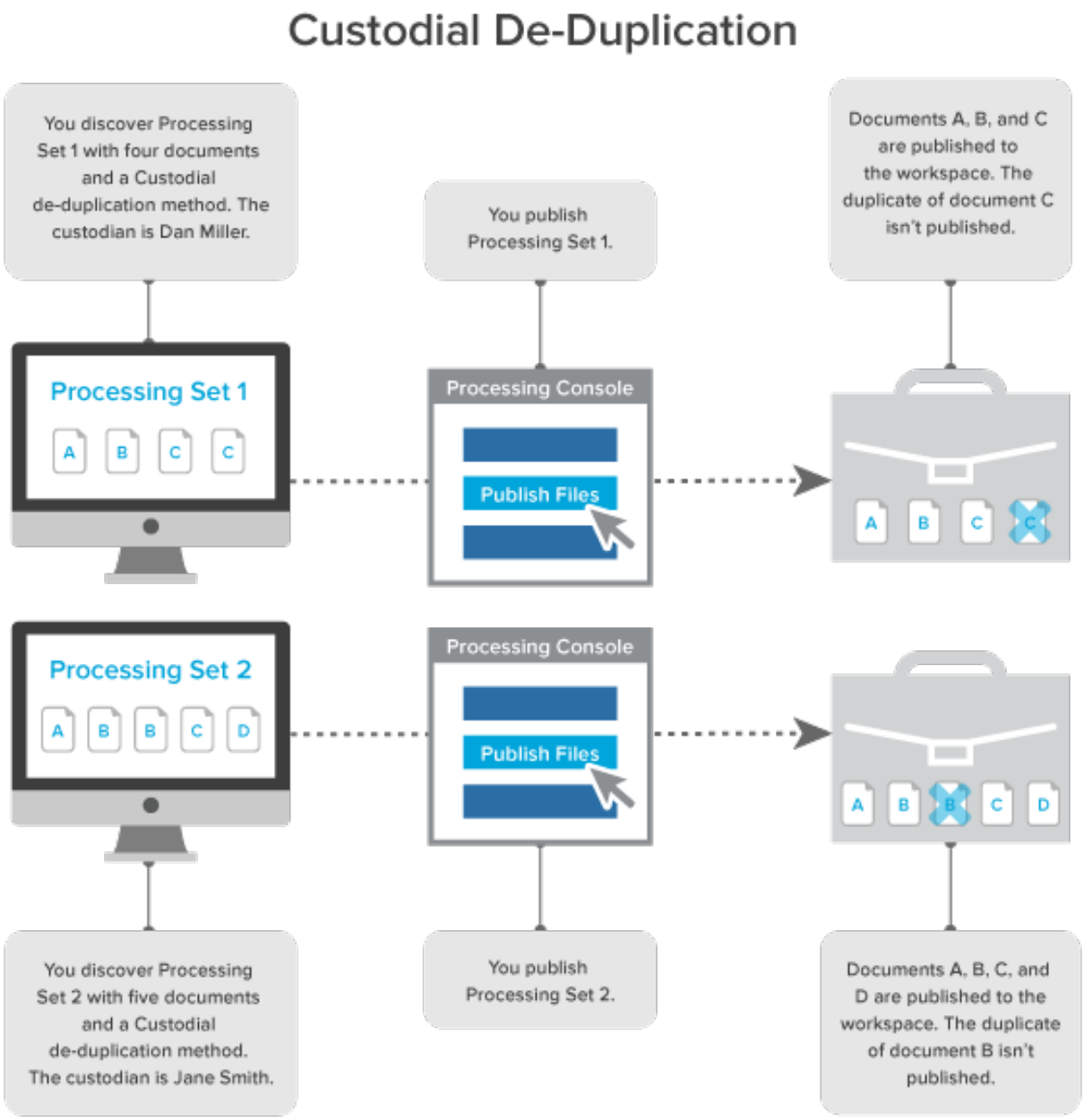
8.1 Global deduplication

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.



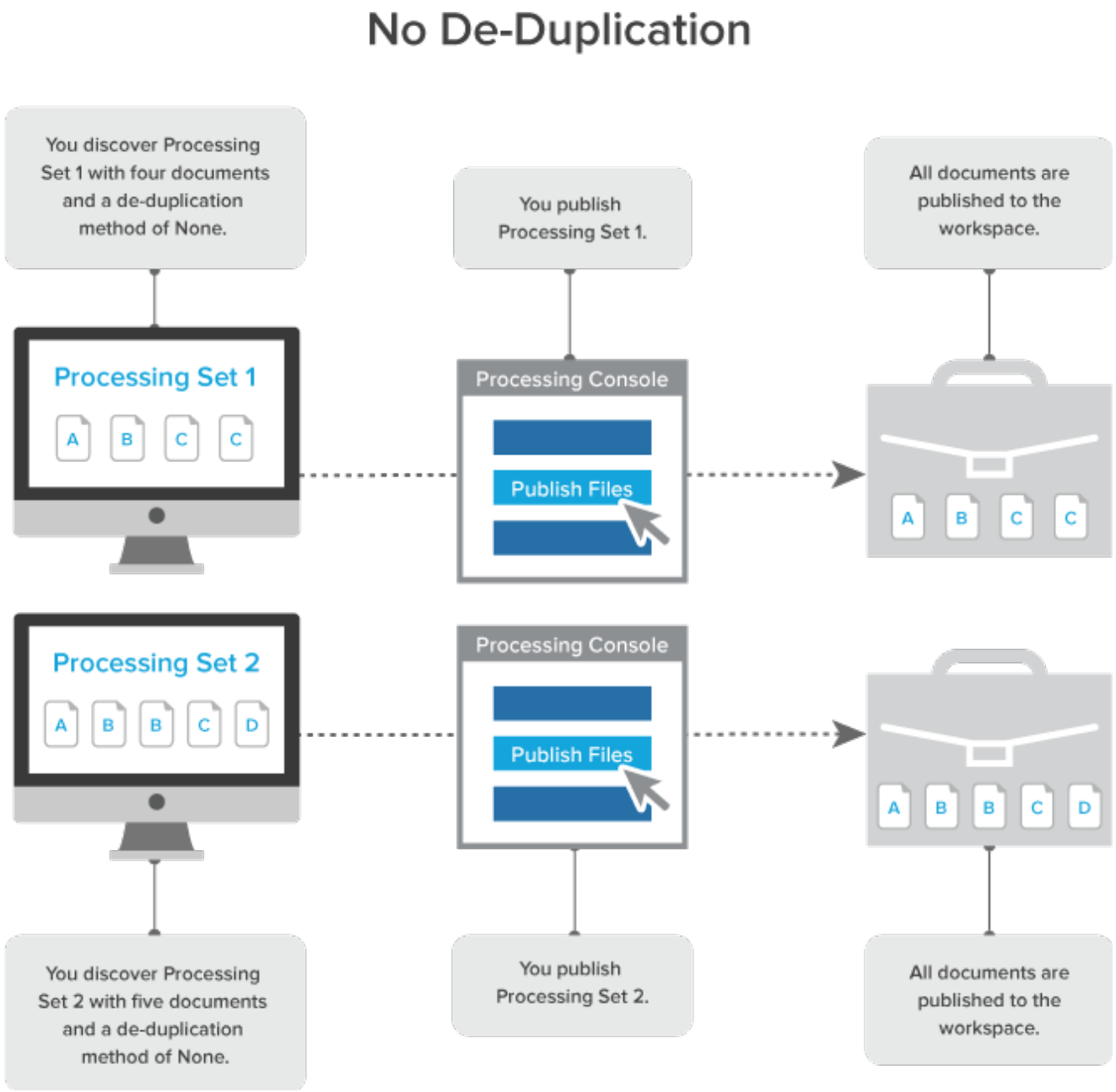
8.2 Custodial deduplication

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.



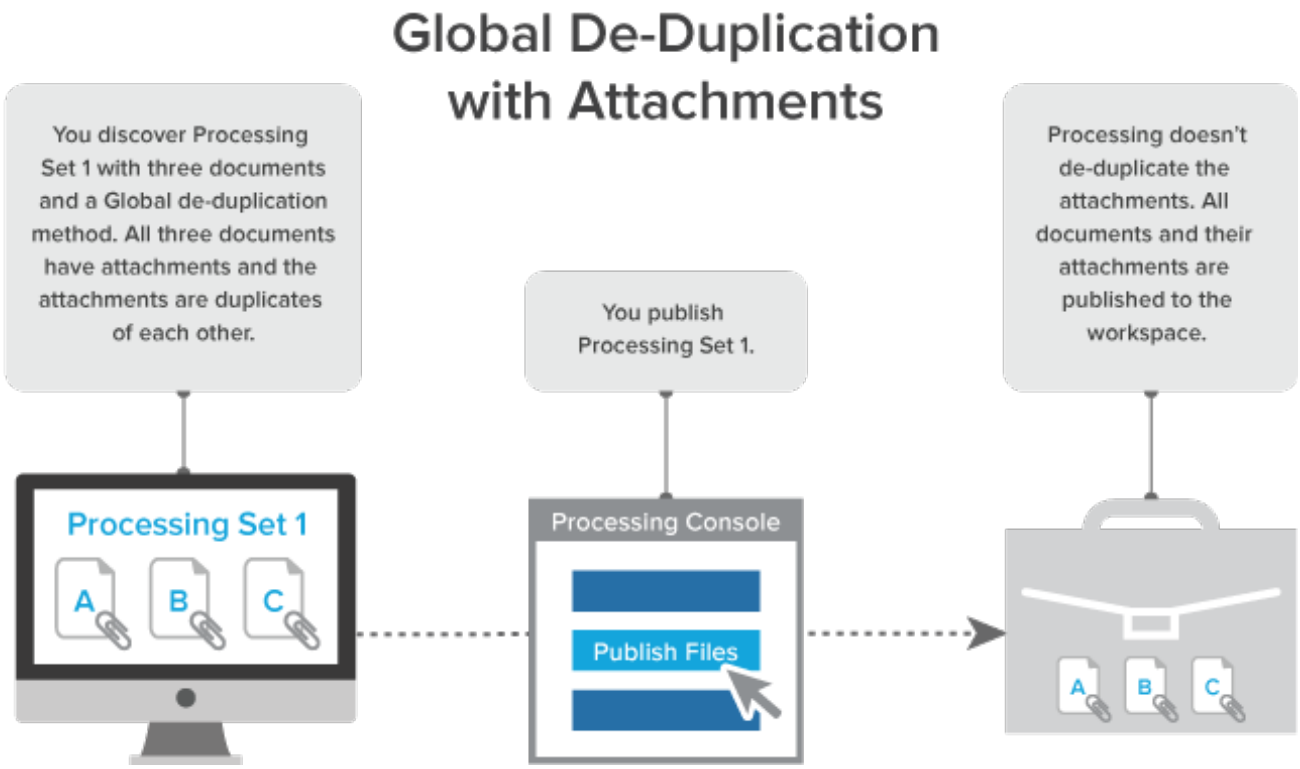
8.3 No deduplication

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



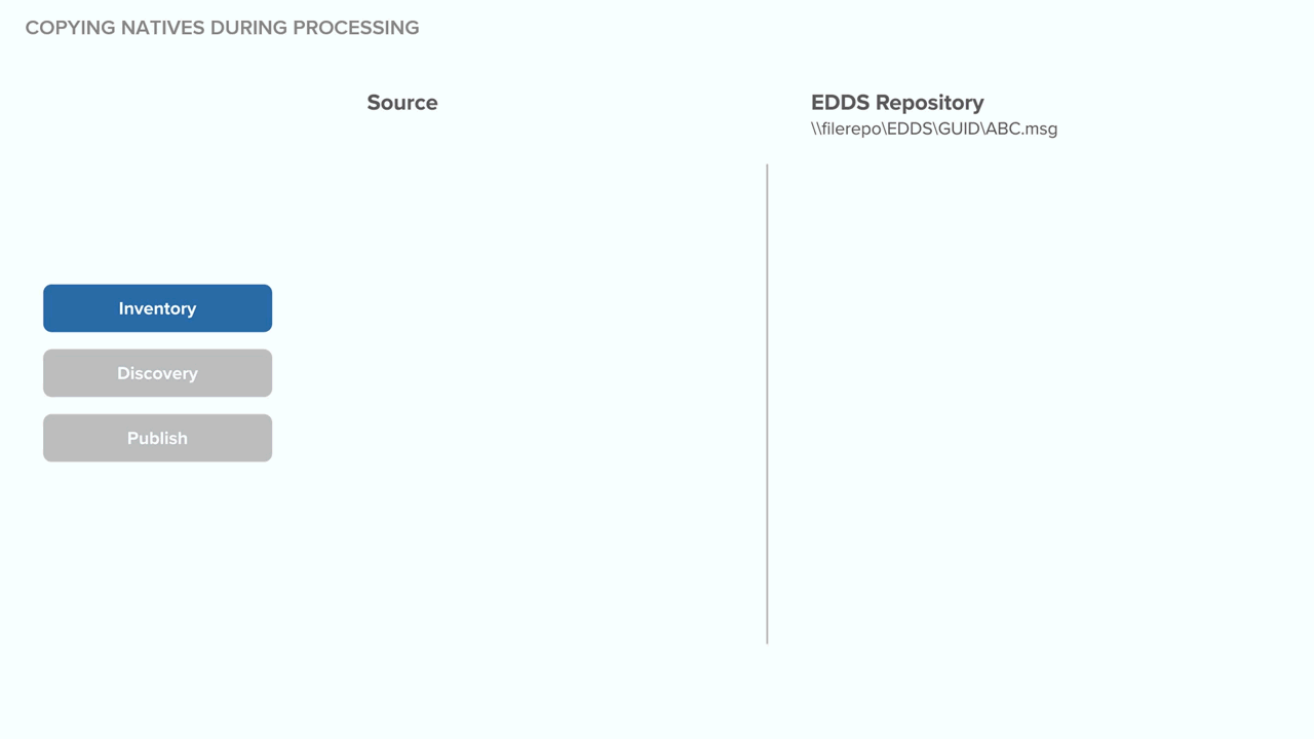
8.4 Global deduplication with attachments

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.



8.5 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.



8.6 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.

8.6.1 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.

8.6.2 Calculating deduplication hashes for emails

8.6.2.1 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.

8.6.2.2 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

8.6.2.3 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

8.6.2.4 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 on the previous page to generate all four de-dupe hashes. Then, these hashes are combined, as described in Calculating the Relativity deduplication hash below, 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

Beginning in Relativity 9.5.342.116, 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. For a detailed list of all the metadata values populated for ICS and VCF files, see the Processing user guide.

8.6.3 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 on the previous page.
   -
     ```
     `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.

9 Quick-create set(s)

With the quick-create set(s) feature, you can streamline the creation of data sources, custodians, and processing sets based on a specific file structure or PST from a single object. For example, if you need to create 73 data sources with a single custodian per data source, quick-create set(s) eliminates the need for you to do so on an individual basis.

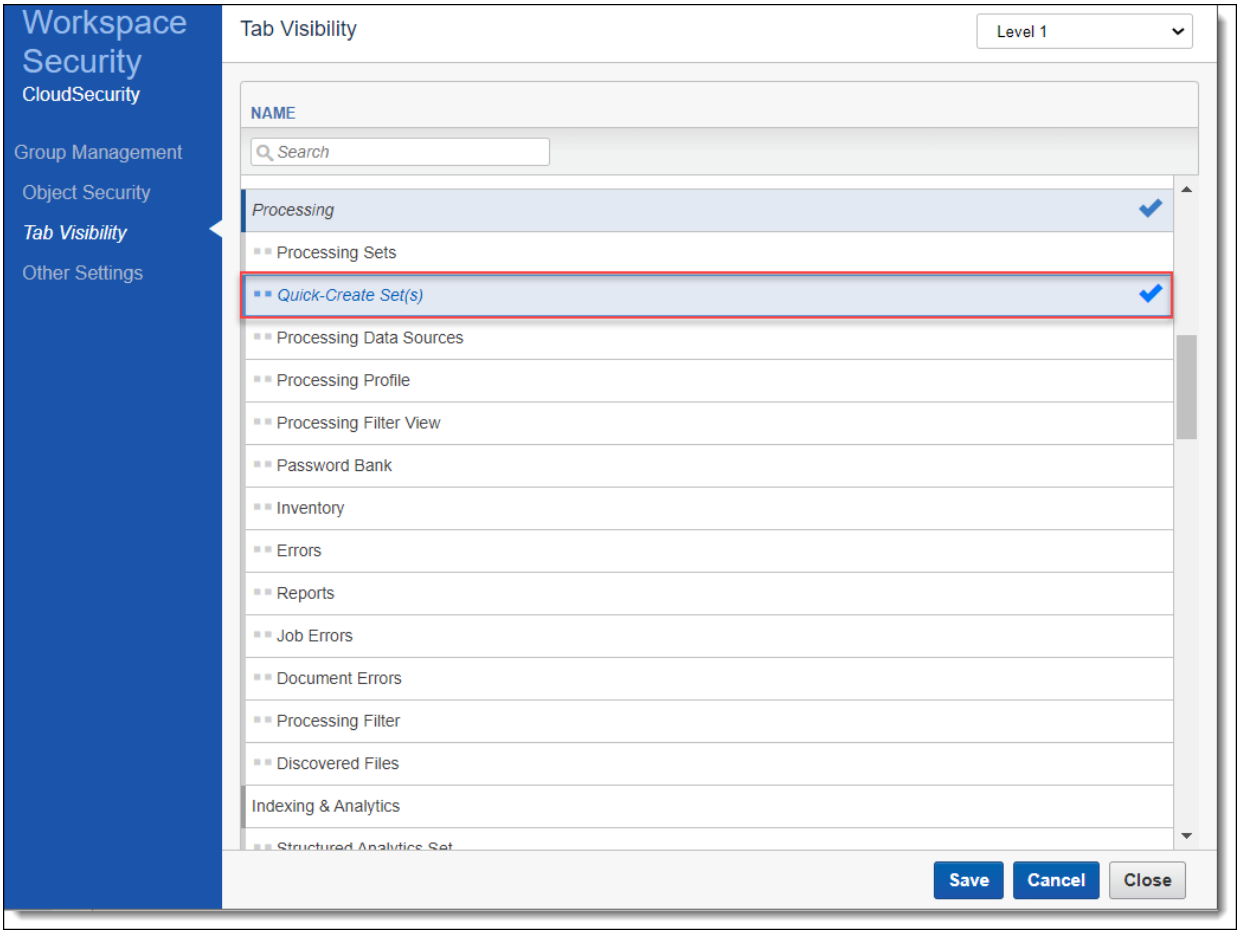
You can also elect to create a single or multiple processing sets based on those custodians and data sources and specify what action you want Relativity to take, such as starting inventory or discovery, for each processing set created.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

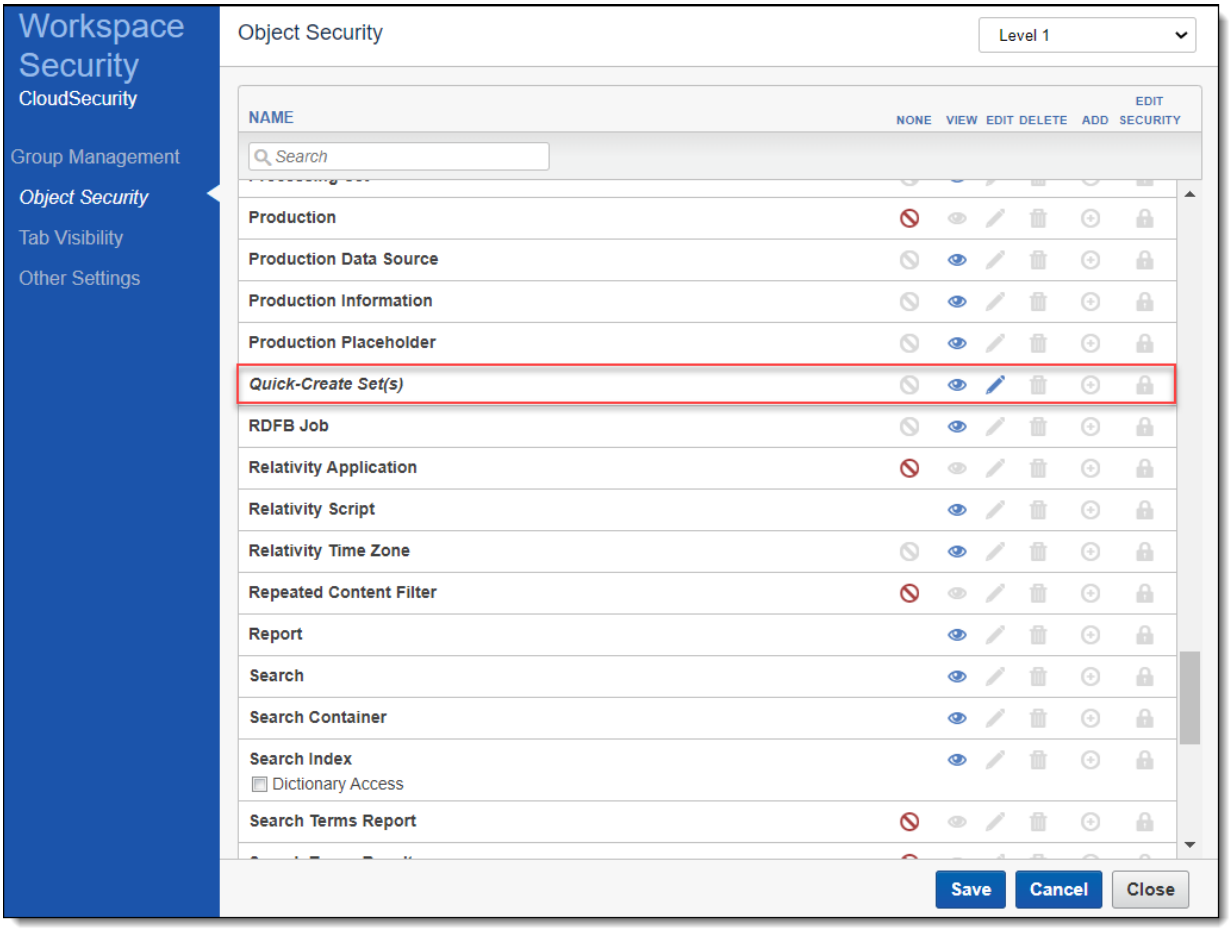
9.1 Required security permissions

In addition to the permissions required to use Processing, you need the following permissions enabled in the Workspace Security console in order to use quick-create set(s).

- The **Quick-Create Set(s)** entry checked in the Tab Visibility section.



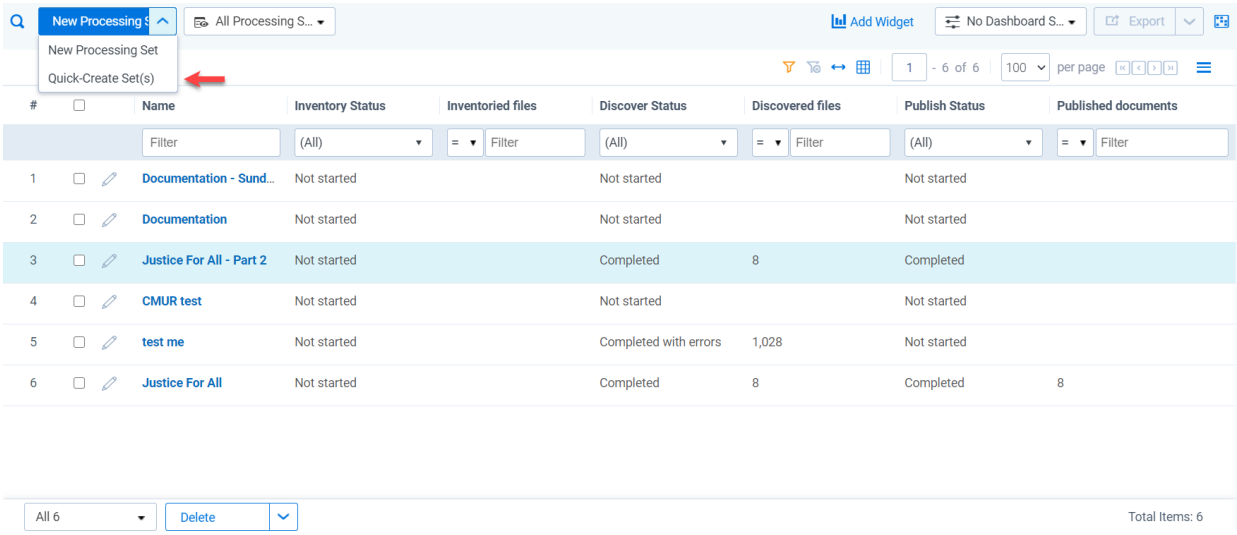
- The Edit permission on the corresponding **Quick-Create Set(s)** entry in the Object Security section.



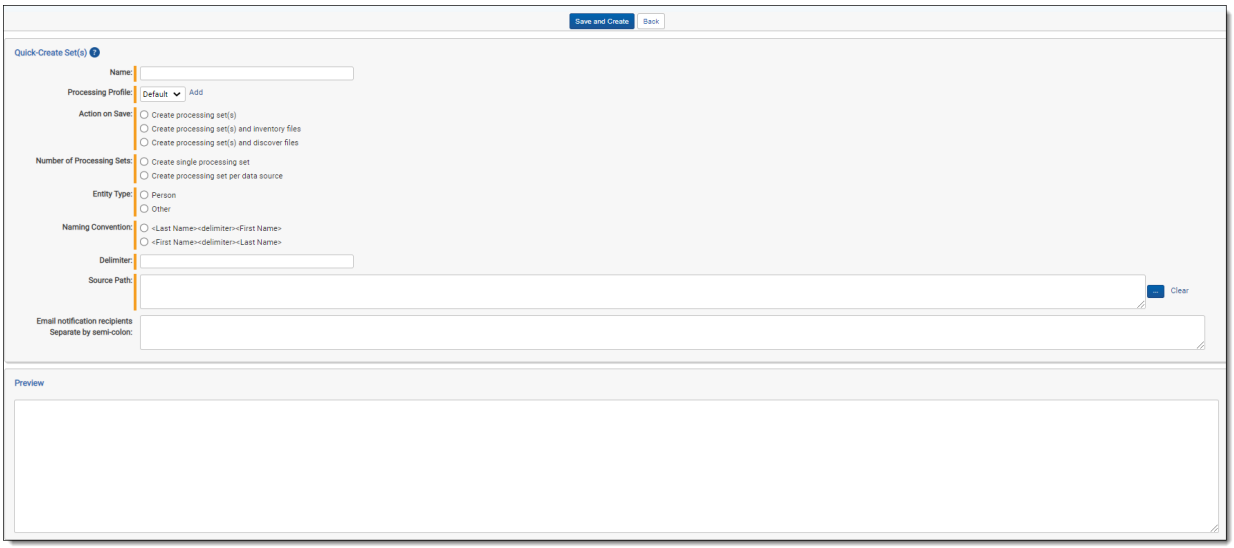
9.2 Using quick-create set(s)

To use quick-create set(s), perform the following steps:

1. Navigate to the **Processing Sets** sub-tab, click the **New Processing Set** drop-down menu, and select the **Quick-Create Set(s)** option.

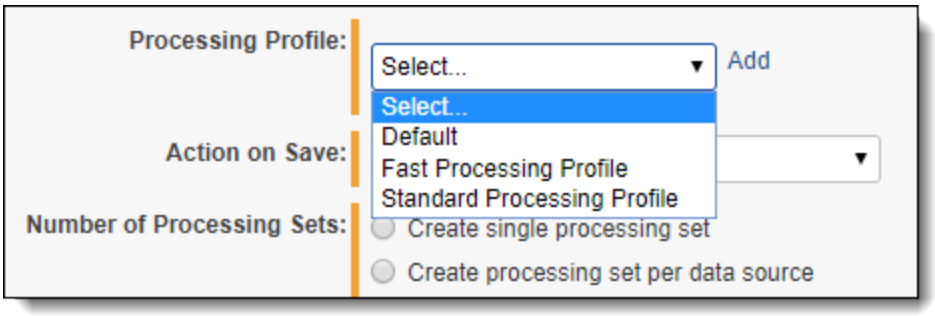


2. Complete the following fields on the Quick-Create Set(s) layout:



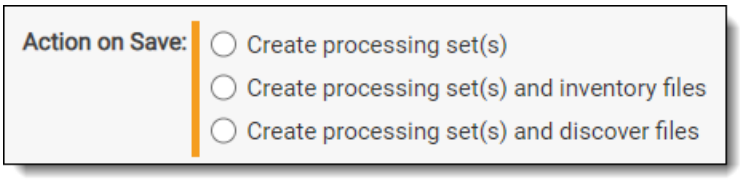
- **Name** - enter the name you want to appear on the processing set(s) created by this quick-create set. If you're creating a single processing set, it will display only the name you enter here. If you're creating multiple sets, they will display the name you enter here appended with - *<Custodian Name>*. The following are examples of processing set names:
  - Processing Set A - Doe, John
  - Processing Set A - Jones, Dan
- **Processing Profile** - select the profile you'd like to use to create the processing sets or click Add to create a new profile. These options are the same as you'd find in the Processing Profile field on the Processing Set layout. The Default, Fast Processing Profile, and Standard Pro-

cessing Profile are available by default.



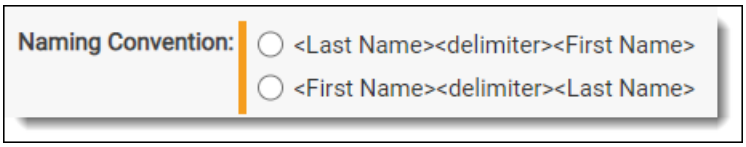
Note: Quick-create Sets do not support the use of Level based numbering. See Processing profiles for details.

- **Action on Save** - select what you'd like to occur when you save this quick-create instance. The options are:



  - **Create processing set(s)** - creates the processing set without starting an inventory or discovery job. This means you'll navigate to the processing set layout and manually start inventory or discovery.
  - **Create processing set(s) and inventory files** - creates the processing set and automatically starts the inventory job, which you can monitor on the processing set layout. Once inventory is complete, you can then manually start the discovery job.
  - **Create processing set(s) and discover files** - creates the processing set and starts a discovery job without first running inventory on the set.
- **Number of Processing Sets** - determine the number of processing sets you wish to create through this quick-create instance. The options are:
  - **Create single processing set** - creates a single processing set.
  - **Create processing set per data source** - creates a processing set for every data source you selected for the source path field.
- **Entity Type** - select either **Person** or **Other**.
  - **Person** - Relativity matches the FirstName and LastName fields, including the Delimiter value entered and the Naming Convention selected. The matching process excludes the FullName field.
  - **Other** - Relativity matches the folder selected as the Source Path to the FullName field of the Entity object. The matching process excludes the FirstName, LastName, Delimeter, and Naming Convention values.

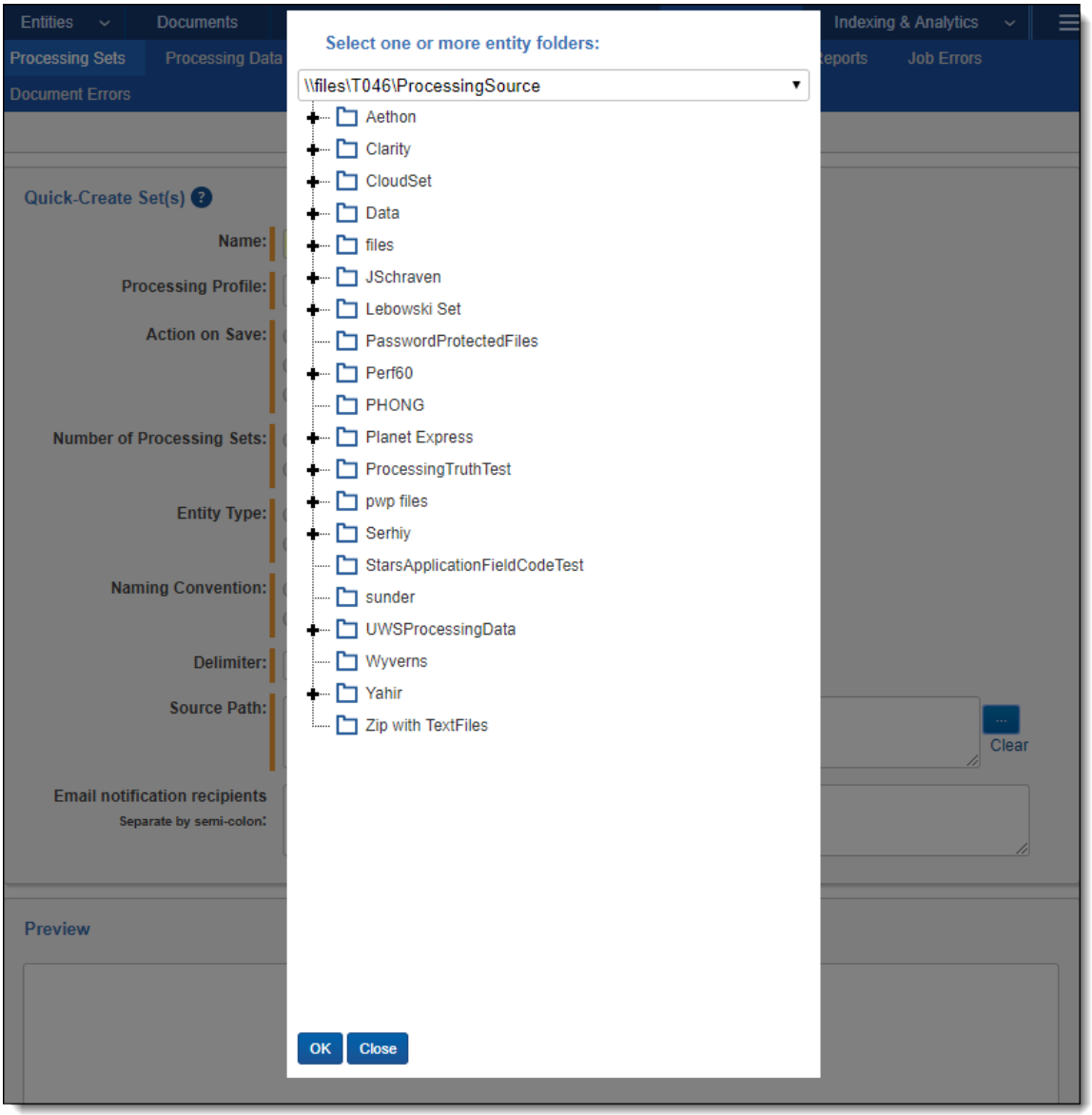
- **Naming Convention** - select the naming convention you like to use for the custodians in the folder structure. This field is only available if you selected Person as the custodian type above. The options are:

  - **<Last Name><delimiter><First Name>** - specifies that the folder names are formatted to contain the custodians last name, followed by the delimiter entered above, followed by the custodian's first name.
  - **<First Name><delimiter><Last Name>** - specifies that the folder names are formatted to contain the custodians first name, followed by the delimiter entered above, followed by the custodian's last name.
- **Delimiter** - enter the character you'd like to act as the delimiter between custodians in the folder structure. This field is only available if you selected Person as the custodian type above. If you enter something different than the delimiter contained in the folder path you selected for the Source Path field, you will encounter an error when you attempt to save. The following are examples of delimiters you could enter here:
  - - (hyphen)
  - . (period)
  - _(underscore)
  - Relativity treats a delimiter of <> as a space.
- **Source path** - the location of the data you want to process. Click the ellipsis to see the list of folders. The source path you select determines the folder tree below, and the folder tree displays an icon for each folder within the source path. You can specify source paths in the resource pool under the Processing Source Location object. Each folder you select here will act as a data source on the Processing Set.

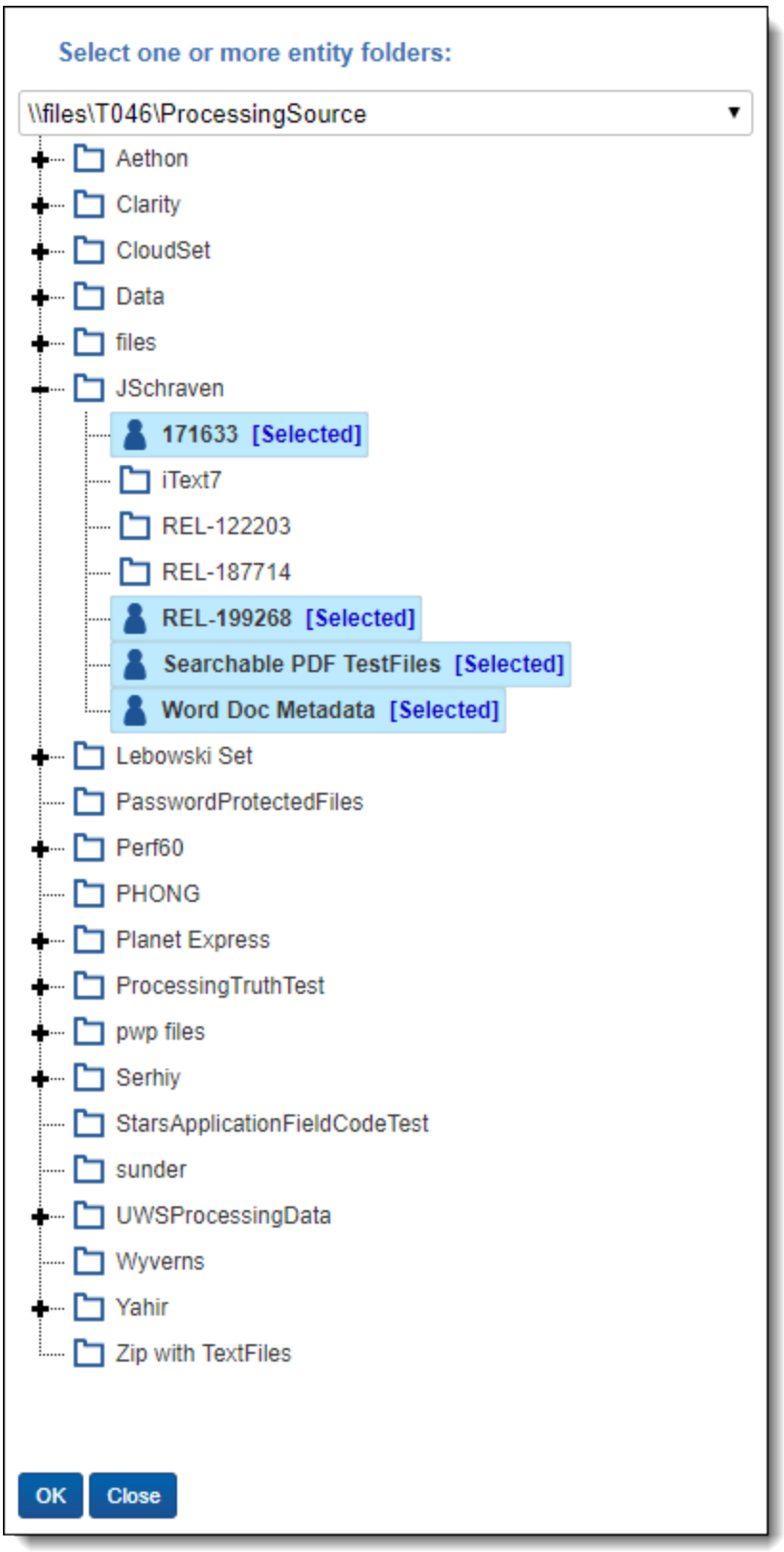
  Note: You can select up to 100 folders.

  Each data source will have a corresponding Custodian, with the folder name serving as the Custodian name. Click **OK** after you select the desired folders; ensure that the folder corresponds to the person or entity you select for the Entity Type field described below.



- You can select any single folder by clicking on it once. Clicking it again clears that selection.
- When you right-click a folder, you see the following options:
  '
  - **Expand** - expands all the sub-folders under the folder you've right-clicked.
  - **Collapse all** - collapses all the expanded folders under the folder you've right-clicked.
  - **Set Children as Entities** - marks the folder's children as custodians. When you select this option, each child folder is marked as **[Selected]** in the folder tree. These folder names then appear as data sources in the Preview window in the Quick-Create Set(s) layout. They also show up as data sources on the processing

set layout once you save the quick set.

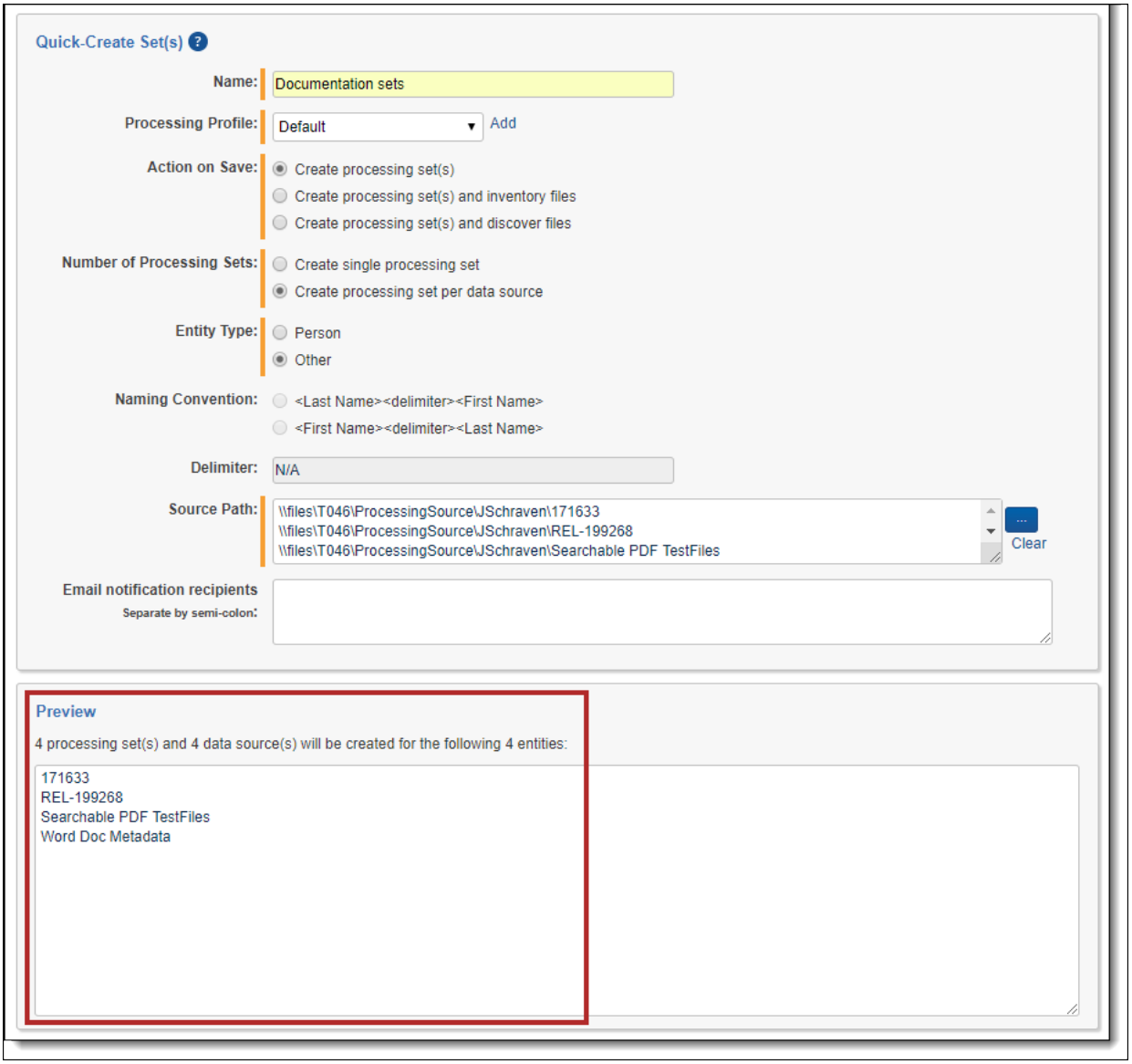


- **Clear** - clears any selections you made.

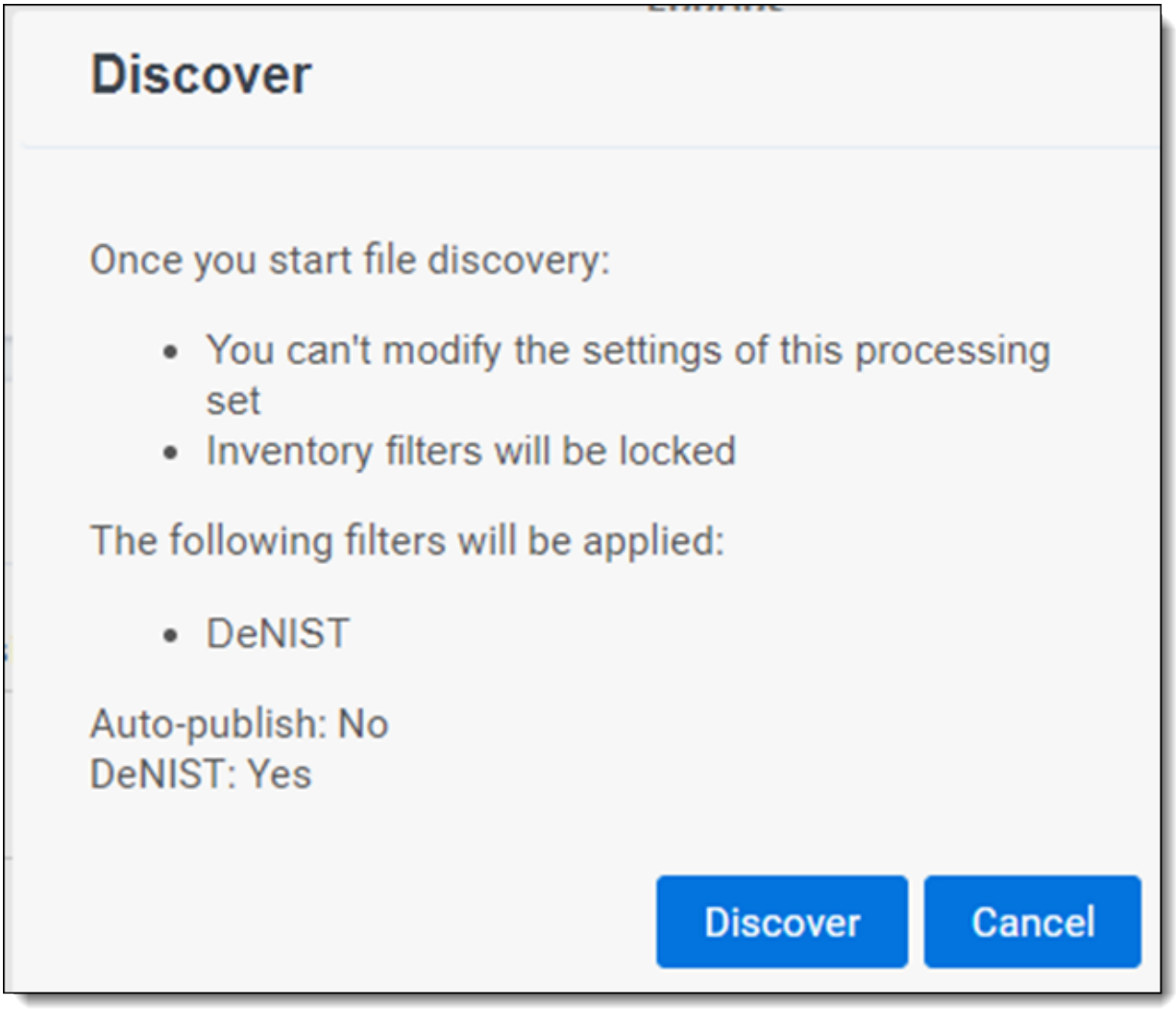
  - **Email notification recipients** - enter the email addresses of those who should receive notifications of whether the processing set creation succeeded or failed. Use semi colons between addresses.

3. (Optional) View the number of data sources you're about to create in the Preview window. If this window is empty, it's most likely a sign that you've selected an invalid source path and/or that you've entered an invalid delimiter; in this case, you'll most likely receive an error when you click Save and Create.

4. Click **Save and Create**.



5. If your Action on Save selection above was to either inventory or discover files, then you must click **Discover** or **Inventory** on the subsequent confirmation message to proceed with saving the quick-create instance.



Once you save and create the quick-create instance, the layout displays a success message and directs you to the Processing Sets tab.

9.2.1 Validations and errors

Note the following details about validations and errors:

- Relativity performs the same validations for quick-create set creation as it does for processing profile and set creation. This means that, for example, if the queue manager happens to go down while you're creating the quick -create instance, you'll see an error message at the top of layout and you

won't be able to save and create it.

- If the quick-create instance fails to create a set, custodian, or data source, the resulting error will go to the Errors tab in Home, and if the error occurs after the creation of any of those objects, it will be considered a processing error and will go to either the Document Errors or Job Errors tab in the processing object.
- If your entity type is set to Person, Relativity validates that the names of the selected folders adhere to the selected delimiter and naming convention values.
- If you enter **,** (comma) as your delimiter, and you have one or more folder names that don't contain a comma, you will receive an invalid source path error after clicking Save and Create. In this case, the Preview window would also not list any entity names.

10 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: Never upgrade your Relativity version while there are jobs of any type currently in progress in your environment. Doing this leads to inaccurate results when you attempt to finish those jobs after your upgrade is complete. This is especially important for imaging and processing jobs.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

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.

Note: When you upgrade from Relativity 8.1 to Server 2022 with processing sets that are in an agent error state, the status section of the upgraded processing set doesn't display the agent error. This is because there is no job in the queue for the data source that contains the error.

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.

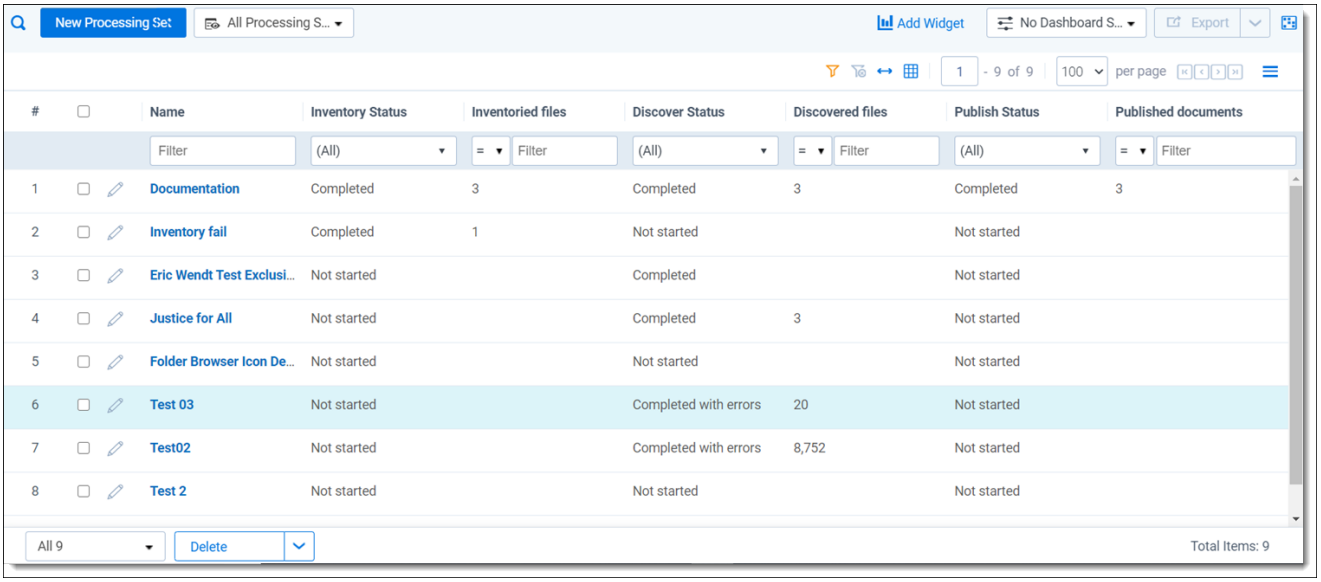
For the details of creating a processing set, see Creating a processing set on page 135.

10.1 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:



- **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.

10.2 Creating a processing set

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

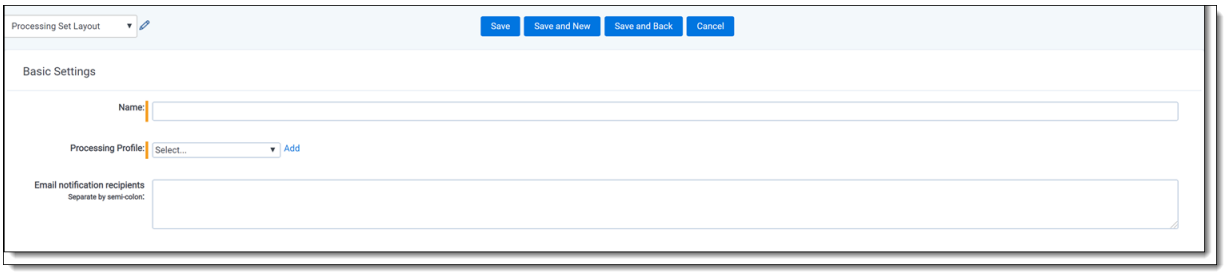
To create a processing set:

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 below.
4. Click **Save**.
5. Add as many Processing Data Sources to the set as you need. See Adding a data source on page 137.

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.

10.3 Processing Set Fields

To create a processing set, complete the following fields:

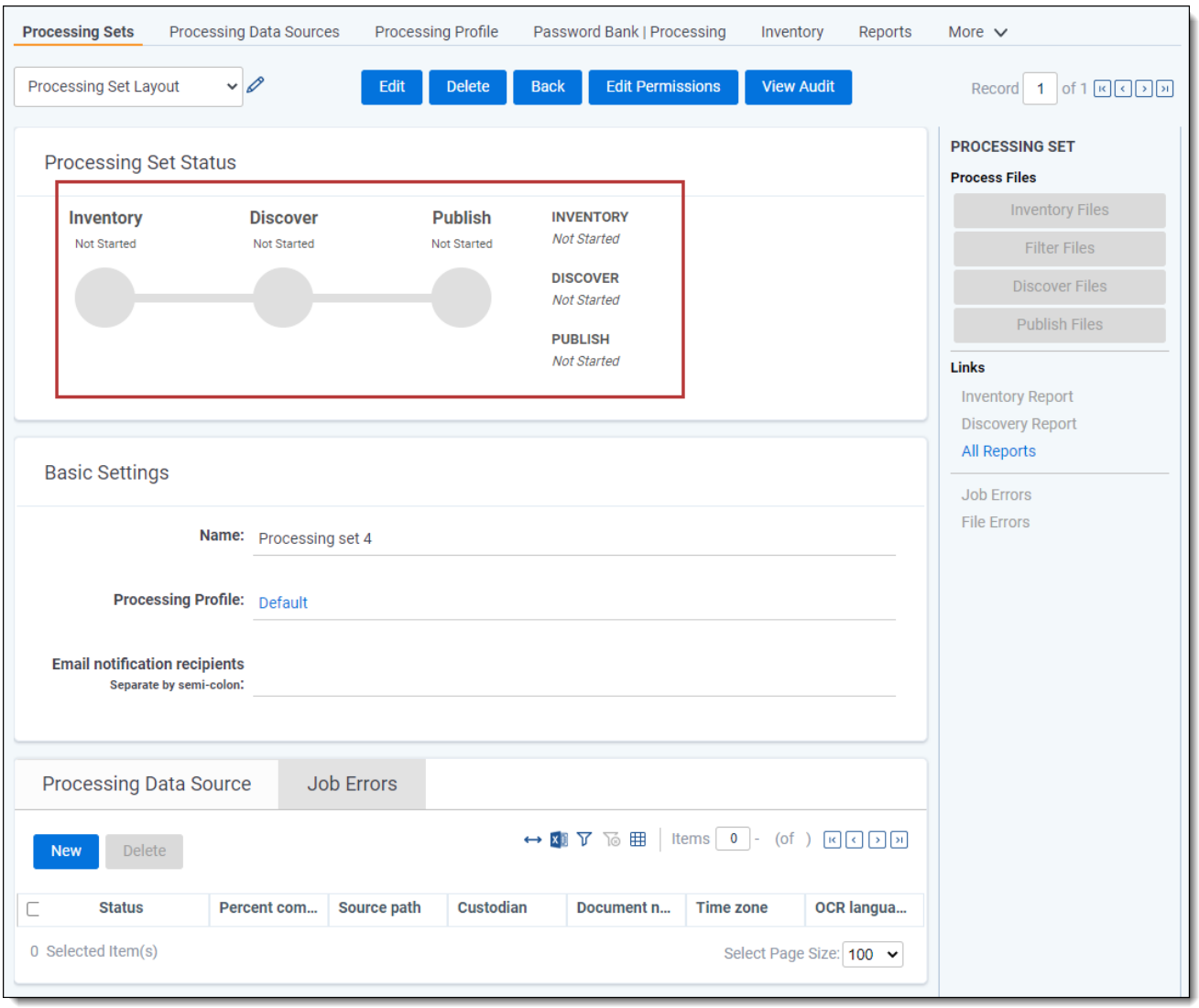


- **Name** - the name of the set.
- **Processing profile** - select any of the profiles you created in the Processing Profiles tab. If you haven't created a profile, you can select the Default profile or click **Add** to create a new one. If there is only one profile in the workspace, that profile is automatically populated here. See Processing profiles on page 85.
- **Email notification recipients** - the email addresses of those whom you want to receive notifications while the processing set is in progress. Relativity sends an email to notify the recipient of the following:

- Inventory
  - Successful inventory completed
  - Inventory completed with errors
  - First discovery job-level error
  - Inventory error during job submission
- Discovery
  - Successful discovery completed
  - Discovery completed with errors
  - First discovery job-level error
  - File discovery error during job submission
- Retry - discovery
  - First discovery retry job-level error
  - Discovery retry error during job submission
- Publish
  - Successful publish completed
  - Publish complete with errors
  - First publish job-level error
  - Publish error during job submission
- Retry - publish
  - First publish retry job-level error
  - Publish retry error during job submission

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

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



The 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.

10.4 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.

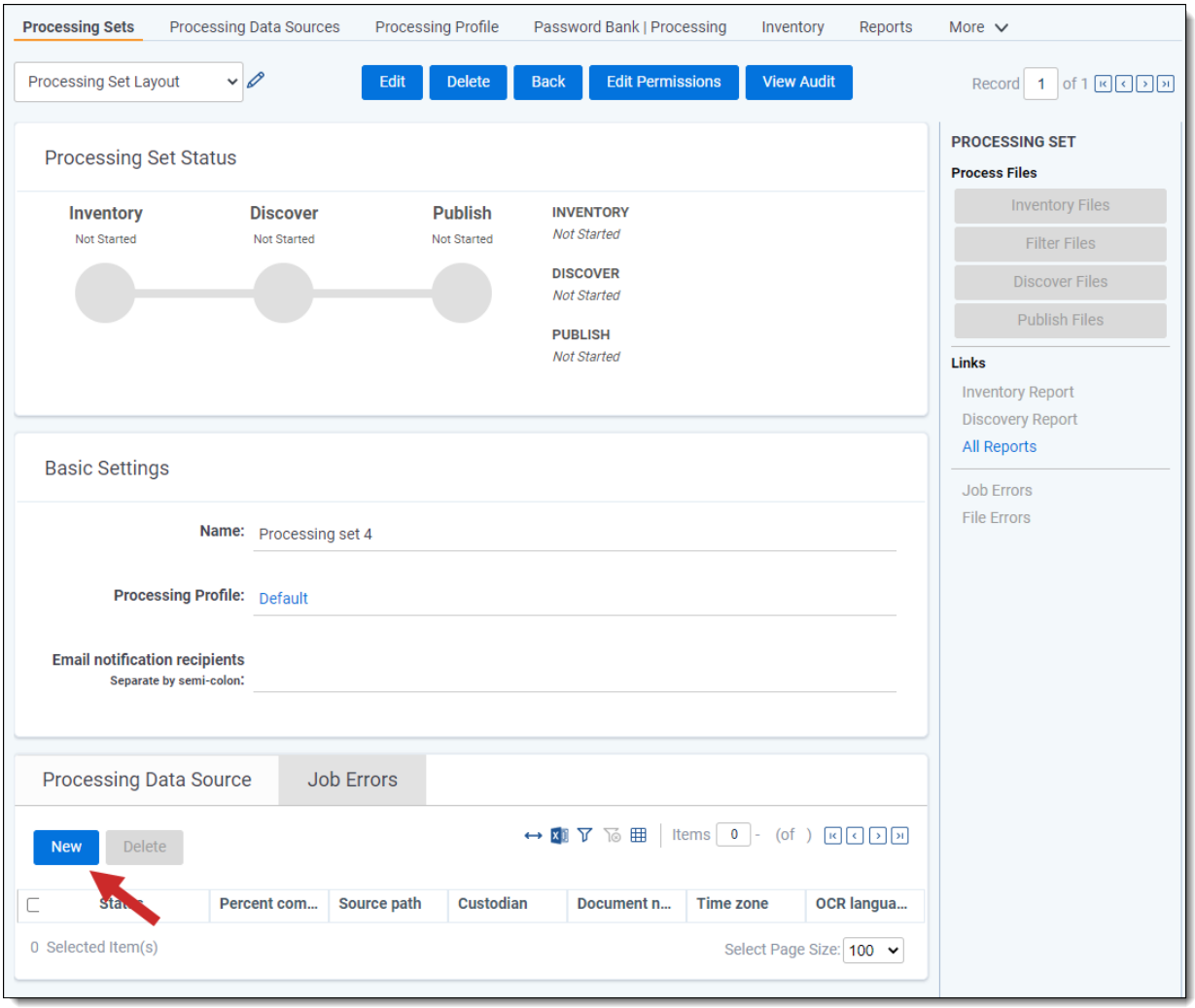
Note: You have the option of using Integration Points to import a list of custodians from Active Directory into the Data Sources object. Doing this would give you an evergreen catalog of custodians to pick from when preparing to run a processing job.

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 135.
2. On the **Processing Data Source** object of the processing set click **New**.



3. Complete the fields on the **Add Processing Data Source** layout. See Data Source Fields on the next page.

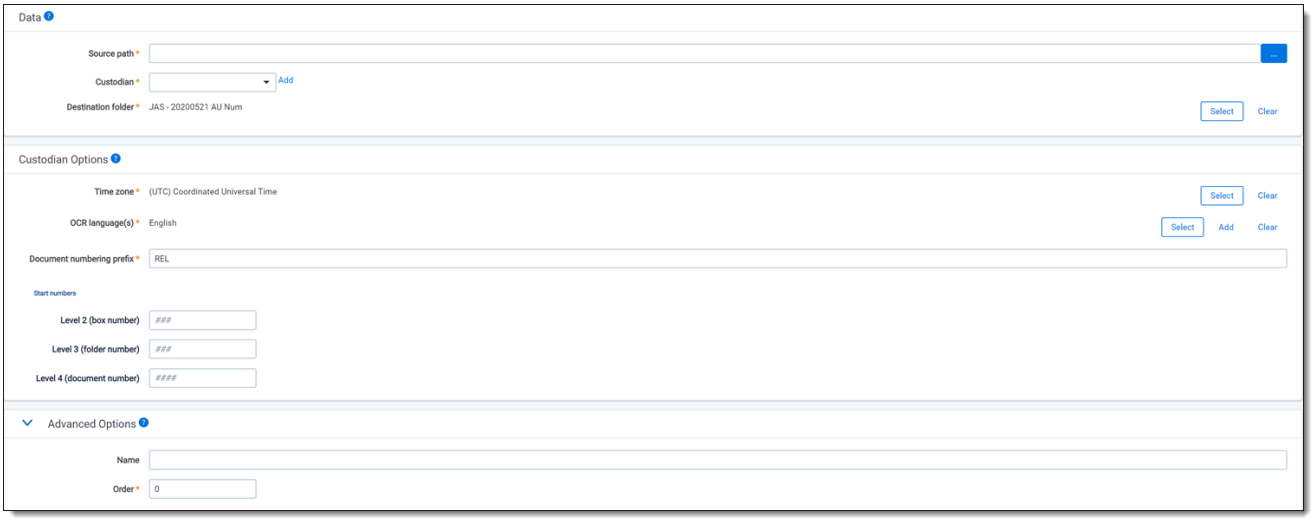
4. Click **Save**. When you save the data source, it becomes associated with the processing set and the console on the right side is enabled for inventory and file discovery.

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 145.

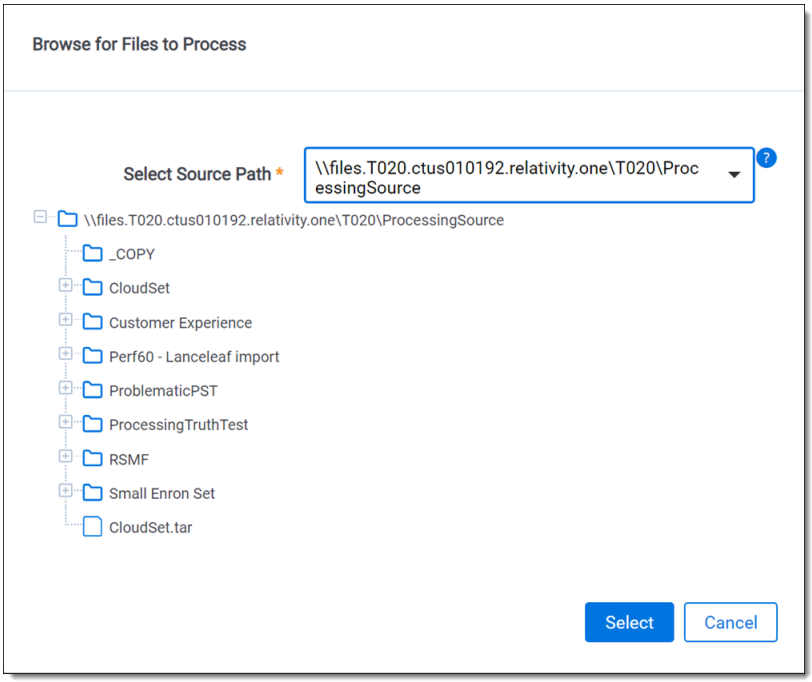
Note: If you add, edit, or delete a data source associated with a processing set that has already been inventoried but not yet discovered, you must run inventory again on that processing set. You can't add or delete a data source to or from a processing set that has already been discovered or if there's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

10.5 Data Source Fields

To add a data source, complete the following fields:



- **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. For processing and imaging data sets containing CAD files, you can configure the timeout value in the AppSettings table.

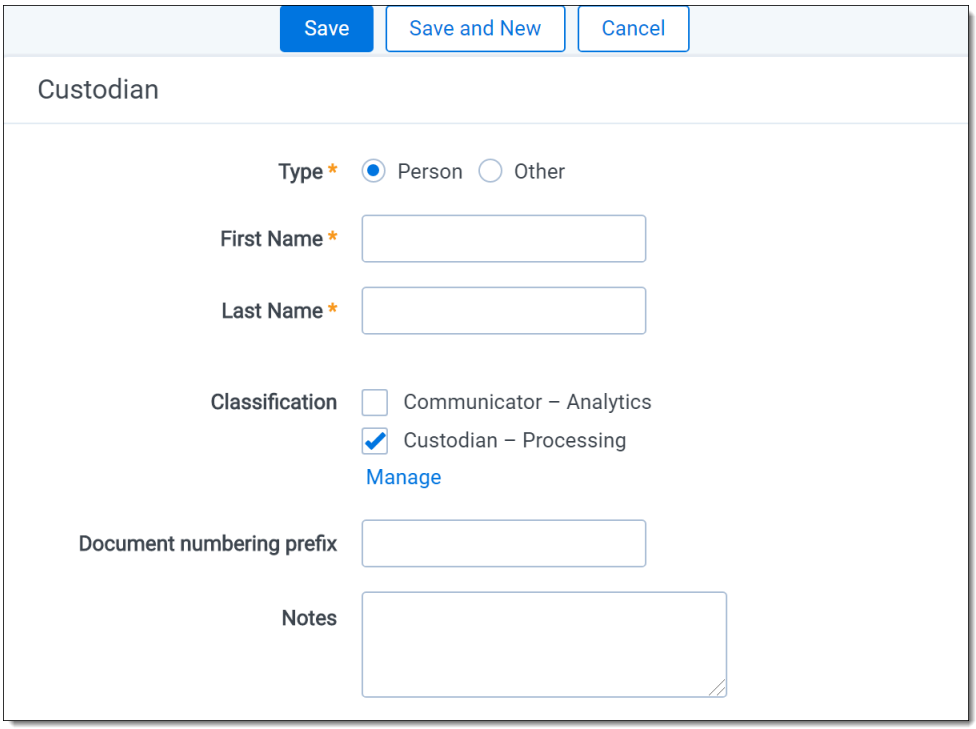


- The processing engine processes all the files located in the folder you select as your source as one job. This includes, for example, a case in which you place five different .PSTs from one custodian in a single folder.
- You can specify source paths in the resource pool under the Processing Source 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 is finished. 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 155.

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.



- **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.
- **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 master 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 master record are published. Additionally, there may be other mapped fields available that can describe additional fields of the duplicates.

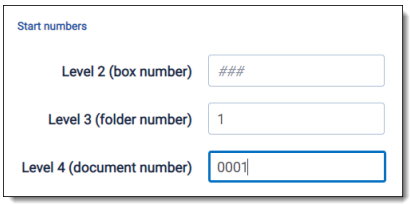
Note: One of the options you have for bringing custodians into Relativity is Integration Points (RIP). You can use RIP to import any number of custodians into your environment from Active Directory and then associate those custodians with the data sources that you add to your processing set.

- **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.
- **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.



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.

10.5.1 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 worker manager 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.

10.5.2 Edit considerations for data sources

Note the following guidelines for modifying data sources:

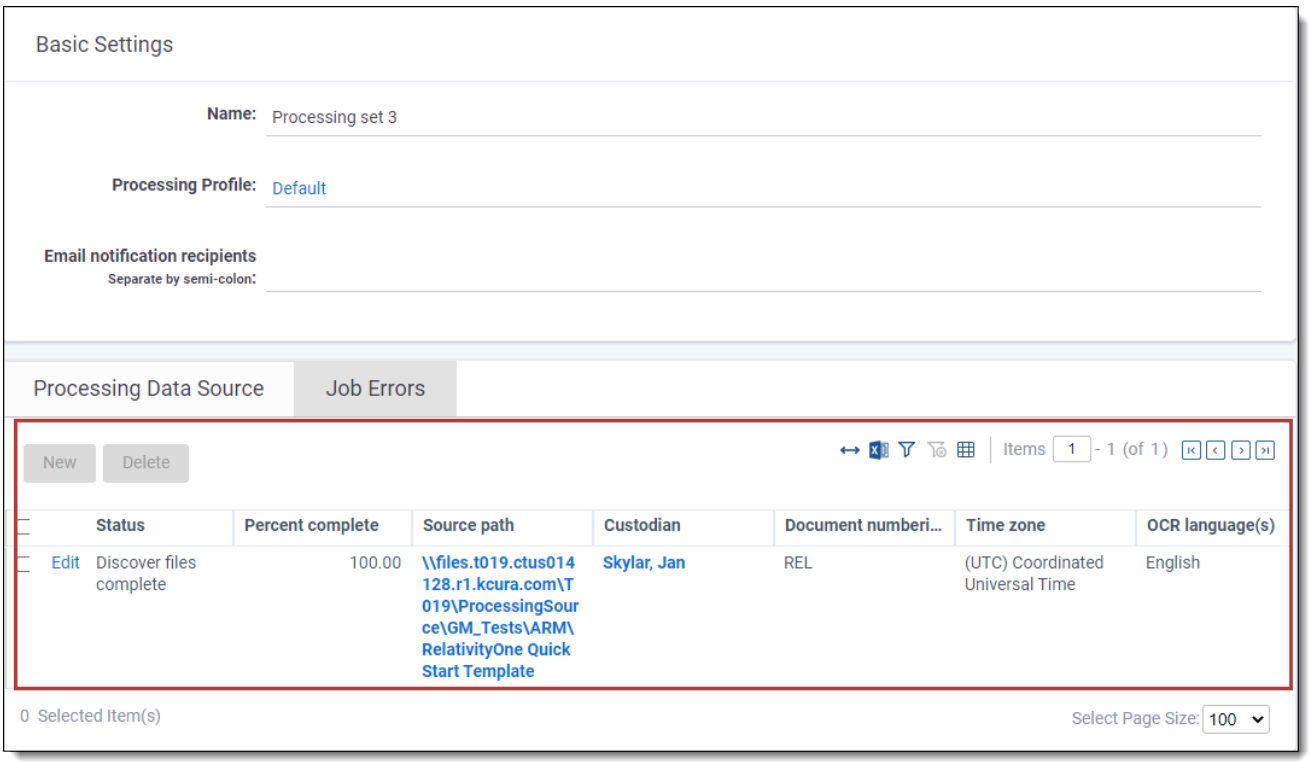
Note: If you've started a processing job with a priority value that is higher than 1, and you want to start and finish a Mass PDF job before that processing job completes, you must go to the Worker Manager Server and manually change the priority of the **Single Save as PDF** choice to be lower than any of the processing choices (**Inventory**, **Discovery**, and **Publish**). Setting the priority of a Mass PDF job must be done before the job begins for it to finish before other processing jobs. For details, see the Admin Guide.

- 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.

10.5.3 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.



This view provides the following fields:

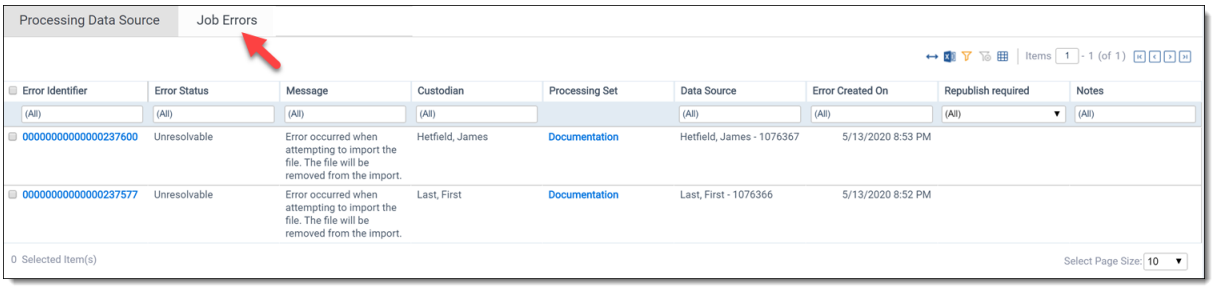
- **Status** - the current state of the data source as inventory, discovery, publish, or republish runs on the processing set. This and the Percent Complete value refresh automatically every five seconds. This field was introduced in Relativity 9.5.162.111. The status values are:
  - **New** - the data 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 Relativity 9.5.253.62 as part of the distributed publish enhancement.
  - **DeDuplication and Document ID Generation** - the master 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 Relativity 9.5.253.62 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 master 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. This field was introduced in Relativity 9.5.162.111.
- **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.

10.5.4 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.



The Current Job Errors view in the Job Errors tab displays all unresolved job errors while the All Job Errors view displays any job error that has occurred throughout the lifecycle of the matter. Both views contain the following fields:

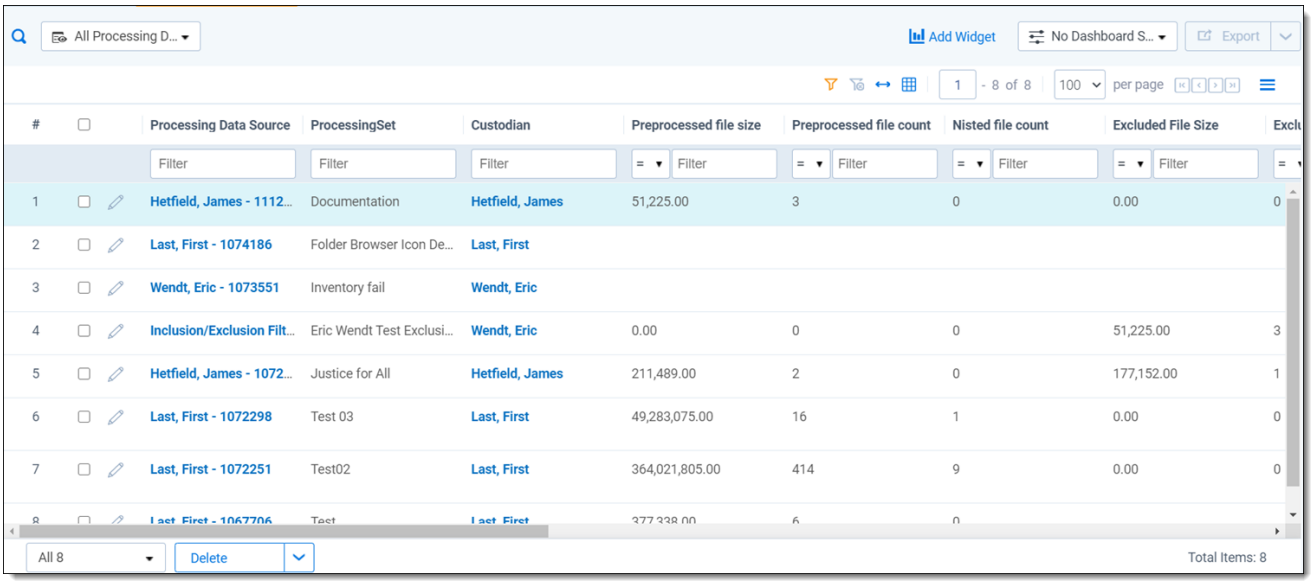
- **Error Identifier** - the unique identifier of the error as it occurs in the database. When you click this message, you are taken to the error details layout, where you can view the stack trace and other information. Note that for Unresolvable errors, the console is disabled because you can't take any actions on that error from inside Relativity. For more information, see Processing error workflow.
- **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 workflow on page 239.

10.6 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.



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 thread 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.

10.7 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.

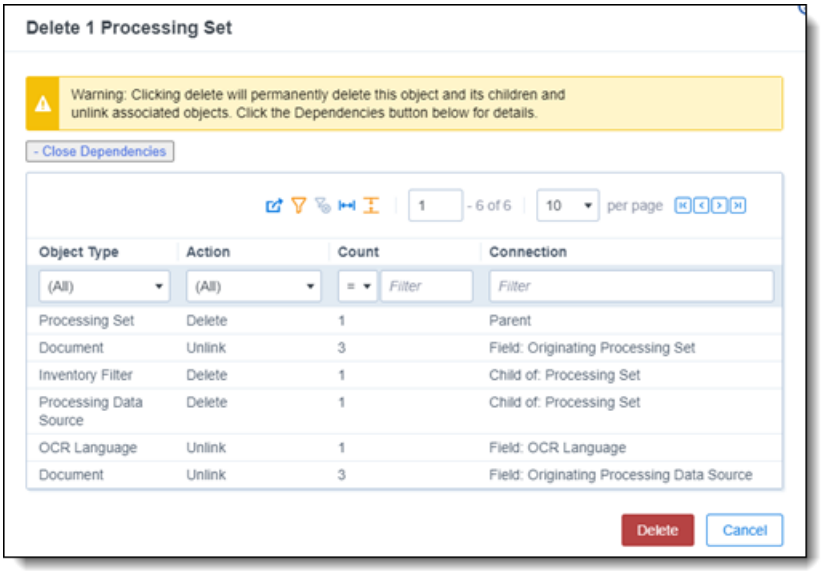
- Object Security
  - Edit permissions for Field, with the Add Field Choice By Link setting checked
  - (Optional) Delete permissions for OCR Language
  - Delete permissions for Processing Data Source, Processing Error, Processing Field, and Processing Set

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.



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.

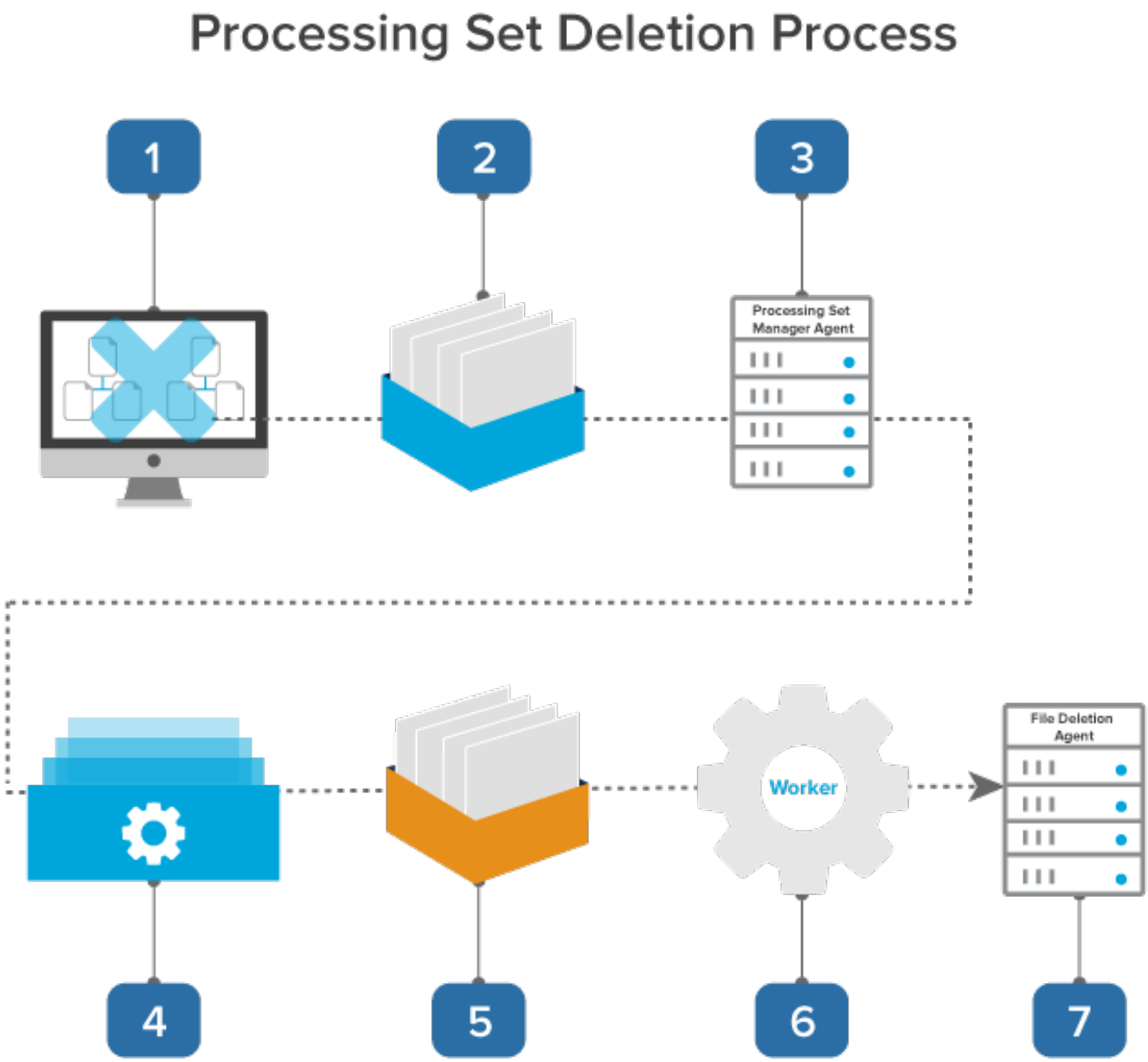
The following table breaks down what kinds of data is deleted from Relativity and Invariant when you delete a processing set in certain phases.

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

Phase deleted	From Relativity	From Invariant
Discovered processing set	Processing set object - errors, data sources	Discovered metadata

When you delete a processing set, the file deletion manager deletes all physical files and all empty sub-directories. Files that the system previously flagged for future deletion are also deleted.

The following graphic and accompanying steps depict what happens on the back end when you delete a processing set:



1. You click Delete on the processing set.
2. A pre-delete event handler inserts the delete job into the worker manager queue while Relativity deletes all objects associated with the processing set.

3. A processing set agent picks up the job from the worker manager 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.

10.8 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:

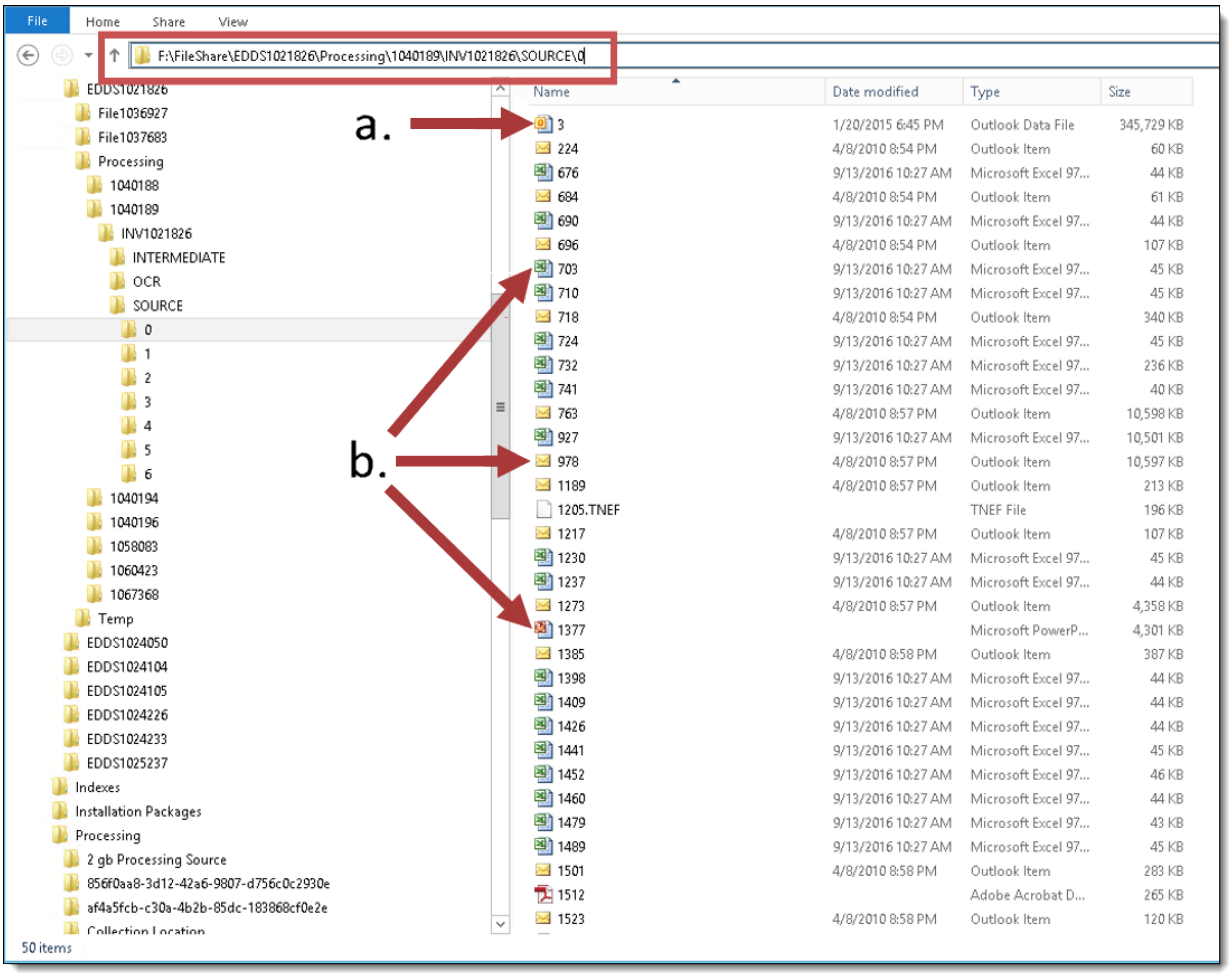
1. Run a processing set.
2. After the processing set is complete, import a small amount of data using the RDC so that you can keep one steady stream of control numbers and pick up manually where the previous processing set left off.
3. After importing data through the RDC, run another processing set,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

10.9 Copying natives during processing

To gain a better understanding of the storage implications of copying natives during processing, note the behavior in the following example.

When you process a PST file containing 20,000 unique total documents while copying natives:

1. You copy the PST from the original source to your Processing Source Location, as this is the identified location where Relativity can see the PST. Note that you can make the original source a processing source by opening the original source to Relativity.



Note: If you run Inventory on this set, Relativity will identify all parents and attachments, but it will only extract metadata on the parent email.

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.

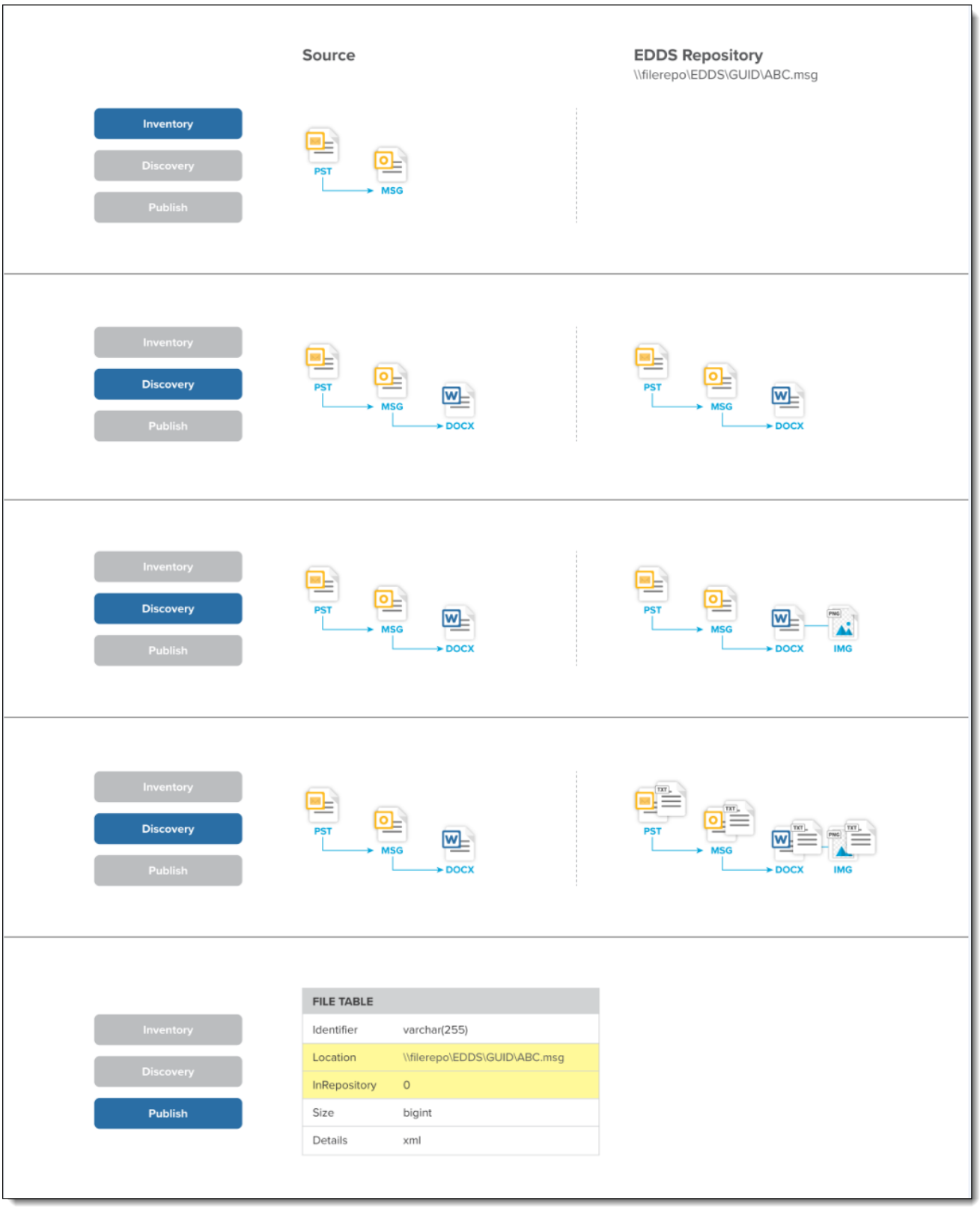
4. Once processing is complete:
   - You can delete the processing source PST.
   - You can delete the PST file in the EDDS folder, assuming there are no errors.

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.

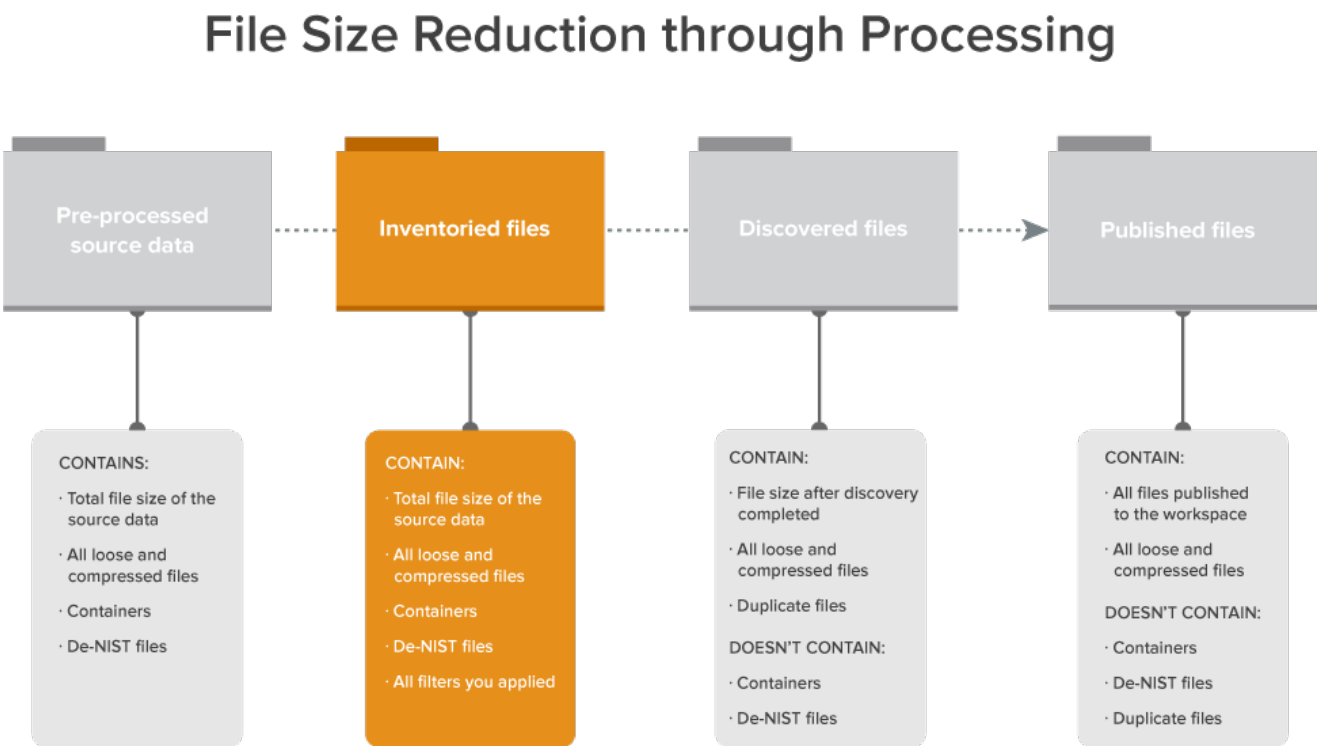
This graphic is designed for reference purposes only.


Source
EDDS Repository
\\filerepo\EDDS\GUID\ABC.msg
Inventory
Discovery
Publish
PST
MSG
DOCX
IMG
FILE TABLE
Identifier varchar(255)
Location \\filerepo\EDDS\GUID\ABC.msg
InRepository 0
Size bigint
Details xml

11 Inventory

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.

The following graphic depicts how inventory 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.



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 aren't 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.

Note: Inventory isn't available in the Relativity Processing Console.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are

based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

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.

Note: You can't use Inventory (in Processing) on IE 8. You can only use Inventory on IE 9 and 10.

Using Inventory and file filters

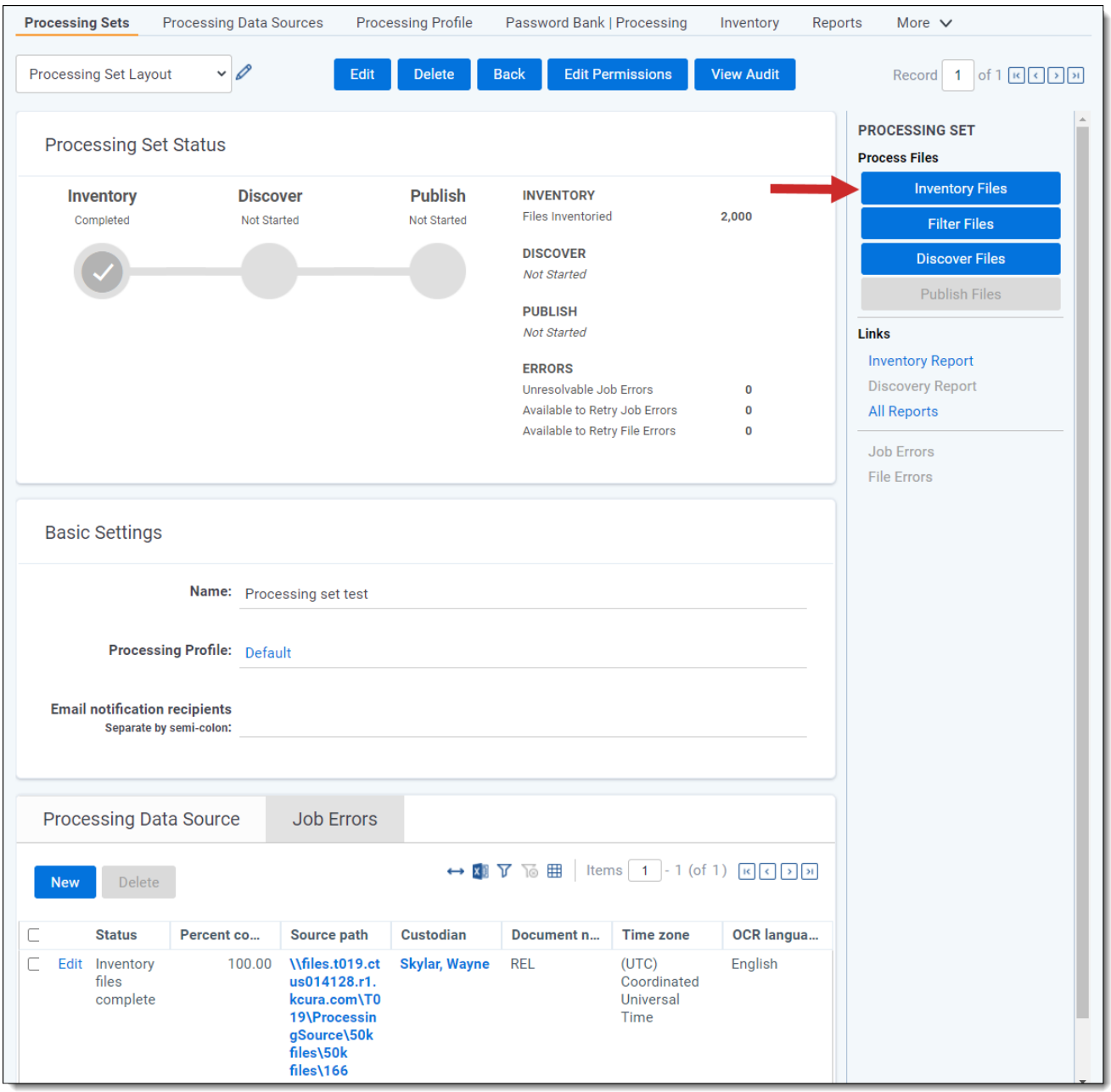
You're a project manager, and your firm requests that you create a processing set that includes a purposefully large data set from a custodian, with loose files and multiple email PST files. They then want you to eliminate all emails from a repository created in 2012, because those pre-date the case 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.

11.1 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've added at least one data source to the processing set.



Note: The default priority for all inventory jobs is determined by the current value of the ProcessingInventoryJobPriorityDefault entry in the Instance setting table.

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

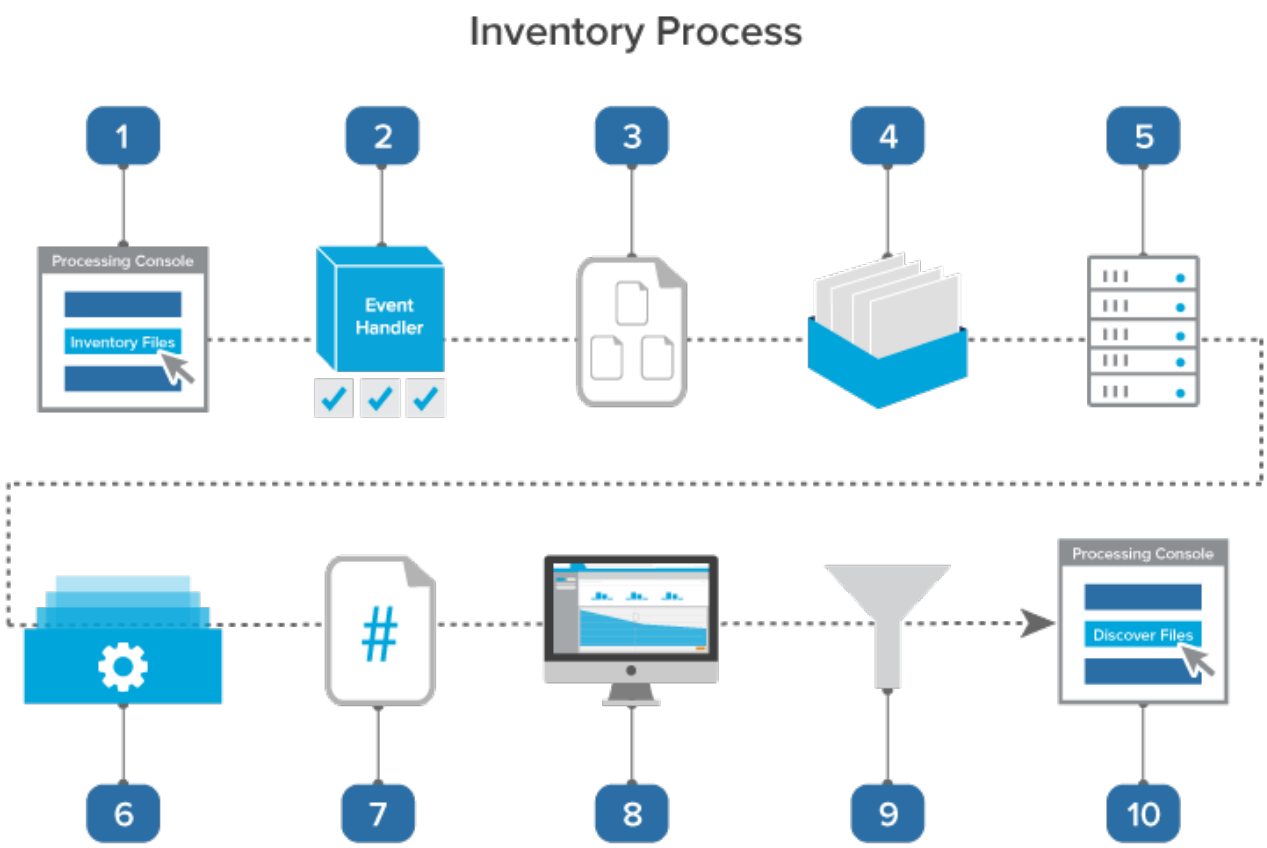
When you start inventory, the Inventory Files button changes to Cancel. You can use this to cancel the processing set. For more information, see Canceling inventory.

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.

11.1.1 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.

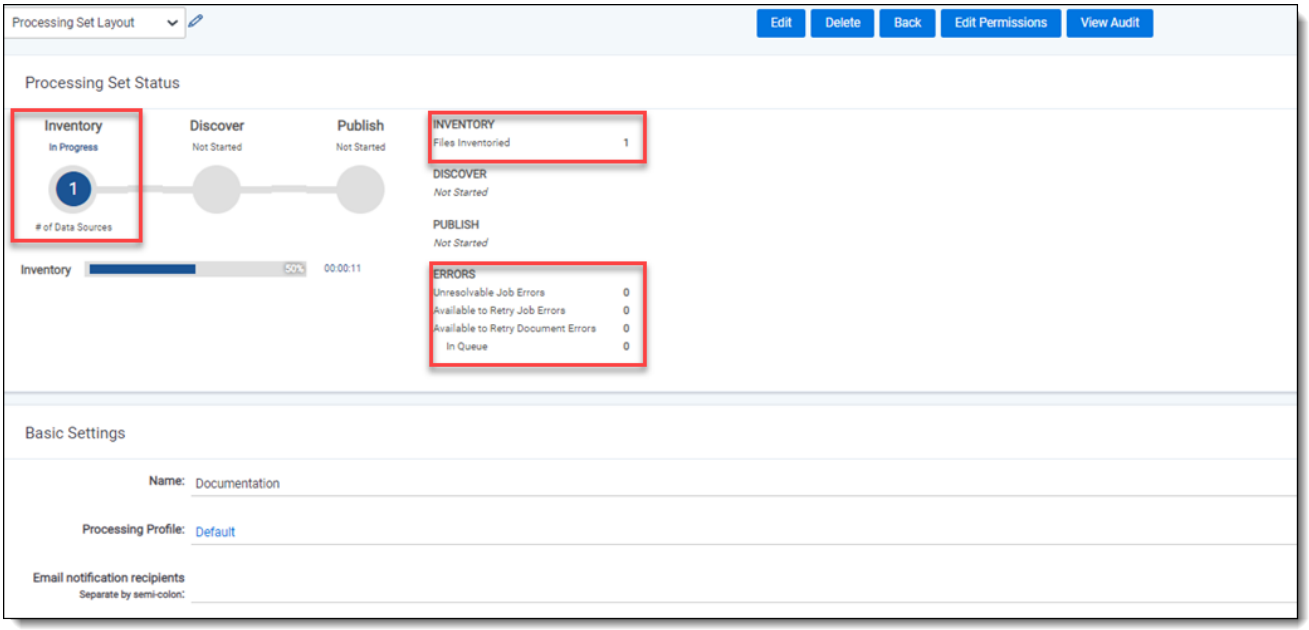


1. You click Inventory Files on the processing console.
2. A console event handler checks to make sure the processing set is valid and ready to proceed.
3. The event handler inserts the data sources to the processing queue.
4. The data sources wait in the queue to be picked up by an agent, during which 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.

11.1.2 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.



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** - 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 workflow for details.

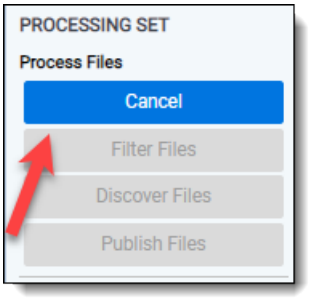
If you skip inventory, the status section displays a Skipped status throughout the life of the processing set.

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 on the next pageand Discovering files on page 184

11.1.3 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**.



Consider the following regarding canceling inventory:

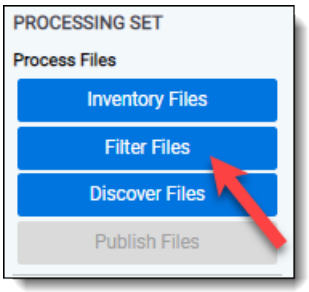
- If you click Cancel while the status is still Waiting, you can re-submit the inventory job.
- If you click Cancel after the job has already been sent to the processing engine, the entire processing set is canceled, meaning all options are disabled and it is unusable. Deduplication isn't run against 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 can't be retried.

11.2 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've inventoried. Everything else that cascades down from the files that were discovered is not subject to the inventory filters that you set.

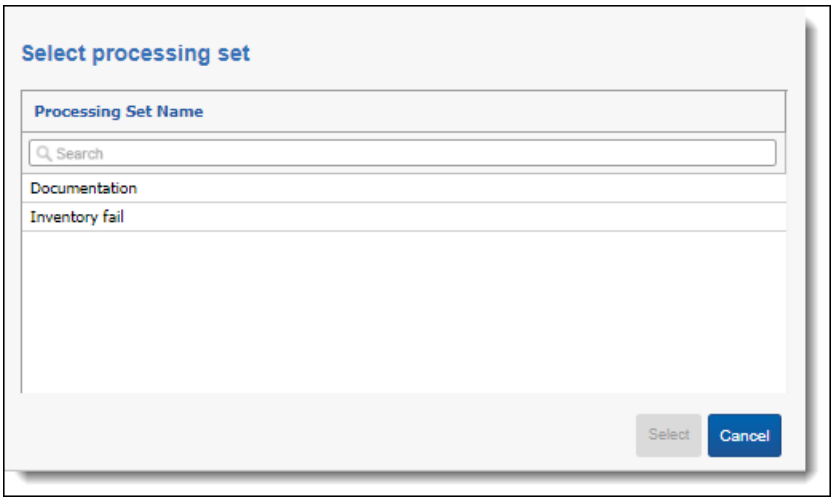
To do this, click **Filter Files** on the console.



When you click this, you're redirected to the Inventory tab, which loads your processing set.

When you click the Inventory tab for the first time from anywhere else in Relativity, no processing set is loaded by default, and you're presented with a list of sets that are eligible for filtering.

Click on a set and click **Select** to load the set on the Inventory layout.



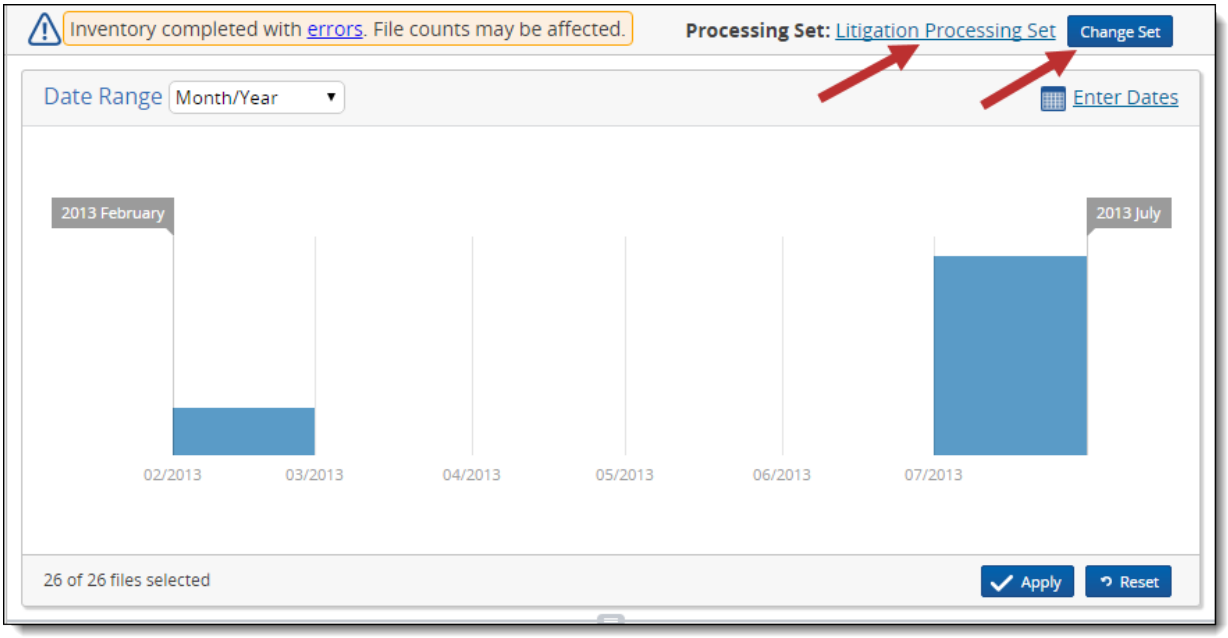
A processing set is not eligible for use in the Inventory tab if:

- You canceled the set.
- You already discovered or published the set.
- You haven't 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'll be directed to the Processing Sets tab to create a new set or check on the progress of an existing set.

The following considerations apply to all processing sets in Inventory:

- If you need to load a different processing set, click **Change Set** to display a list of other available sets.
- You can click the processing set name link in the top right to return to that set's layout.

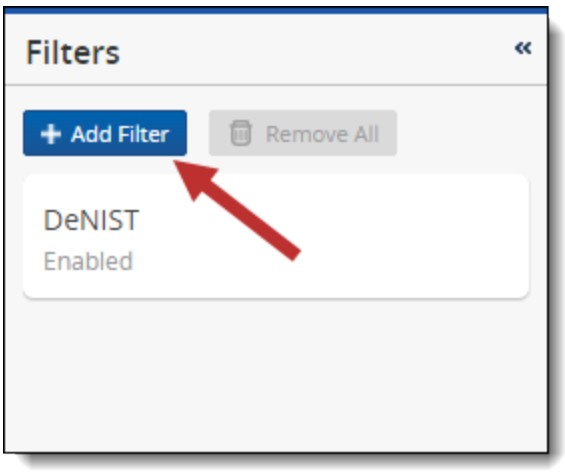


Note: If you leave the Inventory tab after having loaded a processing set, that set and any filters applied to it are preserved for you when you return to the Inventory tab.

You can add filters to the inventoried data and see how those filters affect the data in your processing set. You can't add filters if inventory is not complete or if the processing set has already been discovered.

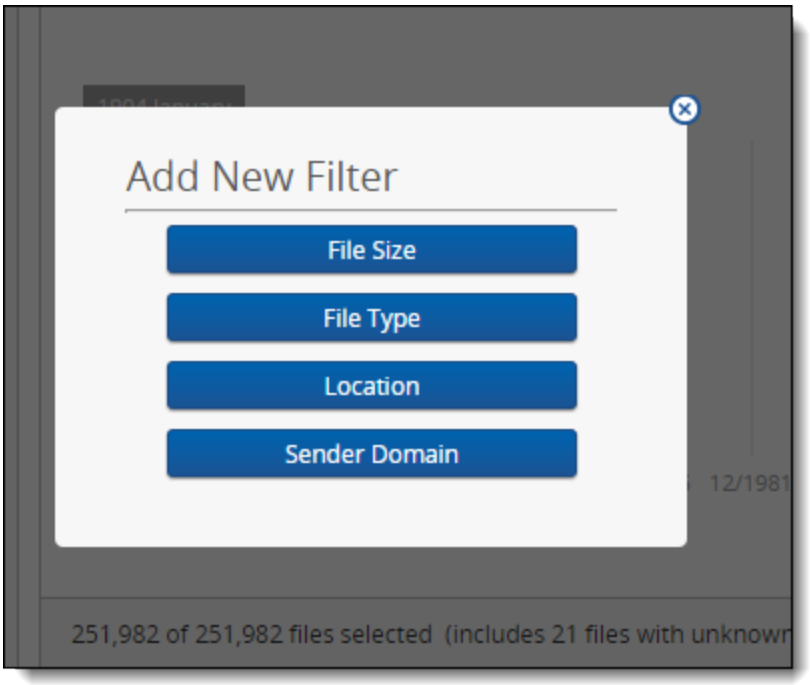
There are six different filters you can apply to a processing set. You can apply these filters in any order; however, you can only apply one filter of each type. This section describes how to apply file location, file size, file type, and sender domain filters. See Applying a Date range filter on page 168 or Applying a deNIST filter on page 172 filter for instructions on filtering inventoried files by those properties.

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



Note: Filters affect the data set only at the time at which you apply them. This means that if you apply a filter to exclude a certain file type from your data but someone from your organization adds more files to the set, including instances of the excluded type, then the recently added files aren't 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 work-space. 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.

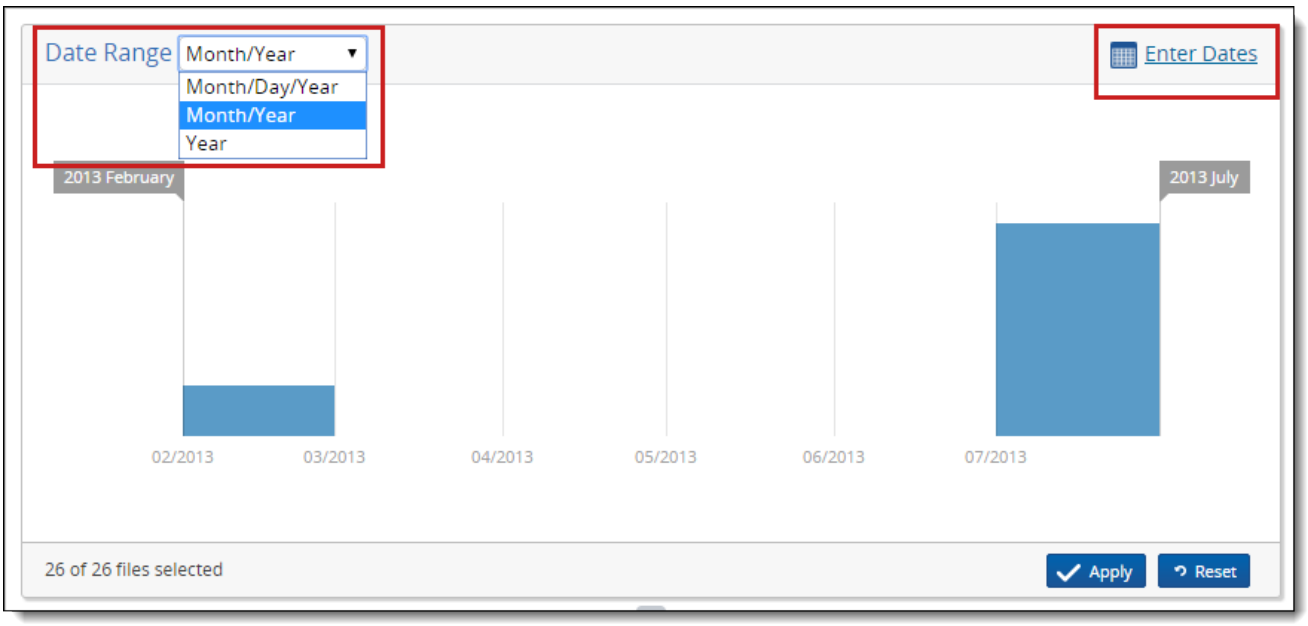
The following considerations apply to all filter types:

- If the applied filter conditions have excluded all files from the set, then there are no results for you to interact with and you can't 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 can't save and reuse filters from one inventory set to another.

11.2.1 Applying a Date range filter

When the selected processing set loads, no filters are applied to the files by default; however, a graph displays the date range for all files in the processing set.

> **Note:** The deNIST filter is applied by default if your processing profile has deNIST field set to Yes.



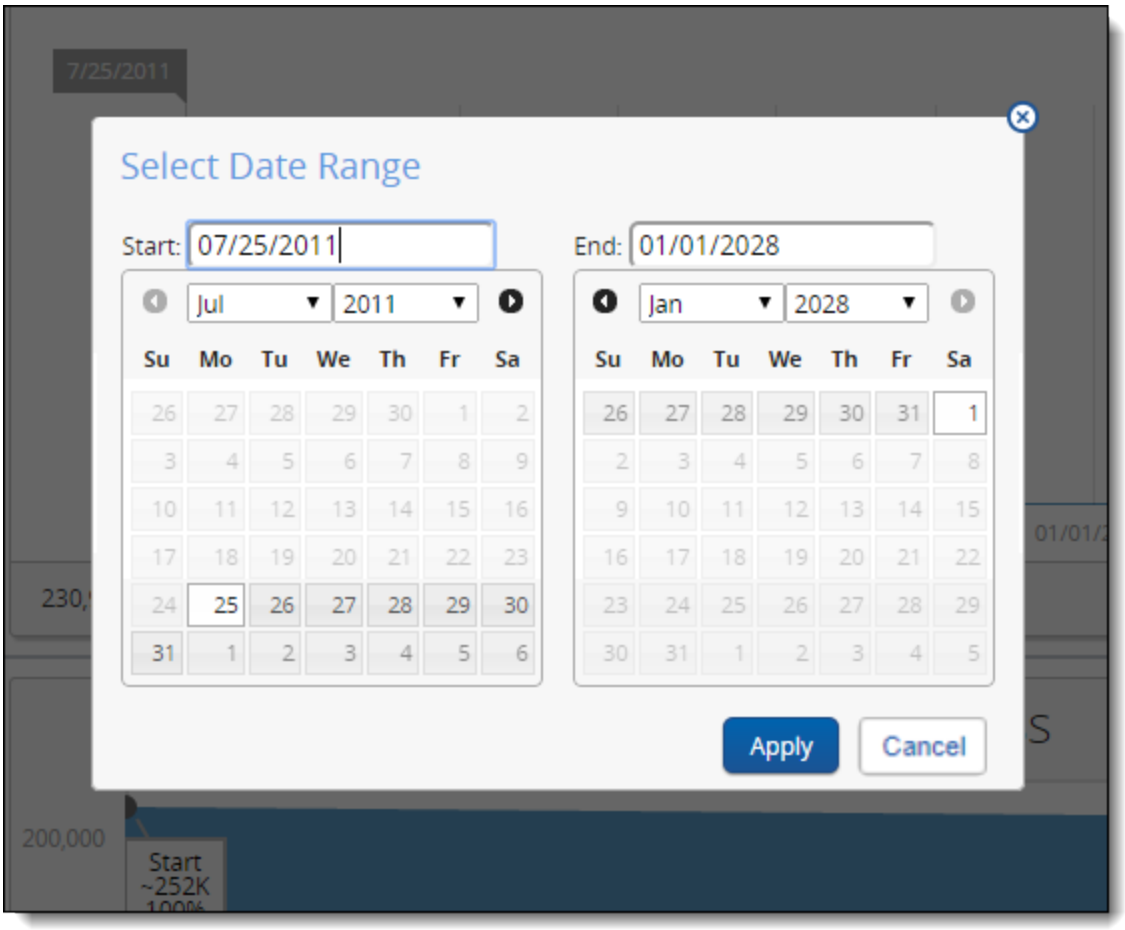
Note: When you filter for dates, you're filtering specifically on the Sort Date/Time field, which is taken from the file's Sent Date, Received Date, and Last Modified Date fields in that order of precedence. This happens on email messages repeated for the parent document and all child items to allow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



- 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.

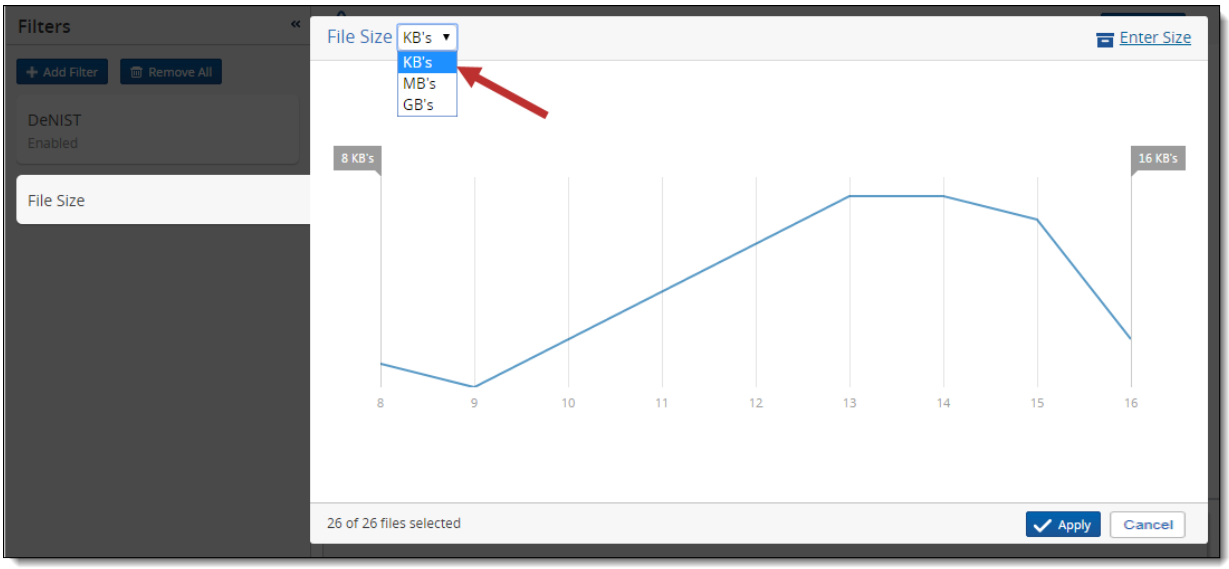
- To reset the parameters of a filter after you apply it, click **Reset**.

Note: If you run a re-inventory job on a processing set to which you've already added the date range filter, the date range display doesn't 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.

11.2.2 Applying a File Size filter

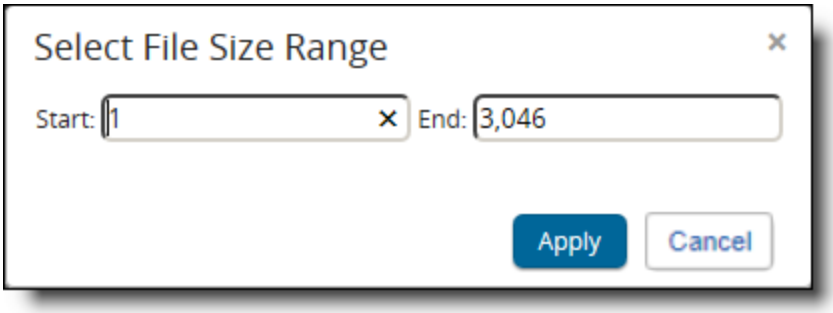
To filter your processing set files by size:

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



- Use the **File Size** menu in the top left of the graph to select from **KB's**, **MB's**, and **GB's**. If all files in the current data set are from the same file size, for example 0 GB's, you can't 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.



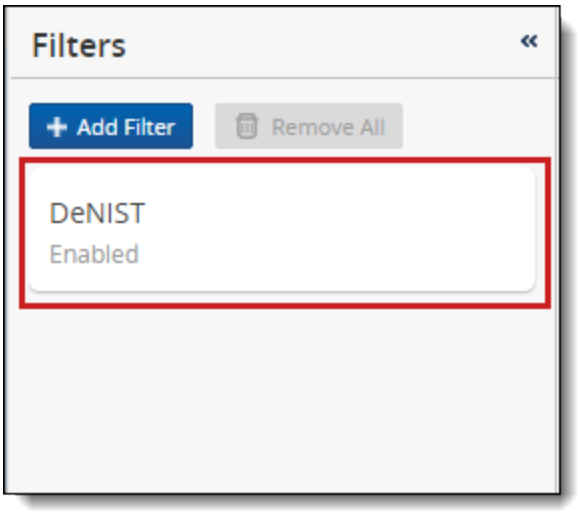
4. Click **Apply** once you've designated all the file sizes you want to exclude. The Inventory Progress pane reflects the addition of the file size filter, as well as the percentage and number of files that remain from the original data set. For more information, see Inventory progress on page 178.

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.

11.2.3 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'll 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 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.

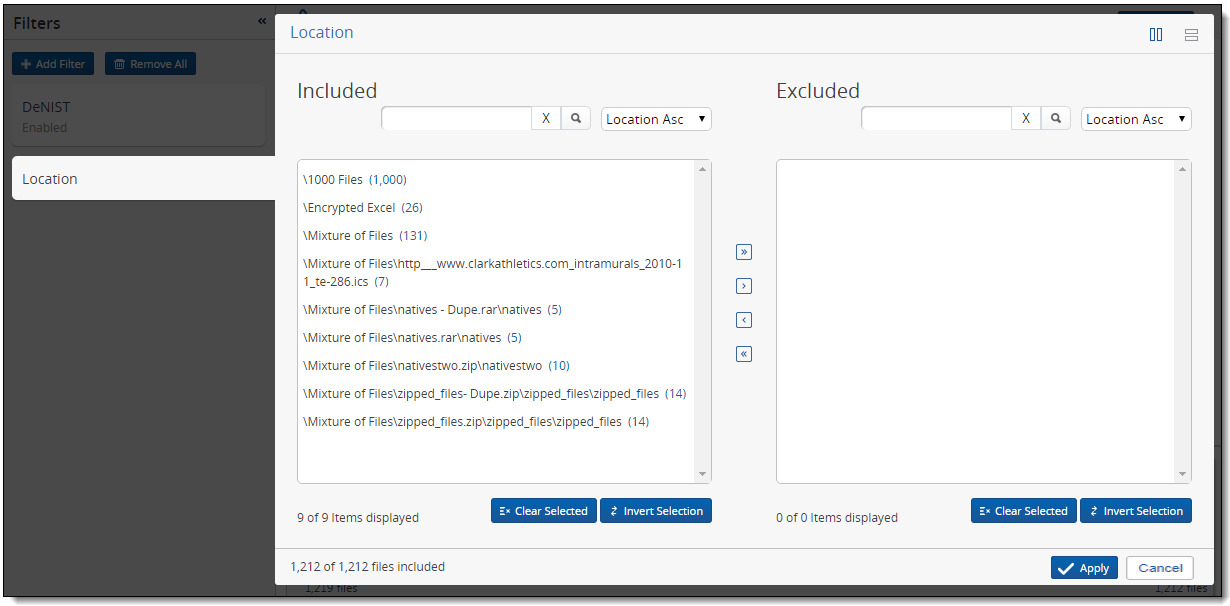


11.2.4 Applying a Location filter

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 a two-list filter on page 175.



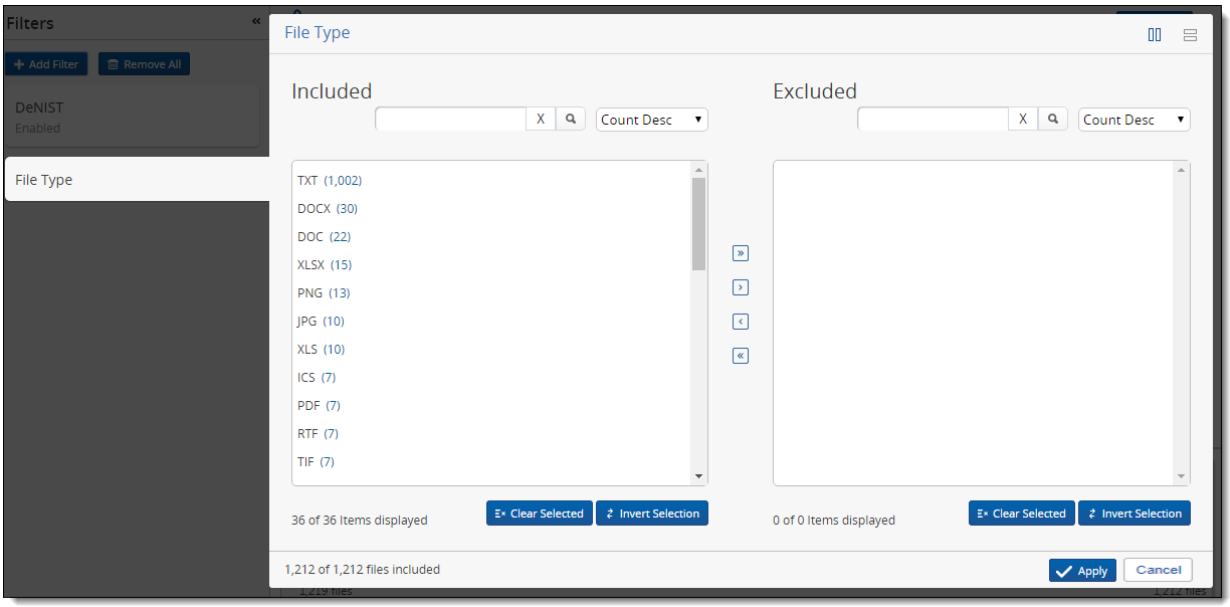
4. Click **Apply** once you've designated all the file locations you want to exclude. The Inventory Progress pane reflects the addition of the location filter, as well as the percentage and number of files that remain from the original data set. For more information, see Inventory progress on page 178.

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

11.2.5 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 a two-list filter on page 175.



4. Click **Apply** once you've 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 on page 178.

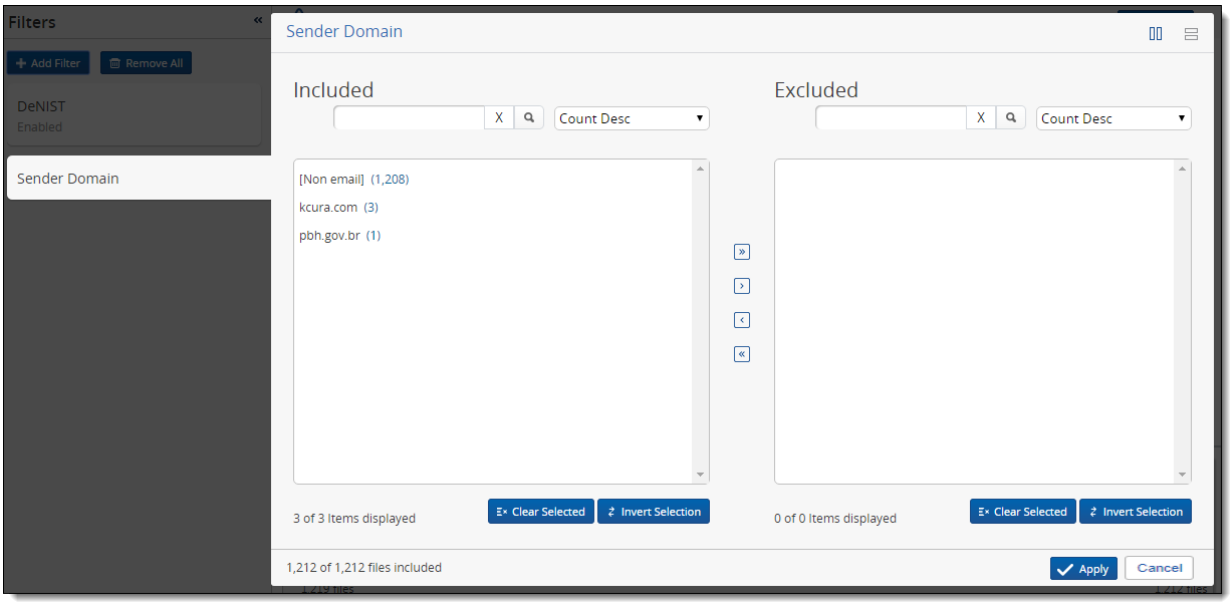
You can now apply an additional filter to further reduce the data set, or you can discover the files.

11.2.6 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 a two-list filter below.



4. Click **Apply** once you've 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 on page 178.

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

11.2.6.1 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 couldn't find a sender domain.
- **[Internal]** - Relativity parsed an LDAP-formatted email address because there was no other valid domain available. When the system can't identify the domain, it attempts to extract the organization or company name from the address.

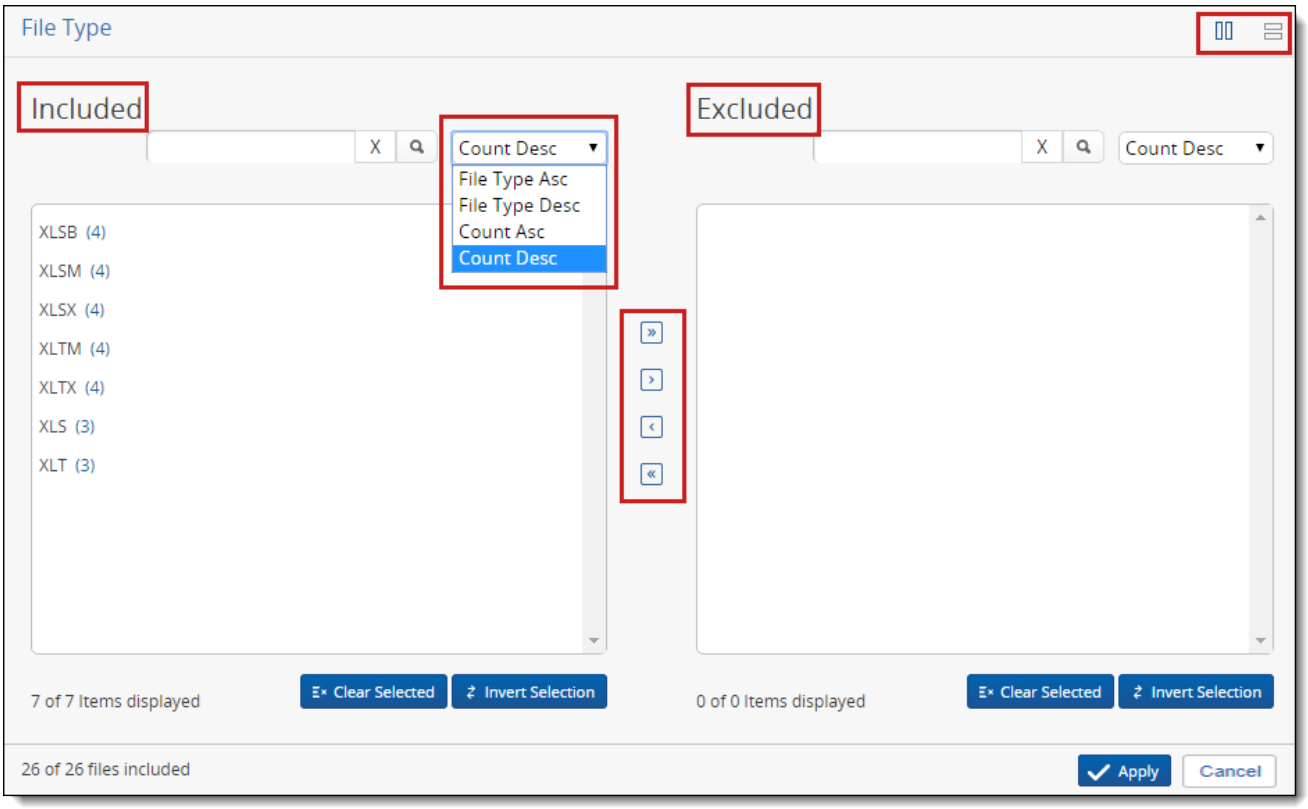
11.2.6.2 Applying a two-list filter

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.



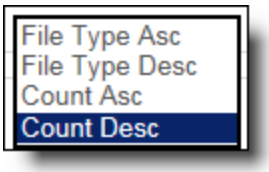
Note: Items removed from the data by edits to a previously applied filter are displayed in later filters with a value of (0) next to them. For example, if you apply the file type filter and then later narrow the date range to the extent that it filters out all files of the PDF type, then the next time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

You can use any of the following options in the two-list filter:

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

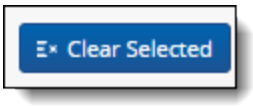
- Toggle the two-list filter to display vertically or horizontally with the parallel line icons in the top right.

  - The vertical lines display all files in the left column, and those designated for exclusion in the right column.
  - The horizontal lines display all files in the top window, and those to be excluded in the bottom window.
- Double-click on any item to move it to the other list.
- Select multiple items in either list by clicking on the items, or select all items between two values in a list with the Shift key.
- Sort the Included and Excluded lists based on the following settings, depending on the filter type:

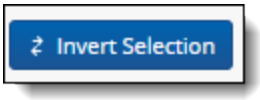


  - **Location Asc** - sorts a list of file locations in alphabetical order.
  - **Location Desc** - sorts a list of file locations in reverse alphabetical order.
  - **Sender Domain Asc** - sorts a list of sender domains in alphabetical order.
  - **Sender Domain Desc** - sorts a list of sender domains in reverse alphabetical order.
  - **File Type Asc** - sorts a list of file types in alphabetical order.
  - **File Type Desc** - sorts a list of file types in reverse alphabetical order.
  - **Count Asc** - sorts a list of variables from the smallest count to the largest.
  - **Count Desc** - sorts a list of variables from the largest count to the smallest.

- **Clear Selected** - marks the previously selected items in the Included or Excluded list as unselected.



- **Invert Selection** - marks the previously selected items in the Included or Excluded list as unselected while selecting the items that weren't selected before.



11.3 Removing filters

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

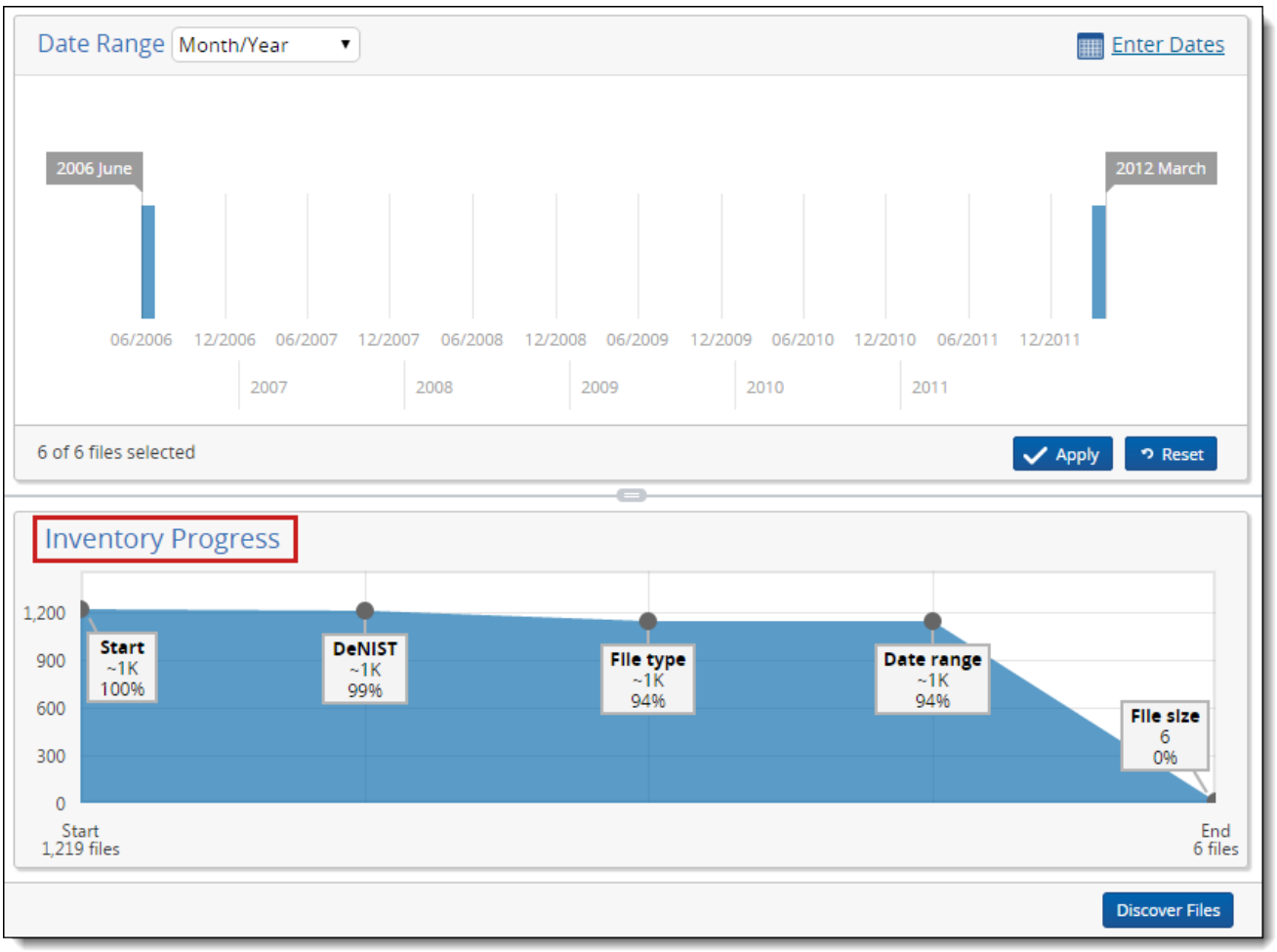
You can also remove filters individually by clicking the **X** on a single filter in the menu. You can't delete a filter if you're currently working with it.

You will be 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

11.4 Inventory progress

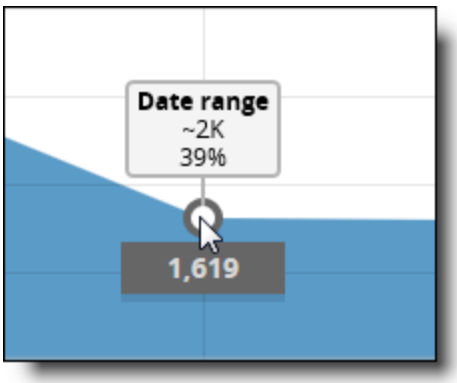
The graph in the Inventory Progress pane reflects all the filters you've applied to the processing set. This graph updates automatically as the inventory job progresses, and provides information on up to six different filters. The vertical axis contains the number of files. The horizontal axis contains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



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've excluded documents by applying filters. This value sits in the bottom right corner of the pane.
- **~#K** - lists the approximate number of files that remain in the data set under the filter type applied.
- **#%** - lists the percentage of files that remain from the data set under the filter type applied. If a filter excludes only a small number of files from the previous file count, this value may not change from the value of the preceding filter type.

You can view the exact number of files that remain in the data set by hovering over the gray dot above or below the file type box.



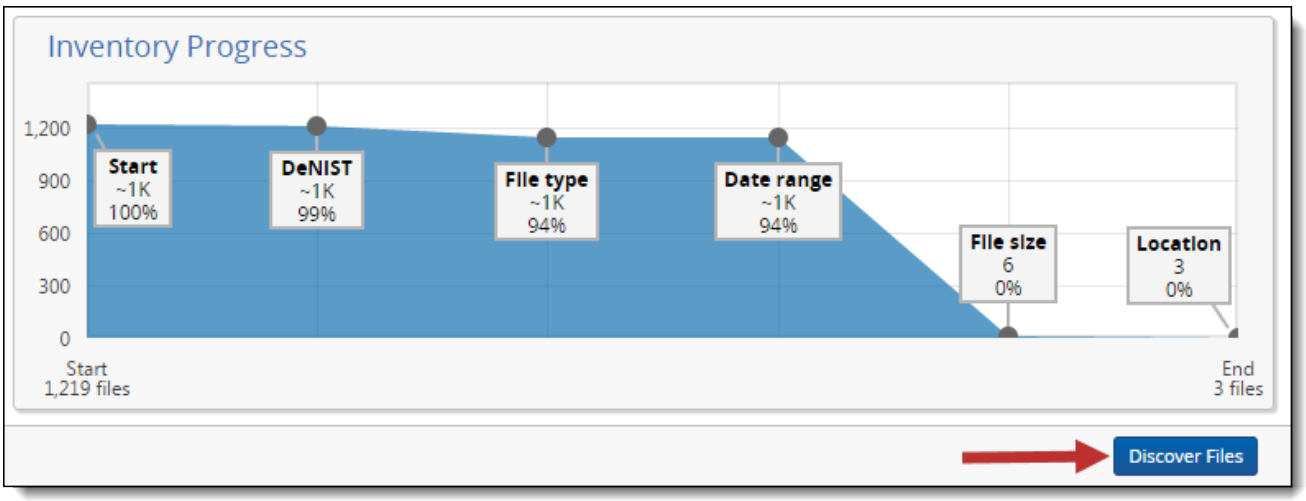
At any time before you discover the files reflected in the Inventory Progress pane, you can reset or delete any filters you already applied.

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

11.5 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 on page 189.

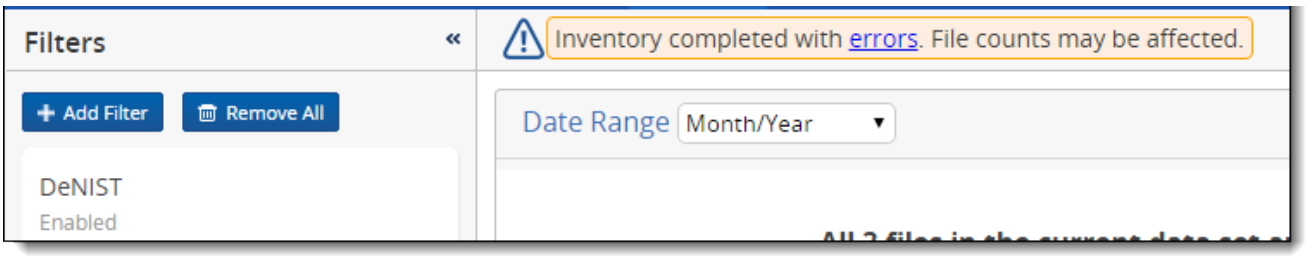


Clicking Discover Files puts the discovery job in the queue and directs you back to the processing set layout, where you can monitor the job's progress.

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

11.6 Inventory errors

If the processing set you select for inventory encountered any errors, the triangle icon appears in the upper left corner of the set. Hover over this icon to access a link to all applicable errors.



Clicking the link to view errors takes you to the Job Errors tab, which contains all errors for all processing sets in the workspace. By default, Relativity applies search conditions to this view to direct you to errors specific to your inventory data. Click any error message in the view to go to that error's details page, where you can view the stack trace and cause of the error.

All inventory errors are unresolvable. If you need to address an error that occurred during inventory, you must do so outside of Relativity and then re-run inventory on the processing set.

See Processing error workflow for details.

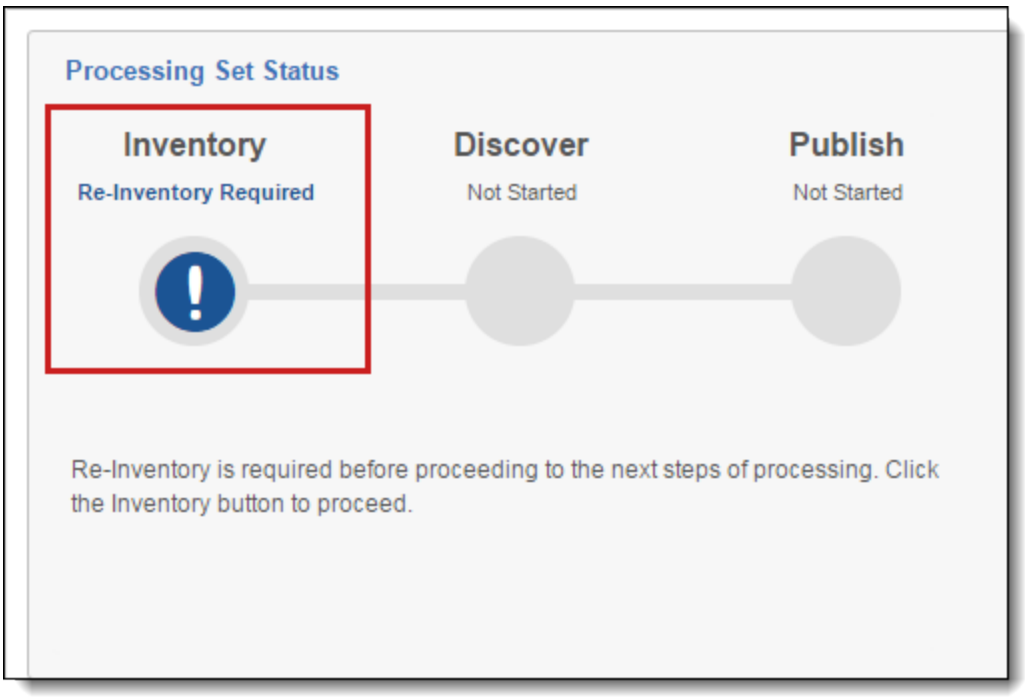
11.6.1 Inventory error scenarios

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.

11.7 Re-inventory

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



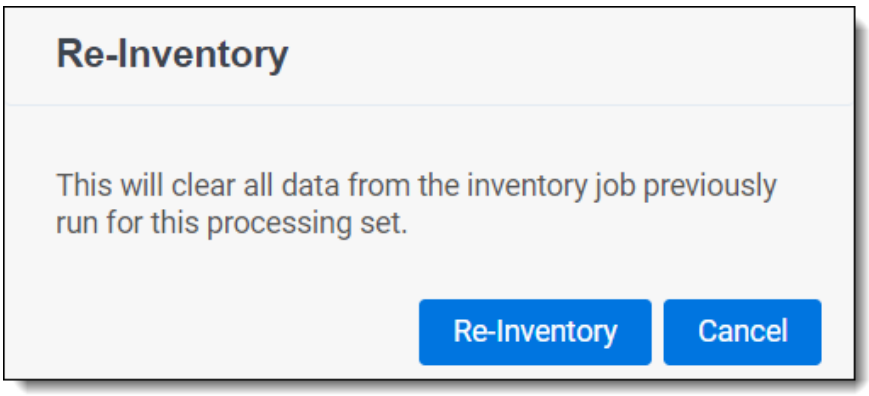
You must run inventory again on a processing set if:

- You've added a data source to processing set that has already been inventoried but not yet discovered.
- You've edited a data source that is associated with a processing set that already been inventoried but not yet discovered.
- You've 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're presented with a confirmation message containing information about the inventory job you're about to submit. Click **Re-Inventory** to proceed with inventory or **Cancel** to return to the processing set layout.



When you re-inventory files:

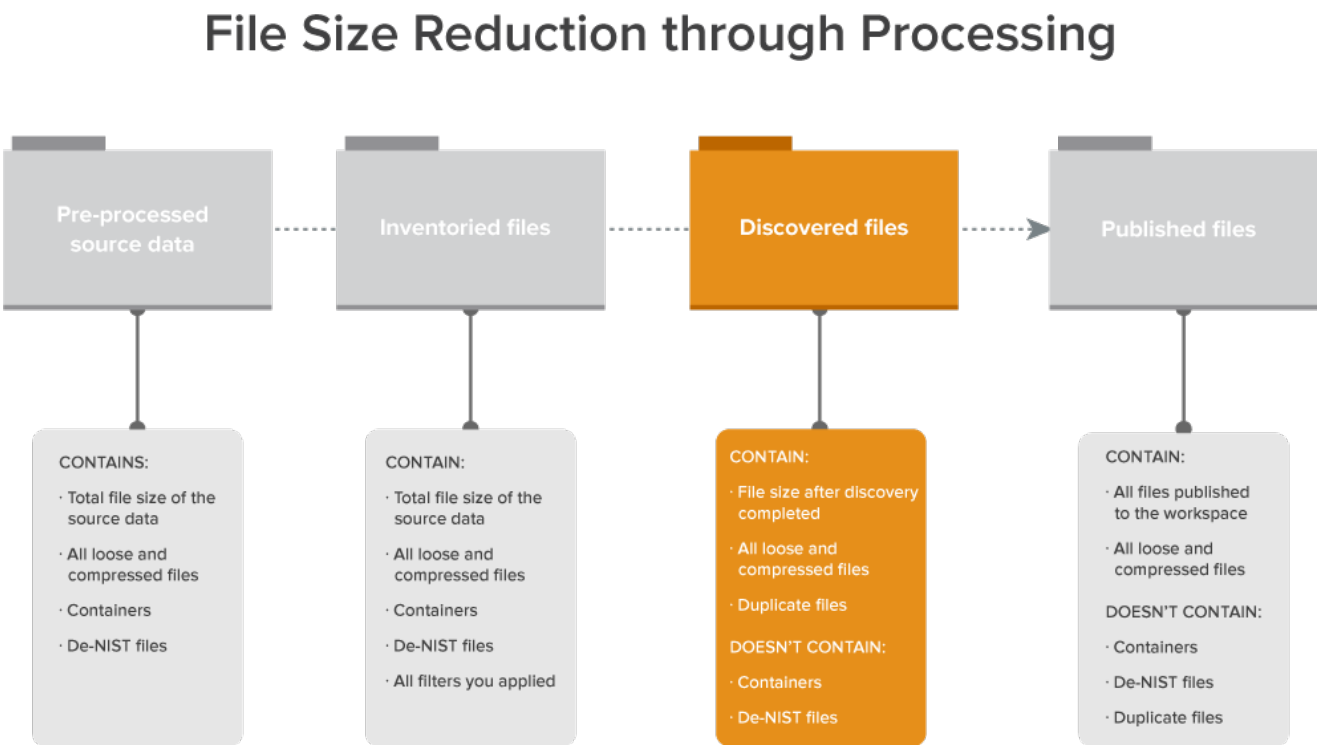
- Filters that you previously applied in the Inventory tab do not get cleared.
- Errors that you encountered in a previous Inventory job are cleared.

12 Discovering files

Discovery is the phase of processing in which the processing engine retrieves deeper levels of metadata not accessible during Inventory and prepares files for publishing to a workspace.

The following graphic depicts how discovery fits into the basic workflow you'd use to reduce the file size of a data set through processing. This workflow assumes that you're applying some method of deNIST and deduplication.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.



The following is a typical workflow that incorporates discovery:

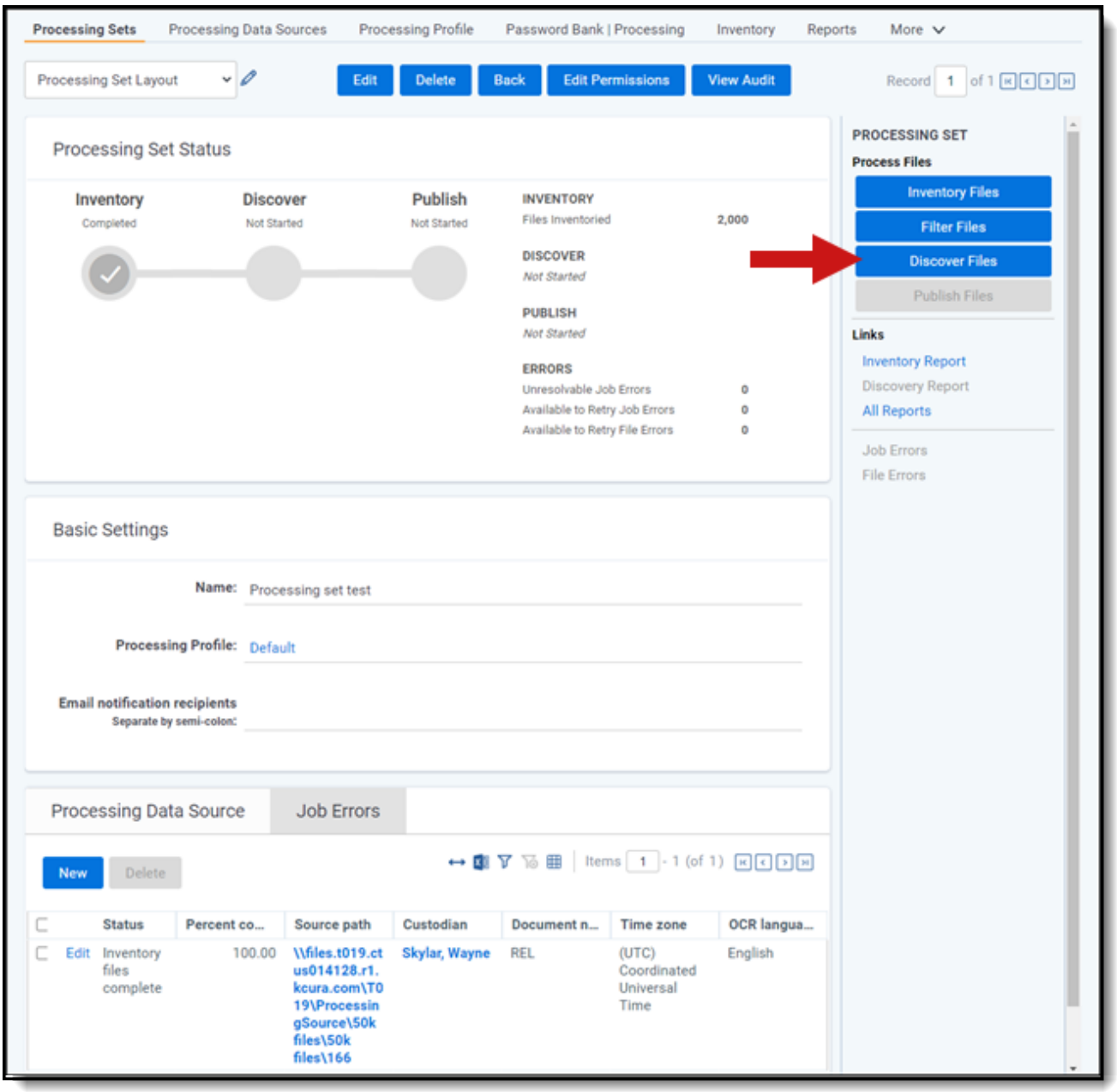
1. Create a processing set or select an existing set.
2. Add data sources to the processing set.
3. Inventory the files in that processing set to extract top-level metadata.
4. Apply filters to the inventoried data.

5. Run discovery on the refined data.
6. Publish the discovered files to the workspace.

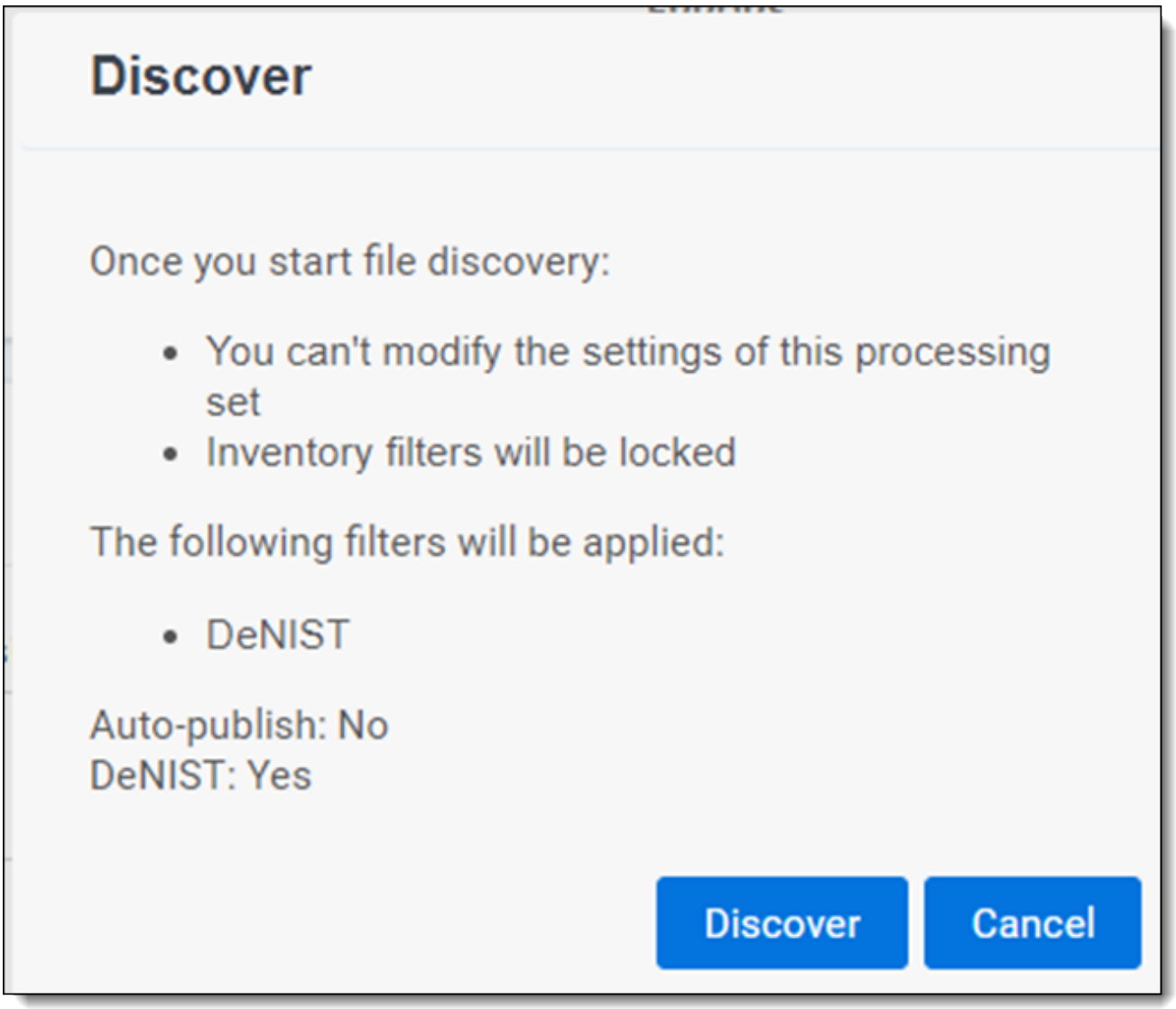
12.1 Running file discovery

To start discovery click **Discover Files** on the processing set console. You can click this whether or not you've inventoried or filtered your files.

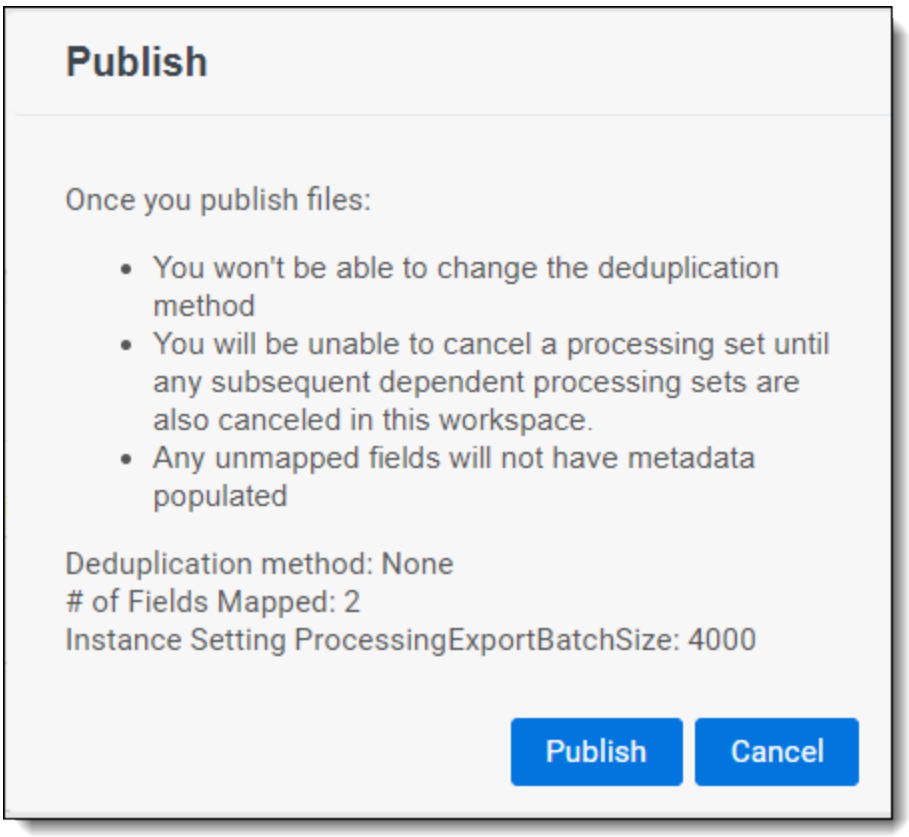
Note: When processing documents without an actual date, Relativity provides a null value for the following fields: Created Date, Created Date/Time, Created Time, Last Accessed Date, Last Accessed Date/Time, Last Accessed Time, Last Modified Date, Last Modified Date/Time, Last Modified Time, and Primary Date/Time. The null value is excluded and not represented in the filtered list.



A confirmation message pops up reminding you of the settings you're about to use to discover the files. Click **Discover** to proceed with discovery or **Cancel** to return to the processing set layout.



If you enabled auto-publish, the confirmation message will provide an option to **Discover & Publish**. Click this to proceed with discovery and publish or **Cancel** to return to the processing set layout.



Note: The default priority for all discovery jobs is determined by the current value of the ProcessingDiscoverJobPriorityDefault entry in the Instance setting table. See the Instance setting guide for more information.

Consider the following when discovering files:

- 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/master document and the original text stored in the database.
- If you've 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'll 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, don't 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 can't start discovery while inventory is running for that processing set.

- When you start discovery or retry discovery for a processing job, the list of passwords specified in the password bank accompanies the processing job so that password-protected files are processed in that job. For more information, see Password bank on page 1.

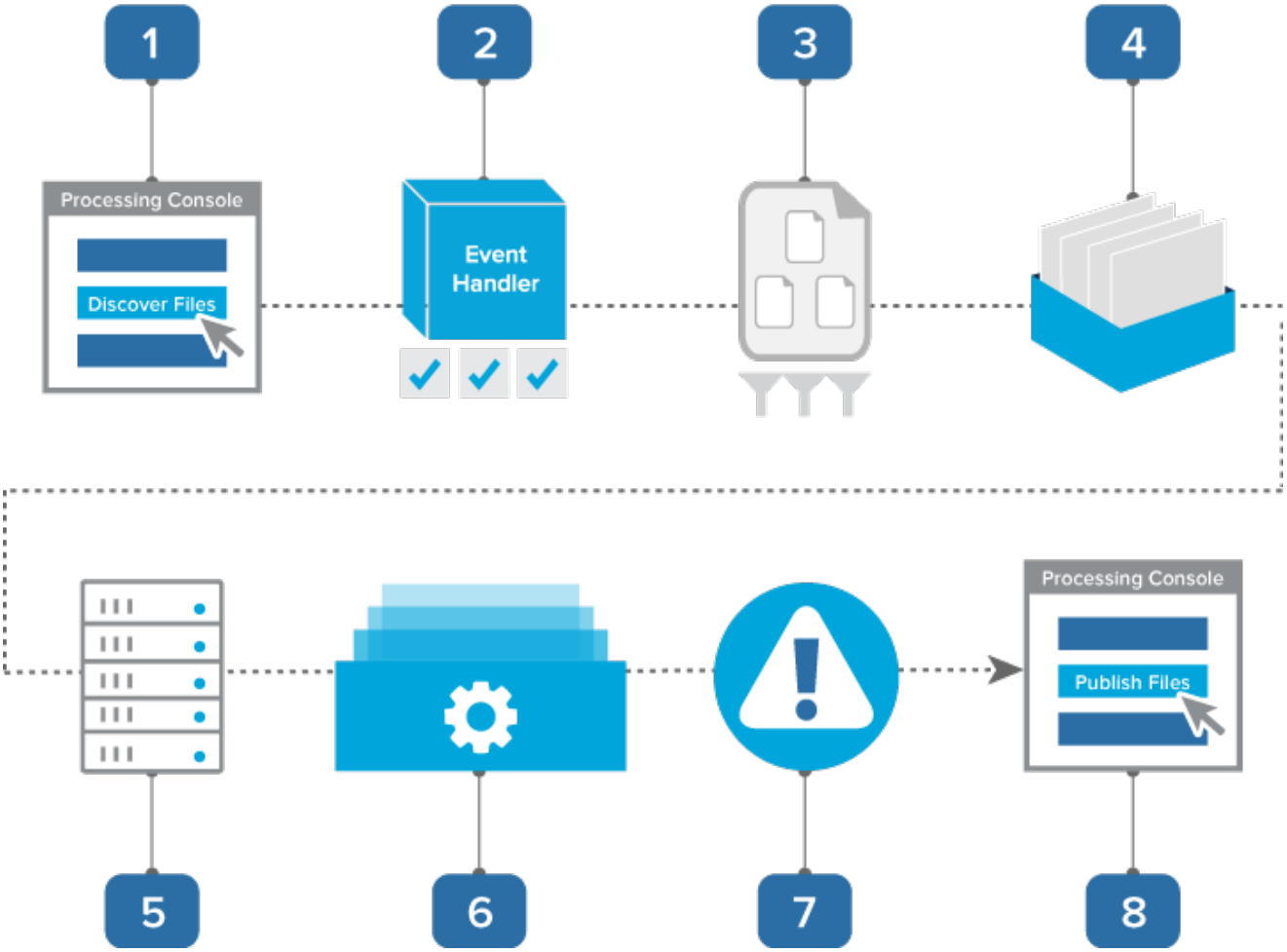
Note: Relativity prioritizes application metadata over operating system file properties where possible. For example, if a file type stores application metadata, such as Date Created and Date Modified, Relativity retains those values for the file. If the application metadata fields are empty or the file type does not store application metadata, Relativity uses the operating system's file properties instead. Application metadata is more reliable since it is stored in the file itself. Operating system file properties can often change. For example, moving a file from one folder to another may change property values. Examples of file types that store application metadata include Microsoft Office files such as Word or Excel.

When you start discovery, the Discover button changes to **Cancel**. Click this to stop discovery. See Canceling discovery for details.

12.1.1 Discovery process

The following graphic and corresponding steps depict what happens behind the scenes when you start discovery. This information is meant for reference purposes only.

Discovery Process



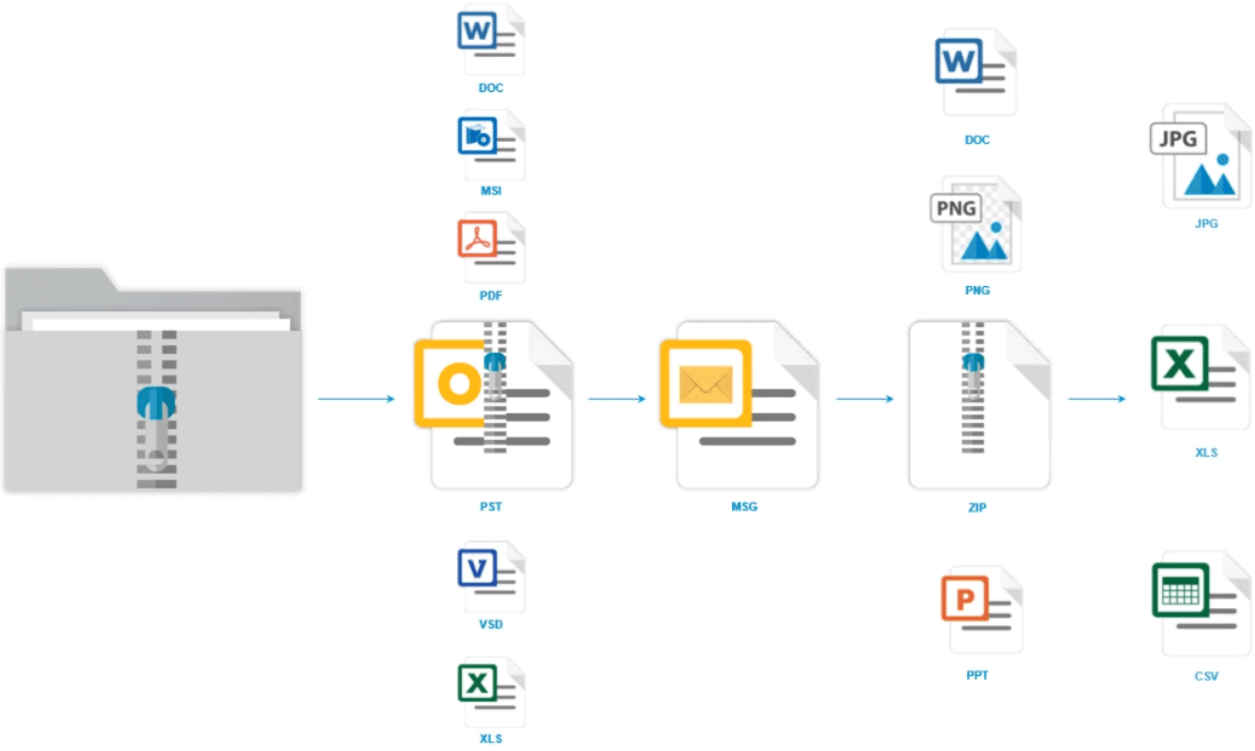
1. You click Discover Files on the processing set console.
2. A console event handler copies all settings from the processing profile to the data sources on the processing set and then checks to make sure that the set is valid and ready to proceed.
3. The event handler inserts all data sources into the processing set queue .
4. The data sources wait in the queue to be picked up by an agent, during which time you can change their priority.
5. The processing set manager agent picks up each data source based on its order, all password bank entries are synced, and the agent submits each data source as an individual discovery job to the processing engine. The agent then provides updates on the status of each job to Relativity, which then displays this information on the processing set layout.
6. The processing engine discovers the files and applies the filters you specified in the Inventory tab. It then sends the finalized discovery results back to Relativity, which then updates the reports to include all applicable discovery data.

7. Any errors that occurred during discovery are logged in the errors tabs. You can view these errors and attempt to retry them. See Processing error workflow for details.For more information, see Processing error workflow.
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.

12.1.2 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.



12.2 Special considerations - OCR and text extraction

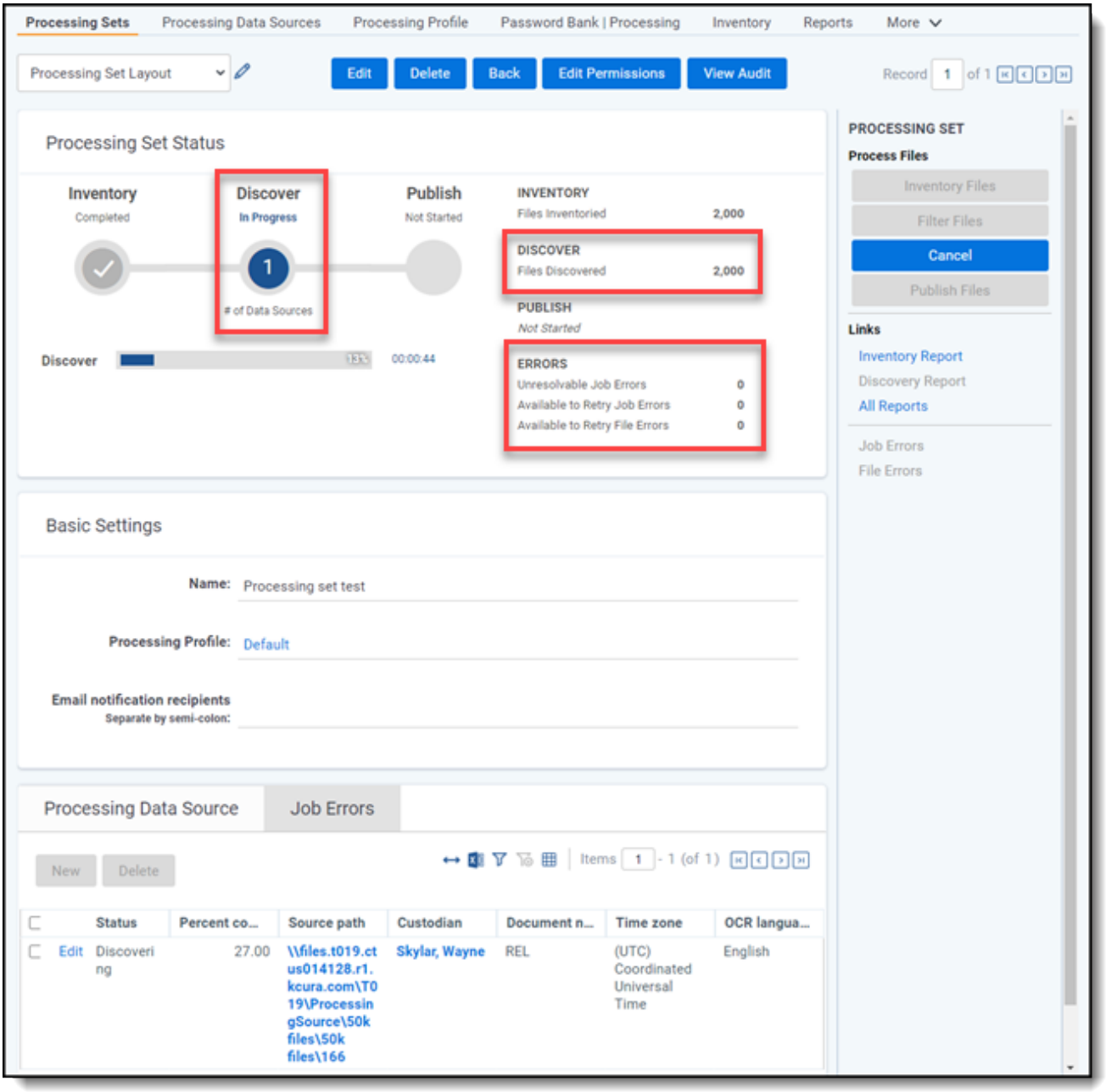
Consider the following regarding OCR and text extraction during discovery:

- 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 aren't 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.

12.3 Monitoring discovery status

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

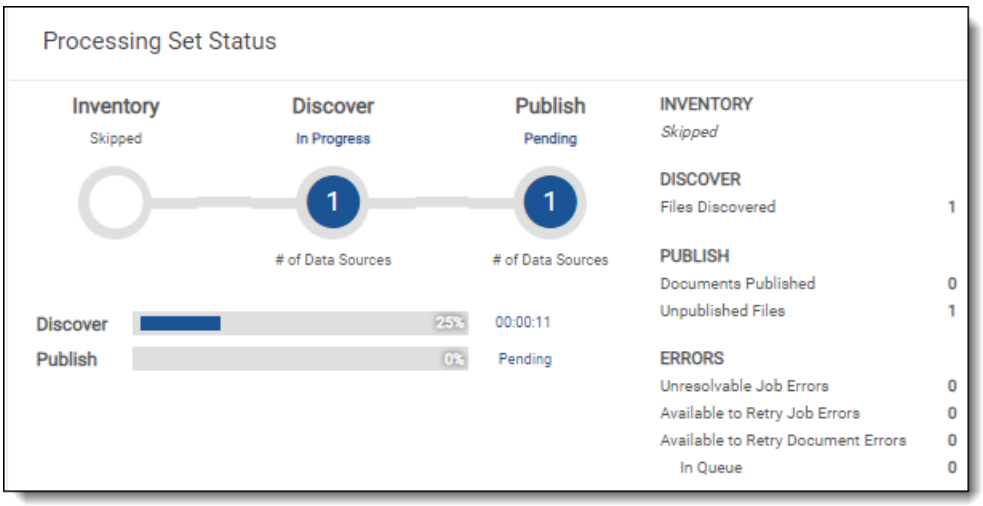


Through this display, you can monitor the following:

- **# 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 inventoried.
- **Inventory | Filtered Inventory** - the number of files you excluded from discovery by applying any of the available filters in the Inventory tab. For example, if you applied only a Date Range filter and excluded only 10 .exe files from your data after you inventoried it, this will display a value of 10. If you applied no filters in the Inventory tab, this value will be 0. This value doesn't include files that were excluded via the DeNIST setting on the processing profile associated with this set.
- **Discover | Files Discovered** - the number of files across all data sources submitted that the processing engine has discovered.
- **Discover | Files with Extracted Text** - the number of files across all data sources submitted that have had their text extracted. This value will only be displayed while the discovery jobs are still in progress. If the value is 0, text extraction has not started yet.
- **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.

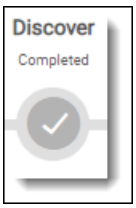
Note: Overall progress is calculated based on the number of data sources and percentage complete for each source.

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.



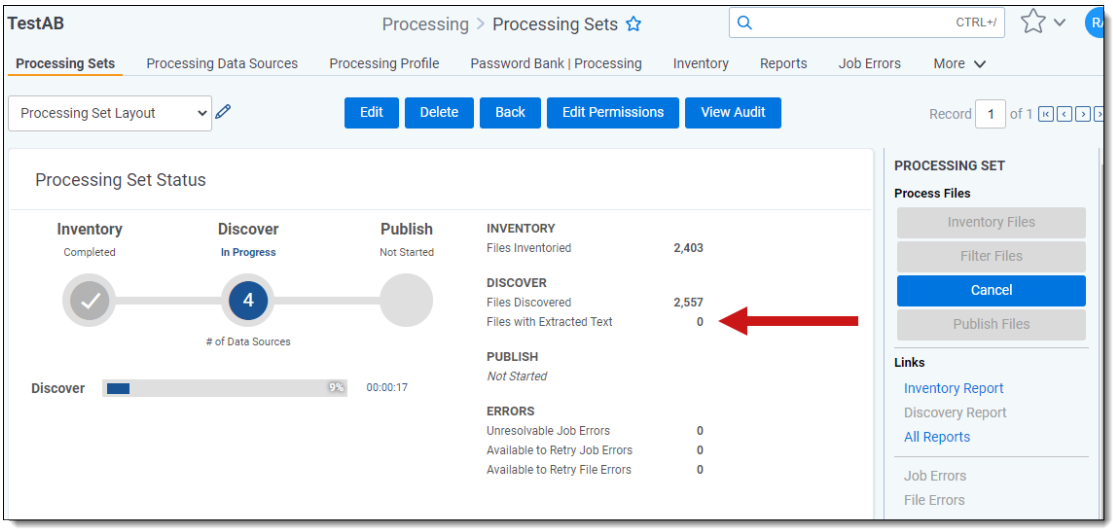
See Processing error workflow for details.

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

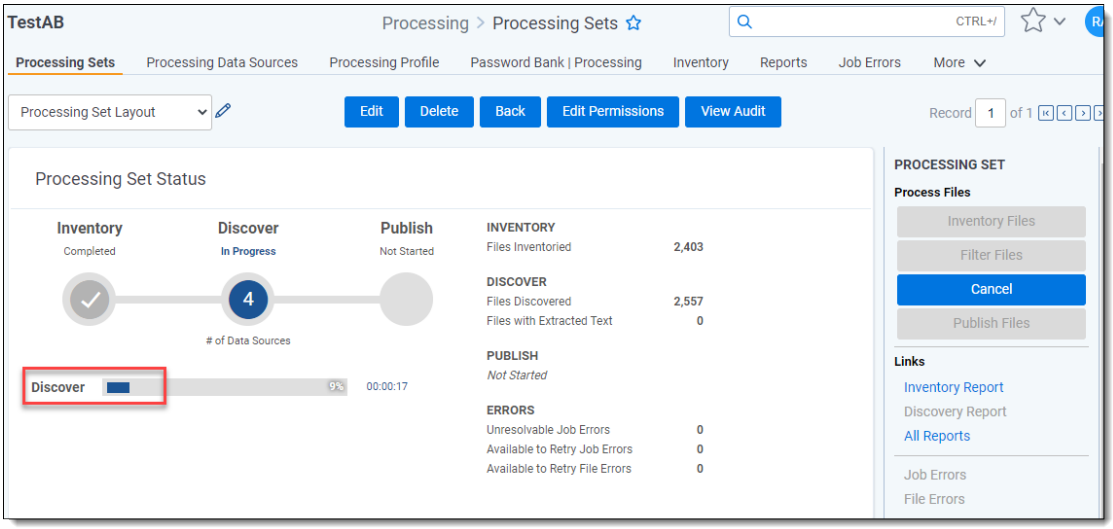


12.4 Viewing text extraction progress in processing sets

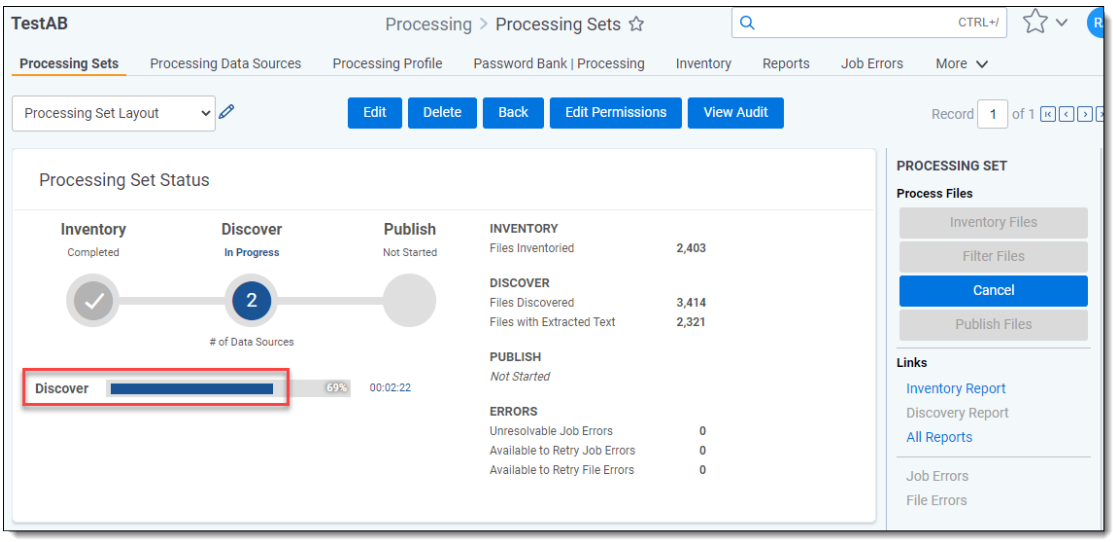
This feature is located below the **Discover | Files Discovered** count and shows the progress of text extraction by displaying an incrementing count of the number of files containing text within the processing sets.



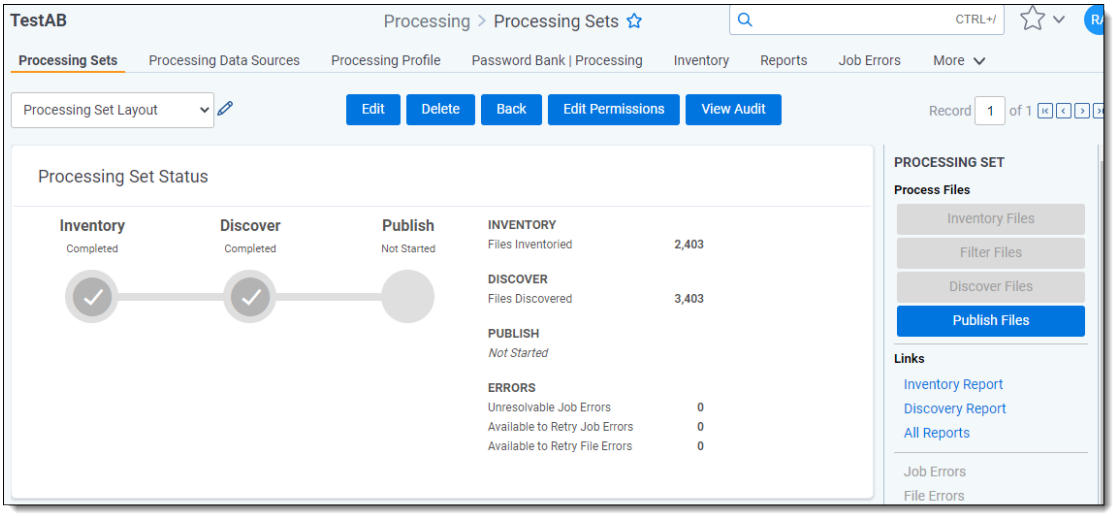
Initial Discovery occurs when the percentage bar displays 0-25%.



Text Extraction occurs when the percentage bar displays 25-50%.



Finalization occurs at a much faster rate (50-100%).

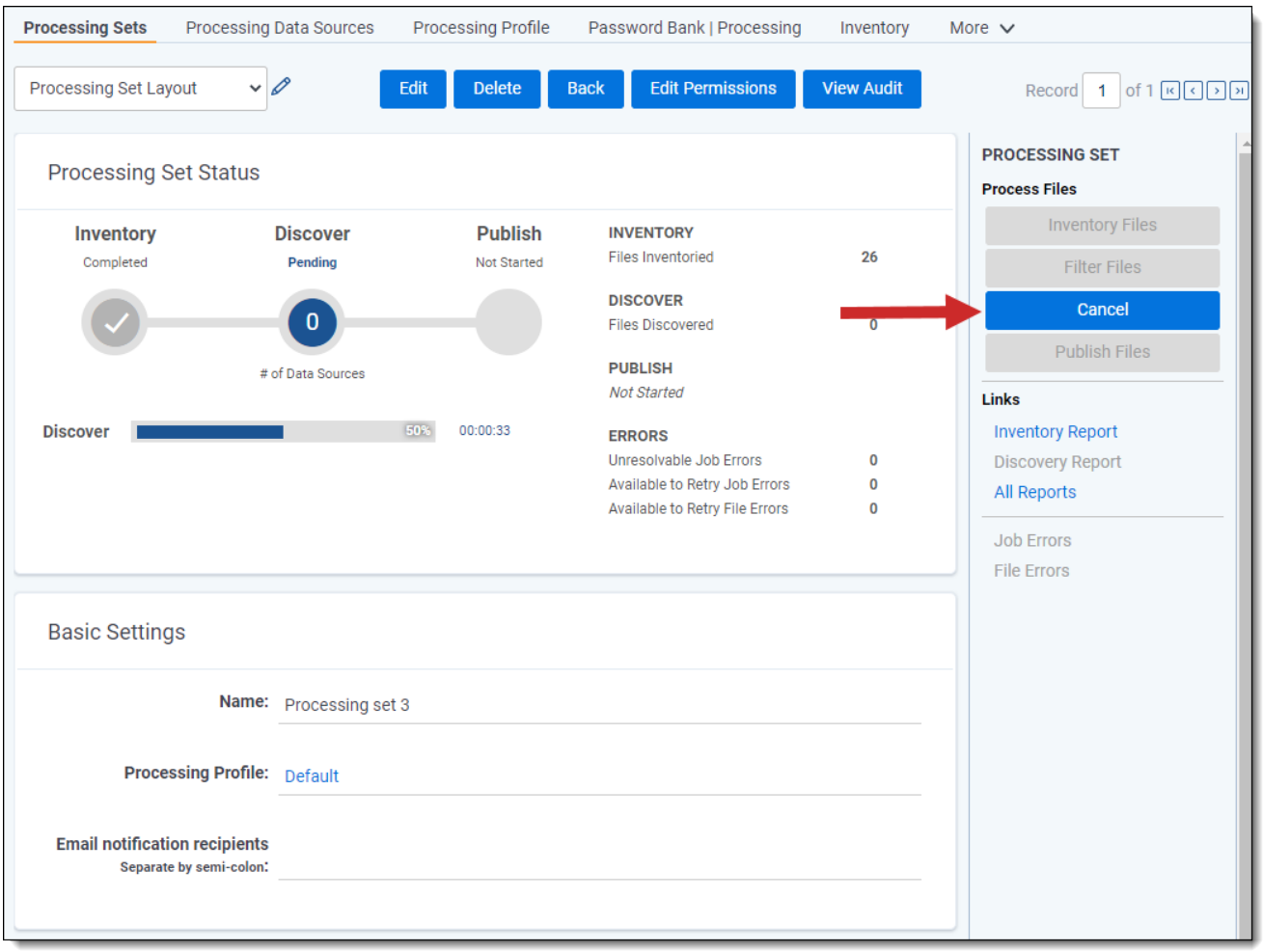


12.5 Canceling discovery

12.5.1 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**.



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 can't be retried.
- Once you cancel discovery, you can't resume discovery on those data sources. You must create new data sources to fully discover those files.

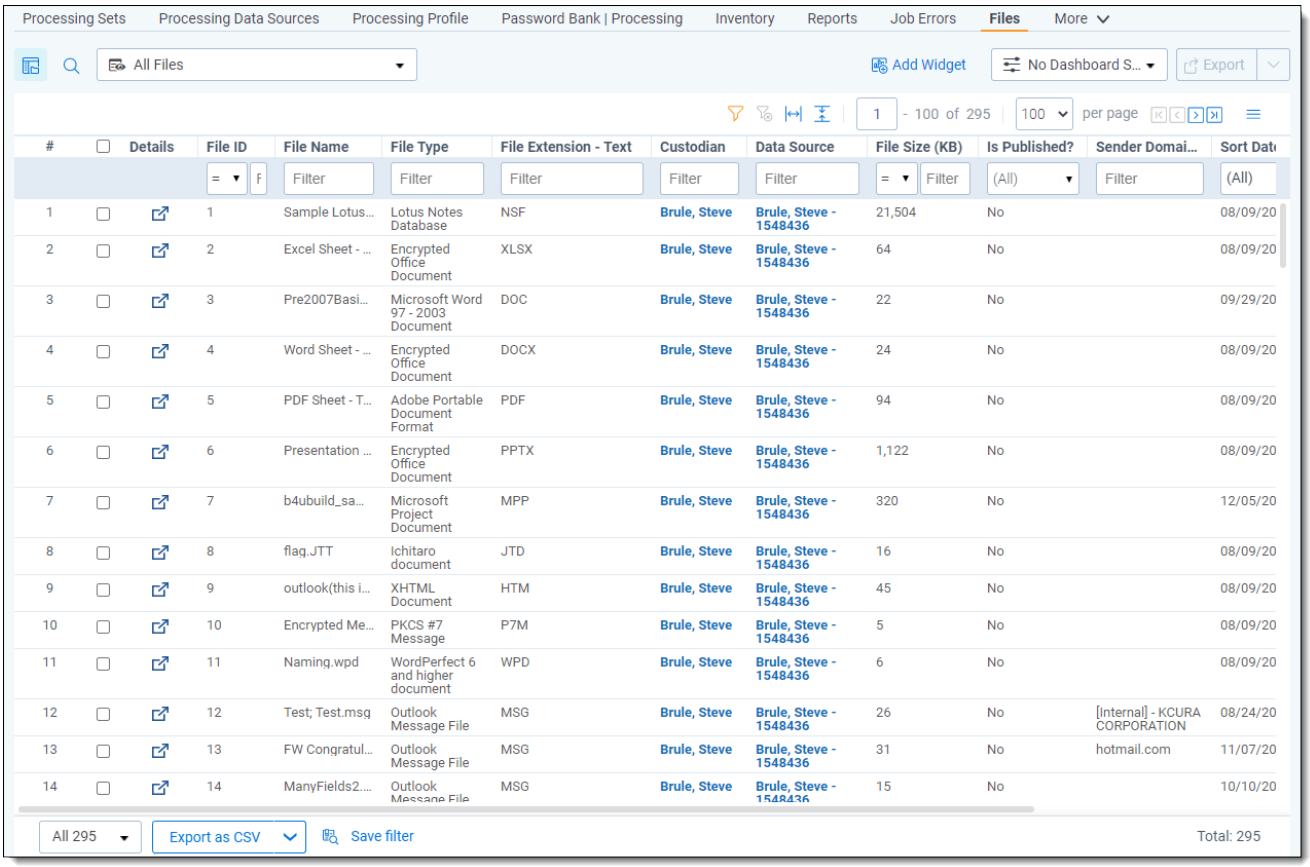
Once you cancel discovery, the status section is updated to display the canceled state.

Processing Set Status
Inventory
Skipped
Discover
Canceled
Publish
Canceled
This processing set has been canceled and can no longer be processed.
INVENTORY
Skipped
DISCOVER
Files Discovered 1
PUBLISH
Documents Published 0
Unpublished Files 1
ERRORS
Unresolvable Job Errors 0
Available to Retry Job Errors 0
Available to Retry Document Errors 0
In Queue 0

13 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.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.



#	Details	File ID	File Name	File Type	File Extension - Text	Custodian	Data Source	File Size (KB)	Is Published?	Sender Domai...	Sort Dat
1		1	Sample Lotus...	Lotus Notes Database	NSF	Brule, Steve	Brule, Steve - 1548436	21,504	No		08/09/20
2		2	Excel Sheet - ...	Encrypted Office Document	XLSX	Brule, Steve	Brule, Steve - 1548436	64	No		08/09/20
3		3	Pre2007Basi...	Microsoft Word 97 - 2003 Document	DOC	Brule, Steve	Brule, Steve - 1548436	22	No		09/29/20
4		4	Word Sheet - ...	Encrypted Office Document	DOCX	Brule, Steve	Brule, Steve - 1548436	24	No		08/09/20
5		5	PDF Sheet - T...	Adobe Portable Document Format	PDF	Brule, Steve	Brule, Steve - 1548436	94	No		08/09/20
6		6	Presentation ...	Encrypted Office Document	PPTX	Brule, Steve	Brule, Steve - 1548436	1,122	No		08/09/20
7		7	b4ubuild_sa...	Microsoft Project Document	MPP	Brule, Steve	Brule, Steve - 1548436	320	No		12/05/20
8		8	flag.JTT	Ichitaro document	JTD	Brule, Steve	Brule, Steve - 1548436	16	No		08/09/20
9		9	outlook(this i...	XHTML Document	HTM	Brule, Steve	Brule, Steve - 1548436	45	No		08/09/20
10		10	Encrypted Me...	PKCS #7 Message	P7M	Brule, Steve	Brule, Steve - 1548436	5	No		08/09/20
11		11	Naming.wpd	WordPerfect 6 and higher document	WPD	Brule, Steve	Brule, Steve - 1548436	6	No		08/09/20
12		12	Test; Test.msg	Outlook Message File	MSG	Brule, Steve	Brule, Steve - 1548436	26	No	[Internal] - KCURA CORPORATION	08/24/20
13		13	FW Congratul...	Outlook Message File	MSG	Brule, Steve	Brule, Steve - 1548436	31	No	hotmail.com	11/07/20
14		14	ManyFields2....	Outlook Message File	MSG	Brule, Steve	Brule, Steve - 1548436	15	No		10/10/20

13.1 Views on the Files tab

The Files tab contains the following views:

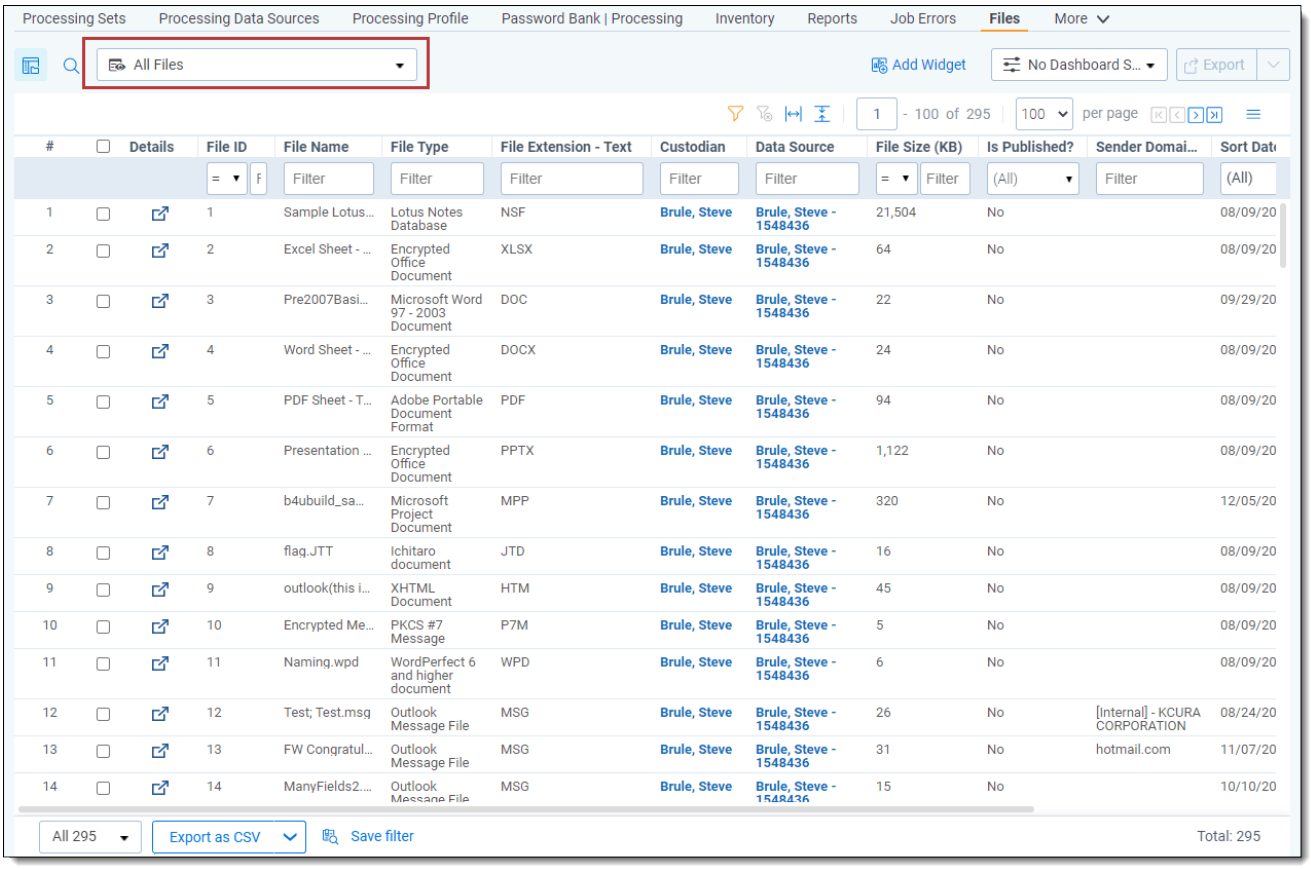
- **All Files** - contains all the files in your workspace.
- **Current Errored Files** - contains all the documents that yielded errors in your workspace that currently have an Error Status value of Not Resolved.

- **All Errored Files** - contains all the documents that yielded errors in your workspace, including those with a current Error Status value of Resolved and Unresolved.
- **Deleted Documents** - contains all the documents you deleted from your workspace.

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

13.1.1 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.



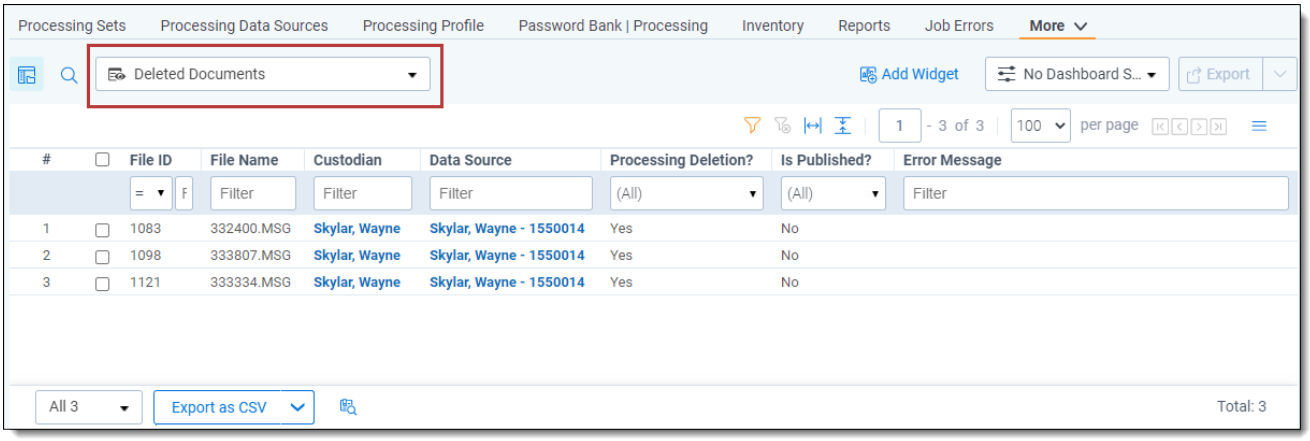
This view contains the following fields:

- **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.

- **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 indicating 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.

13.1.2 Deleted Documents view

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

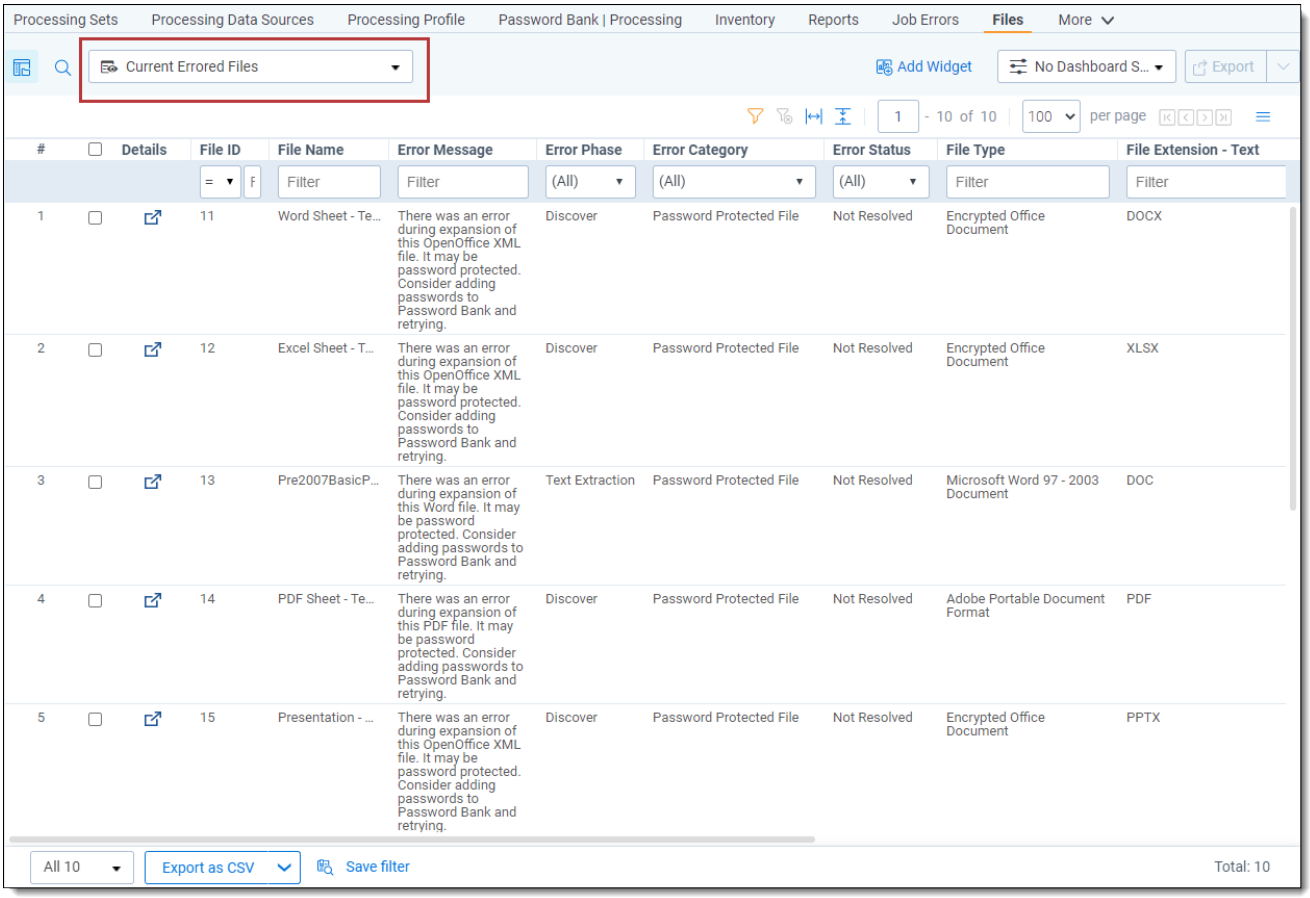


This view contains the following fields:

- **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

13.1.3 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.



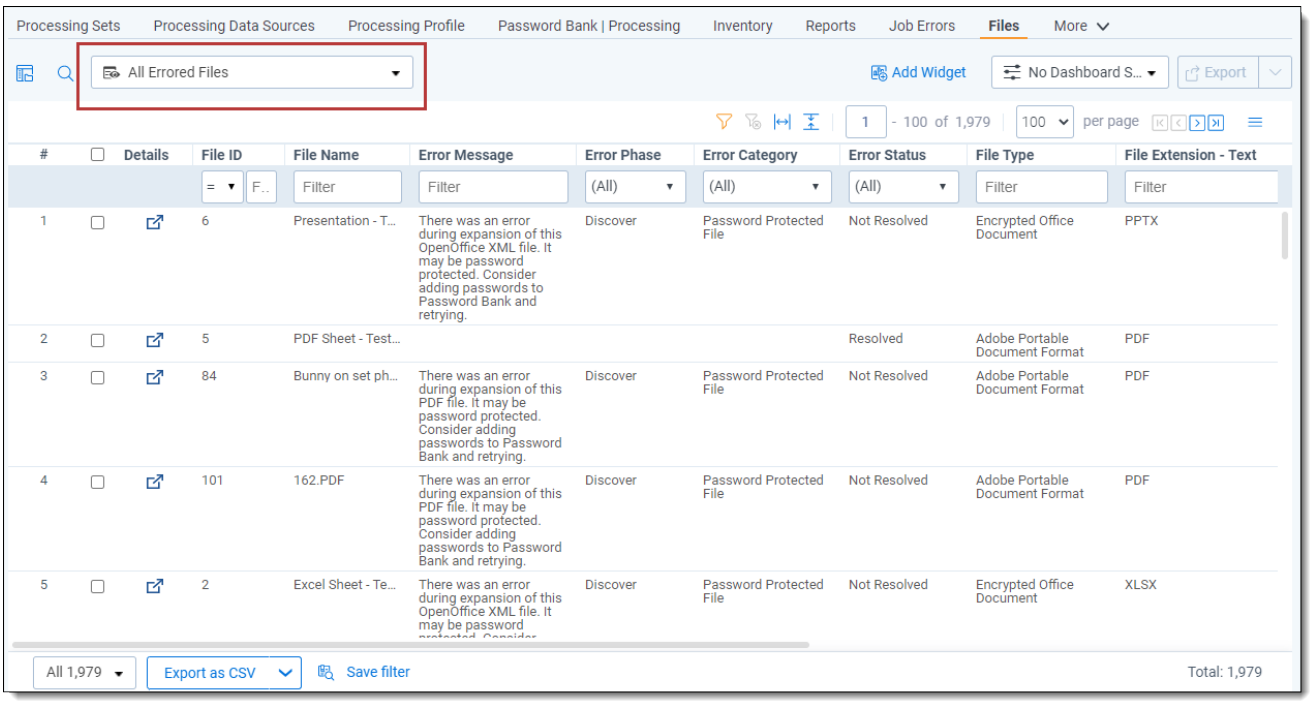
#	Details	File ID	File Name	Error Message	Error Phase	Error Category	Error Status	File Type	File Extension - Text
1		11	Word Sheet - Te...	There was an error during expansion of this OpenOffice XML file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved	Encrypted Office Document	DOCX
2		12	Excel Sheet - T...	There was an error during expansion of this OpenOffice XML file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved	Encrypted Office Document	XLSX
3		13	Pre2007BasicP...	There was an error during expansion of this Word file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Text Extraction	Password Protected File	Not Resolved	Microsoft Word 97 - 2003 Document	DOC
4		14	PDF Sheet - Te...	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
5		15	Presentation - ...	There was an error during expansion of this OpenOffice XML file. It may be password protected. Consider adding passwords to Password Bank and retrying.	Discover	Password Protected File	Not Resolved	Encrypted Office Document	PPTX

- **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.
- **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 Category** - provides insight into the nature of the errors that have occurred on your processed files. For details, see Error category list.
- **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.

13.1.4 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.



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.
- **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.

13.2 Details modal

You can open the Details modal of a file by clicking to see uncompressed file and content metadata not visible by default in the Files view.

The Details modal provides you with supplemental information about errors that have occurred on records during discovery and publish.

Details Processing Errors

File Data

File Name	Presentation - Test1234!.pptx
File ID	6
Custodian	Brule, Steve
Data Source	Brule, Steve - 1548436
Container Extension	
Container ID	0
Container Name	
Contains Embedded Items?	No
Error Message	There was an error during expansion of this OpenOffice XML file. It may be password protected. Consider adding passwords to Password Bank and retrying.
File Extension - Text	PPTX
File Size (KB)	1,122
File Type	Encrypted Office Document
Folder Path	\CloudSet\Password Protected files
Is Container?	No
Is Embedded?	No
Is Published?	No
Processing Deletion?	No
Import Source	\\files.t019.ctus014128.r1.kcura.com\T019\ProcessingSource\CloudSet\Password Protected files\Presentation - Test1234!.pptx
Discovery Group ID	39039462

Close

You can also see a summary and history of all processing errors and retries in this modal. When you click the **Processing Errors** tab, you're presented with the following breakdown of the current errors and error history of the selected file:

Details **Processing Errors**

Error Summary

Current Errors	2
Discover Error(s)	Password Protected File (1)
Text Extraction Error(s)	Password Protected File (1)
Publish Error(s)	none
Delete Error(s)	none

Error History

Error total: 4

Password Protected File (Text Extraction) - Not Resolved
Mon, 29 Mar 2021 16:50:00 UTC
There was an error during expansion of this PowerPoint file. It may be password protected. Consider adding passwords to Password Bank and retrying.

Password Protected File (Discover) - Not Resolved
Mon, 29 Mar 2021 16:49:48 UTC
There was an error during expansion of this OpenOffice XML file. It may be password protected. Consider adding passwords to Password Bank and retrying.

Password Protected File (Text Extraction) - Retried
Fri, 26 Mar 2021 18:59:08 UTC
There was an error during expansion of this PowerPoint file. It may be password protected. Consider adding passwords to Password Bank and retrying.

Password Protected File (Discover) - Retried
Fri, 26 Mar 2021 18:58:27 UTC

Close

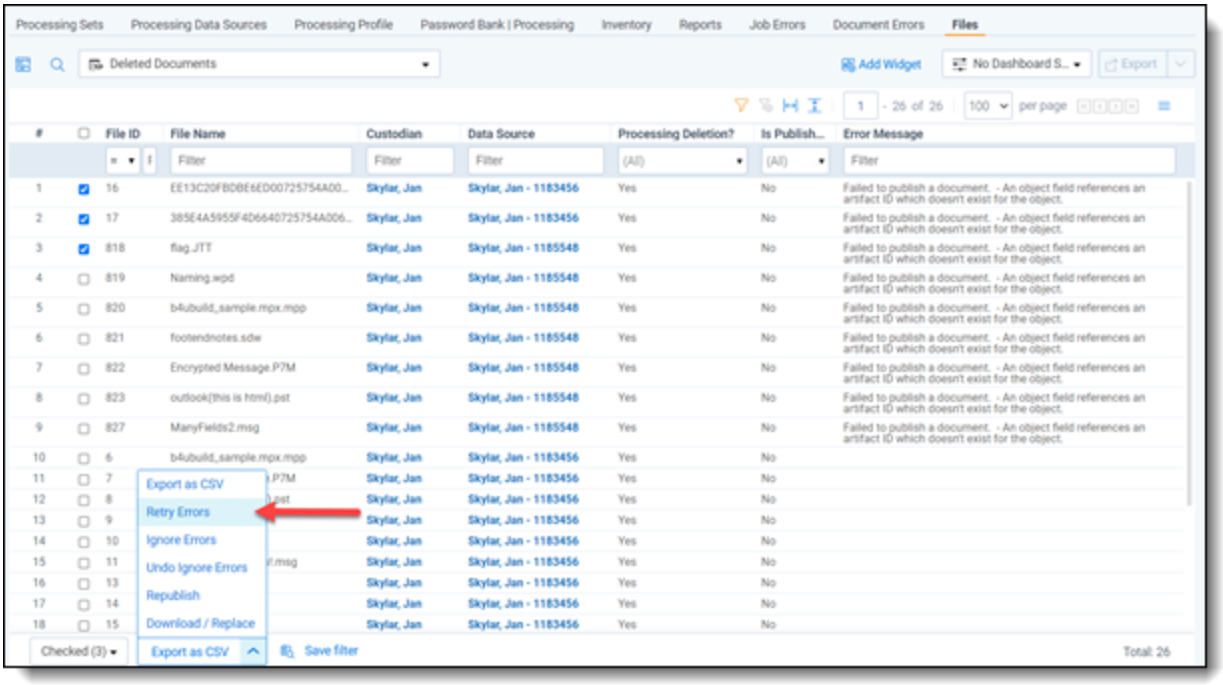
- The Error History section represents all errors that have ever occurred on a file. This acts as a timeline of the record's errors, showing when they occurred, what they were about, and if any are still

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.

13.3 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. These errors can be retried using the **Retry Delete** mass operations option. Once deleted, these documents will be excluded from further processing operations (e.g., deduplication, retry, and republish) and the next duplicate will be published as the new master document, if available. To see a summary of master documents that have been replaced, see the Master Document Replacement Summary report in Processing Reports. See Mass Delete for more information on deleting documents.



13.4 Republishing files from the Files tab

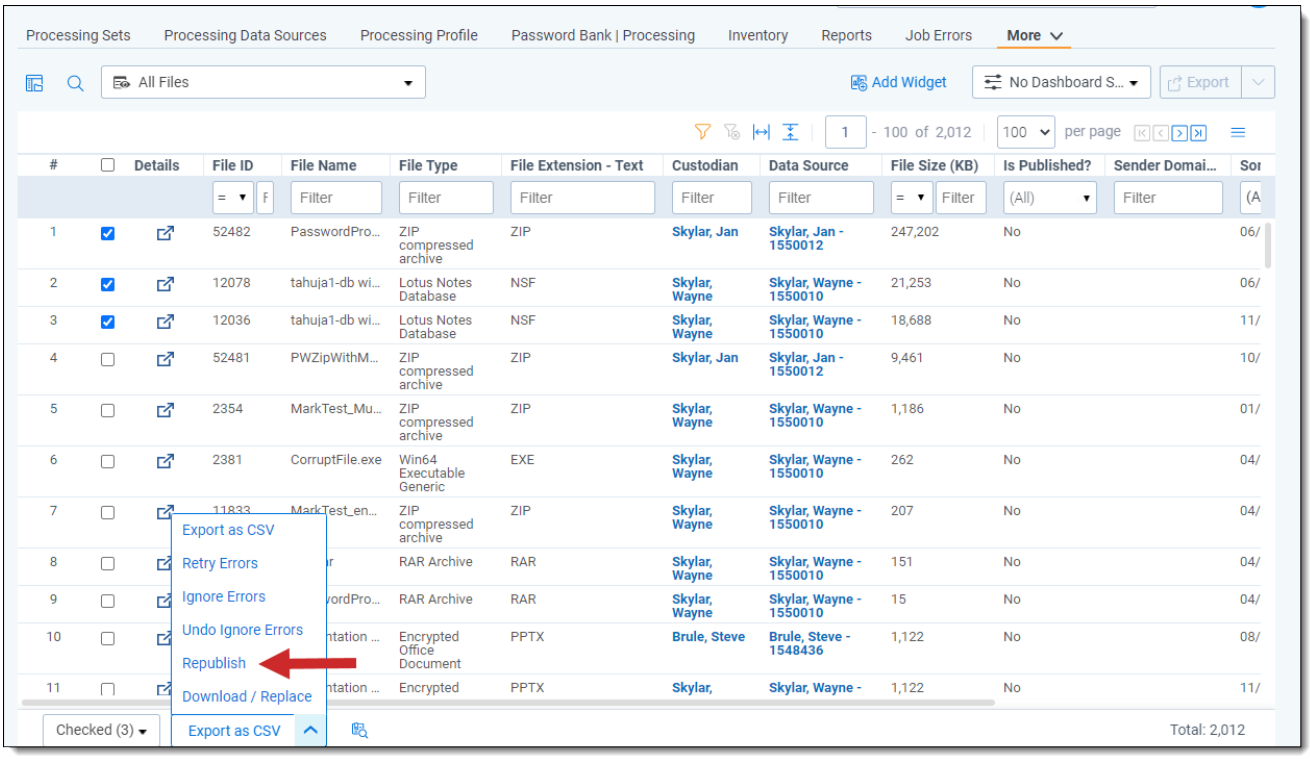
Note: The following information is specific to republishing files at the file level via the Files tab and is distinct from republishing files via the processing set console. For details on republishing via the processing set console, see Republishing a processing set on page 228 on the Publishing files 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.

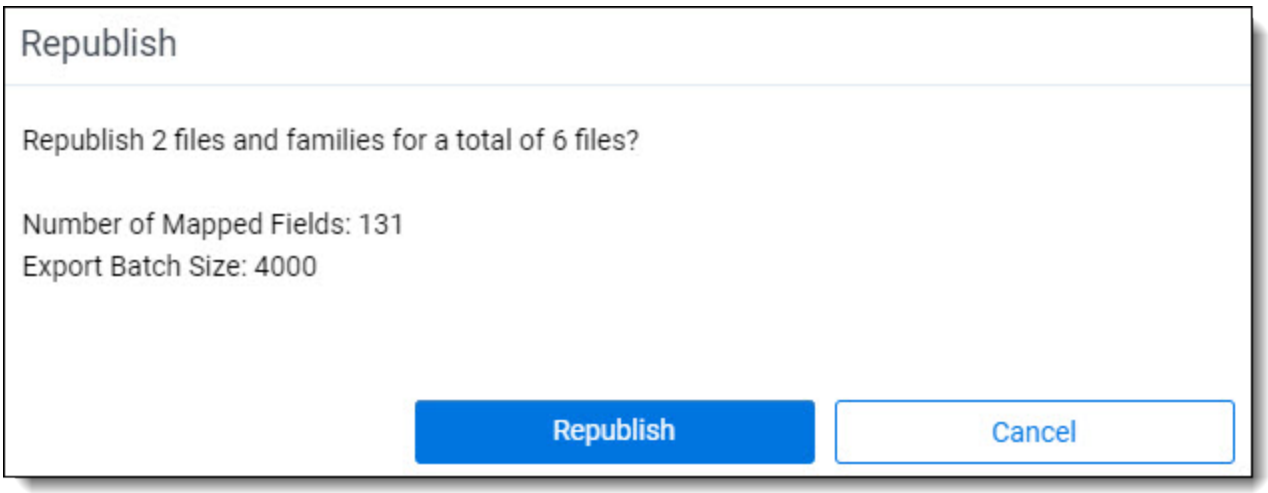
Navigate to the All Files view to see a record of all discovered files and filter to published files via the Is Published? field published. These files can be republished at the file level using the **Republish** mass operation.



When you click **Republish**, you're presented with a confirmation modal containing the following information about the job you're 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've selected files ineligible for republish, the confirmation message will reflect this by stating that there are 0 files to be republished. Ineligible files include files from unpublished processing sets or data sources, containers (e.g., PST, ZIP), duplicate files, and files where the Processing Deletion? field status is Yes.

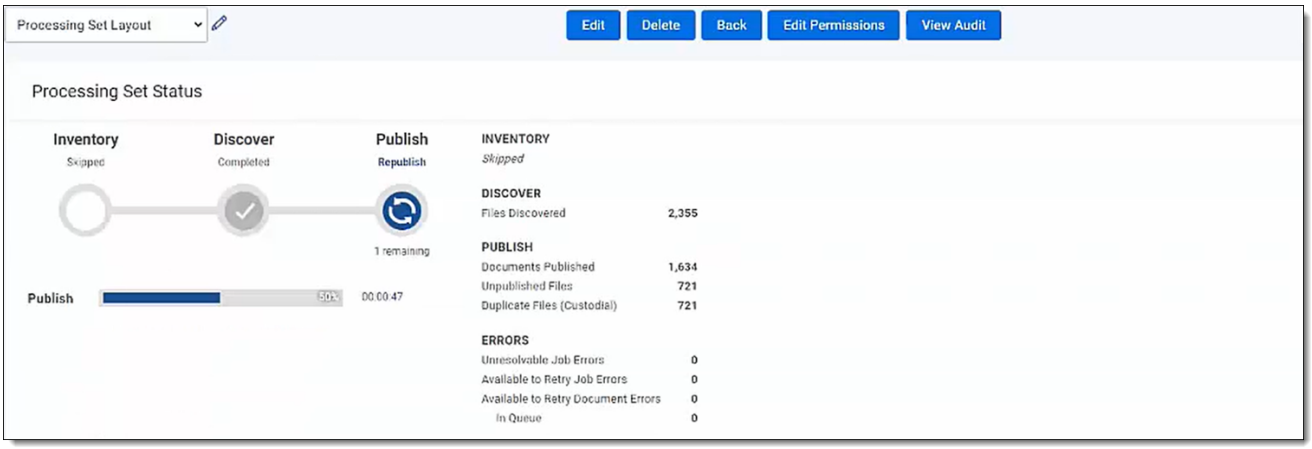


If you've selected eligible files, click **Republish** to proceed or **Cancel** to return to the All Files view.

Note: Once you republish, you will be unable to cancel this job.

To monitor the republish job, check the Status field in the Processing Data Sources tab and Worker Monitoring and Processing Sets tabs.

(Click to expand)



13.4.1 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:

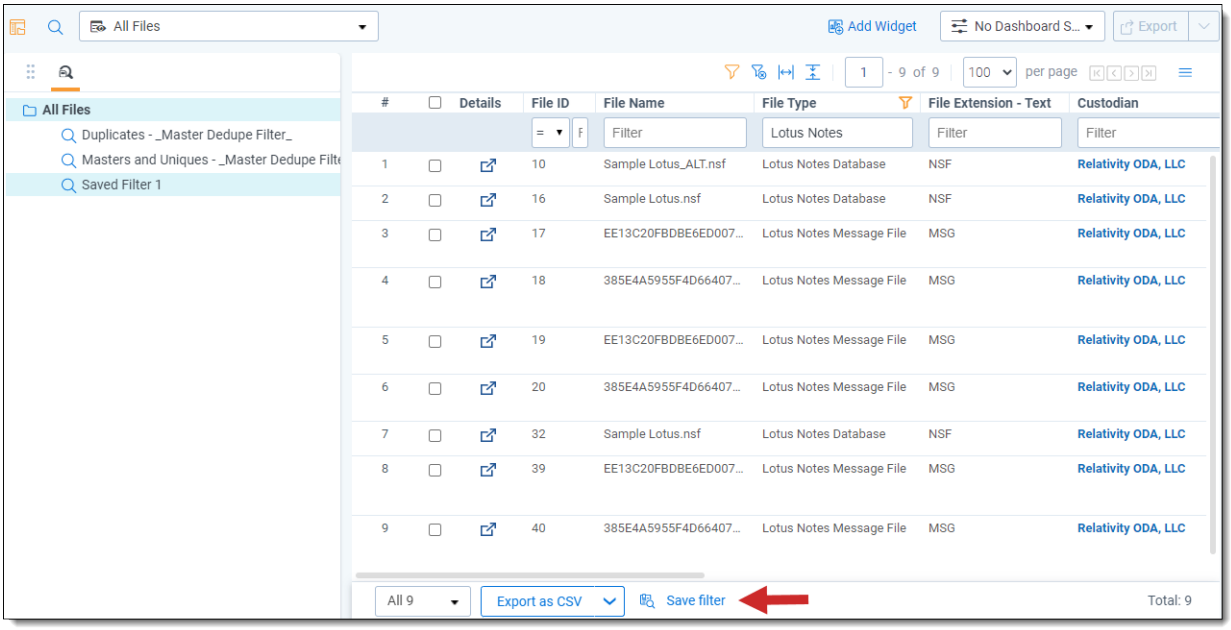
- Additional metadata fields were mapped after the initial publish of the processing set/data source completed. For example, you did not map the File Name field during the initial publish, which resulted in no metadata being populated for the documents. Now, you can map the File Name fields, go to the 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.

13.5 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:

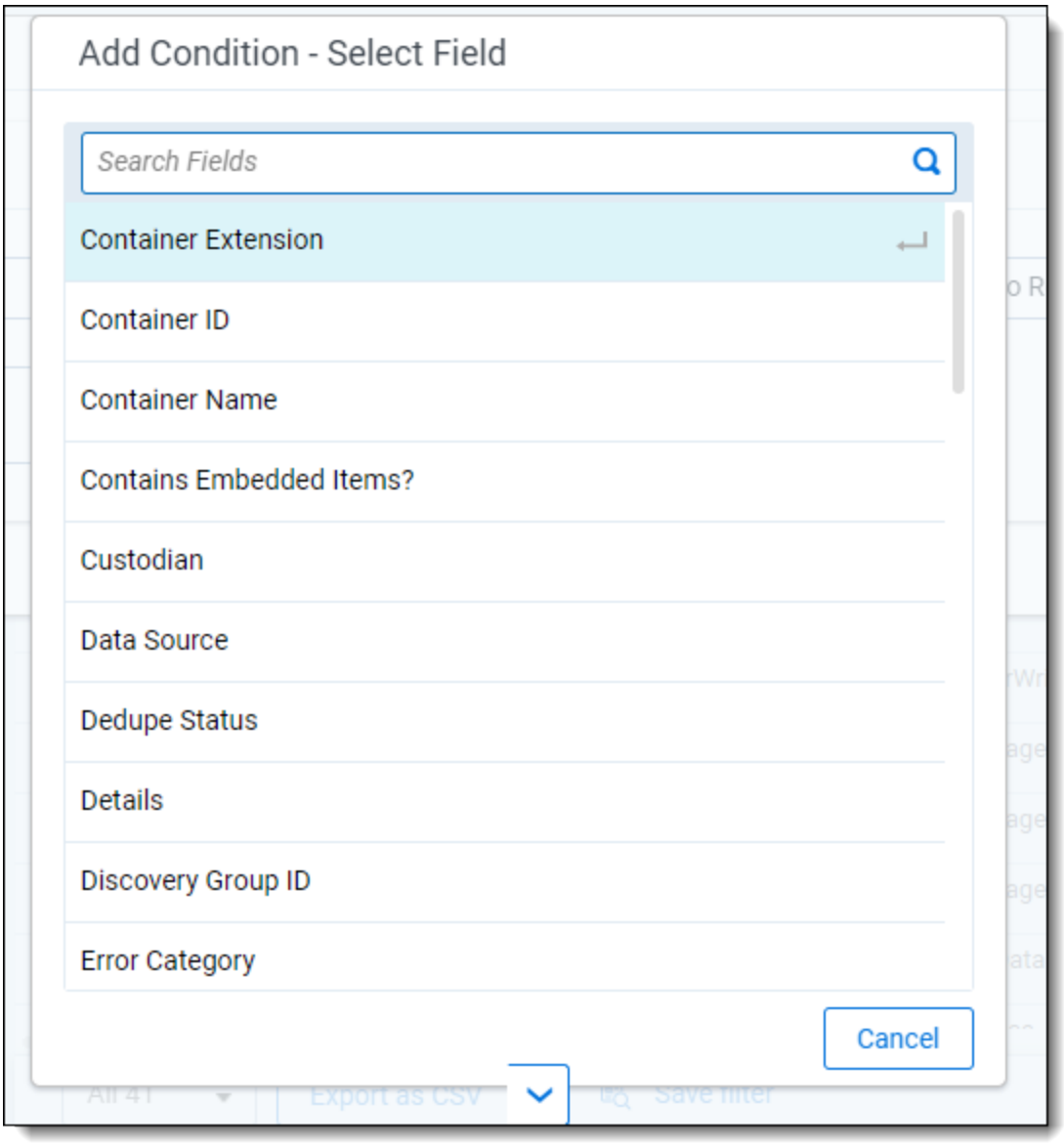
1. Filter on any of the fields in the view and click **Save Filters** at the bottom of the view.



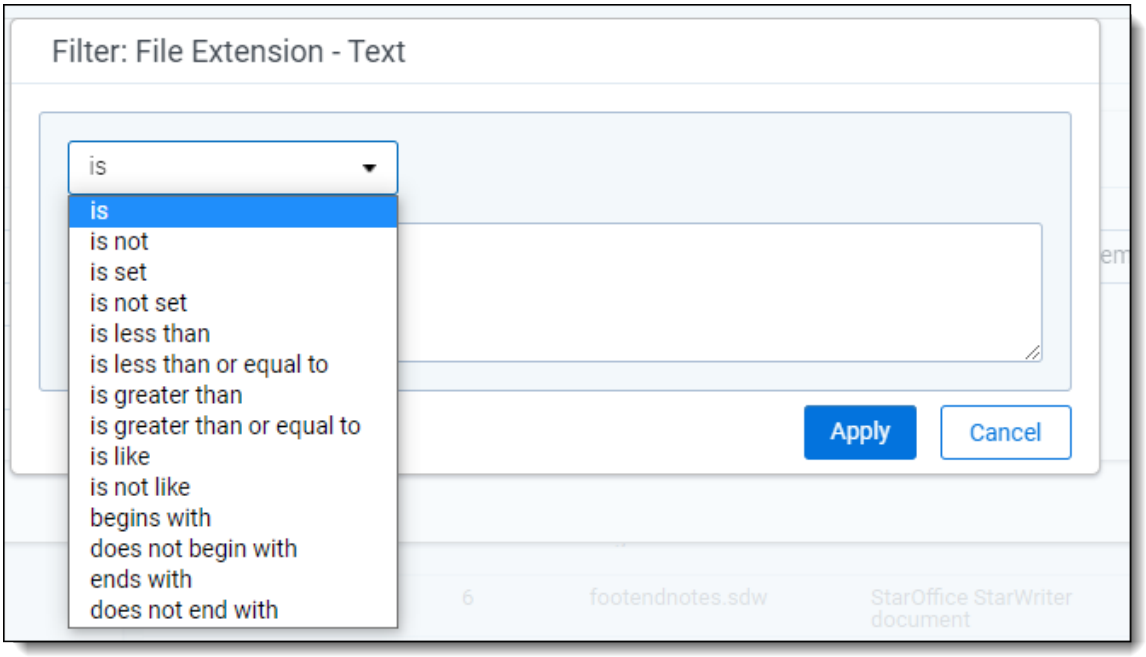
2. In the Saved Filter modal complete following fields and click **Save**.

- **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.

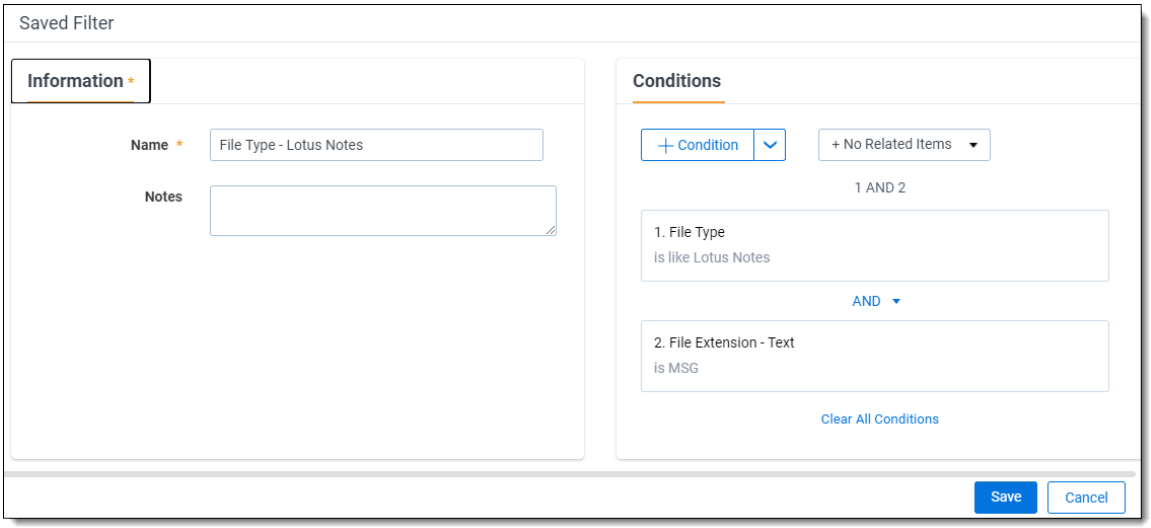
- **Conditions** - a display of the filter conditions you already set on each field in the view you were previously working in. For example, if you'd just filtered for Lotus Notes files on the File Type field, that filter is displayed. Here, you can start over by clear the conditions you already set, or you can add more conditions by clicking **+ Conditions**. Doing this brings up the Add Condition - Select Field modal, in which you can select additional fields.



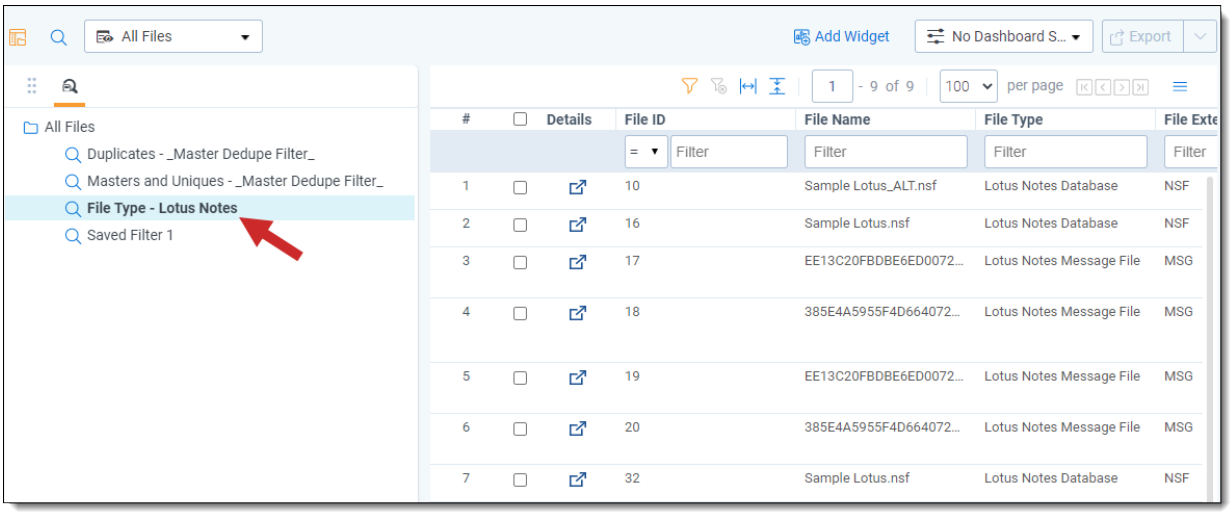
Once you select the additional fields you want to add to the saved filters set, specify the conditions you want to apply to those fields and click **Apply**.



The new field(s) and conditions are then visible in the Saved Filter modal and you can click Save to further refine the documents returned by this filter set.

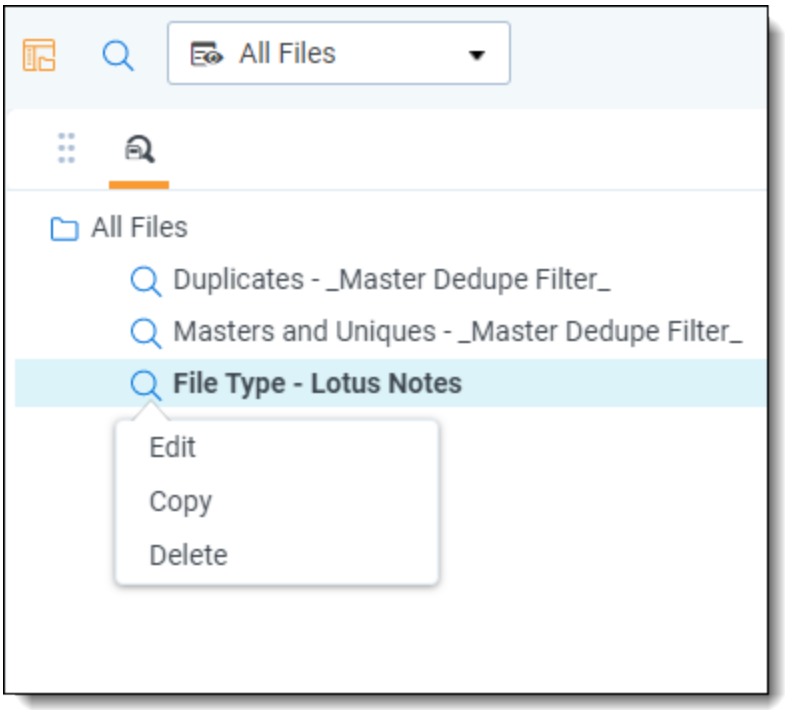


3. Once you save the filter set, return to the saved filters pane and confirm that the new set has been added. You can now use this set whenever you want to locate these specific documents.



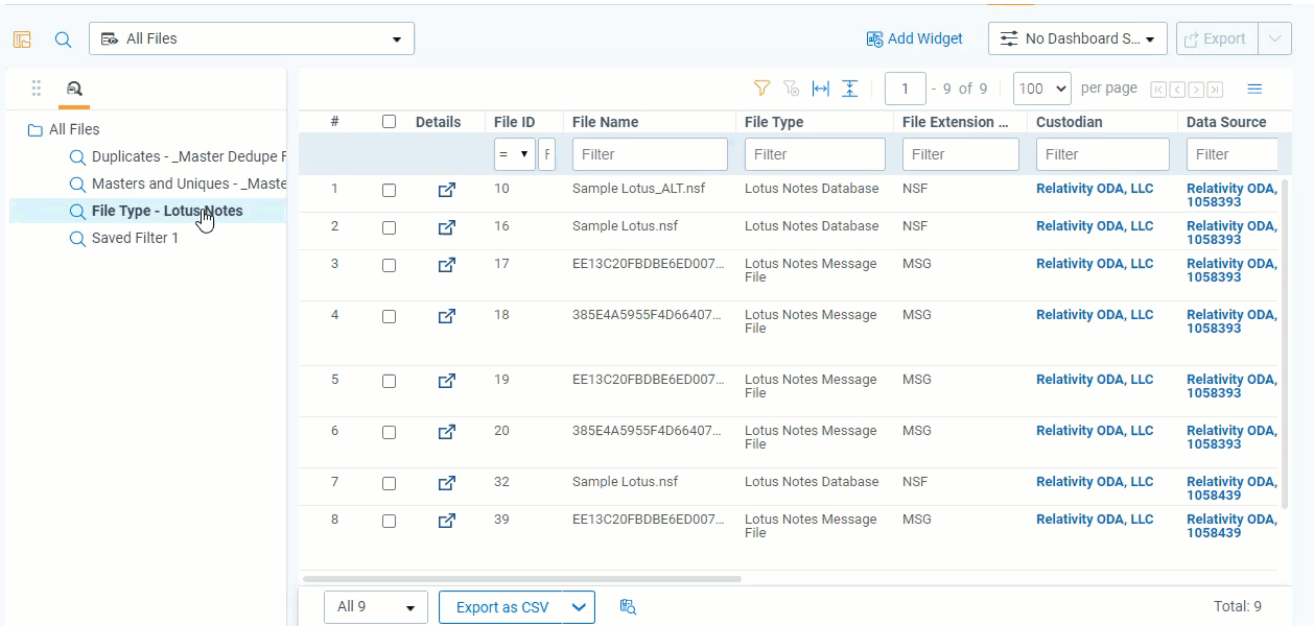
13.5.1 Right-click options

You can right-click on any saved filter set in the saved filters pane, and choose to Edit, Copy, or Delete it.



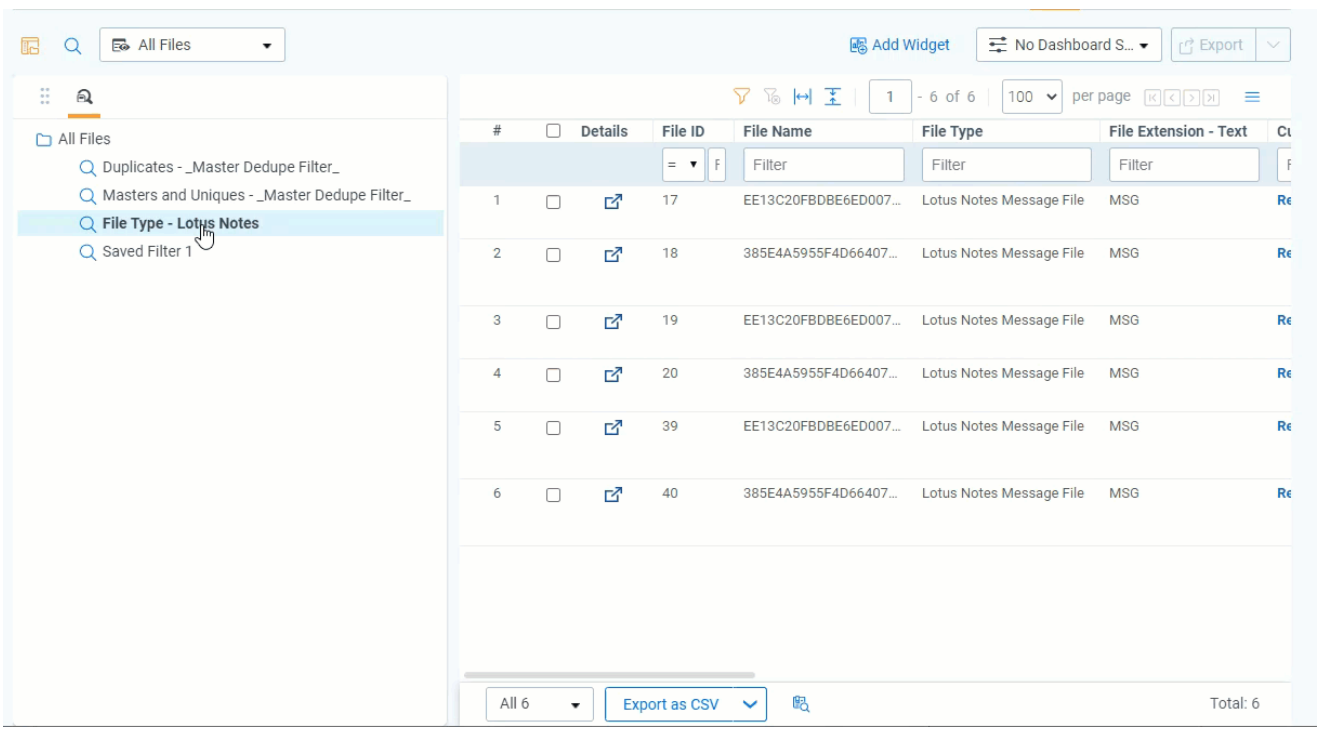
13.5.1.1 Edit

Clicking **Edit** takes you to the Saved Filter modal to add, remove, or modify any fields previously set.



13.5.1.2 Copy

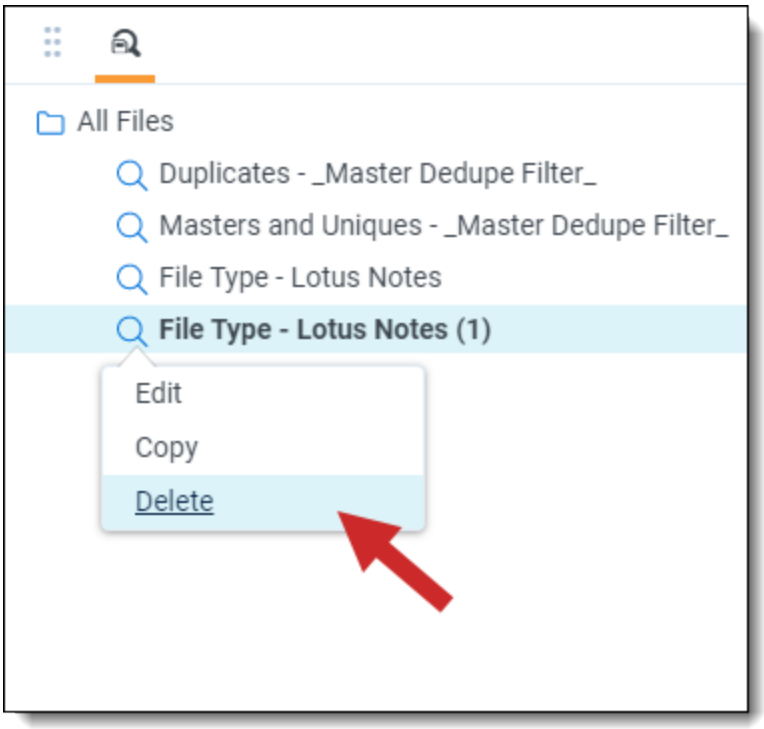
To make a copy of an existing filter set, right-click and select **Copy**.



- This is useful for situations in which you want to retain most of the fields and conditions in an existing filter set, but not all of them. Copying that filter set allows you to quickly make small changes to it and save it as a new set without having to build a new one from scratch. When you select Copy, the new set appears with the same name and a (1) added to the end of it. You can then edit this copied set to give it a new name and different conditions in the Saved Filter modal.

13.5.1.3 Delete

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



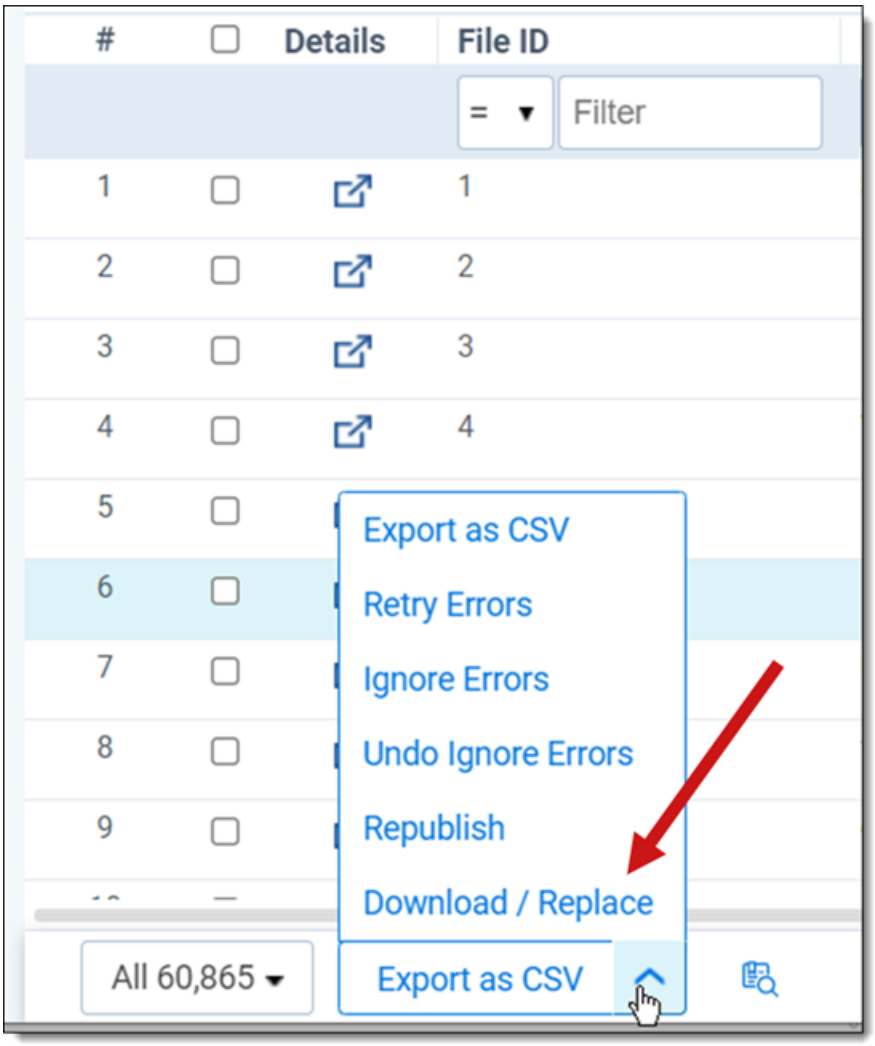
13.6 Download / Replace

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.

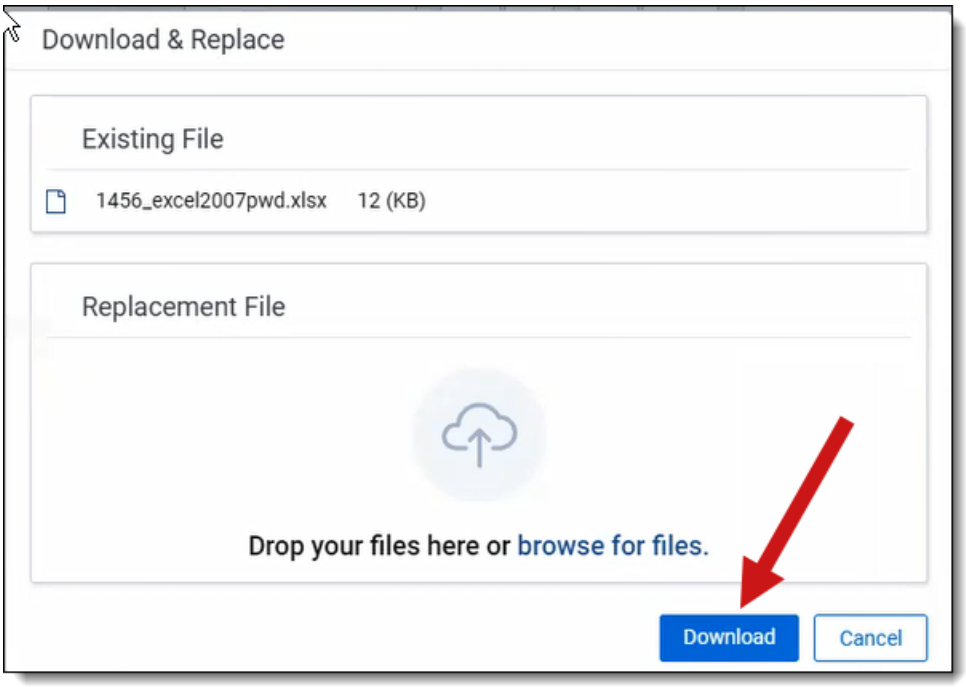
To perform the Download / Replace mass action, perform the following steps:

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

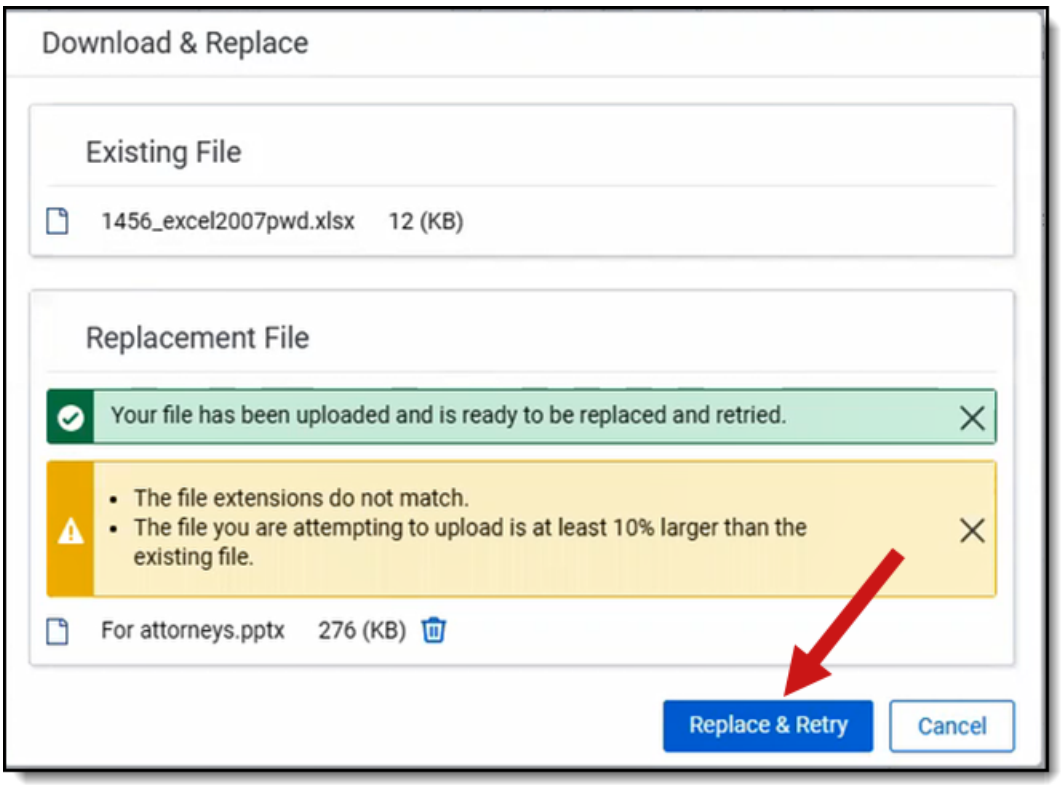
   Note: If you select multiple files, a Download and Replace error displays because this action can only be performed on one file at a time.



2. The Download & Replace dialog box opens allowing you to browse for or drop a replacement file.
3. Once you select or browse and drop the replacement file, click **Download.**



4. Once the file is downloaded, resolve the error, and drag the resolved file back into the 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.
5. Click **Replace & Retry**.



6. 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 will display a status of Resolved.

14 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:

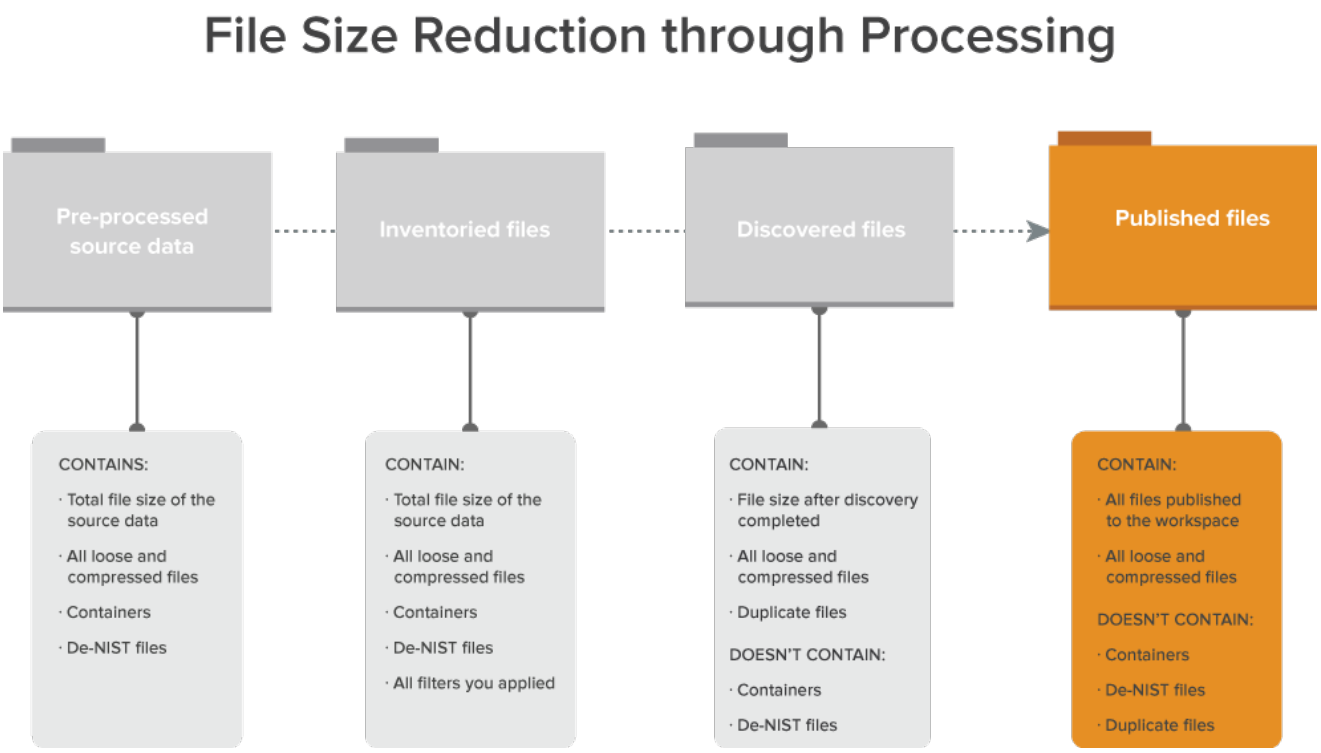
- Applies all the settings you specified on the profile to the documents you bring into the workspace.
- Determines which is the master document and master custodian and which are the duplicates.
- Populates the All Custodians, Other Sources, and other fields with data.

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

Use the following guidelines when publishing files to a workspace:

- If you intend to use both the RDC and Processing to bring data into the same workspace, note that if you select Custodial or Global as the deduplication method on your processing profile(s), the processing engine won't deduplicate against files brought in through the RDC. This is because the processing engine doesn't recognize RDC-imported data. In addition, you could see general performance degradation in these cases, as well as possible Bates numbering collisions.
- Publish includes the three distinct steps of deduplication document ID creation, master 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.



The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

The following is a typical workflow that incorporates publish:

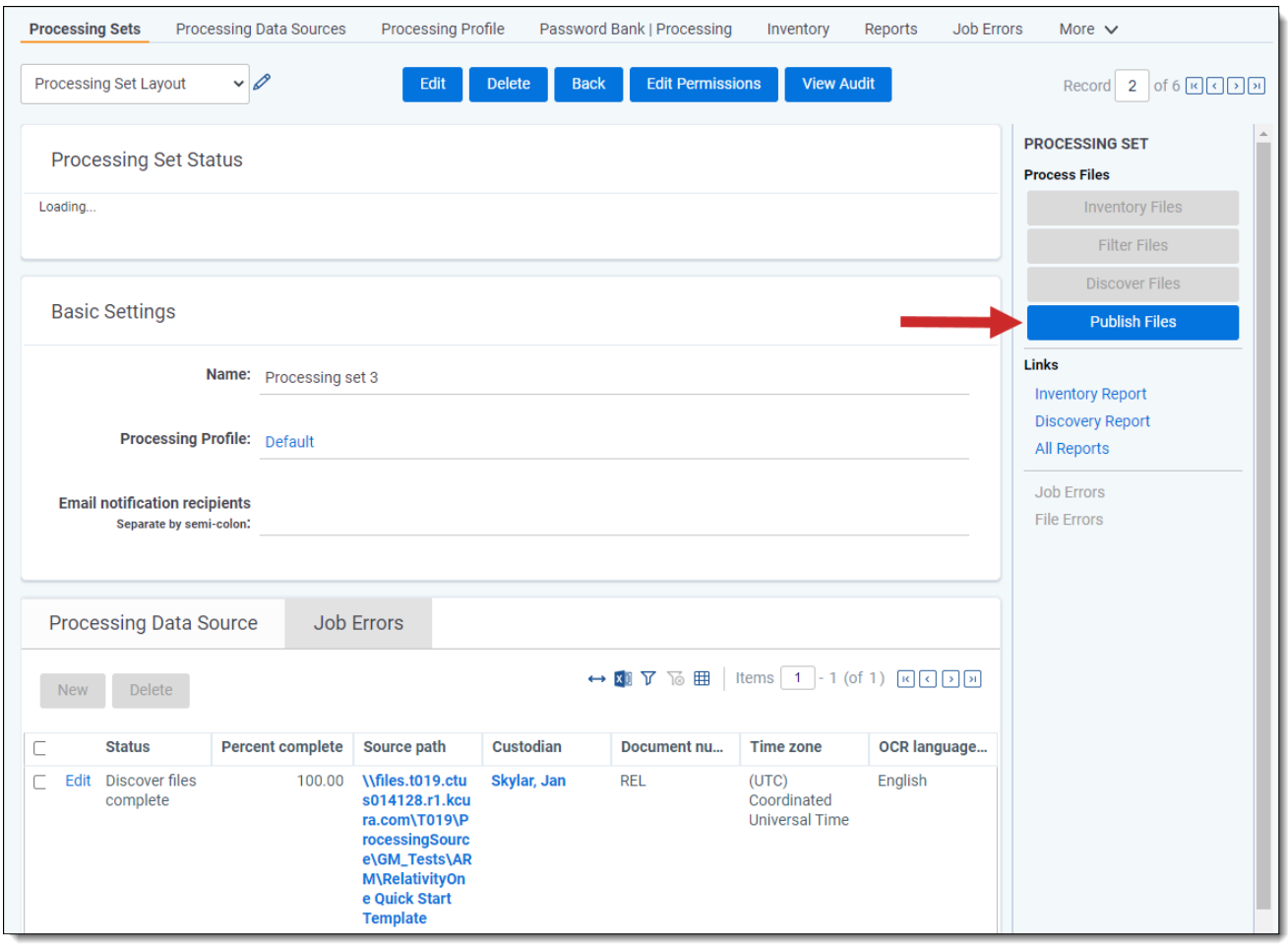
1. Create a processing set or select an existing set.
2. Add data sources to the processing set.
3. Inventory the files in that processing set to extract top-level metadata.
4. Apply filters to the inventoried data.
5. Run discovery on the refined data.
6. Publish the discovered files to the workspace.

Note: The instance setting controls whether the Extracted Text field data is loaded directly from its file path during the publish phase of processing, rather than as part of a client-generated bulk load file.

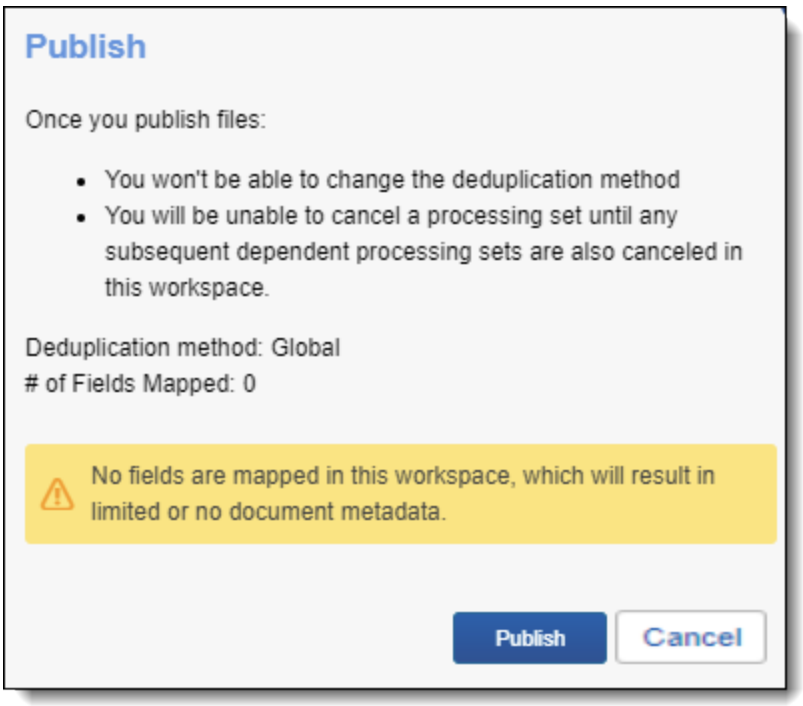
14.1 Running file publish

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.



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.



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, master 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.
- 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.
- 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 masters. 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

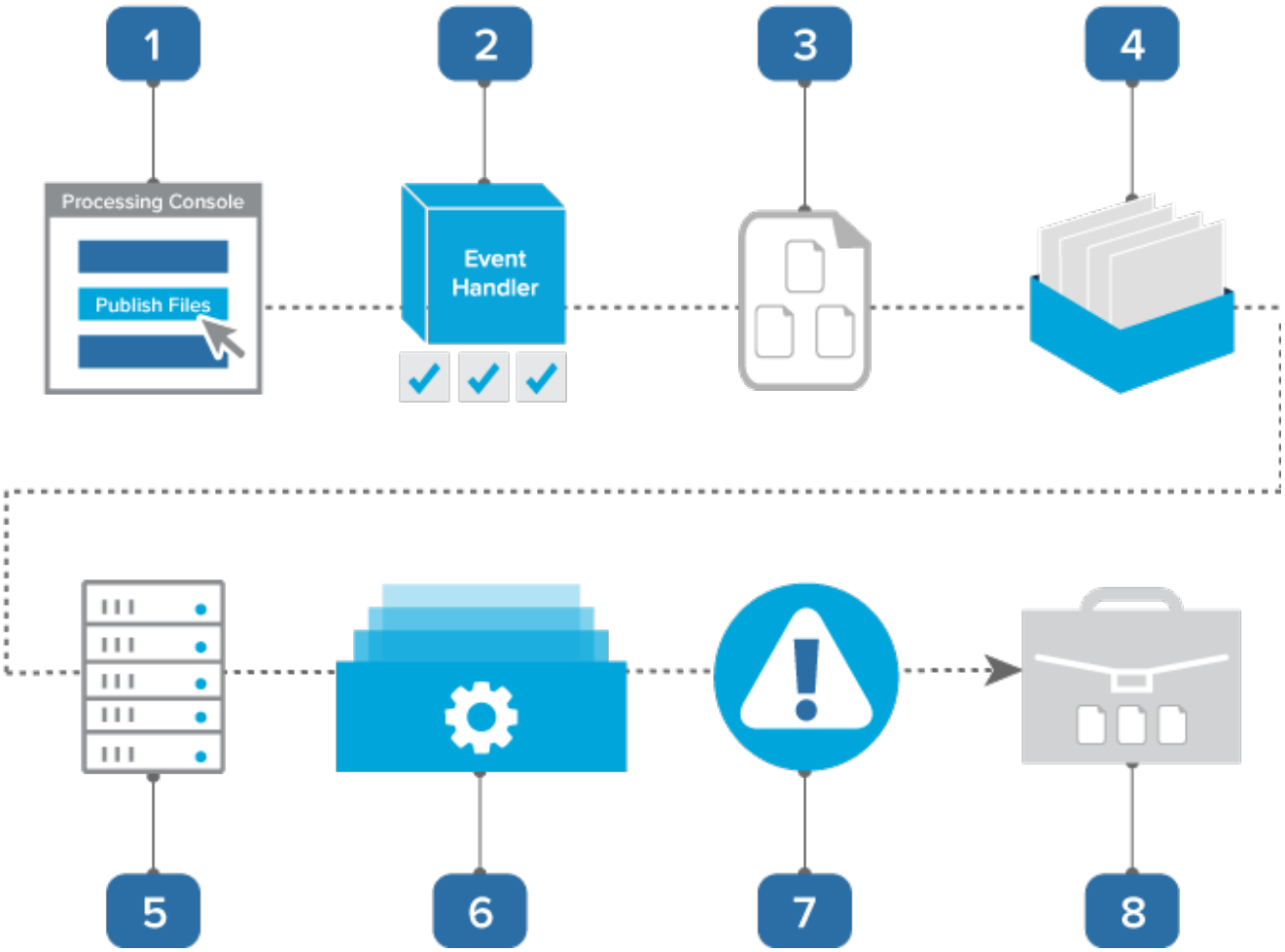
Note: Once you publish data into Relativity, you have the option of exporting it through the Relativity Desktop Client.

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.

14.1.1 Publish process

The following graphic and corresponding steps depict what happens behind the scenes when you start publish. This information is meant for reference purposes only.

Publish Process



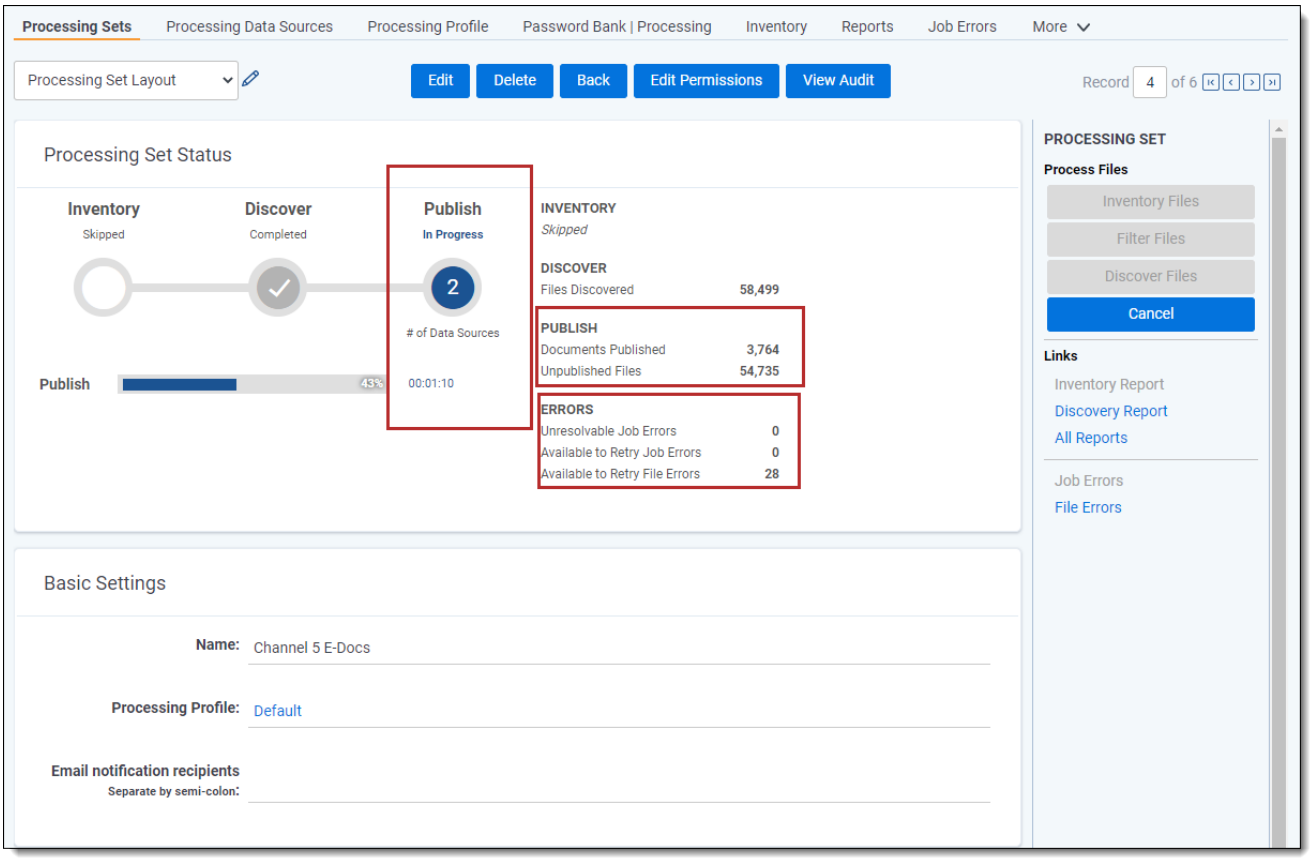
1. You click Publish Files on the processing set console. If you've arranged for auto-publish after discovery, publish will begin automatically and you aren't required to start it manually.
2. A console event handler checks to make sure that the set is valid and ready to proceed.
3. The event handler 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 workflow 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.

14.2 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.



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.

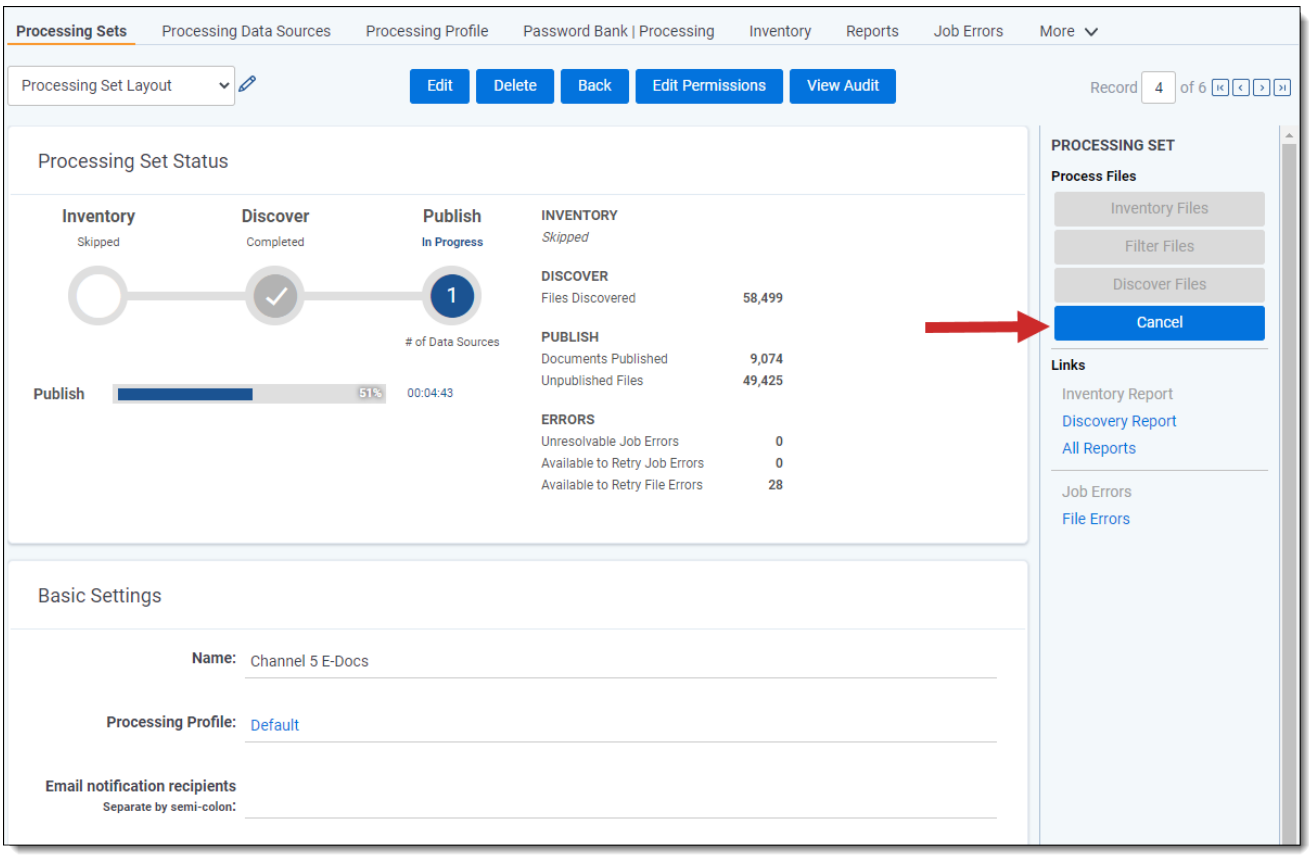
See Processing error workflow for details.

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.

14.3 Canceling publishing

If the need arises, you can cancel your publish job before it completes.

To cancel publish, click **Cancel**.



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.

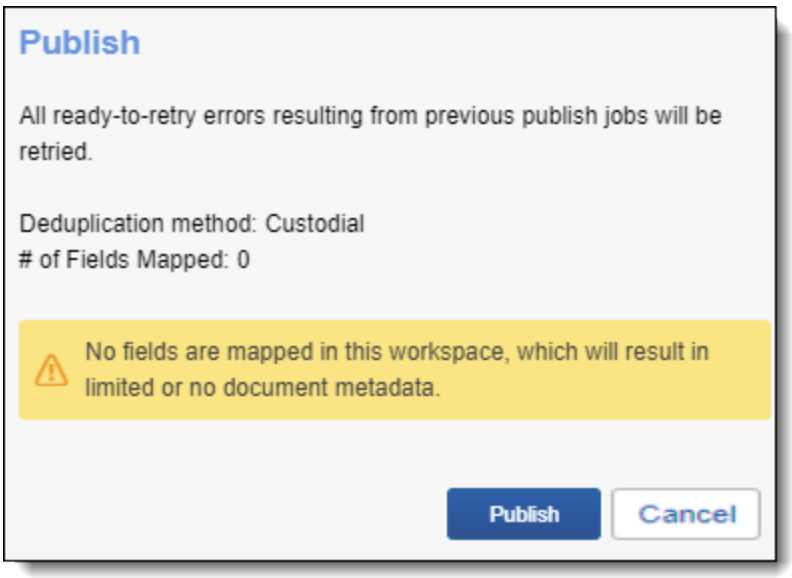
14.4 Republishing a processing set

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 207 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.



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

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.

14.5 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	Retried	Discovering files	
				0000000000000000649	Retried	Extracting data	
				0000000000000000654	Retried	Discovering files	
				0000000000000000656	Retried	Extracting data	
				0000000000000000658	Retried	Discovering files	
				0000000000000000660	Retried	Extracting data	
				0000000000000000662	Retried	Discovering files	
				0000000000000000663	Retried	Extracting data	
				0000000000000000664	Retried	Discovering files	
				0000000000000000665	Retried	Extracting data	
				0000000000000000666	Retried	Discovering files	
				0000000000000000667	Retried	Extracting data	
				0000000000000000669	Retried	Discovering files	
				0000000000000000671	Retried	Extracting data	
				0000000000000000672	Retried	Discovering files	
				0000000000000000673	Retried	Extracting data	
				0000000000000000674	Retried	Discovering files	
				0000000000000000675	Retried	Extracting data	
				0000000000000000676	Retried	Discovering files	
				0000000000000000677	Retried	Extracting data	
				0000000000000000679	Retried	Discovering files	
				0000000000000000680	Retried	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	Retried	Discovering files	
				0000000000000000648	Retried	Extracting data	
				0000000000000000651	Retried	Discovering files	
				0000000000000000653	Retried	Extracting data	
				0000000000000000655	Retried	Discovering files	
				0000000000000000657	Retried	Extracting data	
				0000000000000000659	Retried	Discovering files	
				0000000000000000661	Retried	Extracting data	
				0000000000000000668	Ready to retry	Discovering files	
				0000000000000000670	Ready to retry	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.

15 Post-publish delete

Post-publish delete ensures that Processing stays updated when documents are deleted from review. This topic provides real-world examples of how you can integrate post-publish delete functionality into your processing workflow.

15.1 Post-publish delete overview

If you discover and publish files into review and then Mass Delete them from the Documents tab, the Files tab updates to reflect this deletion, and post-publish delete occurs.

Any files associated with a deleted document are indicated with a Yes value on the Processing Deletion field. All deleted documents can be seen in the Files tab. Note that post-publish delete occurs only if you select the **Document and all associated files** on the delete confirmation modal.

In addition, Processing updates deduplication. Specifically, once the modal closes and the mass delete operation is complete in review, Relativity Processing recalculates deduplication. At this point, the deleted files are removed, and any newly-designated primary documents are automatically published into review.

Note: If a document has duplicates within a single custodian, the document will need to be deleted per occurrence. Duplicate documents are not automatically deleted within a single custodian.

Deleting a file also precludes it from being factored into any future deduplication calculations, including any newly processed data. This occurs regardless of whether the deleted document was a master or unique.

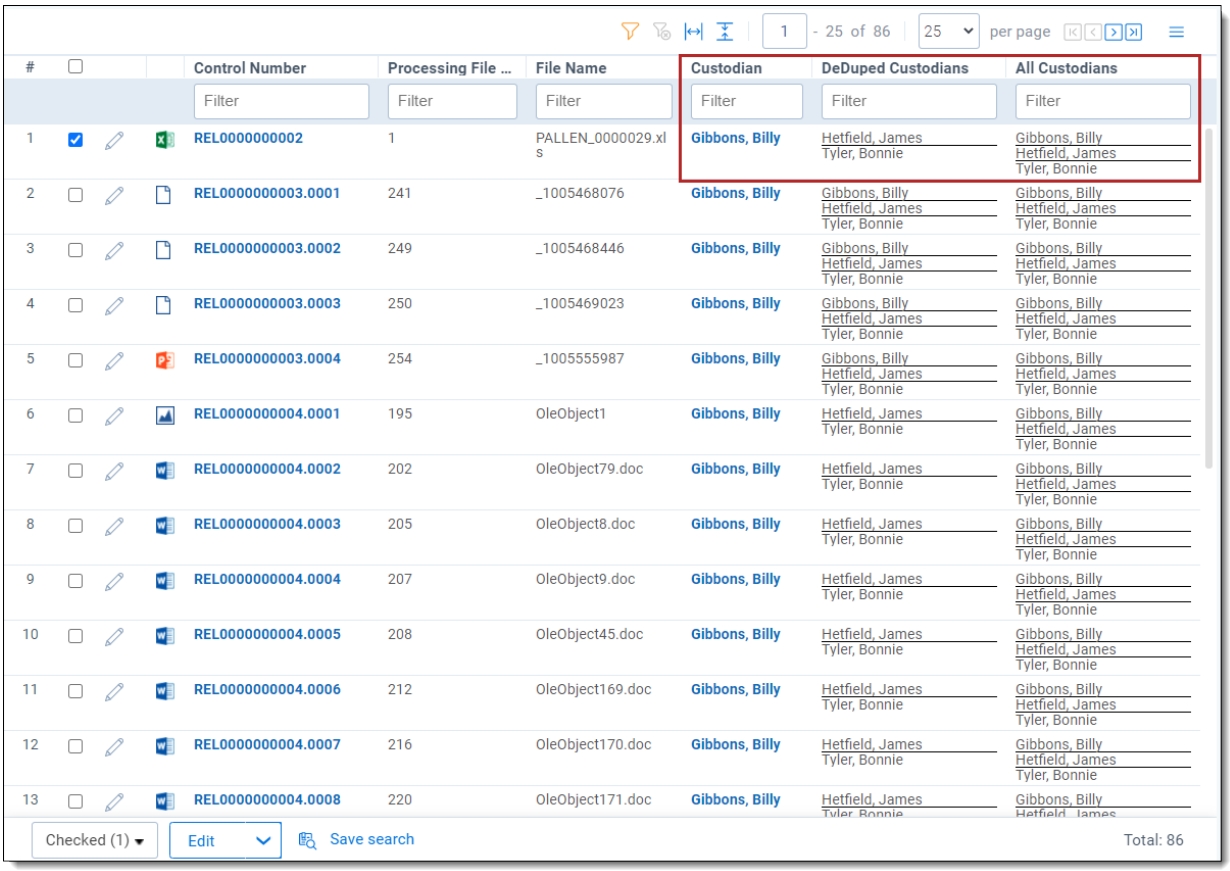
For information on reporting what was deleted, see Master Document Replacement Summary.

15.2 Publishing a new master document

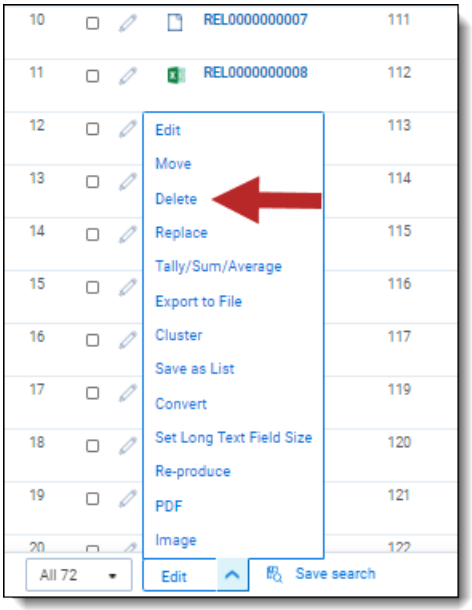
In this example, a user has deleted a primary custodian's document, and Relativity Processing automatically publishes the next deduped custodian's document.

To recreate this example, perform the following steps:

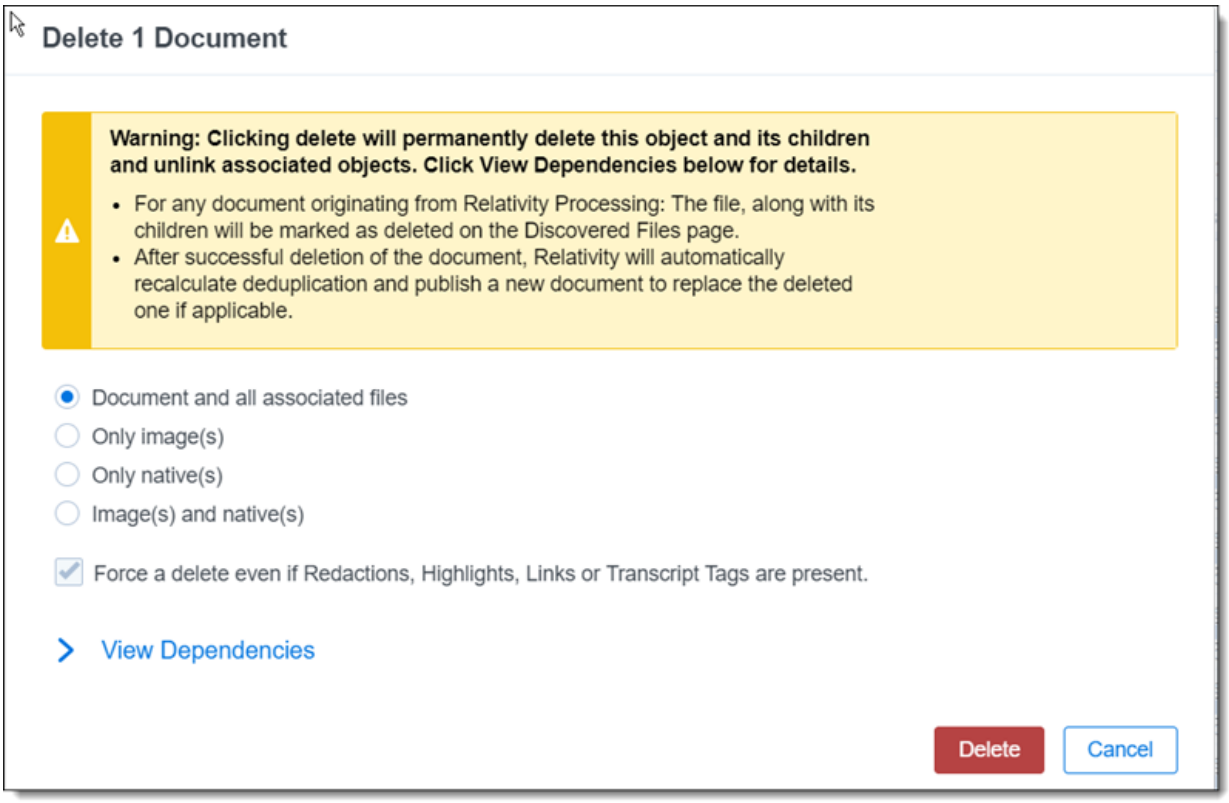
1. In the Documents list select a primary custodian's document.



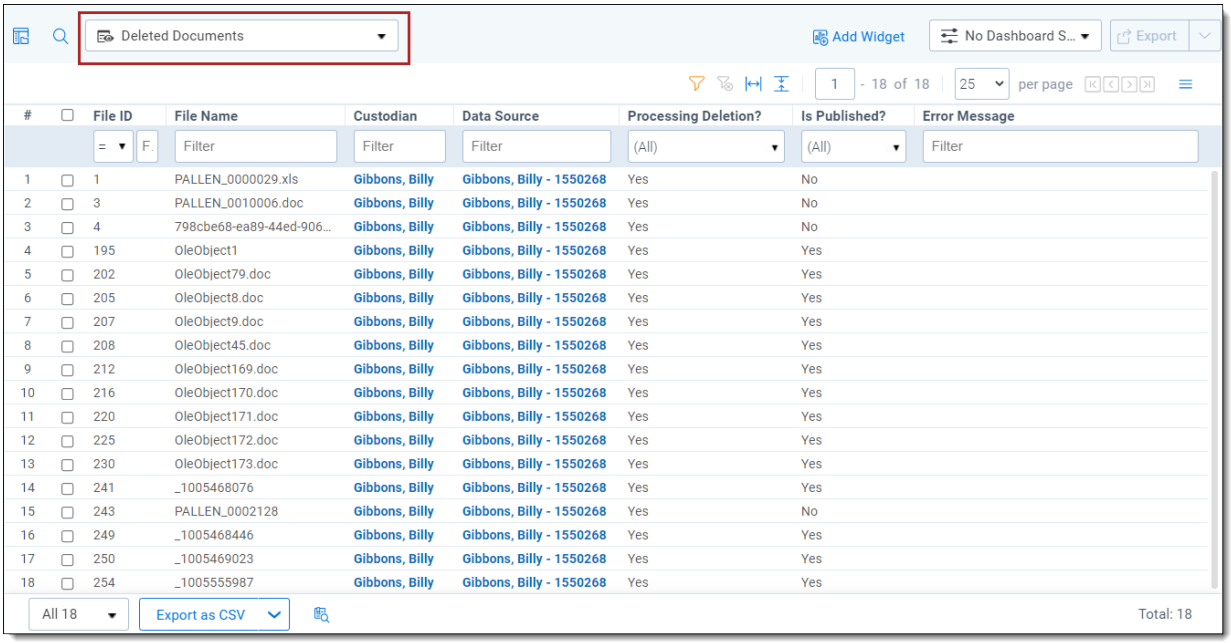
2. Select the **Delete** mass operation.



3. Click **Delete** on the warning message. Once you do this, Relativity deletes the selected document from review, and Relativity Processing flags it for deletion. When this deletion is complete, the deduplication recalculation will begin.

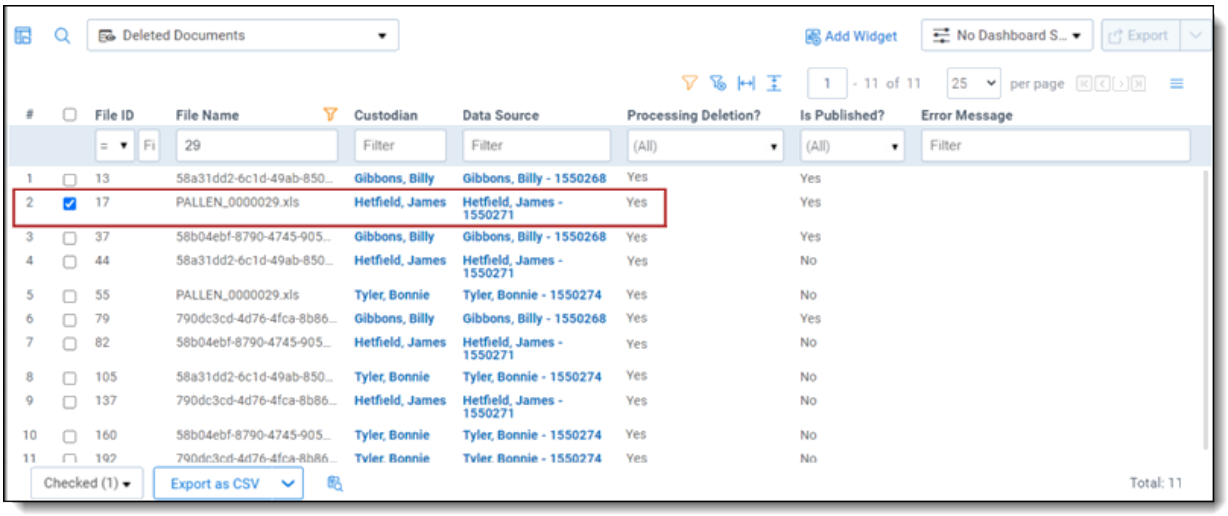


4. Navigate to the **Files** tab and select the **Deleted Documents** view.



#	File ID	File Name	Custodian	Data Source	Processing Deletion?	Is Published?	Error Message
1	1	PALLEN_0000029.xls	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	No	
2	3	PALLEN_0010006.doc	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	No	
3	4	798cbe68-ea89-44ed-906...	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	No	
4	195	OleObject1	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
5	202	OleObject79.doc	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
6	205	OleObject8.doc	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
7	207	OleObject9.doc	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
8	208	OleObject45.doc	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
9	212	OleObject169.doc	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
10	216	OleObject170.doc	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
11	220	OleObject171.doc	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
12	225	OleObject172.doc	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
13	230	OleObject173.doc	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
14	241	_1005468076	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
15	243	PALLEN_0002128	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	No	
16	249	_1005468446	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
17	250	_1005469023	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	
18	254	_1005555987	Gibbons, Billy	Gibbons, Billy - 1550268	Yes	Yes	

5. To confirm that deduplication has been recalculated and a new master identified, locate the document you deleted and note that the following values are displayed:



- **Custodian** - the deduped custodian listed for the document before you deleted it.
- **Processing Deletion?** - Yes

Note: Processing determines the next document in line to be published based on the order in which data sources were originally published in the workspace. If there are multiple copies of the same record in a single data source, Processing will choose the one with the lowest File ID, which means it was discovered first.

15.3 Deleting documents within a family

It's recommended when deleting documents in a family that you delete the entire family at once; however, with post-publish delete you have the ability to delete documents regardless of family status.

In this example, a user accidentally selects only the parent document to delete, which means Relativity Processing also flags all of its children for deletion in the Files tab. The user then needs to identify the child documents that were deleted.

To recreate this example, perform the following steps:

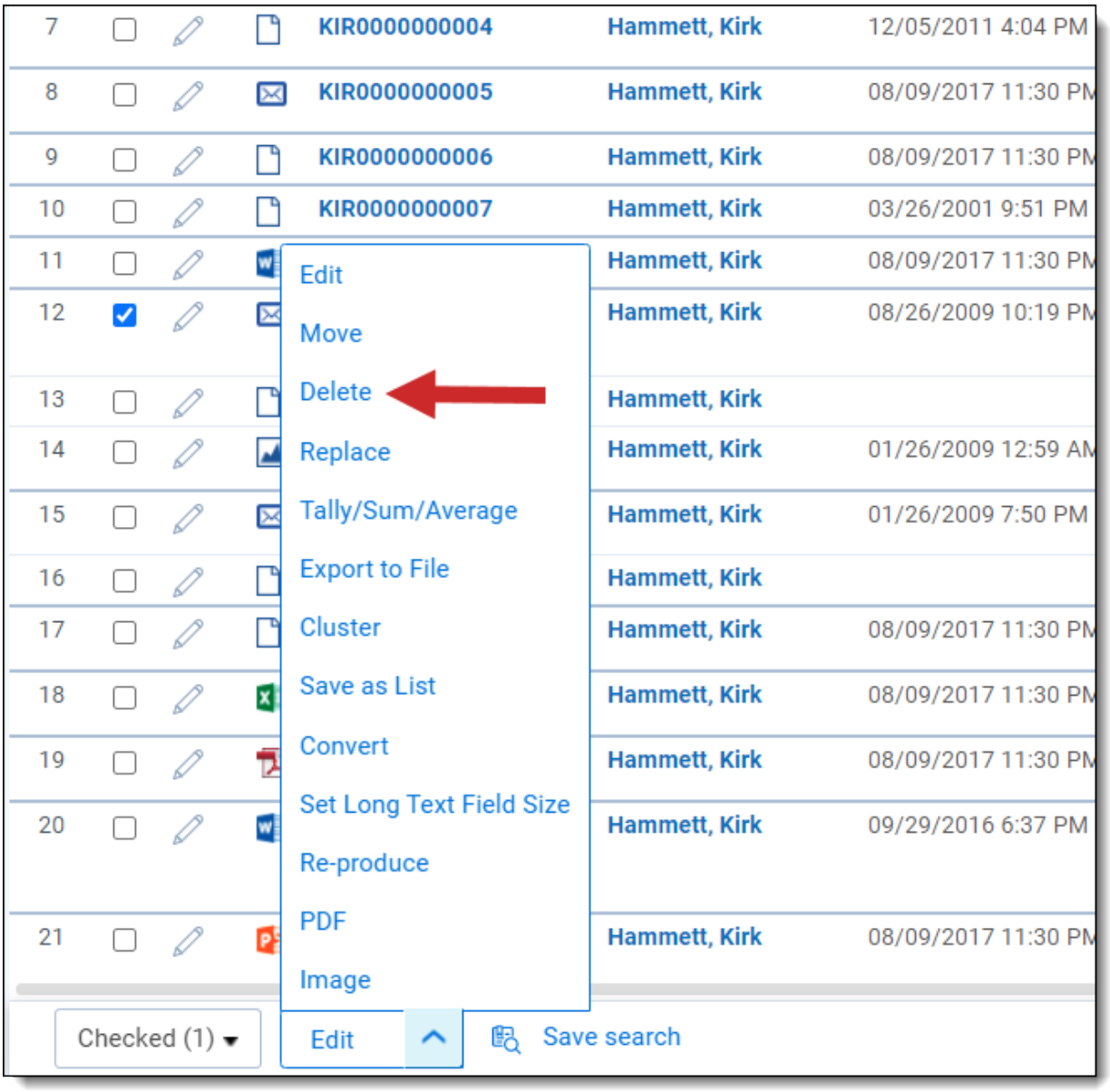
1. Locate a family group of documents in the Document list and select only the parent.

1 - 22 of 22 25 per page

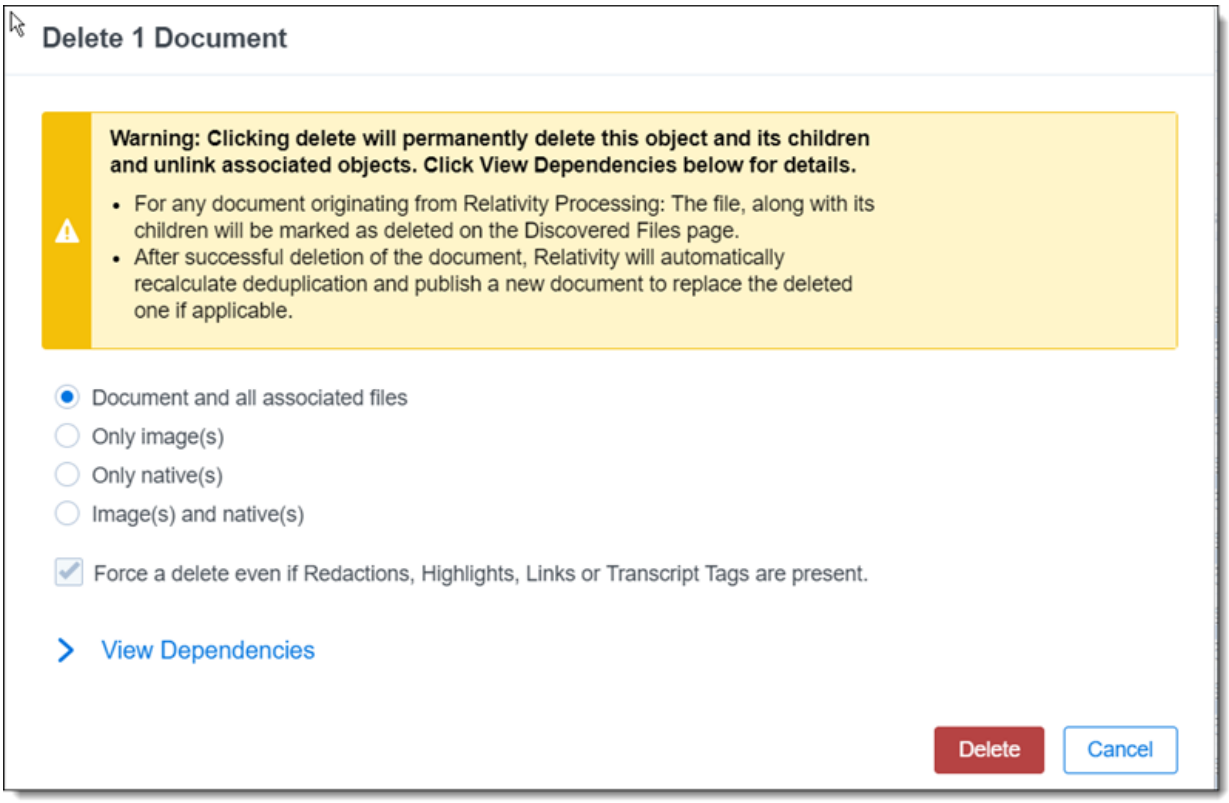
#		Control Number	Custodian	Primary Date/Time	Record Type	Unified Title	Email From
		Filter	Filter	(All)	(All)	Filter	Filter
						Draft.doc	
7		KIR0000000004	Hammett, Kirk	12/05/2011 4:04 PM	Edoc	b4ubuild_sample.mpx.mpp	
8		KIR0000000005	Hammett, Kirk	08/09/2017 11:30 PM	Email	--- Untitled Document ---	
9		KIR0000000006	Hammett, Kirk	08/09/2017 11:30 PM	Edoc	flag.JTT	
10		KIR0000000007	Hammett, Kirk	03/26/2001 9:51 PM	Edoc	footendnotes.sdw	
11		KIR0000000008	Hammett, Kirk	08/09/2017 11:30 PM	Edoc	Naming.wpd	
12	☑	KIR0000000009	Hammett, Kirk	08/26/2009 10:19 PM	Email	Testing	idiscover [CN=idiscover/O=idiscove]
13		KIR0000000009.0001	Hammett, Kirk		Attach	Picture	
14		KIR0000000009.0002	Hammett, Kirk	01/26/2009 12:59 AM	Attach	metadata_sample_001.jpg	
15		KIR0000000010	Hammett, Kirk	01/26/2009 7:50 PM	Email	--- Untitled Document ---	
16		KIR0000000010.0001	Hammett, Kirk		Attach	Picture	
17		KIR0000000011	Hammett, Kirk	08/09/2017 11:30 PM	Edoc	outlook(this is html).pst	
18		KIR0000000012	Hammett, Kirk	08/09/2017 11:30 PM	Edoc	Excel Sheet - Test1234!.xlsx	
19		KIR0000000013	Hammett, Kirk	08/09/2017 11:30 PM	Edoc	PDF Sheet - Test1234!.pdf	
20		KIR0000000014	Hammett, Kirk	09/29/2016 6:37 PM	Edoc	Pre2007BasicPasswordProtectedWordFile - Test1234!.doc	
21		KIR0000000015	Hammett, Kirk	08/09/2017 11:30 PM	Edoc	Presentation - Test1234!.pptx	

Checked (1) Edit Save search Total: 22

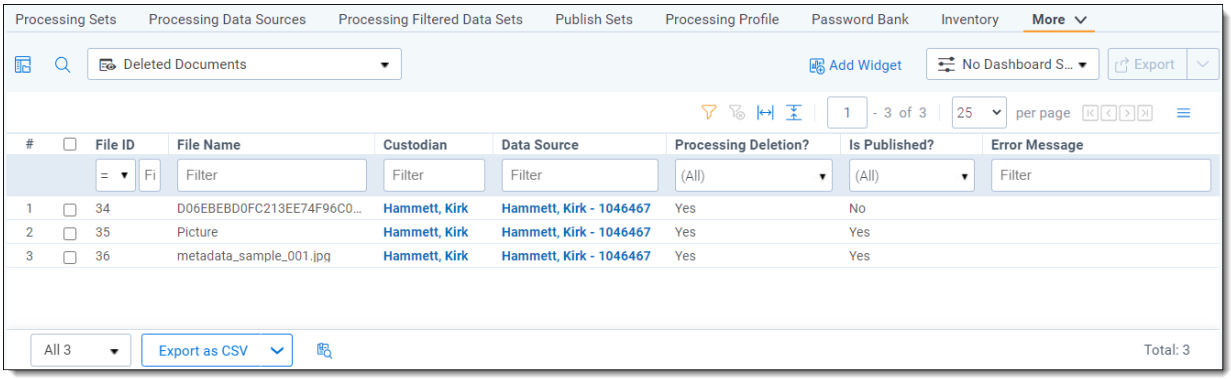
2. Select the **Delete** mass operation.



3. Click **Delete** on the confirmation message. Since this parent document originated from Relativity Processing, it and all of its children will be marked as Yes for the Processing Deletion? field in the Deleted Documents view on the Files tab. The parent will be marked as No for the Is Published? field, and the children will be marked as Yes. In addition, the children will remain visible in the Documents list.

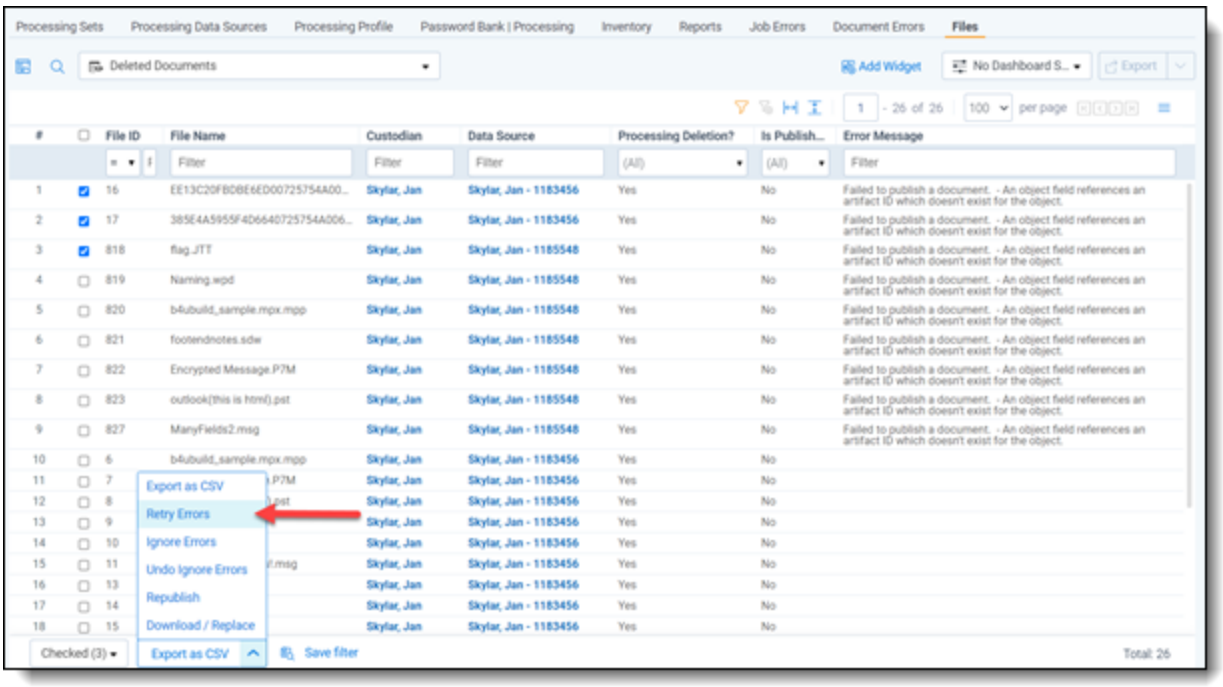


4. Navigate to the **Files** tab and select the **Deleted Documents** view. Confirm that the parent document and its children are present with the expected values for the Processing Deletion? and Is Published? fields.



15.4 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. These errors can be retried using the **Retry Errors** mass operations option. Once deleted, these documents will be excluded from further processing operations (e.g., deduplication, retry, and republish) and the next duplicate will be published as the new master document, if available.



To see a summary of master documents that have been replaced, see the Master Document Replacement Summary report in Processing Reports. See Mass Delete for more information on deleting documents.

16 Processing error workflow

This topic provides information on working with errors that may occur during processing jobs.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

16.1 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

The following security permissions are required to perform actions on Job Errors:

Object Security	Tab Visibility
▪ Job Error - View, Edit ▪ Processing Error - View, Edit	▪ Processing ▪ Job Error

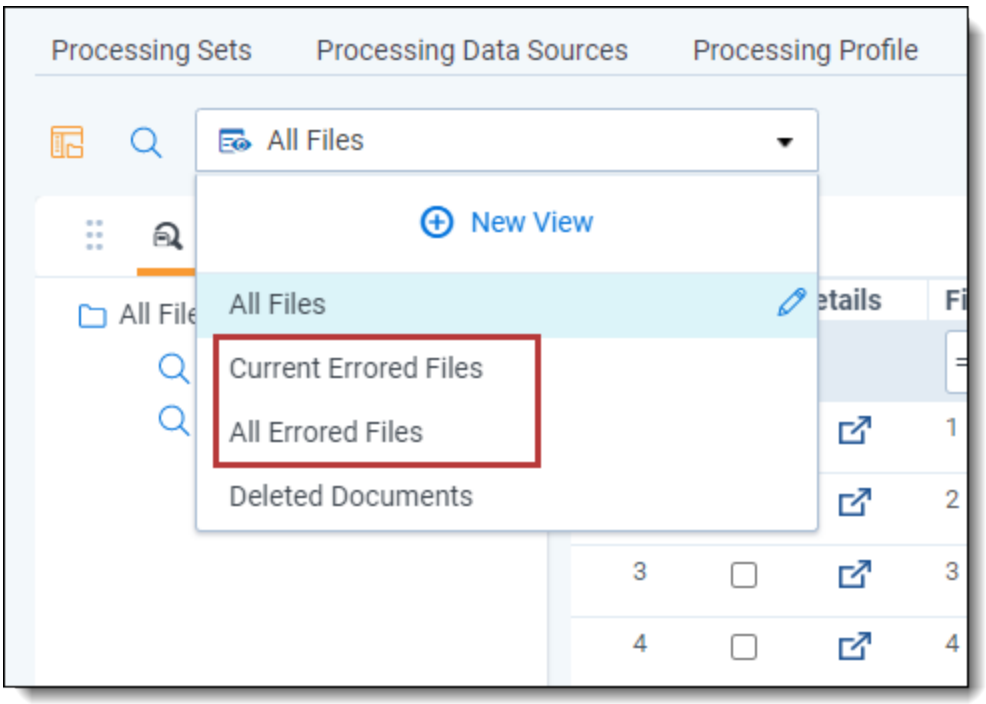
For more information on permissions, see Workspace permissions.

16.2 Processing errors tabs

The **Files** and **Job Errors** tabs allow you to easily locate issues that may have occurred in any processing sets. These errors are notified to you through the processing set page upon job completion.

16.2.1 Files tab

The **Files** tab contains all error information associated to specific files that have occurred during the discovery, publish, and deletion phases of processing. The Current Errored Files and All Errored Files views are tailored to error workflow by only containing information pertaining to the errors that have occurred.

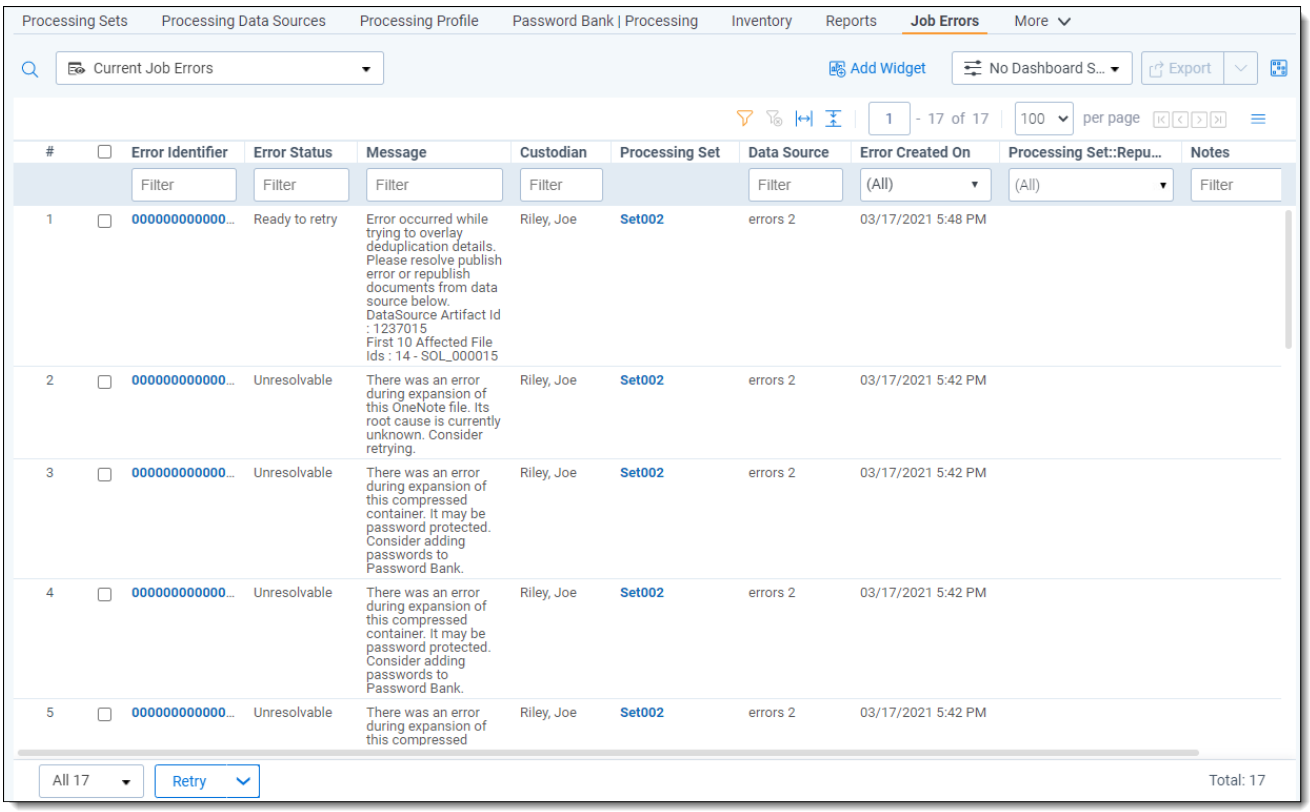


- The Current Errored Files view displays all outstanding errors from processing jobs. This is your primary location for workflows like error retry, ignore, and file replacement.
- The All Errored Files view is primarily utilized for historical reporting of errors from processing sets. This view displays any file that has encountered an error, regardless of whether it was resolved or not. It is a good reference to export an error report out of Relativity for a given collection or set of custodians.

For more information on these views, see Files tab.

16.2.2 Job Errors tab

The **Job Errors** tab contains all 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 to the entire set itself.



16.2.2.1 Job Error views

The Current Job Errors view in the Job Errors tab displays all unresolved job errors while the All Job Errors view displays any job error that has occurred throughout the lifecycle of the matter. Both views contain the following fields:

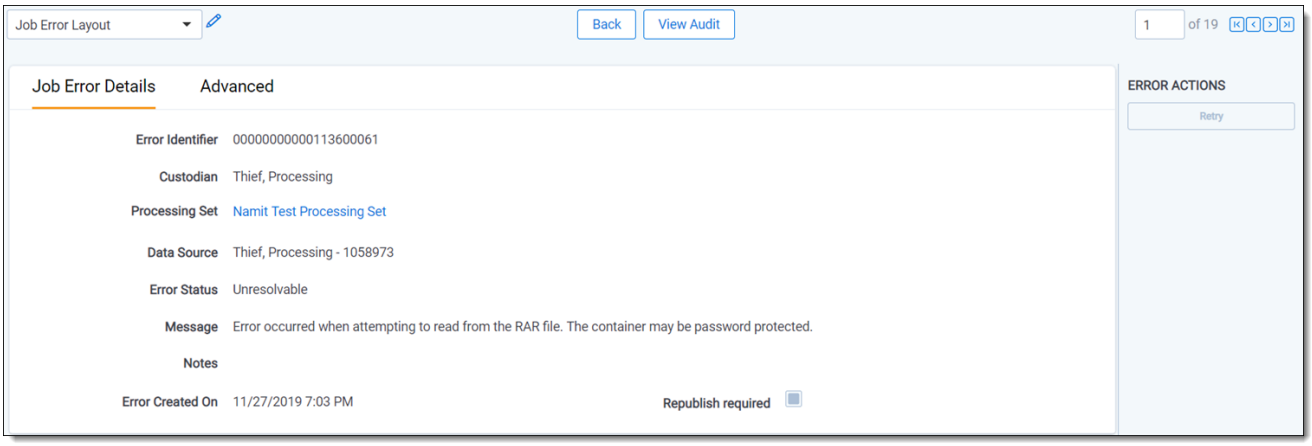
- **Error Identifier** - the unique identifier of the error as it occurs in the database. When you click this message, you are taken to the error details layout, where you can view the stack trace and other information. Note that for Unresolvable errors, the console is disabled because you can't take any actions on that error from inside Relativity. For more information, see Processing error workflow.
- **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.

Note: Errors occurring during inventory are always represented as Job Errors. For more information, see Inventory Errors.

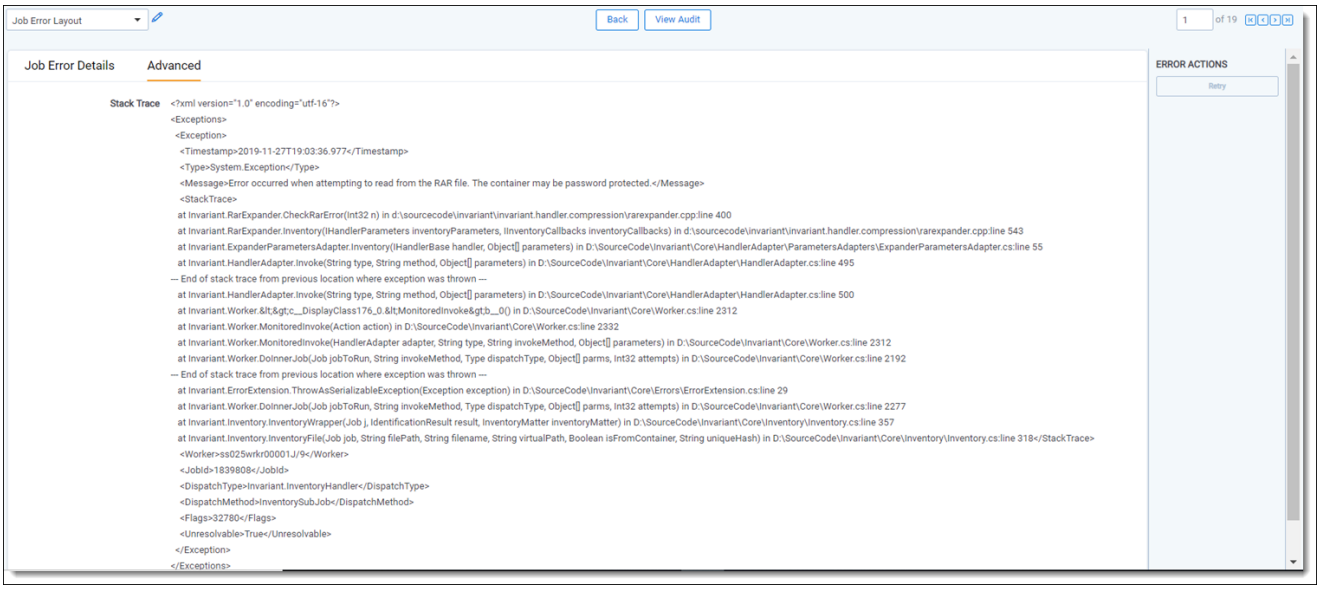
16.2.3 Job Error layout

Clicking on the job error identifier value brings you to the Job Error Details layout.

Note that the Error Actions console is disabled for unresolvable job errors, since you can't retry or ignore job errors the way you can document errors.

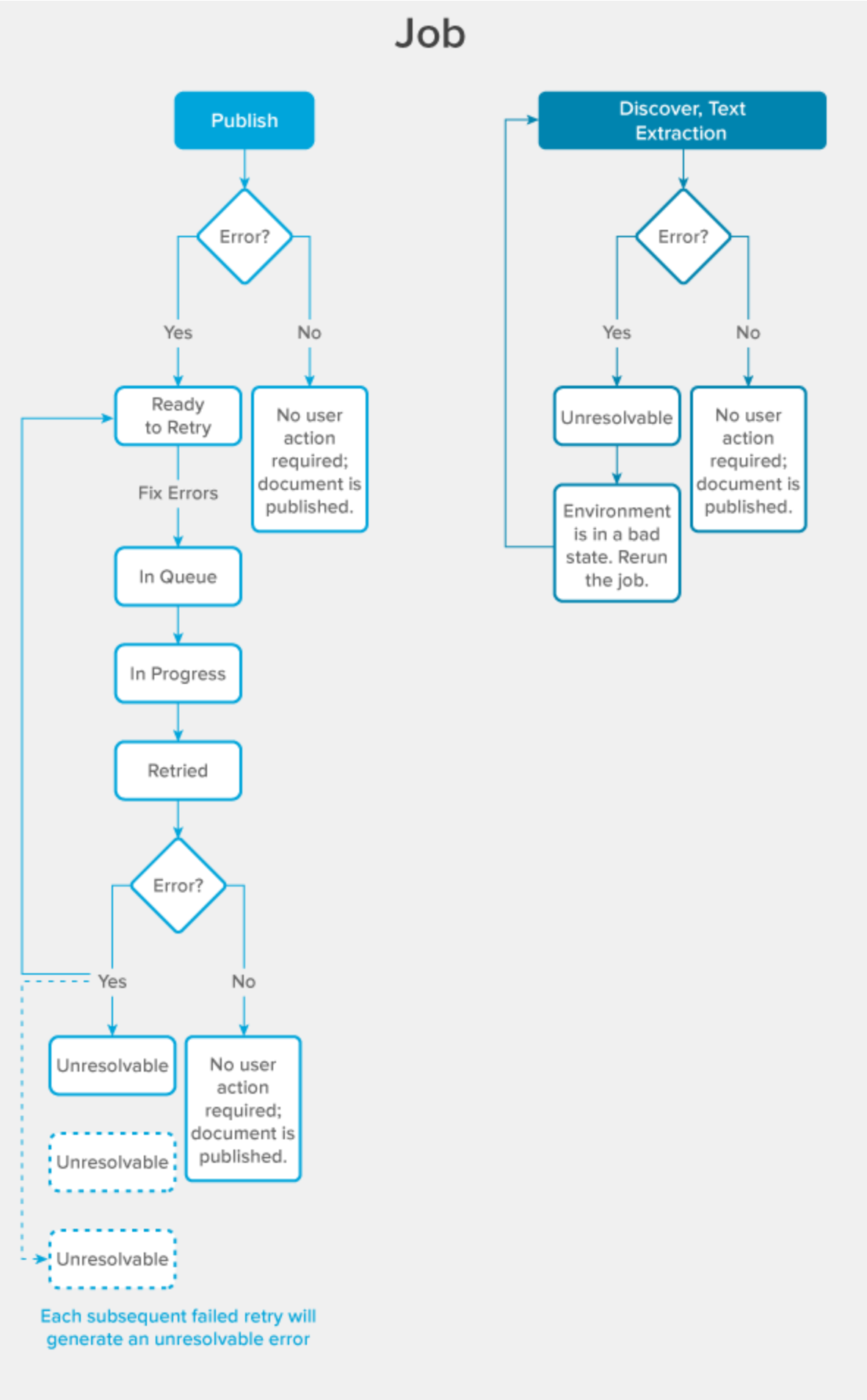


To see the job error's stack trace, click on the **Advanced** tab of the error details layout and view the value in the **Stack Trace** field.



16.2.3.1 Job-level error workflow

See an overview diagram of the job error workflow
The following diagram depicts the standard workflow that occurs when Relativity encounters a job-level error during processing.

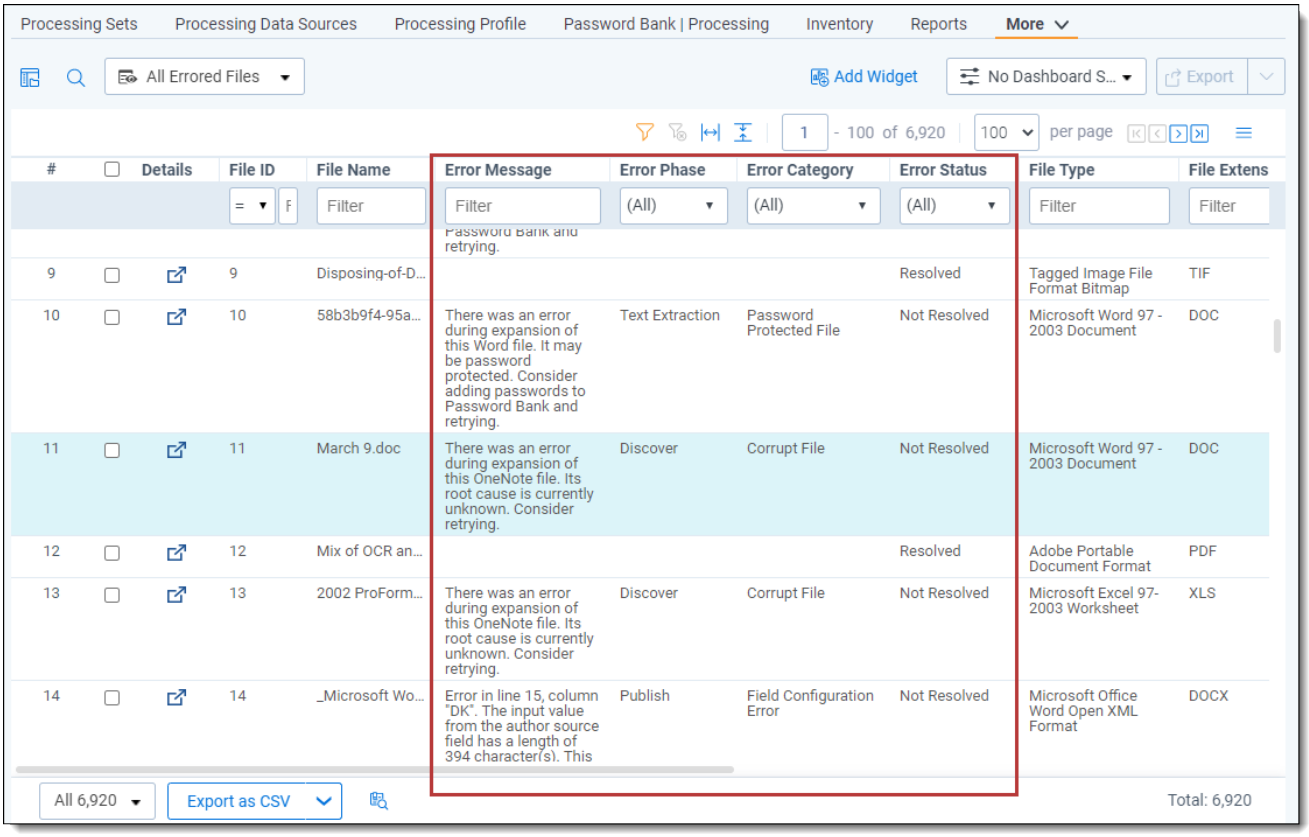

Job
Publish
Error?
Yes
No
Ready to Retry
No user action required; document is published.
Fix Errors
In Queue
In Progress
Retried
Error?
Yes
No
Unresolvable
No user action required; document is published.
Unresolvable
Unresolvable
Each subsequent failed retry will generate an unresolvable error
Discover, Text Extraction
Error?
Yes
No
Unresolvable
No user action required; document is published.
Environment is in a bad state. Rerun the job.

16.3 Useful error field information

The following sections provide information on error-specific fields and views that you can use in your processing workflow.

16.3.1 Combined error fields

The **Files** tab displays a single error associated to a file.

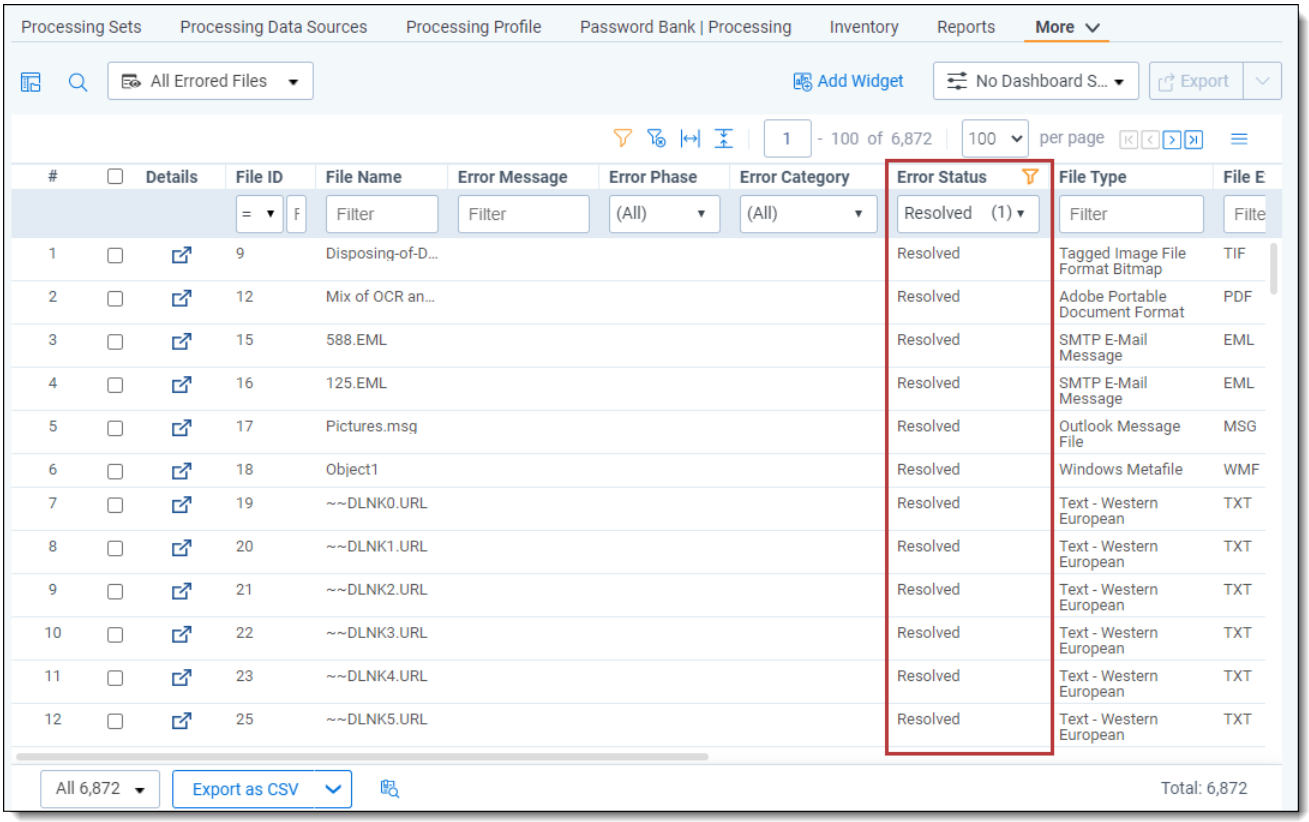


This error displays through the **Error Message**, **Error Category**, **Error Phase**, and **Error Status** fields. However, multiple errors can be associated to a single file at the same time as issues can occur during different phases of Processing. Relativity determines the displayed error based on a set precedence of Processing phases that could potentially block content from being published. The precedence is as follows:

- **Delete** – a document was deleted from Relativity, but encountered an issue, potentially affecting recalculation of deduplication.
- **Publish** – a document was supposed to be promoted to review but encountered an error and was held back.
- **Discover** –a file may have encountered an issue during expansion and may not have extracted a child record and/or associated metadata.
- **Text Extraction** – a file encountered an issue during text extraction and is missing some or all associated text.

16.3.2 Error status information

The **Error Status** field provides information on where the file is in error remediation.

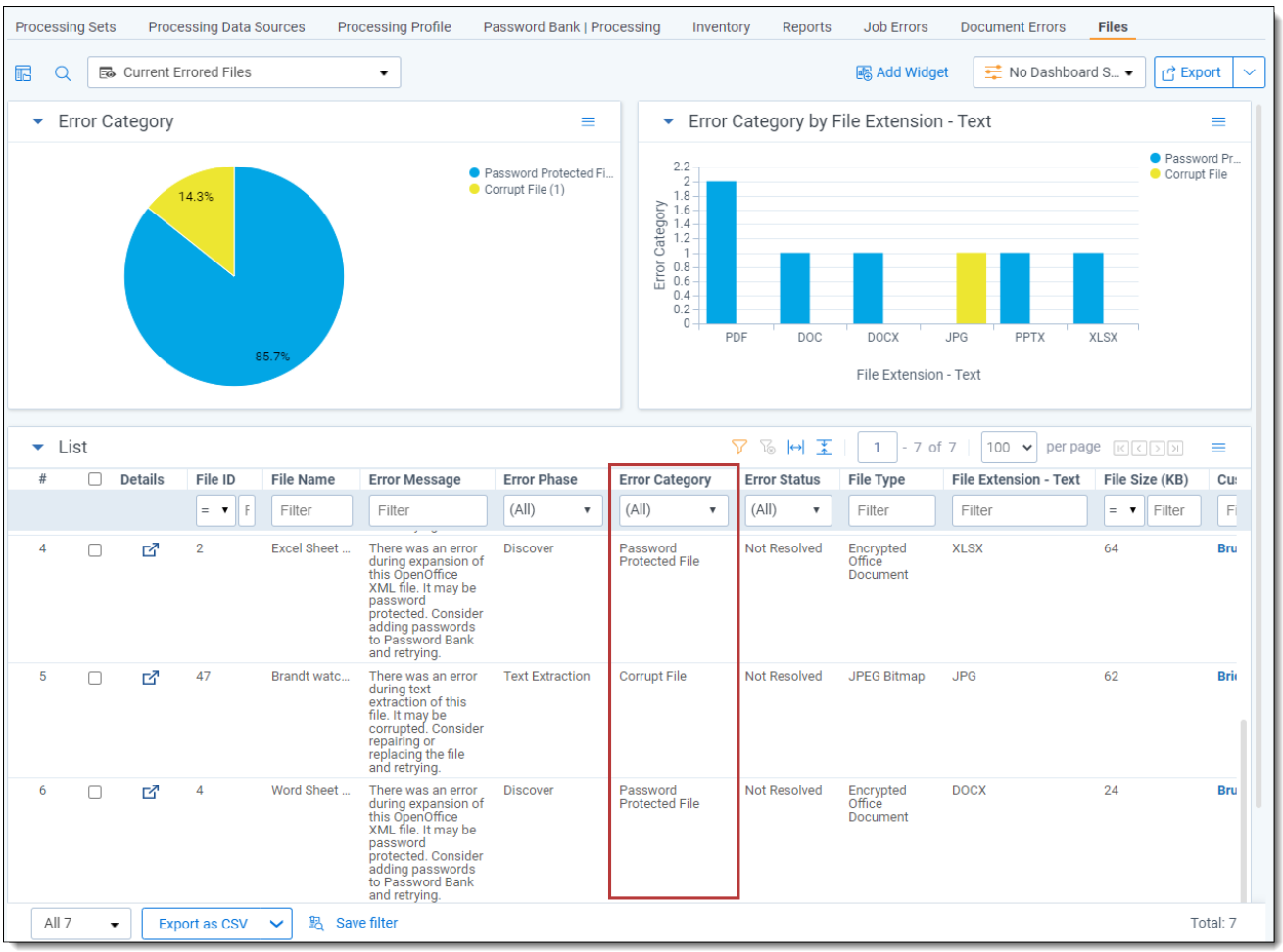


This is helpful to determine if any further actions are required on a file or to see if an error had ever occurred on a record. When a file has all its errors resolved, the **Error Message**, **Error Category**, and **Error Phase** fields no longer display any content, but the **Error Status** field keeps a status of Resolved to indicate that it was a record that initially encountered issues but have since been fixed. The statuses of errors are as follows:

- **Not Resolved** – The error is still outstanding.
- **Resolving** – The error was submitted or in the process of being retried.
- **Resolved** – The error was resolved.
- **Ignored** – The error was ignored. See **File error actions**.

16.3.3 Error Category list

The **Error Category** field provides insight into the nature of the errors that have occurred during processing.



The following table provides a listing of all values on the **Error Category** field, along with a description of what kinds of issues those values bring back if filtered.

Category name	Description
Corrupt Container	These errors are exclusive to container files that have encountered 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. These errors are usually either ignored or downloaded offline for an investigation on whether the corruption can be remediated and then subsequently replaced and retried.
Corrupt File	These errors are exclusive to non-container files that have found elements of corruption during Processing. These errors are either ignored or downloaded offline for an investigation on whether the corruption can be remediated and then subsequently replaced and retried.
Could Not Identify	Relativity Processing was unable to identify the file during Discovery. This may indicate corruption in the file but was unable to be determined at the time of discovery.
Environmental	These errors are caused by issues in the Relativity Processing environment. These should be retried and resolved when encountered.
File Read / Write Failure	These errors are a subset of Environmental issues specifically caused by file system issues. These should be retried and resolved when encountered.

Category name	Description
Missing Attachment	An attachment from a document or email was not able to be extracted 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. A specific root cause was unable to be readily identified, but they should be retried and resolved 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. These errors are usually either ignored or downloaded offline for an investigation on whether the corruption can be remediated and 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. These errors are not resolved unless the proper passwords or encryption keys are placed in the Password Bank. 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 of its contents opens and displays correctly in the viewer, it means a password-protected container was attached to the email.
Password Protected File	These errors are exclusive to non-container files that have encountered some form of password protection or encryption security measures. These errors are not resolved unless the proper passwords or encryption keys are placed in the Password Bank. For more information, see Password Bank.
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. When encountered, the field settings should be resolved according to the error message and resolved.
Unsupported	Relativity Processing has determined that these files are unsupported and was unable to obtain metadata or text from them. These files can be published to your workspace, but they may be inaccessible from the Viewer.

16.3.4 Details modal

You can open the Details modal of a file by clicking to see uncompressed file and content metadata not visible by default in the Files view.

The Details modal provides you with supplemental information about errors that have occurred on records during discovery and publish.

Details Processing Errors

File Data

File Name	Presentation - Test1234!.pptx
File ID	6
Custodian	Brule, Steve
Data Source	Brule, Steve - 1548436
Container Extension	
Container ID	0
Container Name	
Contains Embedded Items?	No
Error Message	There was an error during expansion of this OpenOffice XML file. It may be password protected. Consider adding passwords to Password Bank and retrying.
File Extension - Text	PPTX
File Size (KB)	1,122
File Type	Encrypted Office Document
Folder Path	\CloudSet\Password Protected files
Is Container?	No
Is Embedded?	No
Is Published?	No
Processing Deletion?	No
Import Source	\\files.t019.ctus014128.r1.kcura.com\T019\ProcessingSource\CloudSet\Password Protected files\Presentation - Test1234!.pptx
Discovery Group ID	39039462

Close

You can also see a summary and history of all processing errors and retries in this modal. When you click the **Processing Errors** tab, you're presented with the following breakdown of the current errors and error history of the selected file:

Details **Processing Errors**

Error Summary

Current Errors	2
Discover Error(s)	Password Protected File (1)
Text Extraction Error(s)	Password Protected File (1)
Publish Error(s)	none
Delete Error(s)	none

Error History

Error total: 4

Password Protected File (Text Extraction) - Not Resolved
Mon, 29 Mar 2021 16:50:00 UTC
There was an error during expansion of this PowerPoint file. It may be password protected. Consider adding passwords to Password Bank and retrying.

Password Protected File (Discover) - Not Resolved
Mon, 29 Mar 2021 16:49:48 UTC
There was an error during expansion of this OpenOffice XML file. It may be password protected. Consider adding passwords to Password Bank and retrying.

Password Protected File (Text Extraction) - Retried
Fri, 26 Mar 2021 18:59:08 UTC
There was an error during expansion of this PowerPoint file. It may be password protected. Consider adding passwords to Password Bank and retrying.

Password Protected File (Discover) - Retried
Fri, 26 Mar 2021 18:58:27 UTC

Close

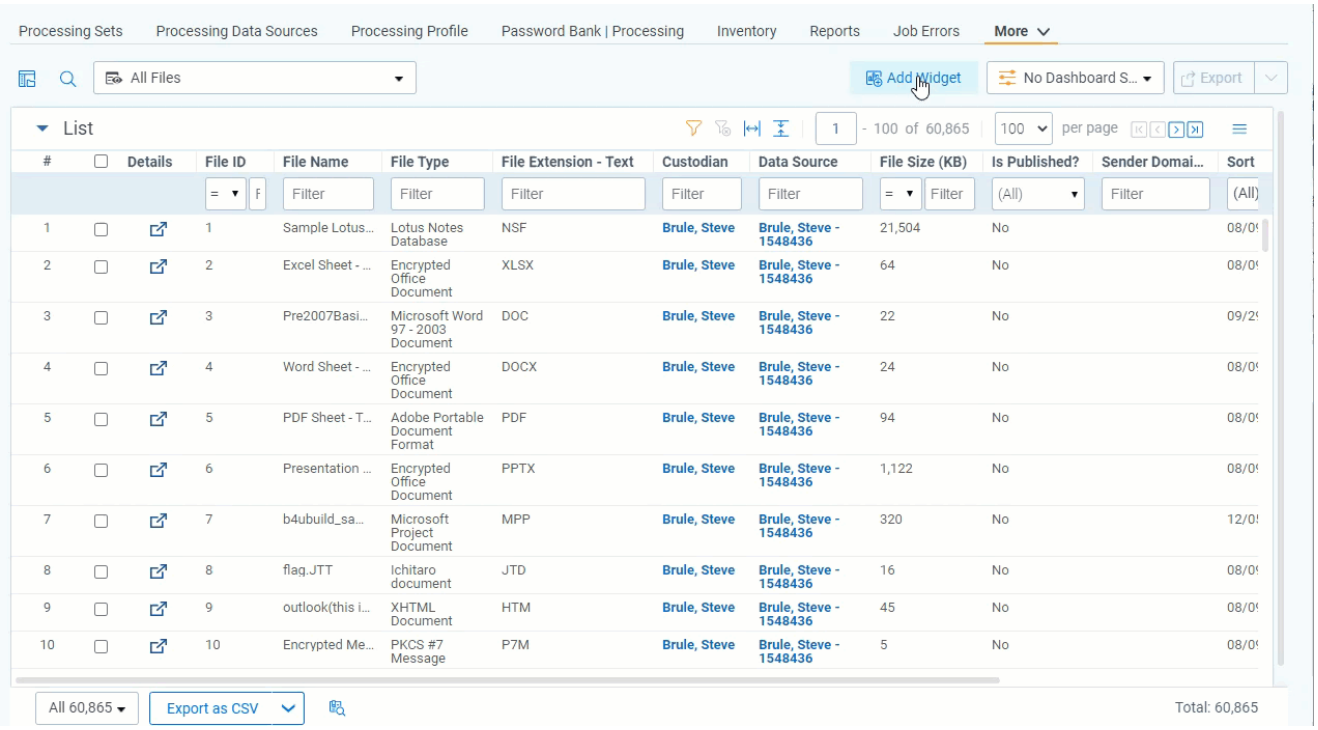
- The Error History section represents all errors that have ever occurred on a file. This acts as a timeline of the record's errors, showing when they occurred, what they were about, and if any are still

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.

16.3.5 Pivotable error fields

By default, all relevant processing error fields are available to group by and pivot on in the Current Errored Files and All Errored Files views of the **Files** tab.



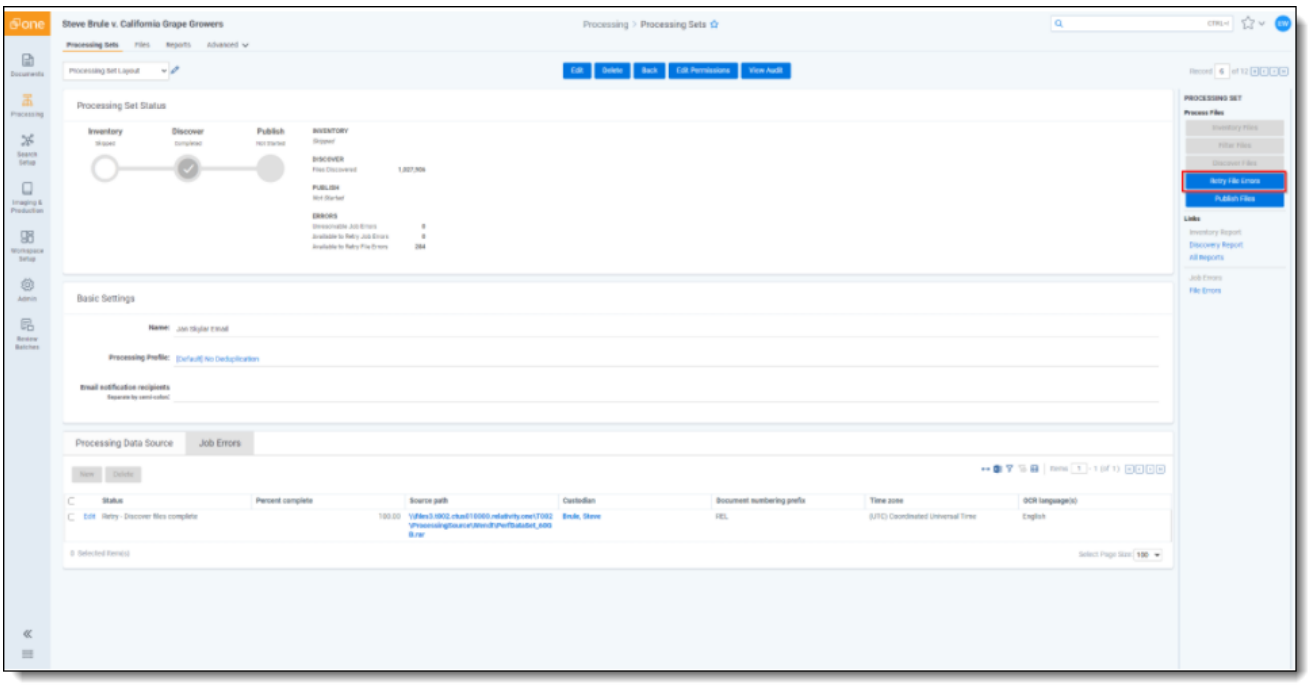
For descriptions of all the fields available for Pivot, see the Files tab.

16.4 File error actions

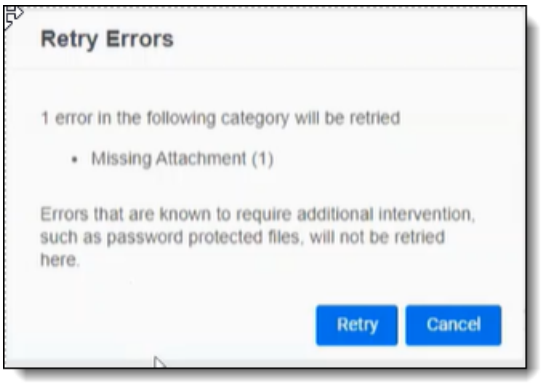
Action can be taken on file errors from the Processing Set page or from the mass operations available on the **Files** tab.

16.4.1 Processing Set error retry

You can retry file errors within the Processing Set by using the **Retry File Errors** button located under the Processing Set console on the right-hand side of the page.



A confirmation message pops up reminding you of the errors you are about to retry. Click **Retry** to proceed or **Cancel** to return to the processing set layout.



Only file errors with a high chance of success will be retried. The probability of success is determined by the error category associated with the file. Error categories such as Corruption or Password Protection will not be retried as they are not likely to be resolved without manual intervention (for example, adding passwords or replacing a corrupt file). A full list of what will and will not be retried can be found below:

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

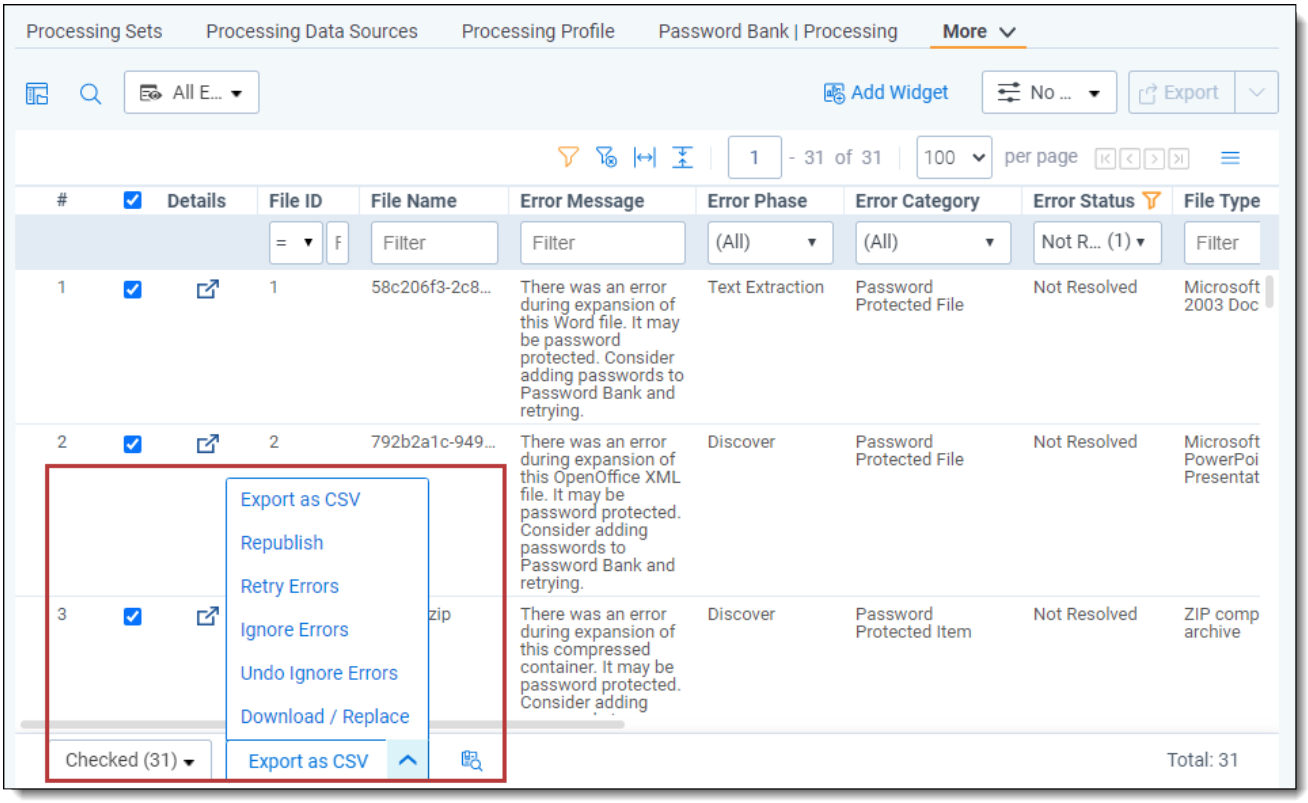
16.5 Files tab error actions

From the Files tab, you can take action on your errored files through the mass operations view.

Mass operation	Description
Export as CSV	This exports the list of processing errors as a CSV file.
Republish	Gives you the option of republishing the files that the errors occurred on. Once you resolve the errors listed, you can use this option, and if the republish is successful, the files will be available in the Documents list, and no errors will be displayed in 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. These issues can be found 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 any discover or text extraction errors are resolved, you will need to manually publish them into your workspace by navigating back to the processing set and clicking the **Retry** button. ■ Not all errors reported in the discovery process can be resolved. This is expected as processing reports all issues it encounters through an error. ■ 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 during

Mass operation	Description
	discovery, that file still has the ability to be published in its current state. The resulting record may have metadata and/or extracted text missing depending on the issue, but it can still be referenced during review. ▪ Relativity automatically retries all publish errors for a set when any error within that set is retried. ▪ Multiple retry attempts cannot be worked on simultaneously. If a secondary retry is submitted while the initial one is still in progress, the second retry will wait in a queue until the first retry is completed. ▪ Only errors with an Error Status of **Not Resolved** can be submitted in a retry job.
Ignore Errors	This provides the ability to set a file's Error Status to Ignored, which will remove it from the Current Errored Files view. The record will still be visible in the All Files and All Errored Files views.
Undo Ignore Errors	This provides the ability to set a file's **Error Status** field back to its original value after it had previously been ignored.
Download / Replace	▪ This provides the ability to download a file to your local machine for investigation. It will also provide the ability to replace an original file with a new version that has been fixed during error remediation. ▪ Note the following regarding downloading and replacing files: ◦ you can only download / 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 limitation for downloading files. ▪ there is a limitation of one gigabyte for uploading replacement files. ▪ performing a replacement of a file will automatically retry its associated errors once completed. ▪ after uploading a new document, when you select Replace & Retry, the native file is updated before you republish. ▪ the retry action for job errors will only retry errors in a Ready to Retry state. For more information on the Download / Replace mass action, see Download and Replace on the **Files** tab page.

The following mass operations are available:



16.6 Common workflows

16.6.1 Identifying and resolving errors

You have completed discovery or publish on your processing set and noticed that it had encountered some errors. You want to investigate and resolve those errors quickly so you can get all possible data into your workspace. Starting from your processing set, perform the following steps:

1. On the right-hand side of the page under Links, select **File Errors** to go directly to the Current Errored Files view on the **Files** tab. Automatic filtering takes you to the files in the previously viewed processing set.
2. On the **Files** tab you can optionally filter down on the errored files that are the most important to resolve. Some common filters are the following:
   - **Error Category** - to group issues of a similar type.
   - **Error Phase** - to group issues that occurred during a particular part of Processing.
   - **Custodian** -if you have a priority Custodian that requires all records to be investigated first.
   - **Sort Date** - to retry files within the matter's relevant date range.
3. Once a group of records is identified to resolve, select the Retry Errors mass action to begin the process. Alternatively, you can retry all errored files without filtering.

4. You can now track your progress of the error retry through the processing set page's progress bar or by navigating to the Worker Monitoring page in Home mode.

For more information on Worker Monitoring, see Processing Administration.

16.6.2 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.

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

Starting from your processing set, perform the following steps:

1. On the right-hand side of the page under Links, select **File Errors** to go directly to the Current Errored Files view on the **Files** tab. Automatic filtering takes you to the errored files in the previously viewed processing set.
2. Locate the file you need to replace. Common techniques are:
   - Filter Error Category for Corrupted File or Corrupted Container.
   - Filter by specific file names.
   - Filter for specific error messages.
3. Select the appropriate checkbox on the left-hand column of the view.
4. Select the Download / Replace option in the mass action menu.
5. From here, two options are available:
   - to inspect and/or repair your container, select the download button
   - once you are in possession of your replacement container, drag and drop it into the modal or select browse for files to locate your container
6. Once the replacement file has been added to the modal, it automatically uploads to Relativity. A quick verification process will let you know if any issues were found or if there were 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 a file had an Author of Steve Bruhle in the original file, but has an Author of Dave Crews in the replaced file, the metadata in Relativity will have Dave Crews filled out.

You can create a field and map it to the 'All Fields - Replaced Extracted Text' non-system field. In this way, you can easily use the field to determine if the document contains an extracted text placeholder. For more information on field mapping, see Mapping processing fields on page 53.

17 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.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

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.

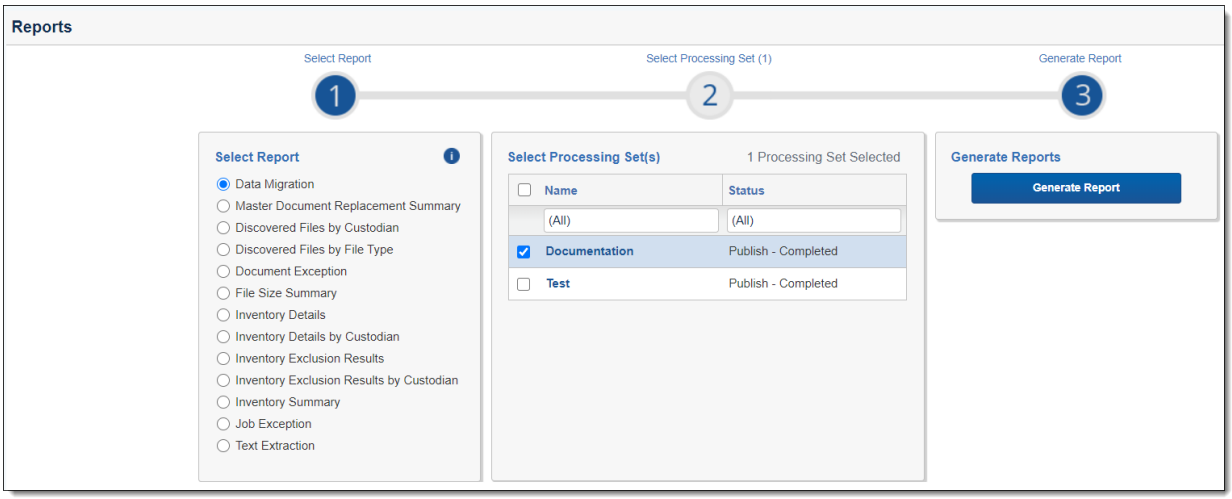
To produce this, go to the Reports tab under Processing and run the Discovered Files by Custodian report for the processing set(s) that your firm suspects are wrong.

17.1 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**.



6. At the top of the report display, you have options to print or save the report. To save, select a file type at the top of the report.

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.

17.2 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.

17.2.1 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.
- **Publishing Errors** - the number of files excluded due to errors during publication.
- **Total Excluded Files** - each custodian's total number of excluded files.

17.2.2 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.

17.2.3 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.

17.3 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.

17.3.1 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.

17.3.2 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.

17.3.3 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.

17.4 Discovery File Exclusion

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.

17.4.1 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.

17.4.2 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.

17.4.3 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.

17.5 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.

17.5.1 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.

17.5.2 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.

17.5.3 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.

17.5.4 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.

17.5.5 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.

17.5.6 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.

17.6 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 processing on page 27 for a list of file types and extensions supported by Relativity for processing.

17.6.1 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.

17.6.2 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.

17.6.3 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.

17.6.4 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.

17.6.5 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.

17.6.6 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.

17.7 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.

17.7.1 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.

17.7.2 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.

17.7.3 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.

17.8 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.

17.8.1 Pre-Processed File Size

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

17.8.2 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

17.8.3 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

17.9 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 256 for more information. You can run this report on inventoried processing sets.

17.9.1 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.

17.9.2 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.

17.9.3 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.

17.9.4 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.

17.9.5 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.

17.10 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.

17.10.1 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:

17.10.2 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.

17.10.3 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.

17.10.4 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.

17.10.5 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.

17.11 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 256 for more information. You can run this report on inventoried processing sets.

17.11.1 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.

17.11.2 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.

17.11.3 Location | Excluded File Count

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

17.11.4 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.

17.11.5 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.

17.12 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:

17.12.1 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.

17.12.2 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.

17.12.3 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.

17.12.4 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.

17.13 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 256 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.

17.13.1 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.

17.13.2 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.

17.13.3 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.

17.13.4 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.

17.14 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.

17.14.1 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.

17.14.2 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.

17.15 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. For more information, see the Admin Guide.

17.15.1 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.

17.15.2 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.

17.15.3 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.

17.15.4 Processing Sets

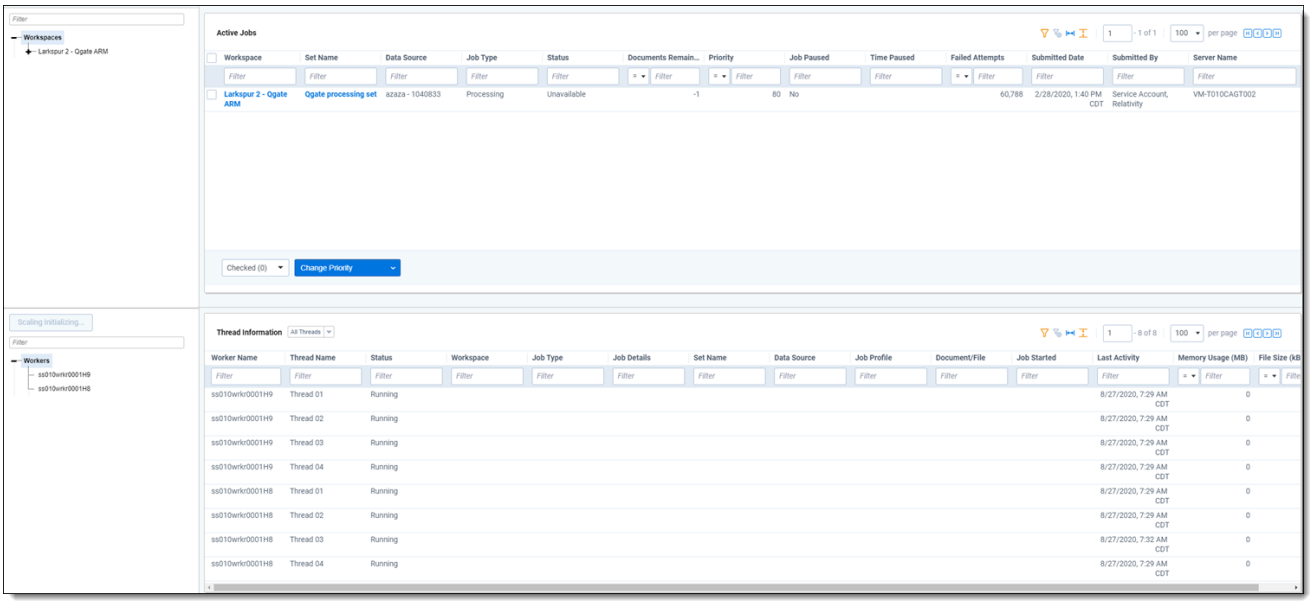
This section lists all processing sets included in this report. Each processing set listed is accompanied by the following information:

- **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.

18 Processing Administration

The Processing Administration tab provides a centralized location for you to access data on active processing and imaging jobs throughout your Relativity environment, as well as the status of all workers assigned to perform those jobs. You can find this information in the Worker Monitoring sub-tab.

You can also use the Processing History sub-tab to identify all actions taken related to processing in your environment.



The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

18.1 Security considerations for processing administration

Consider the following items related to security and client domains (formerly multi-tenancy):

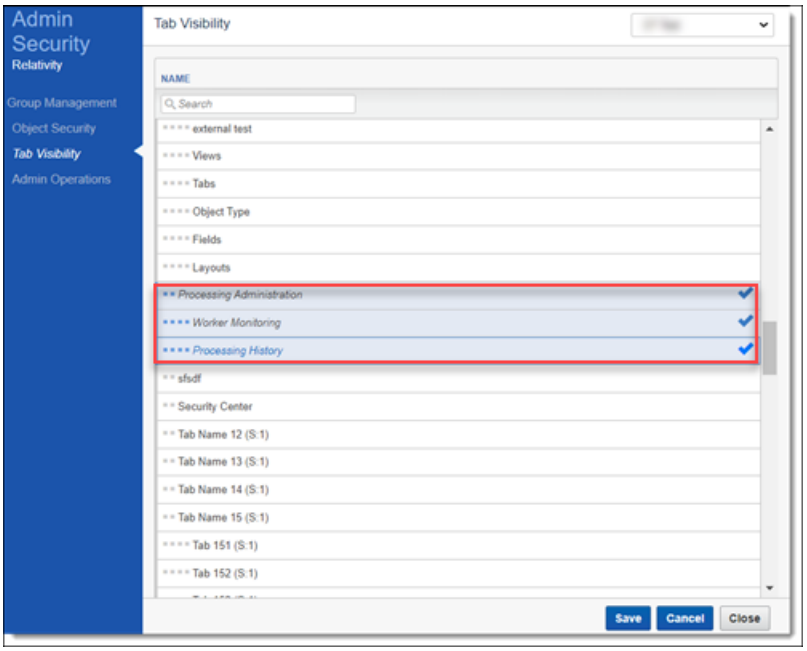
- If you're the system admin for a client domain environment, Relativity makes it so that your tenants can only see jobs in their client domain. This eliminates the possibility of information leaks for workers that don't actually operate within your client domain.
- In client domain environments, users from one client domain can't see any 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: Only System Administrators can modify processing jobs on the Worker Monitoring tab. Other users can see the Worker Monitoring tab with instance level permissions, but will have an error thrown when attempting to modify processing jobs.

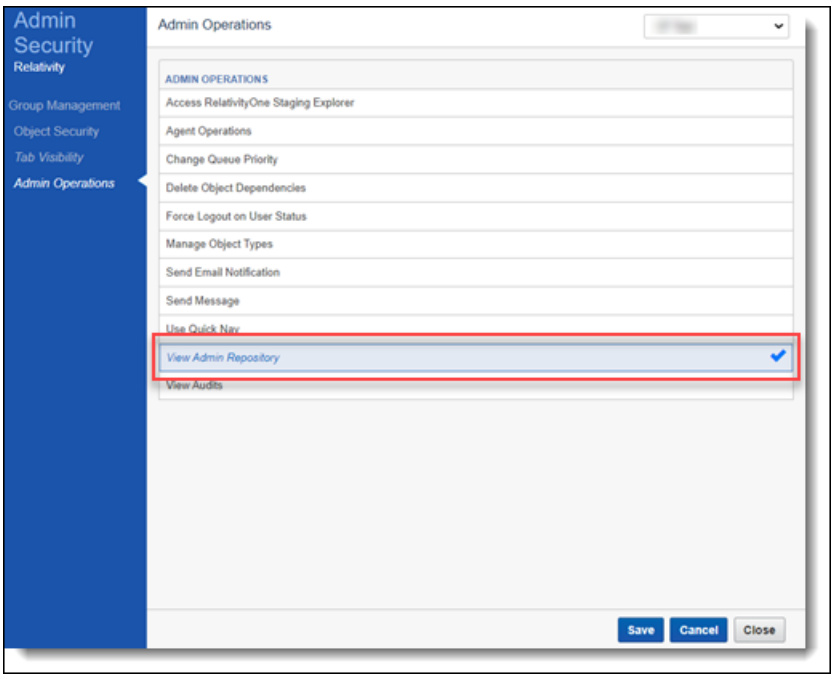
Groups don't have access to the Processing Administration tab or sub-tabs by default. To grant them access, perform the following steps:

1. From Home, navigate to the **Instance Details** sub-tab under kCura Admin.
2. In the Security box, 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 Administration**, **Worker Monitoring**, and **Processing History**.



6. Click **Save**.

You must also have the View Admin Repository permission set in the Admin Operations console in the Instance Details tab to use the Processing Administration tab.

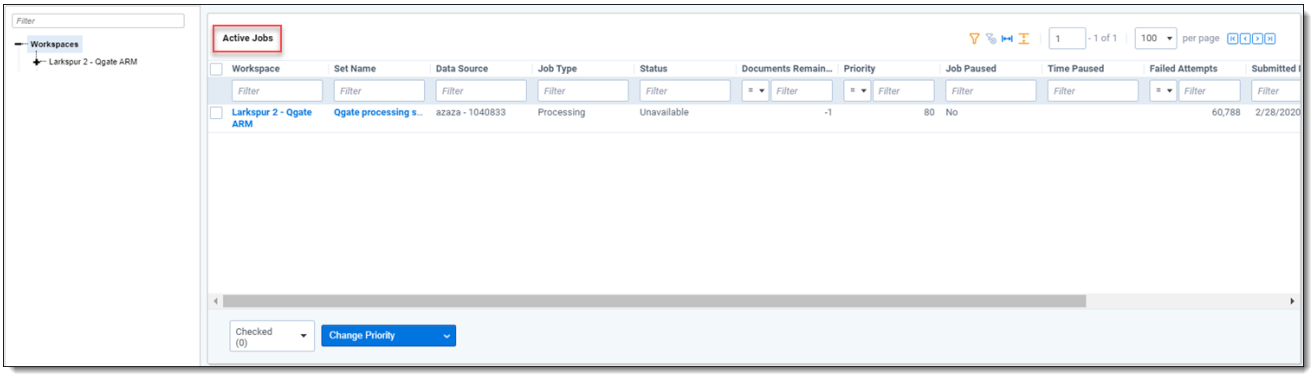


18.2 Monitoring active jobs

To see all active processing and imaging jobs in the environment, view the Active Jobs view in the Worker Monitoring sub-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 will 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** – 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
  - Initialize Workspace
  - 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've clicked **Inventory Files**, **Discover Files**, or **Publish Files** but the job hasn't been picked up yet.

- **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. If you change the priority on a job and there is currently another job in progress, the new job will be picked up and worked on as worker threads become available. Changing the priority only changes the priority for that immediate job.
  - Processing sets are processed in the queue on a first-come, first-served basis.
  - 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 and Single/Mass PDF jobs all have a priority of 100 in the queue. These jobs will be run on a first come, first serve basis with each other).

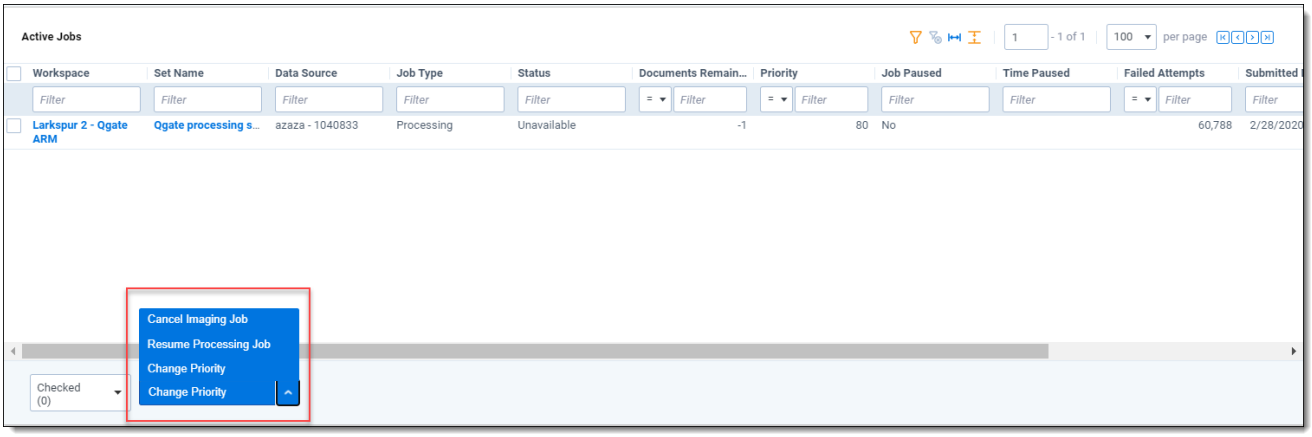
- If you've started a processing job, and you want to start and finish a Mass PDF or Mass Imaging job before that processing job completes, you must go to the Worker Manager Server and manually change the priority of the **Single Save as PDF** choice to be lower than any of the processing choices (**Inventory**, **Discovery**, and **Publish**). Setting the priority of a Mass Save as PDF job or Mass Imaging job must be done before the job begins for it to finish before other processing jobs. For details, see Worker manager server.

- **Job Paused** - the yes/no value indicates whether the job was paused. A paused job typically occurs if there is an issue with the processing agent. You can't 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. You can change this value by adjusting the ProcessingRetryCount value in the Instance setting table.See the Instance setting guide for more information.
- **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.
- **Server Name** – the name of the server performing the job. Click a server name to navigate to the Servers tab, where you can view and edit server information.

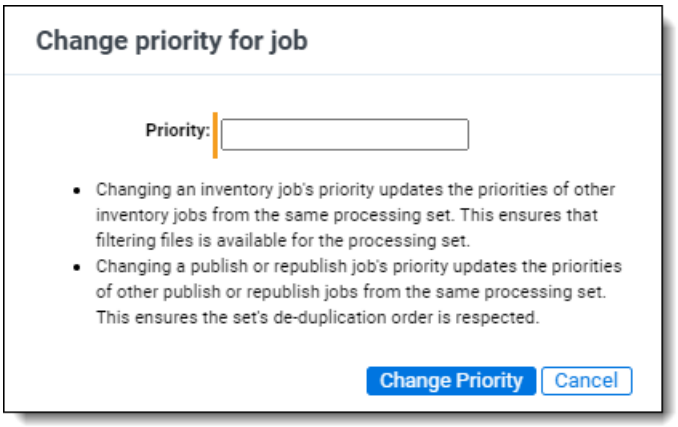
At the bottom of the screen, the active jobs mass operations buttons appear.

18.2.1 Active jobs mass operations

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



- **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.

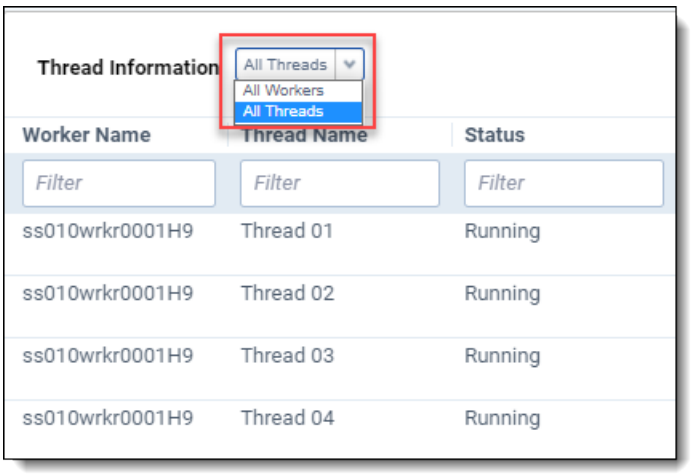


    - If you change the priority of a publish or republish job, you update the priorities of other publish and republish jobs from the same processing set. This ensures that deduplication is performed in the order designated on the set.
    - When you change the priority of an inventory job, you update the 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.

18.3 Checking worker and thread status

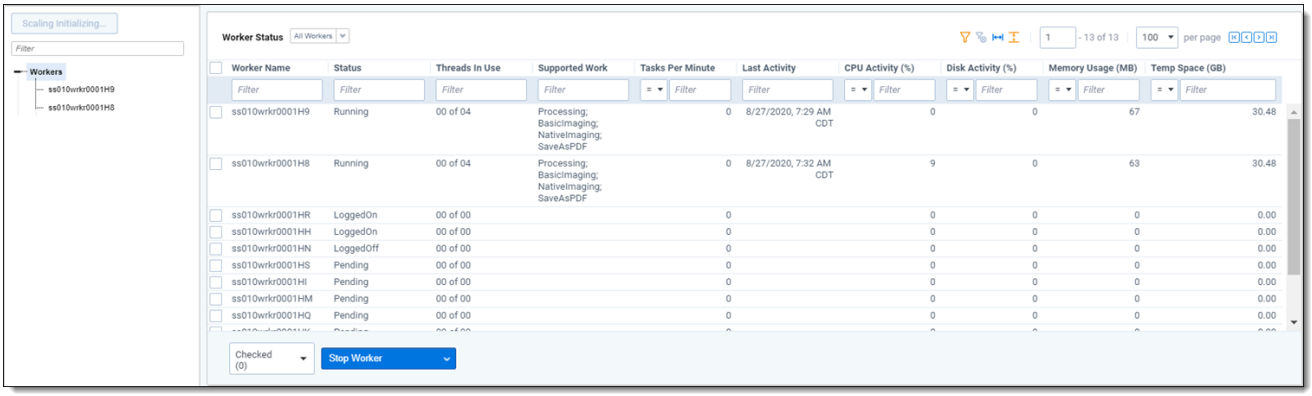
To check on the health of your workers and threads, navigate to the Thread/Worker Information pane at the bottom of the Worker Monitoring tab.

Use the drop-down list at the top of the view to select whether you want to see worker or thread information.



Note: The network on the Utility Server isn't set up to view the status of your workers; therefore, you'll see all workers logged off in the Worker Activity window in the RPC, and you'll need to refer to the All Workers view of the Processing Administration tab in Relativity to see the status of your workers.

The All Workers view contains the following fields:



- **Worker Name**- the name you gave the worker. You can only see workers that are on resource pools to which you have access. For better visibility of a worker's current workload, you can hover over the worker's name to display a call-out containing its CPU activity, disk activity, tasks per minute, temp space (GB), and last activity. These values will reflect those contained in the fields to the right of the worker name.
- **Status** - reflects the current status of the worker.
  - If set to **Service not responding** or **Service call failed**, then the other fields on the view will be empty or display percentages of 0, as they'll have no current job information to report. An idle thread is simply one for which there is currently no active work.
  - If set to **Running**, it means that the worker thread is performing one of its designated jobs.
  - If the queue manager service goes down, the Servers page in Relativity may display a different status value for your workers than what appears on the Worker Status page. This is because the Servers page retrieves the worker's status from the Queue Manager and displays whatever status the worker had before the queue manager went down. When the queue manager is

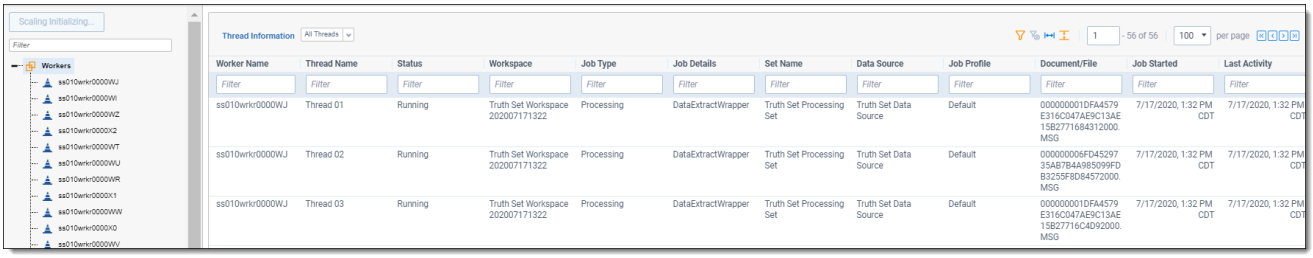
down, there's no way for the Servers page to get an updated status. Meanwhile, the Worker Status page displays a custom error such as Service not responding, and is actually more accurate than the Servers page in this case.

- **Threads in Use** - the number of threads that are busy performing processing or imaging out of the total number of threads available on the worker. This value depends on the configuration of the worker.
  - The maximum total number of threads is 16. The minimum number of threads in use is zero if the worker is idle and not working on anything.
  - The formula for the maximum threads on a worker is 1 thread per 750MB of RAM or 2 threads per CPU up to 16 threads total, whichever thread count is smaller. For more information on worker-thread configuration, see the Admin Guide.
- **Supported Work -** the work for which the worker is designated. This value could be any combination of the following job types:
  - **Processing** - the worker is designated to perform processing jobs.
  - **Imaging** - the worker is designated to perform basic and native imaging jobs.
  - **Save As PDF** - this option is unavailable.

    Note: Relativity performs conversion on designated conversion agents. For more information, see Configuring your conversion agents in the Upgrade guide.
- **CPU Activity** - the amount of CPU resources the worker is using to perform its designated work across all CPU's on the Invariant worker machine. If the actual CPU reading is between 0 and 1, Relativity rounds this value up to 1.
  - In most cases, a high CPU Activity reading with a low Memory in Use is indicative of smoothly running worker that has no stuck jobs.
  - This value can't necessarily tell you whether or not a worker is stuck on a job, but it can tell you whether or not the worker is making a higher-than-normal effort to complete a job relative to other jobs in the environment.
  - If all 16 threads are in use and the CPU is reading at 100% for an extended period of time, you might consider adding additional CPU's to the worker.
- **Disk Activity** - the percentage of disk activity on the worker. If the actual disk activity reading is between 0 and 1, Relativity rounds this value up to 1.
  - If the disk drive is being hit excessively relative to other jobs, it can be an indicator that the worker is either low on memory and burdening the paging file, or it can mean that it is working on a large number of container files.
  - If the disk drive activity is very high relative to other jobs for a sustained period of time, it's probably a sign that you're having an issue with your worker.
  - Low CPU usage coupled with high disk activity can be indicative of low memory or a high container count. You should always address low memory by adding more memory. With a high container count, there's nothing to address, but note that performance can be impacted if the disks are slow.

- **Memory in Use (GB)** - how many GB of RAM the worker is currently using. If the actual memory reading is between 0 and 1, Relativity rounds this value up to 1.
- **Tasks Per Minute** - how many singular units of designated work the worker is performing per minute.
  - Examples of tasks are discovering a document and converting a native.
  - There is no normal range of values for this field, as it depends on what the worker has been designated to do. For slower processes, you'll see a lower number of tasks per minute being performed than for more efficient processes.
- **Temp Space (GB)** - the amount of space in GB that is free on the disk drive assigned to the TEMP environment variable on the worker machine.
  - The value will vary based on the disk's capacity.
  - Only the disk associated with the TEMP environment variable is reflected here.
  - Relativity uses the TEMP folder to temporarily write files while the worker is busy, especially during ingestion.
- **Last Activity** - the date and time at which a thread last communicated to the worker.

The All Threads view contains the following fields:

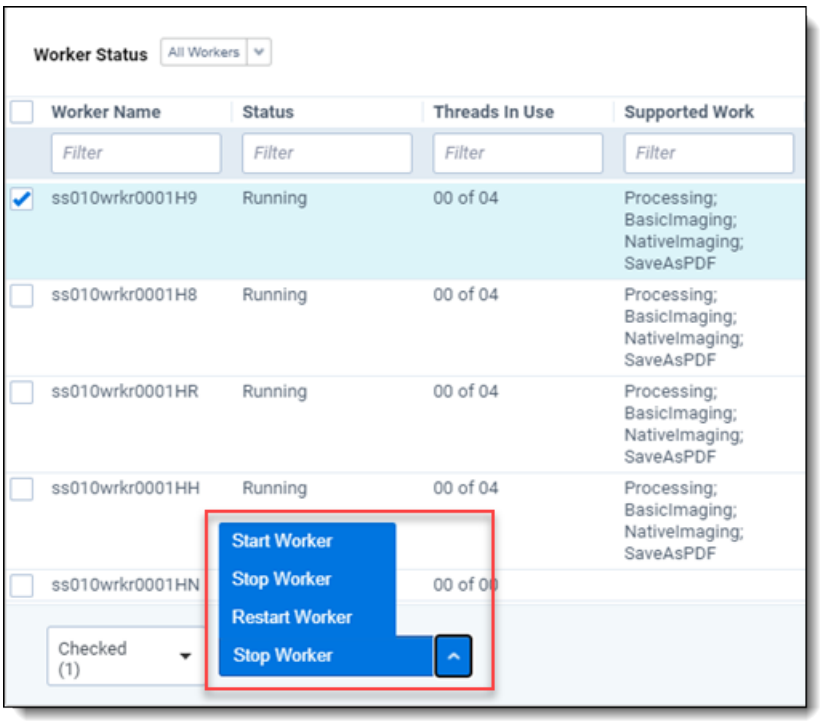


- **Worker Name** - the name of the worker to which the thread belongs.
- **Thread Name** - the name of the thread, as assigned to the worker by the processing engine.
- **Status** - indicates whether the thread is **Idle** or **Running**.
  - If set to Idle, then most of the other fields in the view will be empty, as they'll have no current job information to report. An idle thread is simply one for which there is currently no active work.
  - If set to Running, it means that the worker thread is performing one of its designated jobs.
- **Workspace** - the workspace containing the job that the thread is currently performing.
- **Job Type** - the type of job the thread is performing.
  - This will be some variation of processing or imaging.
  - Depending on the job type, you may not see values populated for other fields in the threads view. For example, an image-on-the-fly job doesn't require an imaging set, so the Set Name field will be empty for this type of job. For more information, see Thread data visibility.
- **Job Details** - the phase or task of the job type that the thread is performing.
  - The value displayed here could be one of many root or sub jobs of the job type running in Invariant.

    - The value displayed here is useful to provide to Relativity support when troubleshooting the worker issue associated with it.
- **Set Name** - the processing or imaging set that the threads are working on. This field isn't populated for image-on-the-fly or mass imaging jobs.
- **Data Source** - the source location of the data being processed or imaged.
    - For processing, this is the data source attached to the processing set that the worker is running.
    - For imaging, this is the saved search selected on the imaging set.
- **Job Profile** - the name of the processing or imaging profile used by the set. This field is blank for image-on-the-fly jobs.
- **Document/File** - the name of the native file that the thread is currently processing, imaging, or converting.
    - This is the name of the file as it exists in its source location.
    - Some imaging jobs may display a value of "Retrieving data" for this field while they gather the data required to display the document name.
    - This field is blank if the status of the thread is Idle.
- **File Size (KB)** - the size of the document that the thread is currently working on. If the actual file size is between 0 and 1, Relativity rounds this value up to 1. This field will be blank if the status of the thread is Idle.
- **Memory Usage (MB)** - how much memory the thread is currently using to perform its work. This field will be blank if the Document/File value reads "Retrieving data" if the status of the thread is Idle or if the system is gathering the data required to display the document name.
- **Job Started** - the date and time at which the processing or imaging job started.
- **Last Activity** - the date and time at which a thread last communicated to the worker.

18.3.1 Worker mass operations

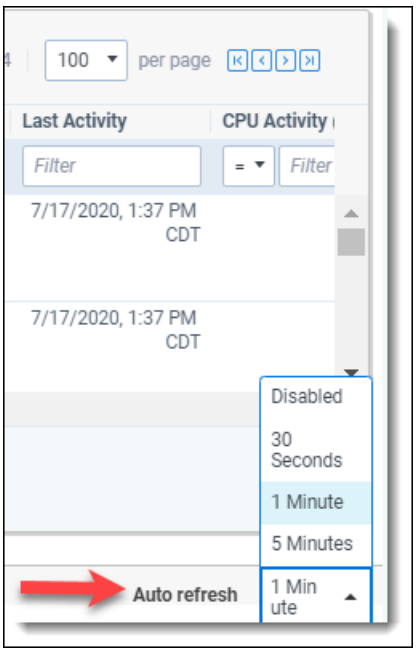
A number of mass operations are available to take on workers from the All Workers view.



- **Start Worker** - starts the selected worker(s), making it available to pick up assigned work from the worker manager server.
- **Stop Worker** - stops the worker(s), preventing the worker from receiving jobs sent from the worker manager server. When you do this, the worker finishes only the tasks it's working on, not the entire job. The remainder of that job is then available to be picked up by another worker.
- **Restart Worker** - restarts a stopped worker, enabling it to receive jobs sent from the worker manager server. Restarting a worker ends the current functional thread and cycles the worker from an offline to an online state.

18.3.2 Auto refresh options

The Active Jobs and All Threads/Workers views receive worker information when loaded and update every time the page refreshes. To configure the rate at which these views automatically refresh, select a value from the **Auto refresh** drop-down list at the bottom right of the view.



- **Disabled** - prevents the automatic refresh of the view and makes it so that job and worker/thread information only updates when you manually refresh the page. This option is useful at times of heavy worker 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.
- **30 seconds** - arranges for the views to automatically refresh every thirty seconds.
- **1 minute** - arranges for the views to automatically refresh every one minute.
- **5 minutes** - arranges for the views to automatically refresh every five minutes.

18.3.3 Thread data visibility

When you access a threads view on the worker status page, not all fields are applicable to all jobs. Therefore, you'll find that certain fields are not populated depending on the type of work taking place.

The following table breaks down which types of workers are populated for each thread field at a particular phase of their work.

Designated work - phase	Set Name	Data Source	Job Profile	Document/File	File Size (KB)
Processing - Inventory	✓	✓	✓	✓	✓
Processing - Discovery	✓	✓	✓	✓	✓
Processing - Publish	✓	✓	✓	✓	✓
Imaging - Imaging set	✓	✓	✓	✓	✓
Imaging - Image-on-the-fly				✓	✓
Imaging - Mass image			✓	✓	✓

18.3.4 Errors

Errors can occur when Relativity attempts to communicate worker information to you as it receives that information from Invariant.

- **Service not responding** - the queue manager service on the worker manager server is down or not accessible.
- **Service timed out** - the call to Invariant timed out.

  Note: The WorkerStatusServiceTimeout entry in the Instance setting table determines the number of seconds before calls from the worker status API service to Invariant are timed out. If you encounter an error related to the service timing out, it means that the value of this Instance setting table entry has been reached. By default, this is set to five seconds.

- **Service call failed** - an unspecified failure that is most likely due to an old version of Invariant that doesn't have worker status API call being installed on your machine. This gets logged in the Errors tab in Relativity.

Note: For optimal performance, the processing engine caches worker and thread data for 30 seconds. If you refresh the page within the 30 second time period, the same cached data will display until the cache expires and new data is retrieved from Invariant upon refresh or when loading the page.

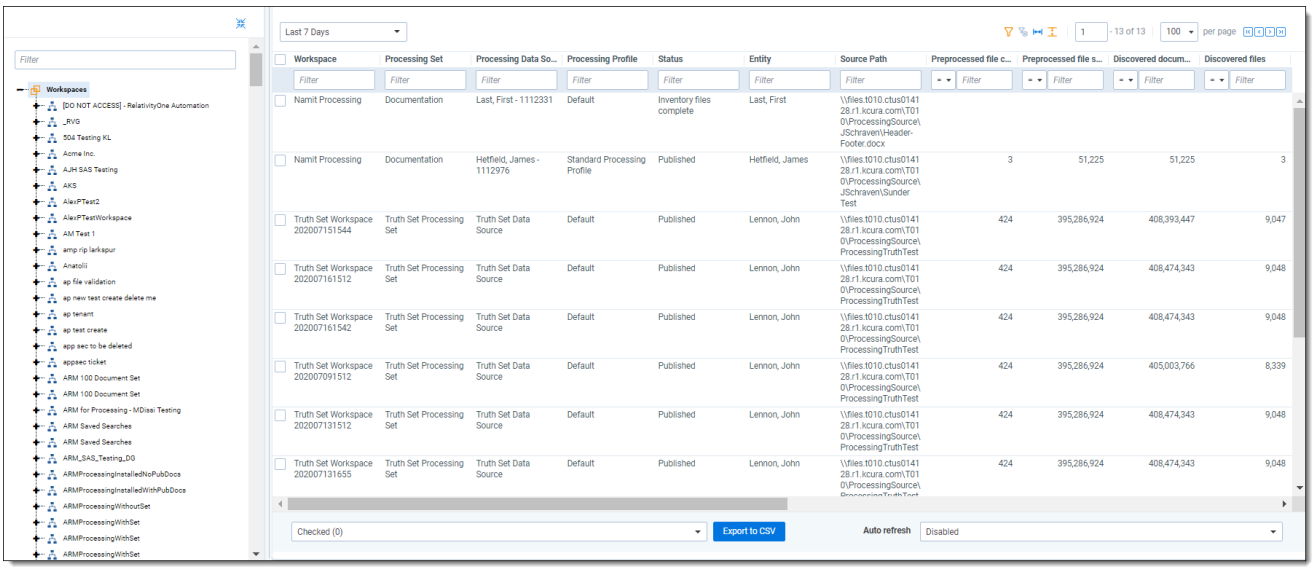
18.4 Using the Processing History sub-tab

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.

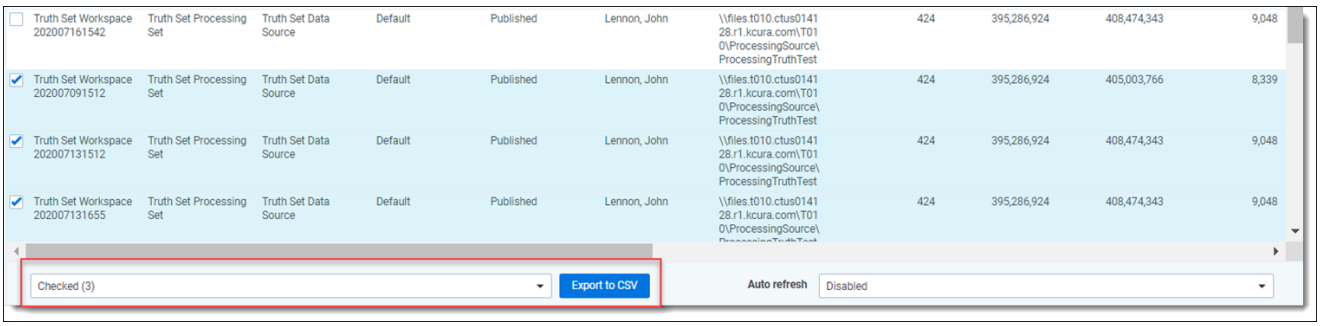
The Processing History view provides the following fields:



- **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.

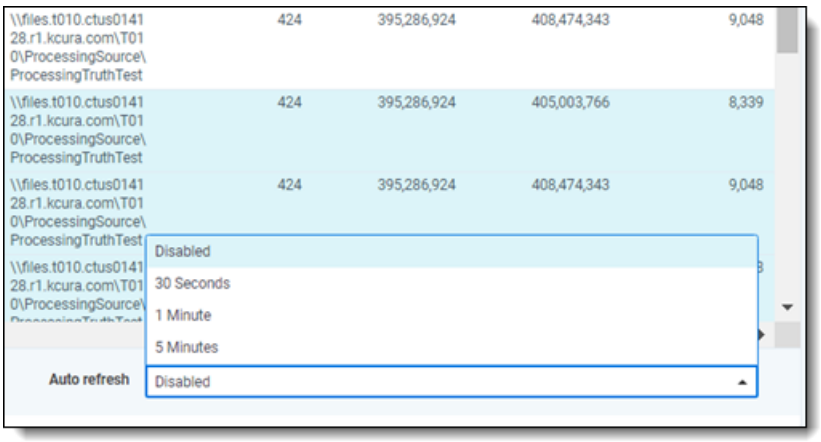
You have the option of exporting any available processing history data to a CSV file through the Export to CSV mass operation at the bottom of the view.



18.4.1 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.



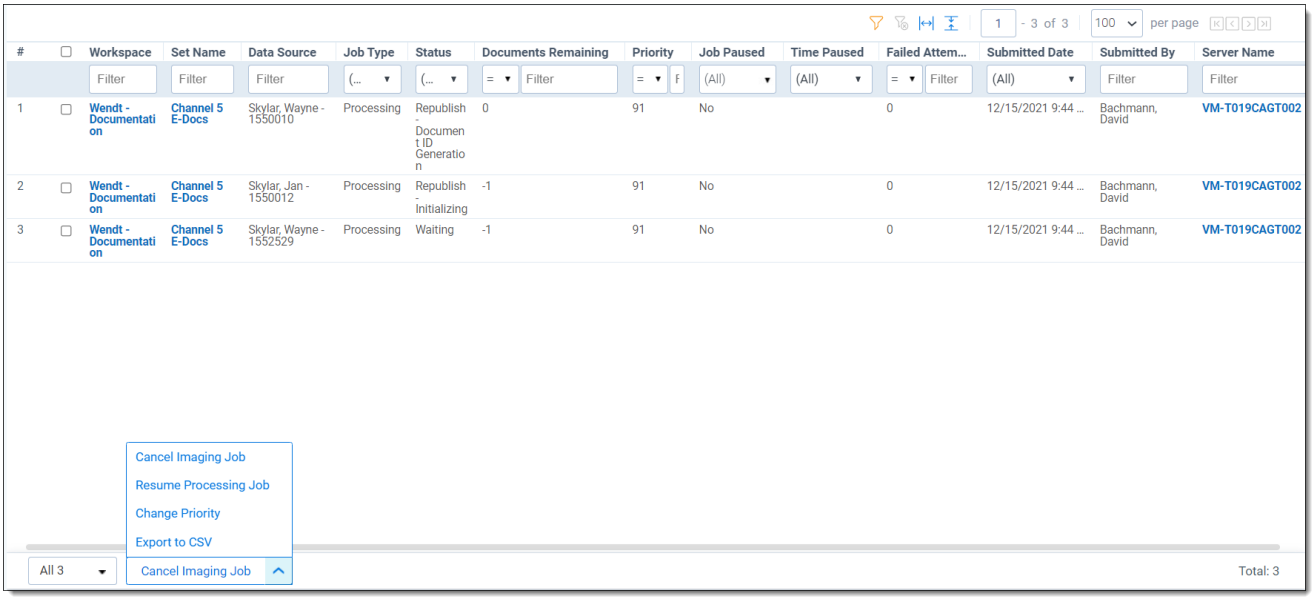
- **Disabled** - prevents the automatic refresh of the view and makes it so that processing history information only updates when you manually refresh the page. This option is useful at times of heavy processing usage, in that it offers you more control over the refresh rate and prevents the information 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.

19 Managing processing jobs in the queue

When you start a processing job, that job eventually goes to the Worker Manager Queue, which you can access through the Queue Management tab to view the status of your processing jobs and change the priorities of jobs that you know need to be completed before others in your environment.

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

Note: Processing jobs get the same priority in the queue as native imaging jobs. TIFF-on-the-fly jobs, however, take precedence over both processing and native imaging.



The following columns appear on the Worker Manager Queue sub-tab:

- **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 as *<Custodian Last Name>, <Custodian First Name> - < Artifact ID>* if you left this field blank when creating the data source.

- **Job Type** – the type of job running. The worker manager server handles processing and imaging jobs.

  Note: When you click Filter Files on the processing set console, you're performing an intermediate step that is not considered an actual job type by Relativity. For that reason, filtering files is not displayed in the Worker Manager 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.
- **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've clicked **Inventory Files**, **Discover Files**, or **Publish Files** but the job hasn't been picked up by a worker yet.

- **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.
  - Processing sets are processed in the queue on a first-come, first-served basis.
  - Discovery, publishing, and imaging jobs are multi-threaded and can run in parallel, depending on the number of agents available.
  - Processing jobs get the same priority in the queue as native imaging jobs. TIFF-on-the-fly jobs, however, take precedence over both processing and native imaging.
  - Publishing jobs take priority over discovery jobs by default.
- **Job Paused** - the true/false value indicates whether the job was paused.
- **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. You can change this value by adjusting the ProcessingRetryCount value in the Instance setting table. See the Instance setting guide for more information.
- **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.
- **Server Name** – the name of the server performing the job. Click a server name to navigate to the Servers tab, where you can view and edit server information.

At the bottom of the screen, the following buttons in the drop-down appear:

- **Cancel Imaging Job** - cancel an imaging job. Only imaging jobs can be canceled from the Processing Queue sub-tab. If you have processing jobs selected and you click **Cancel Imaging Job**, the processing jobs are skipped.
- **Resume Processing Job** - resumes any paused processing jobs that have exceeded the failed retry attempt count. To resume a paused processing job, check the box next to the data source(s) that you need to resume and click **Resume Processing Job**. You can resume multiple jobs at the same time.

- **Change Priority** - change the priority of processing jobs in the queue.
  - 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.

If you click **Discover** or **Publish** on the Processing Set Layout, but then cancel the job before the agent picks it up, you can return to the set and re-execute the discovery or publish job.

20 Processing FAQs

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 do you fix an error after a document is published?

In Relativity 8 and above, you can retry the file again.

For a version of Relativity prior to 8.0, fix the file and then ingest it in its own processing set.

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.

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.