Relativity®

Relativity Processing Console Guide

December 5, 2025 | Version 25.xxx.x

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

Table of Contents

1 Relativity Processing Console

The Relativity Processing Console is a desktop application that lets you control complex processing jobs on a granular level. The RPC offers power processing users to extend the full capabilities of Processing.

1.1 RPC features

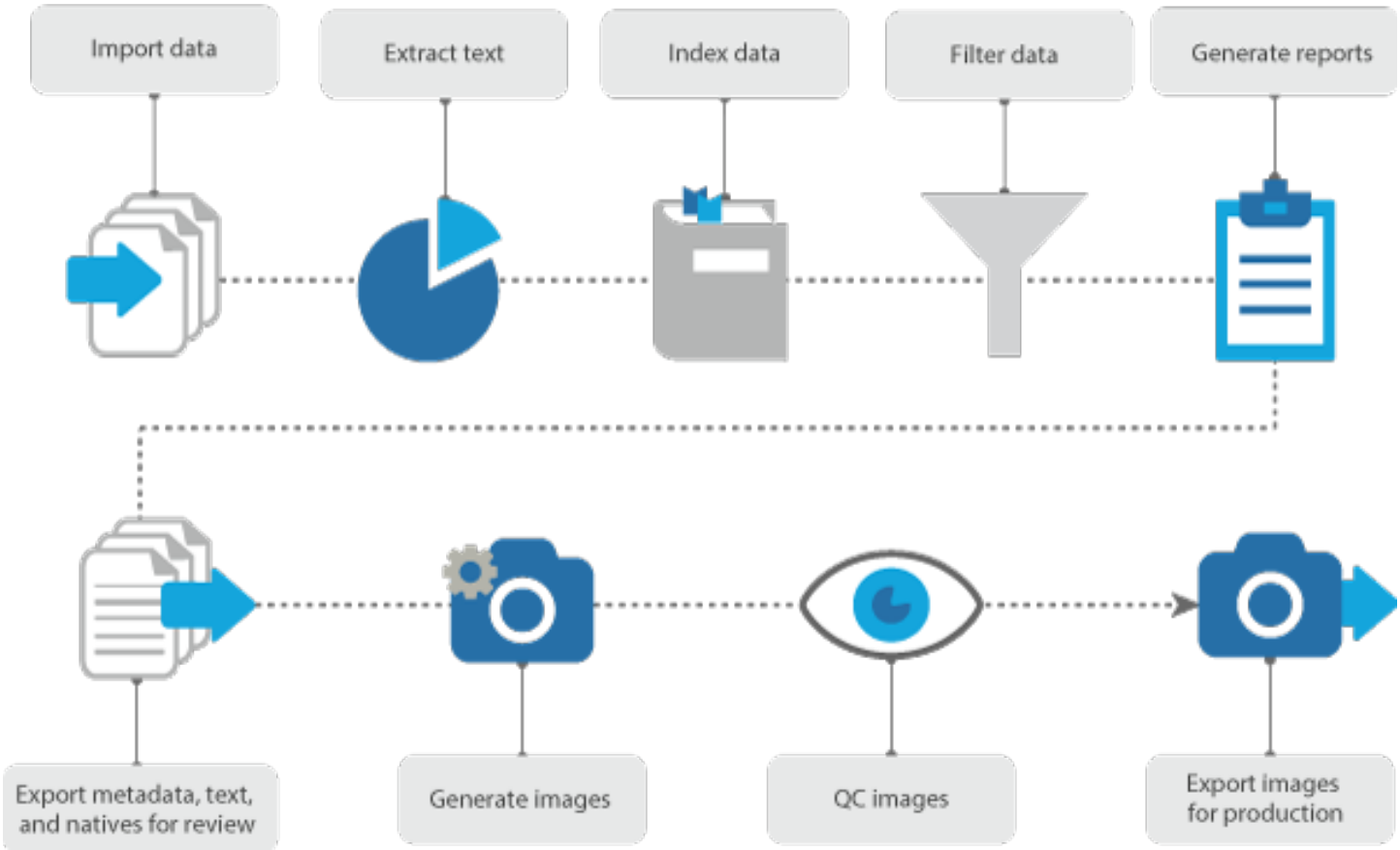
The RPC includes, but is not limited to the following features:

- **Granular worker and job control** – prioritize worker servers into work groups, which you can allocate to work on specific jobs or processing functions.
- **Selective import of data** – choose specific folders and files to import for a given custodian, import or exclude specific files by extension, De-NIST documents.
- **Extract text** - able to extract text directly from documents, images, or files with embedded images OCR'd automatically.
- **Indexing options** - uses the dtSearch as its full-text search engine, automatically indexes all metadata along with extracted text to ensure nothing is overlooked.
- **Filter options** - multiple filtering options, including date/time, de-dupe with extremely flexible options, cross reference, expression, and full text.
- **Reporting** – create several reports, including de-dupe filter, error, summary, search filter.
- **Generate images** – render documents as PDFs without the need for print drivers, significantly improving performance. Imaging jobs are also multi-threaded, and you can monitor worker activities.
- **QC Functionality** – preview all images and extracted text - or a random sample - track progress, auto-advance images, and flag them for further review or replacement.
- **Full metadata access** – all extracted metadata fields are identified and stored in individual fields for potential use.
- **Robust export options:**
  - Numerous filtering options, including date/time, cross reference, expression, and full text.
  - Customize and brand/endorse the images for export.
  - Select metadata fields.
  - Format fields – for example, specify date or date/time for date fields.
  - Variable substitution – for example, specifying the return of a variable in a specific field.
  - Build switch statements – for example, similar to a v-lookup, "when x is found return y".
  - Track status.

1.2 Running an RPC job

There are many ways to run a job using the RPC. Below is a typical workflow. Note that you're free to deviate from it as necessary to meet your own needs.

RPC Workflow



Depending on your needs, you can re-order or skip some of these steps completely. For example, if the filtering you require doesn't include keyword searching, you have the option of skipping indexing. You can also skip extracting text, indexing, and filtering and jump directly to imaging, and you'll still be able to export text. Image generation also creates page-level text by default, which you can export either on a page or document level, as needed.

It's important, however, to know the prerequisite steps for the tasks you intend to perform. The following table breaks down those prerequisites.

Desired RPC task	Prerequisite
Extracting text	Imported data • Can be done immediately after import
Generating images	Imported data • Can be done immediately after import
Indexing data	Either text extraction or image generation
Full-text search filtering	An indexed import
All other filtering	Imported data • Can be done immediately after

Desired RPC task	Prerequisite
	import • No text extraction or imaging is required
Exporting metadata	Imported data • Can be done immediately after import
Exporting images or page level text	Image generation
Exporting doc level text	Either text extraction or image generation
Exporting native files	Imported data • Can be done immediately after import
Exporting PDF files	Image generation

1.3 Special considerations

You can use the RPC in conjunction with Processing, or as a standalone tool. For example, you might use the RPC to monitor and troubleshoot jobs that you kick off from Processing, or run disparate jobs solely in the RPC.

However, consider the following when using the RPC in conjunction or disparately from Processing.

- The Inventory feature only exists in Processing.
- The ingestion filtering feature only exists in Processing because Inventory only exists in Processing. However, it is still possible to De-NIST an import or include/exclude specific file extensions from an import job using the RPC.
- You can't direct export to Relativity for review at this time.
- Reports are different than Processing reports.
- You can't cancel a job in the RPC.
- You can use the RPC to track a processing job in Processing, however, the RPC doesn't capture the following in Relativity:
    - Re-leveling
    - Retrying and/or clearing errors
    - Altering document extracted text
    - Changing the parent/child document relationship
- Changing the following RPC settings on a processing job may cause unexpected behavior in Processing as there may be conflicting settings on the same data sets:
    - Adding store-level settings
    - Adding project-level settings

1.4 Logging in to the RPC

To provide visibility into who has accessed the RPC and for how long, Relativity requires you to provide user authentication to log in to the RPC. Any active Relativity user can log in to the RPC with their Relativity credentials.

This functionality was introduced in Relativity 9.4.398.62.

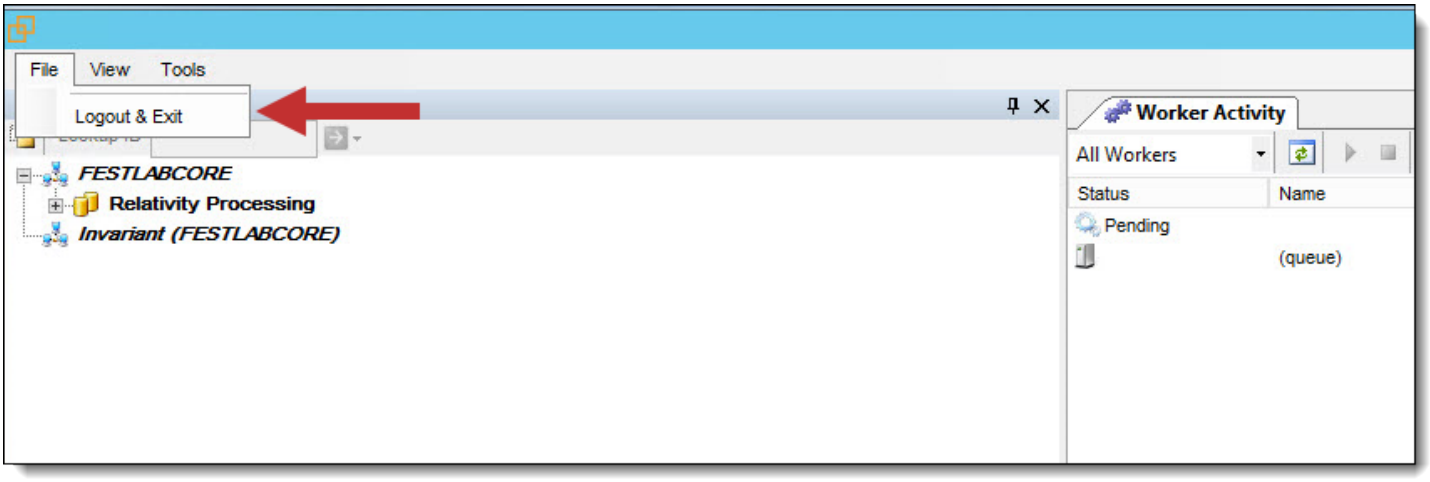
To log in, click on your instance of the RPC to bring up the login screen, enter your Relativity username, and click **Continue**.

Enter your password and click **Login**. If you use RSA authentication to log in to the RPC, make sure that the server on which the RPC is installed is configured for RSA, or you won't be able to log in.

Note: If you don't see the login screen, or your login attempt(s) fail, go to the machine on which the RPC is installed, get the identity server URL, open a browser on the machine where the RPC is installed, enter the identity server URL, and attempt to log in again. The identity server URL is case sensitive.

You now have access to all RPC functionality.

To log out of the RPC, navigate to the **File** tab and click **Logout & Exit**.



Note: Relativity records RPC logins and logouts in the user's audit history; however, it doesn't designate those actions as being RPC-specific. In this case, you could see duplicate records for "Login" and "Logout" in your audit history, one for Relativity and another for the RPC.

2 Installing the RPC

Before you begin installing the Relativity Processing Console (RPC), you must update your environment with the required security and other configuration settings. See the System requirements guide.

2.1 Licensing

You must have a Processing license and be running Server 2025 or above in order to use the Relativity Processing Console. See the Licensing guide.

2.2 Distribution

For RPC distribution, contact Customer Support.

2.3 Pre-installation requirements

The RPC is an application that you can install to your desktop on a local machine or to the same server on which the Queue Manager is installed. The machine must be on the same domain as the processing worker and file servers, and the SQL Server.

Before running the RPC installer, verify that the items in the following checklists have been completed. This ensures that you won't encounter any disruptions during the RPC installation process:

Web UI	Done
You're running Relativity 8.2 or above.	
Processing is installed to at least one workspace.	
You have a valid Processing license.	
The RPC version matches the Invariant version.	
At least one Worker is running and is designated for processing work.	
The version of Processing installed via the Servers tab is correct.	
The version of the RPC Installation package corresponds to the Relativity version.	
IE Enhanced Security Configuration is off.	
Authentication process is now mandate client registration.	
Use of a local admin user account, and running RPC with *Run as administrator*.	

Processing SQL Server	Done

In SQL Server Management Studio, the DataFiles location is set in the Invariant.dbo.AppSettings table. This should be a network share accessible to the Relativity Service Account:
`SELECT Value1 from Invariant.dbo.AppSettings`
`WHERE Category = 'DataFiles'`	

In SQL Server Management Studio, the dtSearchPath location is set in the Invariant.dbo.AppSettings table. This should be a network share accessible to the Relativity Service Account: `SELECT Value1 from Invariant.dbo.AppSettings`	

Processing SQL Server	Done
`WHERE Category = 'dtSearchPath'`	

In SQL Server Management Studio, the IdentityServerURL is set correctly in the Invariant.dbo.AppSettings table. This value is case sensitive. If it isn't entered in the proper case, you can run the following statement to capitalize "relativity":
`--begin tran`
`--update Invariant.dbo.AppSettings`
`--set Value1 = 'https://<WebServerName>/Relativity/Identity'`
`--where Category = 'IdentityServerURL'`
`--commit`	

The user (or group) account running the RPC from the desktop has access to SQL Server Management Studio via Windows Authentication and has to have the following server roles assigned: • bulkadmin • dbcreator • public	
All pre-existing and newly created processing server databases (Invariant and Store databases) have RPC users mapped to the SQL Server logins with the db_owner permission set.	

Local machine	Done
The account you're logged in to has a Windows Authentication login on the Processing SQL Server.	
The account running the RPC is a local administrator on the worker servers. This is required to stop a worker and take a worker offline.	
The following four SQL CLR type packages are installed locally, as well as the Microsoft Report Viewer 2012. These must be installed before you install the Microsoft Report Viewer 2012: • CLR Types for SQL 2012 (x86 32 bit) • CLR Types for SQL 2012 (64-bit)	
The Microsoft Report Viewer 2012 is installed locally: • Microsoft Report Viewer 2012	

Before running the RPC installer, you are required to first whitelist and then register any processing-related machines for the Secret Store so that those machines can read and write to and from the store. This includes the Invariant queue manager, all workers, and all machines on which the RPC is installed. See The Relativity Secret Store Guide for more information on configuring servers for the Secret Store.

Once you've verified or completed all of the above items, you can move on to running the RPC installer.

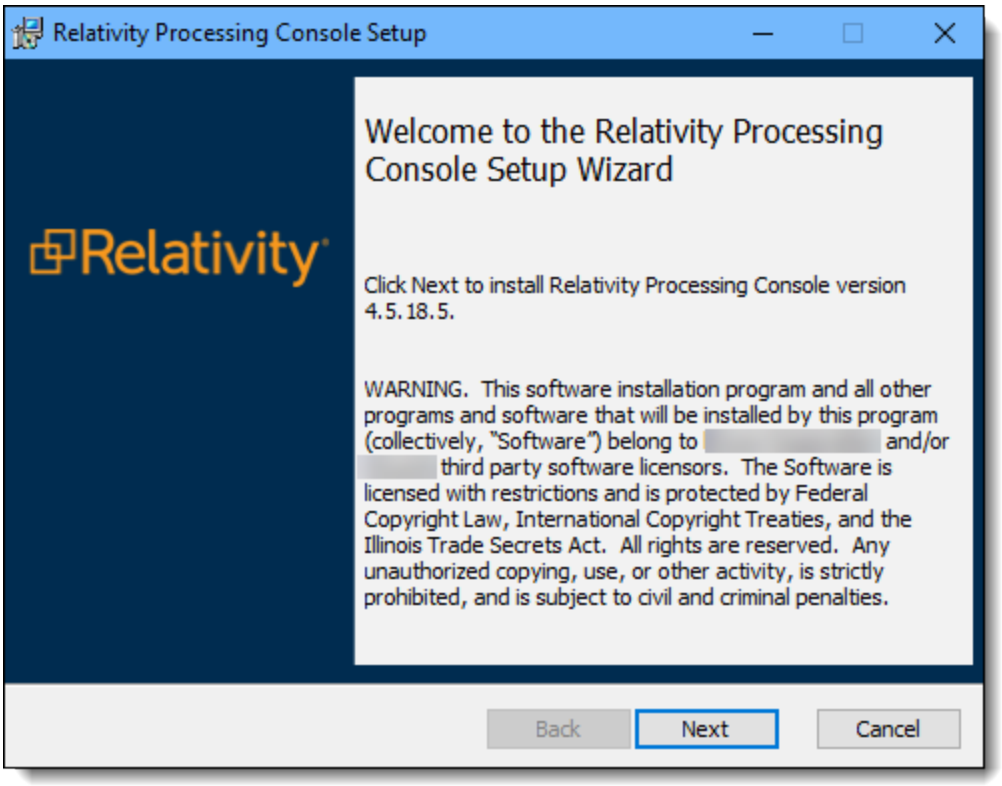
2.4 Installing the RPC

The RPC is available in a 64-bit version and is only available on a local install, not on a network install.

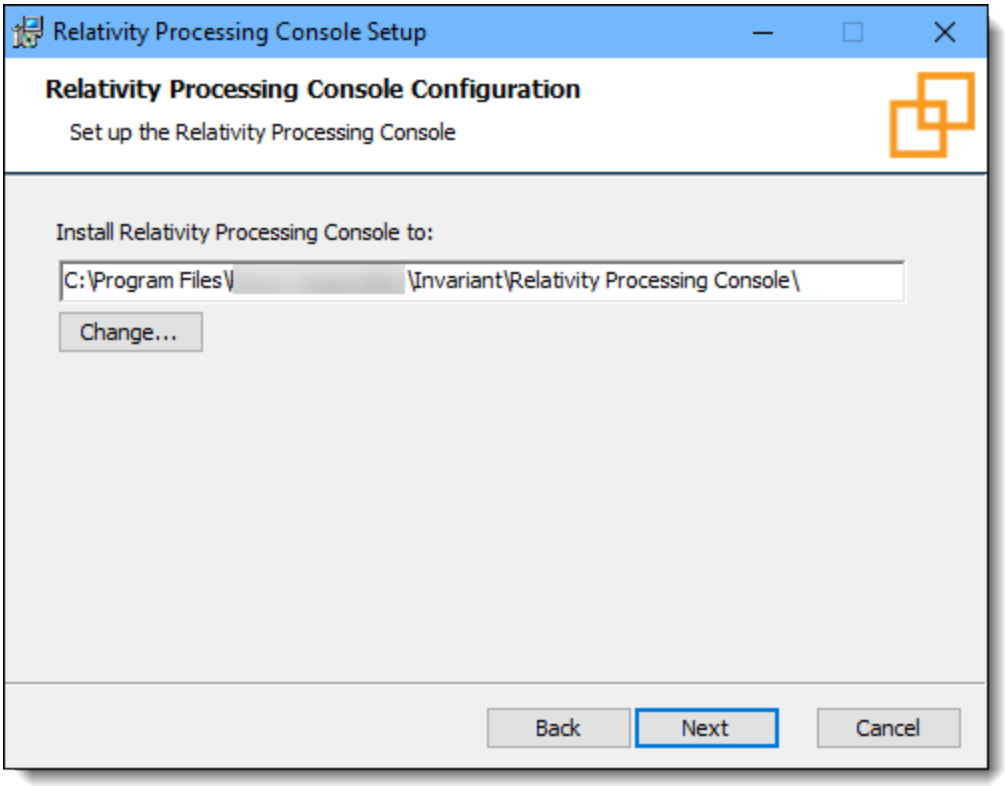
To install the RPC:

1. Log in to the machine with the local admin account where you want to install the RPC.
2. Open the installer executable to launch the setup wizard.

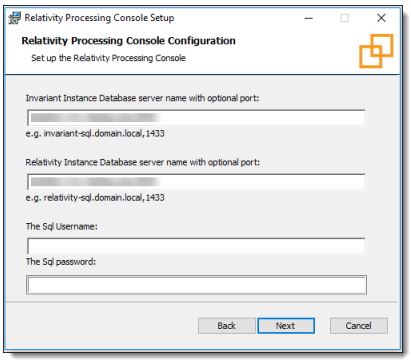
3. Click **Next** on the Welcome screen.



4. Enter the installation directory for the RPC. This path must be 200 characters or fewer.

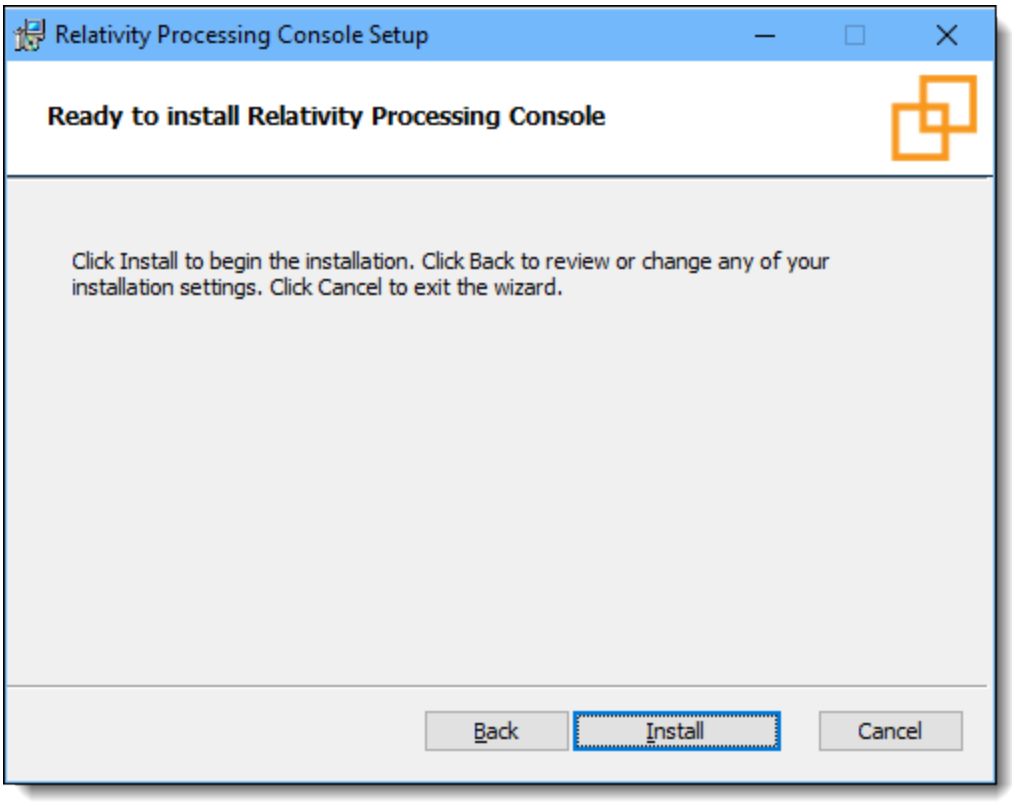


5. Click **Next** and enter the following database information:



   - Invariant instance database server name with the optional port number
   - Relativity instance database server name with the optional port number
   - Sql username must be EDDSDBO
   - Sql password

6. Click **Next** and then click **Install**.



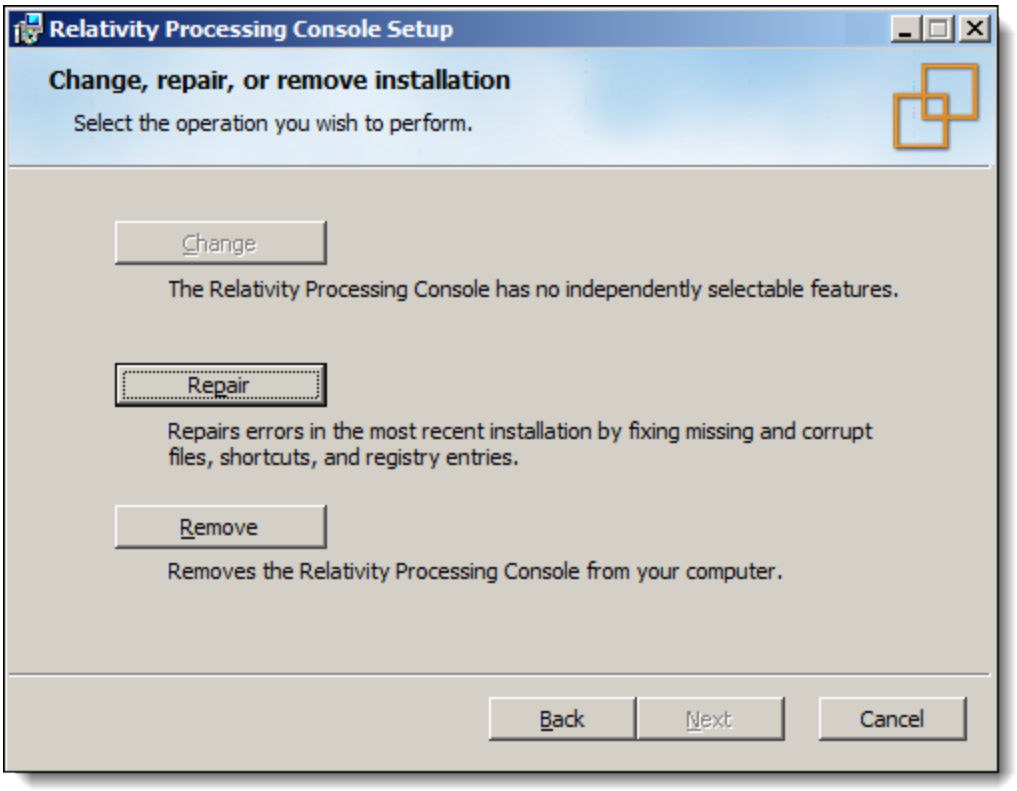
2.5 Validating the RPC installation

Before you can use the RPC for the first time, you must successfully submit one processing or inventory job from Processing.

2.6 Repairing or uninstalling the RPC

Use the installation wizard to repair or uninstall the RPC and the Queue Manager Server.

1. Log in to the machine with the local admin account where the RPC and the Queue Manager are installed.
2. Open the installer executable to launch the setup wizard.
3. Click **Next** on the Maintenance screen.
4. Click **Repair** or **Remove** to modify your installation.



- **Repair** - click this button to repair your RPC installation. The installer adds a fresh copy of any deleted RPC files, but it doesn't make any configuration changes or replace other files that already exist. On the **Ready to Repair the RPC** window, click **Repair**.
- **Remove** - click this button to uninstall all RPC components from your machine. On the **Ready to Remove the RPC** window, click **Remove**.

2.7 Upgrading the RPC

To upgrade the RPC, run the latest version's installer. You don't need to uninstall your current RPC version.

3 Importing data

You can use the RPC to import data into a store. The RPC extensively uses the job queue to distribute the import workload across multiple machines in order to maximize processing resource efficiency.

3.1 Creating a data store

Before you can import data, you must create a data store in the RPC, which adds a new database to the SQL Server. Data stores allow you to organize your data by a store, group, project, and custodian. This organization allows you to optimally use the RPC when you perform text extraction, generate images, filter, and export.

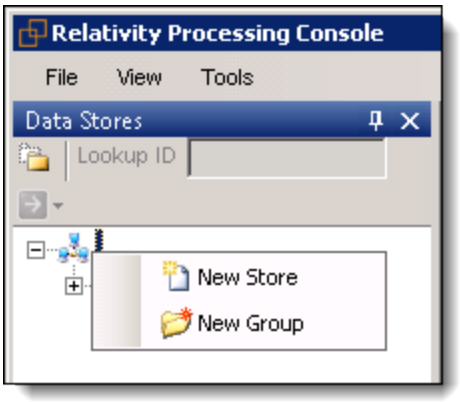
Note: You can't create a data store enabled for DataGrid.

Any databases that you create in Relativity also appear in the RPC. See Working with Relativity-generated data stores below.

Note: The RPC is designed to work in distributed environment across multiple SQL, file, and worker servers. You may see multiple SQL Servers listed in the Data Store window.

To create a data store:

1. Open the RPC.
2. In the Data Stores window, right-click on an SQL Server.



3. Choose one of the following menu options:

- **New Store** - creates a new database on the server.
- **New Group** - creates a group that you can use to organize data stores. For example, you can add all data stores for the same customer to a group.

Note: You can delete groups but not data stores. You can hide a data store by right-clicking the data store and select **Hidden**.

Once you've created a data store, you can import data. See Importing data on the next page.

3.1.1 Working with Relativity-generated data stores

You can use the RPC export feature to generate a load file from a Relativity-generated data store. For example, you may want to create a load file for Relativity import. See Exporting data on page 59.

3.2 Importing data

Once you create a data store, you can add documents to the RPC by importing them to a data store. The RPC extensively uses the job queue to distribute the import workload across multiple machines in order to maximize processing resource efficiency.

The RPC generally uses the following workflow to import data (such as a PST file) from a hard drive:

1. Identifies the file type, and calculates the hash for the file so that it can be copied to the network. These hashes are later used for deduplication filters. See Filter types on page 45.
2. Calls the specific handler for the file type, and passes in the required parameters. Each file type has its own handler (or plugin) that requires its own set of parameters.
3. Adds job to the queue so worker can execute the method for opening the file (such as a PST), count the items in it, and add 100 item groups to the job queue.
4. Extracts first group of, for example, 100 items from a container file and adds them to the queue as a job. Continues with the second group of 100 items, and so on (Multiple workers may be running these jobs.)
    - You may see something different than a 100 items in your queue for a container file. This is because for large container files, the RPC breaks the job into a number of smaller sub jobs, and this number is not fixed. It varies, as it depends on the original size of the container file.
5. Adds jobs for attachments to the queue, jobs for metadata extraction, and so on. This process continues until there are no remaining sub tasks to add to the queue.

The RPC extracts only metadata from the files during an import job. See Extracting text on page 22 for more information.

3.2.1 Import jobs

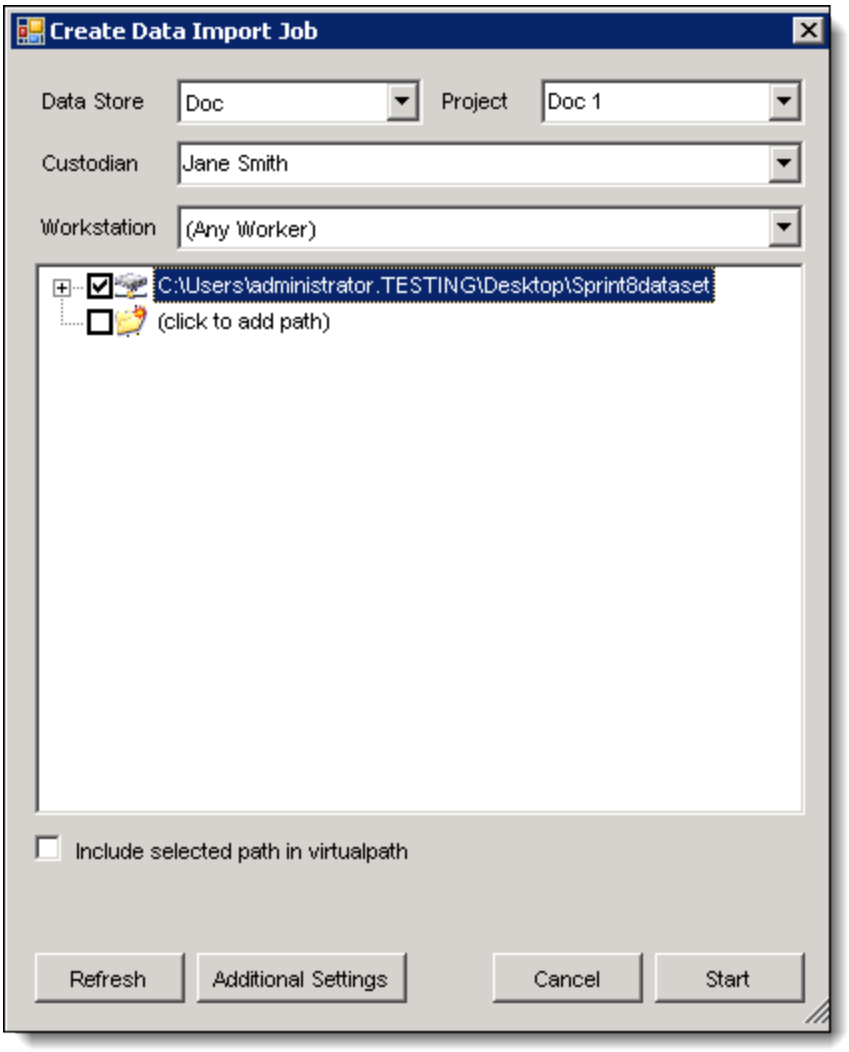
Create an import to organize your data in a single database (data store). For example, you create a project and designate different custodians under that job. Use jobs to determine your export sequence.

Note: Prioritizing custodians for de duplication is not supported in the RPC. Custodian order is determined during ingestion. These settings can be changed at any time including during the import process.

To create an import job:

1. In the Data Stores window, right-click a data store.
2. Navigate to **Tasks** > **Import**.

3. From the Create Data Import Job dialog, enter or select the information for the following fields:



- **Data Store** - the selected data store displays by default. Select another data store from the drop-down list.
- (Optional) **Project** and **Custodian** - enter names for these fields. These names display in the Data Store window.
  - **Project** - the project that you want to specify for the data import job.
  - **Custodian** - the custodian that you want specify in the project.

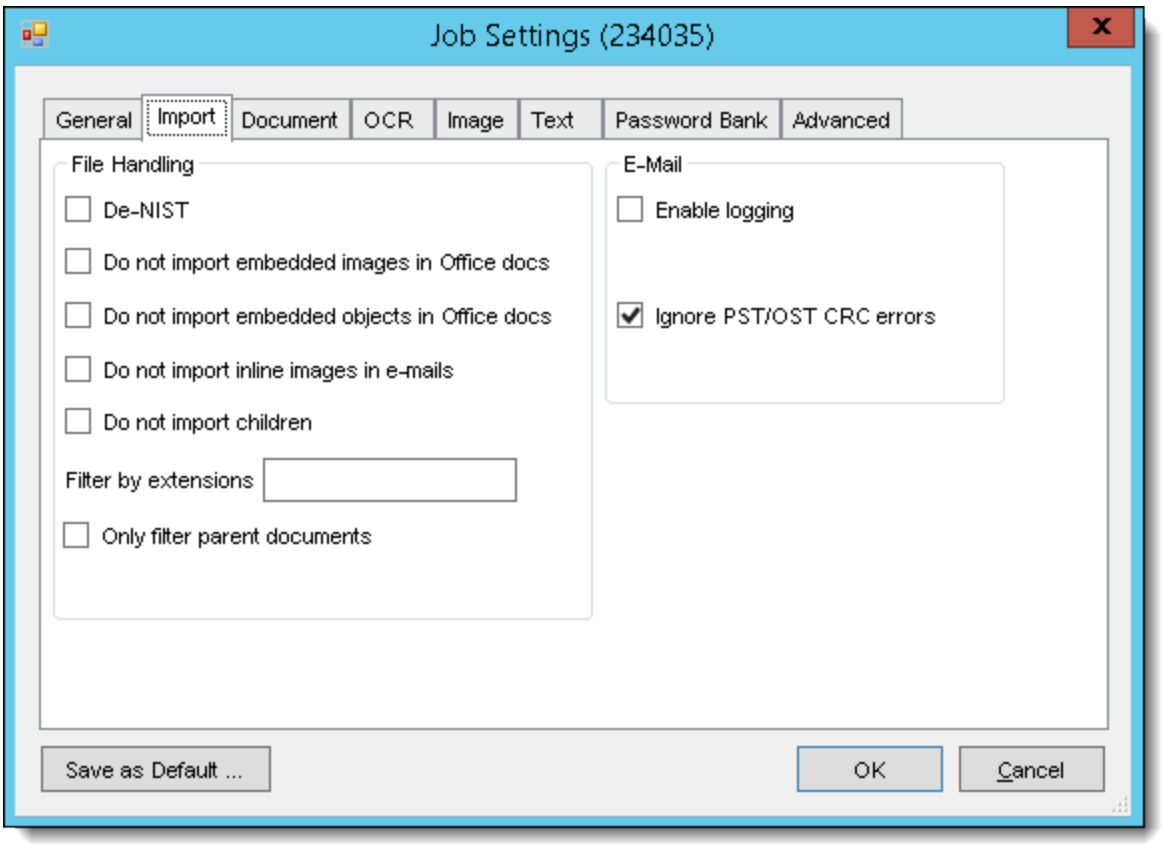
  Note: You can import data without defining a project or custodian. You can change these settings at any time, including during the import process.
- **Workstation** - select the machine with the data that you want to process. When you select a workstation, the List box displays all the drives and CD-ROMS available on that computer, so that you can make them visible to other worker machines. This option is useful if the data resides on a CD-ROM, USB drive, or other physical media that is associated with a specific worker machine.

4. Use one of the following methods to select the folders or files that you want to import:
   - Expand and drill-down in the Server List box to locate the required data. Select the check boxes for the folders or files.

- Select the check box for **Click to add path**. Click this box to clear it, and paste or type the path to the import folder. You can use this option if the data is already on the network.
- If you select a checkbox for a folder, the RPC imports all of its content even though you won't see the individual check boxes for the documents selected when you expand the list.
- You have the option of adding a SQL setting if you have a fixed import location that you would like to select in the folder window. To do this, go to the Invariant database, add an entry to the AppSettings table with a category of MapVolume, and enter the UNC path from which you'd like to import in the Value2 column. Keep the other columns as NULL. You can app multiple MapVolume entries to this table.

Note: It is possible to add a SQL setting if you have a fixed import location that you would like to see as a location to import data from. In the Invariant database add an entry to the AppSettings table with a category of 'MapVolume' and enter the UNC path you would like to import from in the Value2 column. The other columns should stay as NULL. Multiple MapVolume entries can be added to the table.

5. Choose the option **Include selected path in virtual path** if you want the virtual path to start at the server name rather than lower in the path hierarchy. If you don't select this option, the path starts with the folder name.
6. Click the **Additional Settings** button.
7. In the Job Settings dialog, select the **Import** tab.



8. Select any of the options listed in the following table.

Options	Select this option if you want to...
File Handling	
De-NIST	Remove known system files from a set of data. The RPC uses the database provided by the

Options	Select this option if you want to...
	National Institute of Standards (NIST), which includes approximately 14 million unique SHA-1 hashes for known system files, such as kernel32.dll, user32.dll, and others. These files can automatically be eliminated from a forensic collection because they are not custodian created files. (There is discoverable information in files on the NIST list, but none of it is pertinent to the case/matter unless you're using NIST.)
Do not import embedded images in Office docs	Excludes various image file types found inside Microsoft office files. For example, .JPG, .BMP, or.PNG in a Word file.
Do not import embedded objects in Office docs	Excludes various files found inside Microsoft Office files. For example, an Excel spreadsheet in a Word file.
Do not import children	Performs a one for one import of the files selected for import extracting no children or embedded items.
Do not import inline images in e-mails	Excludes various files found inside emails. For example, .JPG, .BMP, or .PNG in an email file.
Filter by extensions box	Exclude or include file types added to this text box during import. You can't select both in the same import. The RPC will filter on detected file extension, rather than the original extension. To include a file type, enter a plus sign, file extension, and a semicolon. Replace the plus with a minus sign to exclude a file type as illustrated below: • Include: +mpp;+msp • Exclude: -mpp; -msp **Note:** You can choose to either include or exclude items on an import. You cannot have both in the same import.
Only filter parent documents	Applies filter only to top-level (or parent) and loose documents. Since this filters only top-level documents, all the associated attachments will be returned.
E-Mail	
Enable logging	Debug or troubleshoot a job that crashed, or other minor performance issues. This log file is created on the C: drive of worker machine processing the file, and log file names use the storage ID of the worker. The file contains subject names, entry IDs, and other information.
Ignore PST/OST CRC errors	Ignores errors generated when a cyclic redundancy check is performed on PST files.

Note: Time zone is applicable to OCR/imaging, text extraction, exporting, and publishing, but not importing. By default, the RPC stores date type metadata in the database in Coordinated Universal Time (UTC). While you can optionally set the time zone on the Job Settings General tab when importing, this information is required prior to running text extraction, imaging or exporting.

9. Click **OK** to redisplay the Create Data Import Job dialog.
10. Click **Start**. The import job moves into the queue, which displays in the Job Activity window.
11. In the Job Activity window, highlight the import job, and select a **Group** for it. After a worker starts the job, you can monitor its progress in this window. See Managing workers and jobs on page 80.

Once you've imported data to a data store, you can extract text. See Extracting text on page 22.

3.3 Troubleshooting import errors

The RPC imports all documents into the system, even those with unsupported file types. This comprehensive import allows you to run exception reports that list any files that the RPC couldn't process.

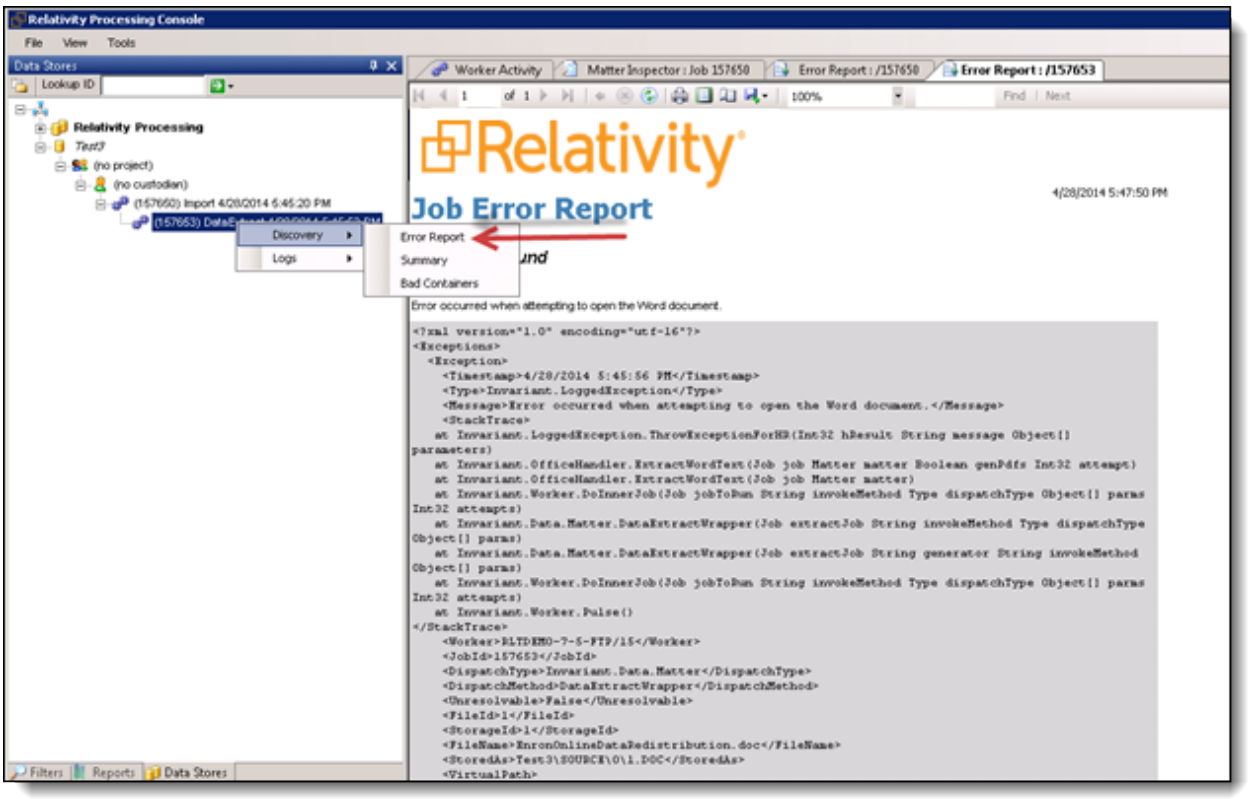
While the RPC can identify most file types, it doesn't have corresponding handlers for all types. The RPC imports unsupported file types but they aren't added to the job queue, and undergo no further processing. You can generate a detailed error report that lists these files as well as any password-protected files.

The RPC also collects the following import information found in various reports:

- **De-Nisted Files** - when the RPC discovers a file, it queries the NIST database for a match. If the file is found, the RPC doesn't import the file, but adds an entry for it to the De-NIST table associated with the job. This table can later be referenced to obtain a list of de-NISTed files.
- **Bad container files** - these files include containers without children, as well as password protected, incomplete, or corrupt containers that documents could not be extracted from. Common container file types include ZIP, RAR, NSF, and PST. You can use this information to troubleshoot these container files.
- **File Details** - the RPC captures detailed information about jobs in standard summary reports. See Running standard reports on page 99.

3.4 Generating job error reports

1. In the Data Stores window, drill down to a completed import job. An import job is directly under a custodian.
2. Right-click the import job > **Discovery**.



3. Click a report type. The RPC generates the report and displays it in a new tab.

The following reports are available for troubleshooting an import job:

- **Error Report** - lists each error that occurred with detailed exception information. You can also run this report to troubleshoot extracted text errors and imaging errors.
- **Summary** - lists the frequency of each file type in the job, the total number of discovered documents and counts of de-NISTed files. It also includes file sizes of all files, de-NISTed files and the size of files remaining after de-NISTIng not counting containers. Additionally, there is a brief entry describing each error encountered during import.
- **Bad Containers** - lists all container files that Invariant was unable to pull any documents from. If a container throws an error on import but Invariant was able to get a single file from it, the container will not appear on this report however it will appear on the error report.

Note: Each error listed in the summary report has a corresponding detailed version in the error report. If you reprocess error files and they don't encounter new errors, they will still appear on the report when you run that report again. If the reprocessed files encounter errors on the retry, these new errors are listed along with the original errors. The bad container report is dynamic, which means if a container was successfully re-imported, it doesn't appear on the bad container report when you run that report again.

See Running standard reports on page 99 for information about other reports.

4 Extracting text

Once you've imported data to a data store, you can extract electronic text from the documents to build indexes used for keyword searches. Depending on the document type, Invariant extracts the text directly, or it may physically layout the document before it can extract the text.

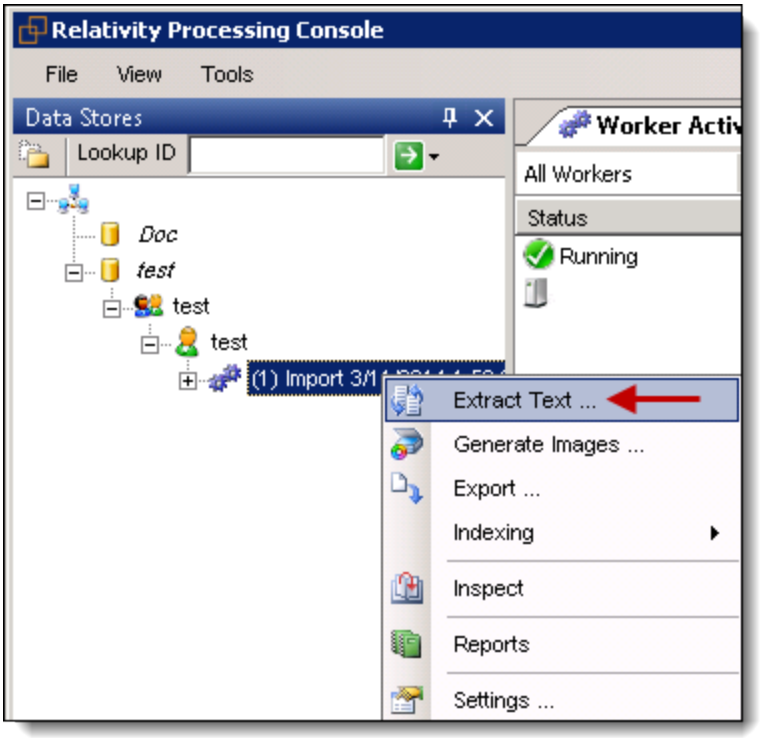
Invariant can directly extract text from file formats that it can open, such as Word, Excel, PowerPoint, and others. For example, Invariant instantiates Word, opens the document, and requests the electronic text from Word. Any comments or markup added with the Track Changes feature are also extracted as electronic text. In addition, Invariant extracts any embedded graphics that the document contains during the import process. Invariant treats the graphics as children of the document, in which they separately undergo OCR. When you run a keyword search on the entire document family, any text matches that Invariant pulls from the images are indirectly linked to the document, since the images are children of the document.

For other file types, Invariant must physically layout the document by converting it to PDF. Invariant has a PDF engine and other functionality to layout the document in memory as text only. It then extracts the electronic text from each page of the PDF, as well as detects any graphics. Invariant handles the graphics by sending them to the OCR engine for electronic text extraction. Some files requiring this type of processing include JPEG, TIFF, PDF, Visio, and JungUm Global (GUL).

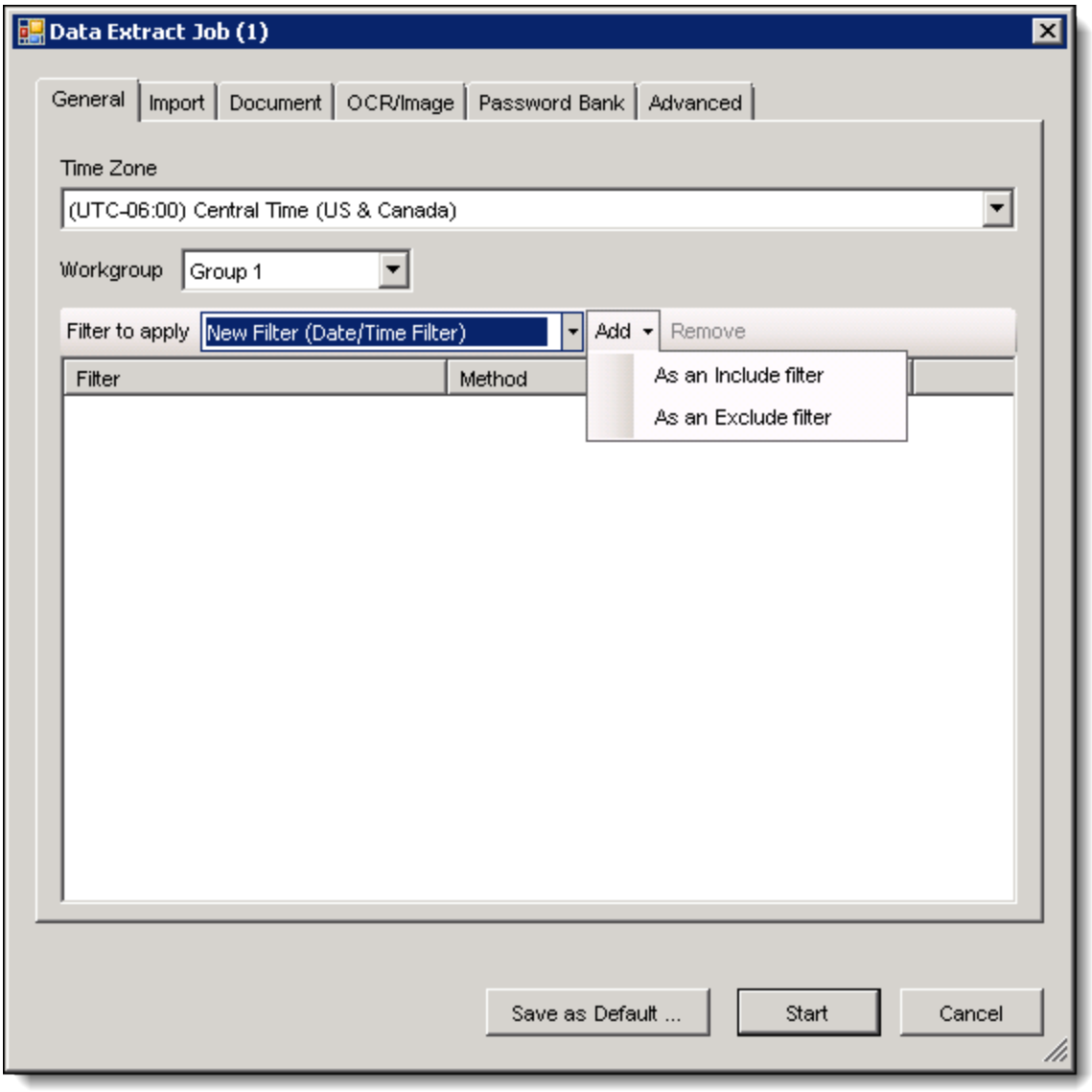
4.1 Extracting text

To extract text from an import job:

1. In the Data Stores window, drill down to a completed import job. An import job is directly under a custodian.
2. Right-click the import job >**Extract Text**.



3. From the Data Extract Job dialog box, on General tab, select options for the following fields:



- **Time Zone** - set time zone required by the data set. This information is especially important for processing emails, as date and time are listed in messages and appointments. Obtain the correct time zone from the client.

  Note: You should specify the time zone setting prior to running text extraction. If you change the time zone after running text extraction, then you should run text extraction again. Also, make sure to change the following settings prior to re-running: on the OCR/Image tab in the General Options section, check **Overwrite intermediate files** and un-check **Preserve existing pages**.

- **Workgroup** - select a group. If you want to start the job manually, leave this field set to Group 0.
- **Filter to apply** - use created filters to control the data that you want included in or excluded from the extraction job. Leave this field blank if you want to extract all the data. See Filtering data on page 42.
- **Add** - If you're using a filter, click **Add**. Select an option to include or exclude the data returned by this filter in the text extraction job.

Note: We recommend not using filters on first-time extracted text. Instead, use filters to retry errors on subsequent text extraction jobs. See Using filters to resolve errors on page 26. If you intend to create a full text search filter on your data, the text will need to be indexed prior to filter creation. Any items filtered out of text extraction will throw an error during the indexing stage.

4. (Optional) Select the **OCR/Image** tab if you want to set options for documents that must undergo OCR before the text can be extracted. See OCR/Image tab settings in Generating images on page 34.
5. Click **Start**. The job is moved into the queue, which displays in the Job Activity window.
6. Monitor the progress of the job in the Job Activity window. See Managing workers and jobs on page 80.

Note: If your workgroup is set to Group 0, highlight the job and select another group to begin processing. If you have workers assigned to Group 0, this is not necessary, and the first available worker will begin the job.

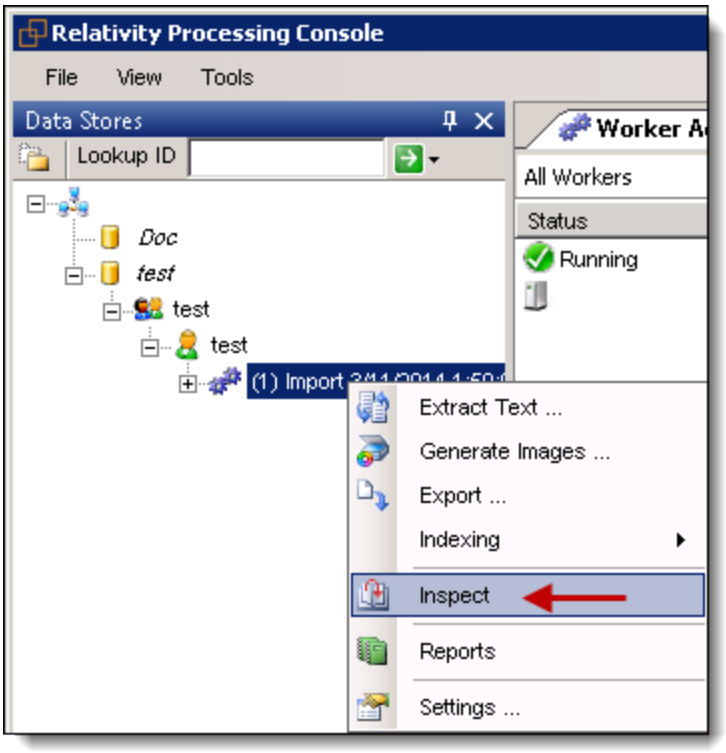
Once the text extraction job completes, you can index the data. See Indexing data on page 30.

4.2 Inspecting the extracted text

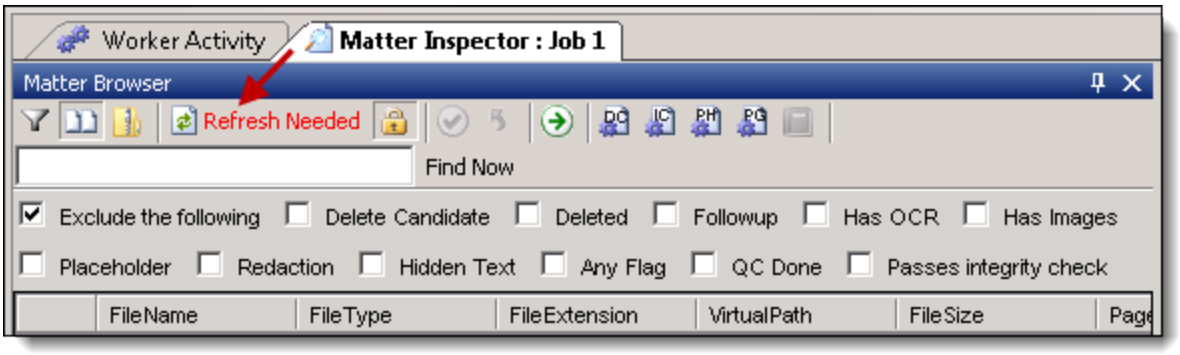
After the job completes, verify that the text is successfully extracted from the files.

To inspect the extracted data:

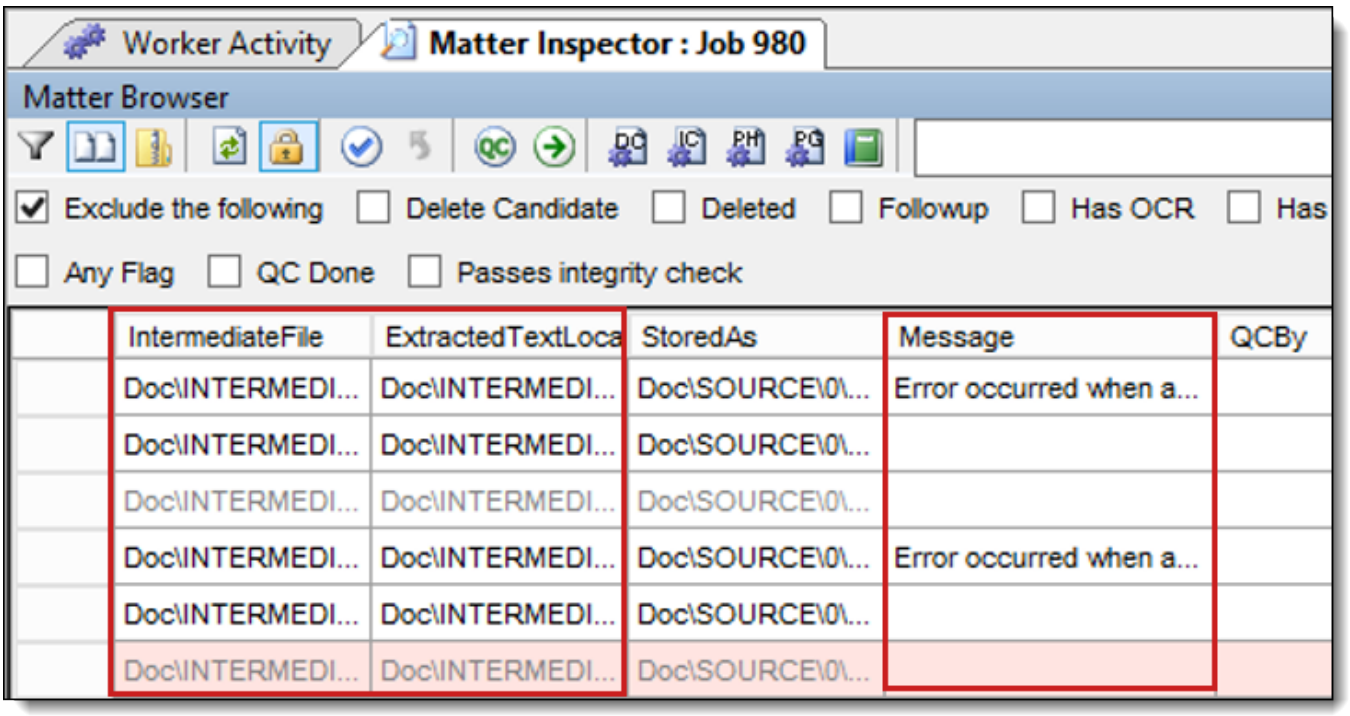
1. In the Data Stores window, drill down to the import job that you used for the extract text job.
2. Right-click on the import job, and click **Inspect**.



3. From the Matter Inspector dialog box, click **Refresh Needed** to display the documents in the text extraction job. If you did not apply a filter, all the documents display.

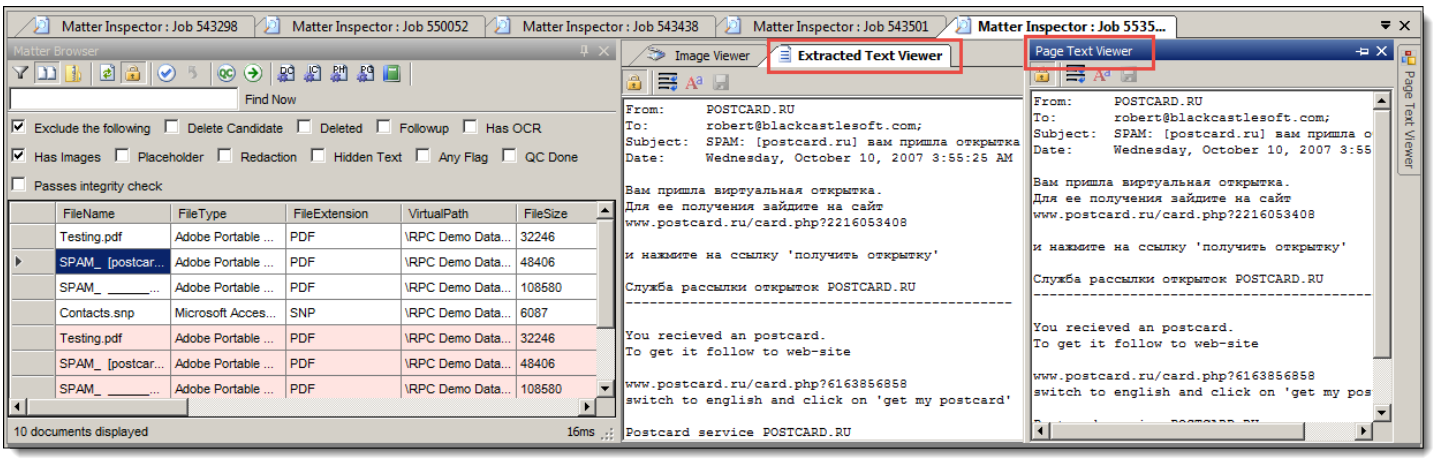


4. In the Matter Inspector window, scroll to the right to display Message, Extracted Text File Name, and Intermediate file columns. You can sort each column in ascending or descending order by clicking on the column header. Review the following columns and their content for troubleshooting information:



- **Message** - lists any errors that occurred during import, text extraction or image generation. You may want to re-run text extraction on documents with errors, depending on the type of error. See Manually re-running single documents with errors on the next page.
- **Extracted Text File Name** - lists the path where the extracted text is stored. If this column is empty, the file may be an image type file, in which case the Intermediate File column will be populated or the file may be of an unprocessable type (audio/video files, system files, etc). If the document fits neither of these categories, you should re-run text extraction on the file. See Manually re-running single documents with errors on the next page.
- **Intermediate File** - lists the path to the intermediate file (PDF) created when text can't be directly extracted from the native file. This intermediate file undergoes OCR to provide text for the document. If there is no extracted text, this column is empty, and the file is not unprocessable, you should re-run text extraction on the file. See Manually re-running single documents with errors on the next page.

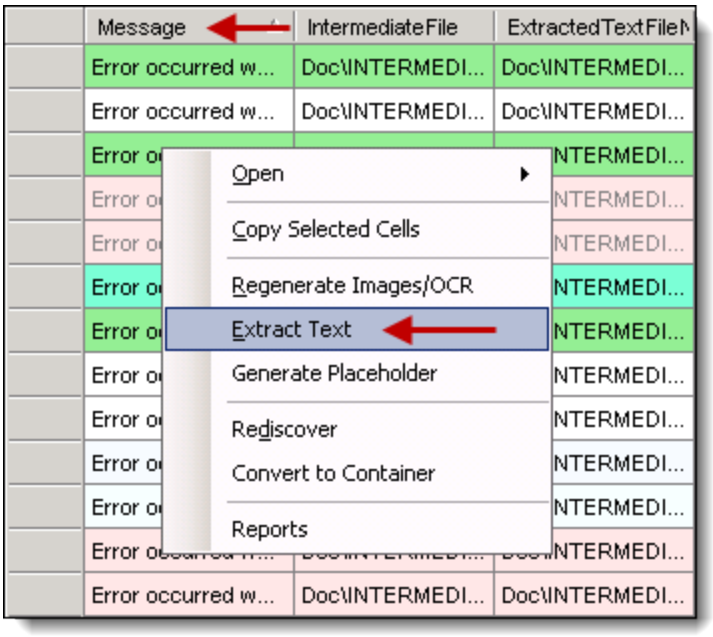
To examine the extracted text for a document, click **Extracted Text Viewer**. If the document only has an intermediate file, click **Page Text Viewer**.



4.3 Manually re-running single documents with errors

Use these steps to manually resolve an error in a single document:

1. Right-click an error in the Message column, and select **Extract Text**. A new extraction job creates.

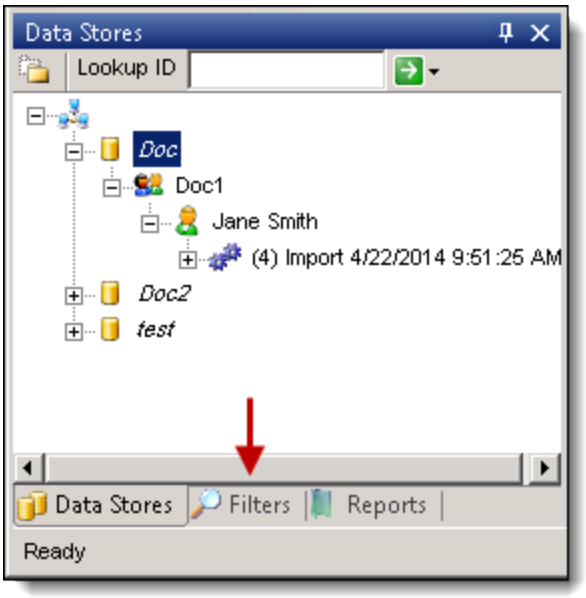


2. In the Job Activity window, highlight the job and select a **Group** to begin processing. See Managing workers and jobs on page 80.
3. After the job processes, repeat the steps in Inspecting the extracted text on page 24.

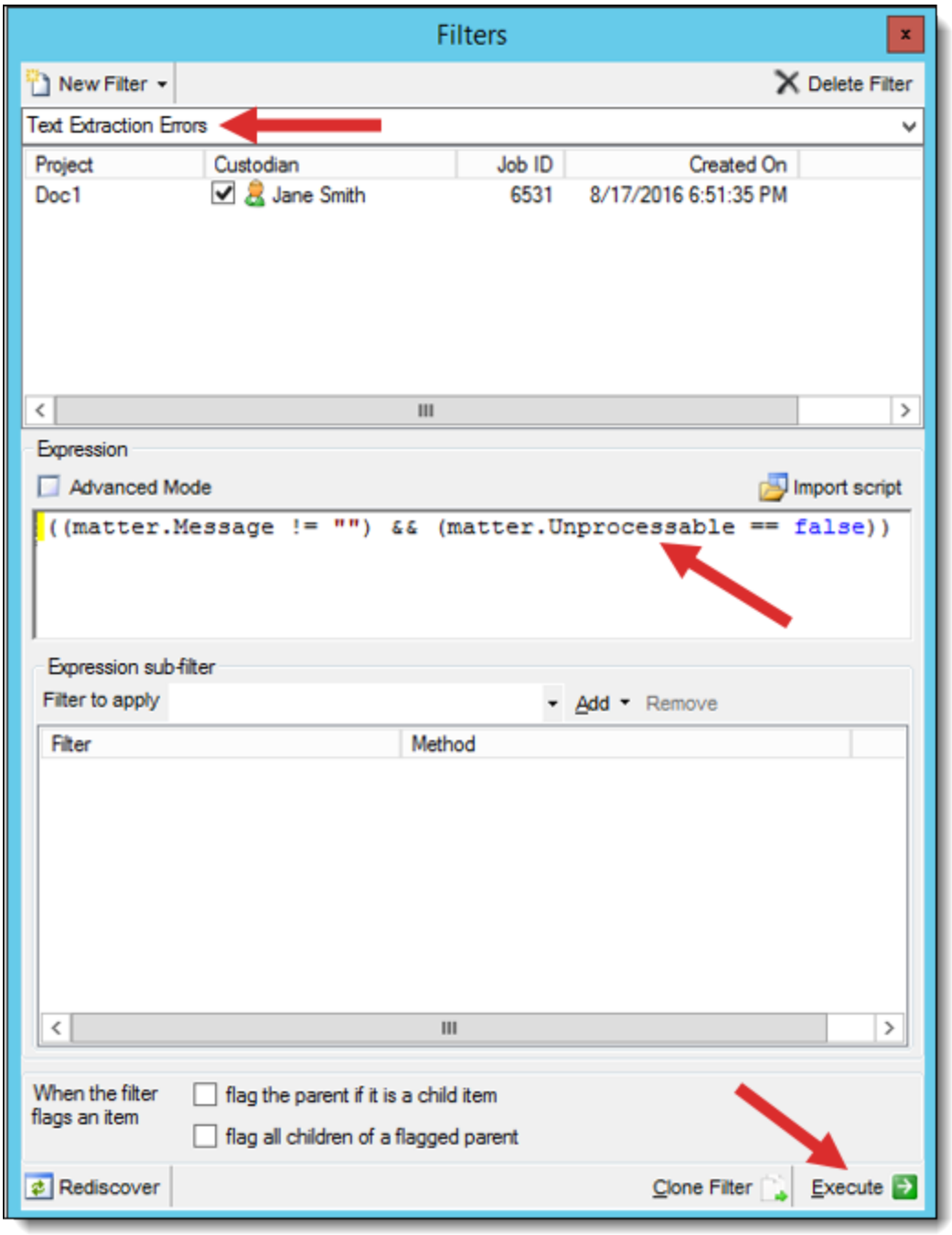
4.4 Using filters to resolve errors

You can resolve multiple errors simultaneously by using filters. You can filter by file ID and rerun only those files which do not have extracted data. For more information, see Filtering data on page 42.

1. Highlight a data store that you want to filter for errors.
2. Navigate to the **Filters** tab in the Data Store window (You can also right-click on a data store, and select **Filters** from the menu.)



3. Click **New Filters** and select **Expression**.
4. Enter a descriptive name in the **New Filter** box and press **TAB**. (If you don't press TAB after entering a filter name, the filter name might not get updated in the database.)

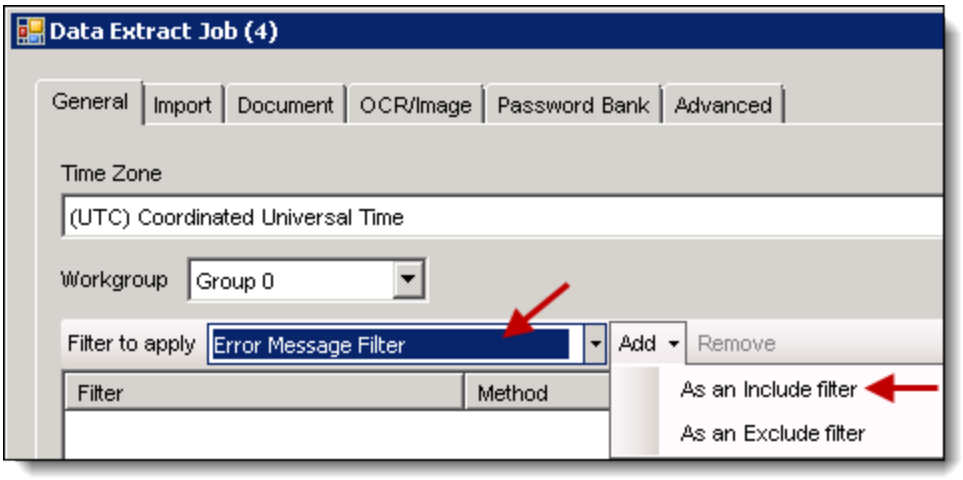


5. Select the checkbox next to the Custodian that you want to run the filter on.
6. In the Expression box, enter the following text: ((matter.Message != "") && (matter.Unprocessable == false)).

   Note: This expression returns any records where the Message column is not blank and the file is not flagged as unprocessable. This means that any files from which the text was extracted successfully are not returned. You must capitalize Message and Unprocessable, and use two sets of double quotation marks in the expression.

7. Click **Execute** at bottom of Filter tab. When the filter is finished click **Done** in the pop-up dialog box.
8. In the Data Stores window, drill down to the import job that you used for the extract text job.
9. Right-click the import job > **Extract Text**.
10. On the General tab, select your filter in the **Filter** to apply box filter.

11. Click **Add** and select **As an include filter**. The RPC will only rerun documents that have an error message and are not flagged as unprocessable.



12. In the Job Activity window, highlight the job and select a **Group** to begin processing. See Managing workers and jobs on page 80.
13. After the job processes, repeat the steps in Inspecting the extracted text on page 24. All messages should be cleared.

5 Indexing data

Before you can run any full text search filters on the imported documents, you must first build an index from your import job's extracted text. You can then use operators and other features provided by a dtSearch index to construct full text search filters.

The RPC breaks indexing jobs into 1000 document chunks, which are then run in parallel using multiple workers and threads to improve performance. The RPC merges the dtSearch indexes after these jobs complete.

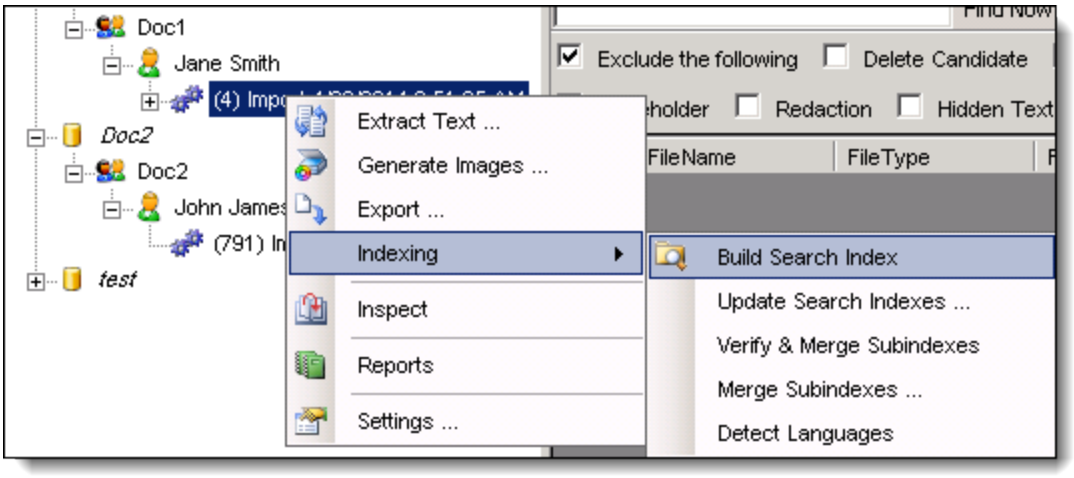
During indexing, errors may occur from documents that don't have any extractable text. The RPC continues building the index when an error occurs, but those documents without extractable text aren't represented in the index. Using SQL Reporting Services, the RPC generates an error report that lists the documents and error details. For example, a document without extracted text or intermediate file is listed in this report.

Note: You can index data at any point after text extraction or image generation. If you choose to skip text extraction and proceed directly to image generation the indexing process will combine the page level text into a doc level text file that will be used in the index.

5.1 Running an indexing job

Before running an indexing job, you must first run text extraction or generate images on the job. See Extracting text on page 22 or Generating images.

1. From the Data Stores window, drill down to the import job that you used for the extract text job.
2. Right-click the import job > Indexing > Build Search Index.



3. In the Job Activity window, highlight the indexing job and select a Group to begin processing. See Managing workers and jobs on page 80.
4. After the job processes, right-click the indexing job in the Data Store window > **Discovery** > **Error Report**.
5. Review the report to determine if the index needs to be rebuilt.

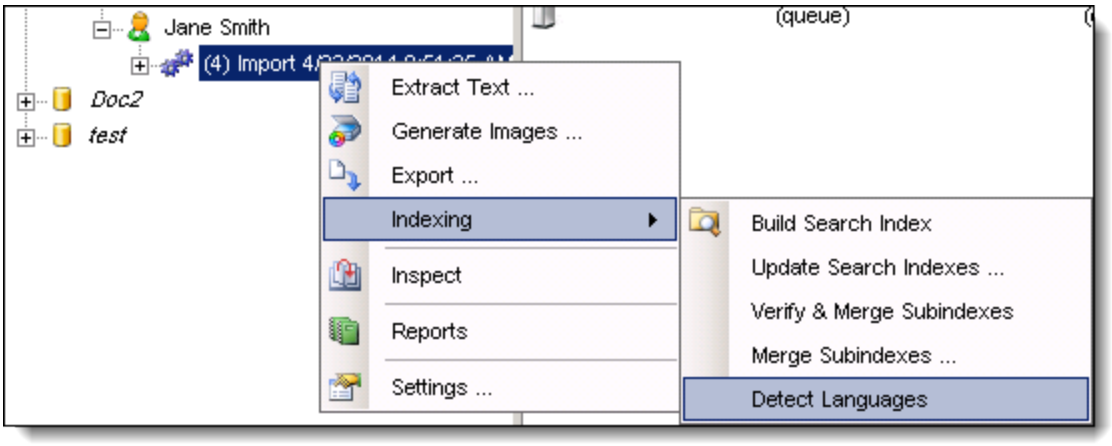
Note: You can also use the Matter Inspector tab to investigate an indexing error. To do this, obtain the storage ID from the error report. In the Data Store window, highlight the job, and select Inspect to open the Matter Inspector. Use the storage ID to locate the document. Determine if the document is an image, needs OCR, is password-protected, or has another issue.

5.2 Language detection

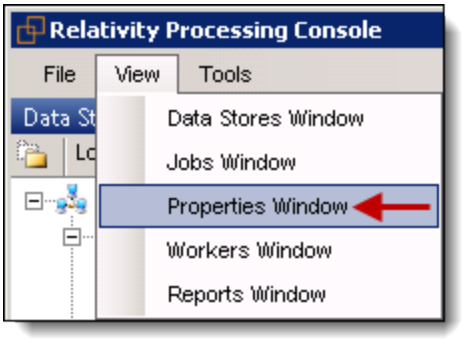
The RPC supports language detection by adding metadata to each document in an import job. Using the metadata, the RPC creates a language map of the document. It then maps out a range in the document for each language that it contains, and provides a score that identifies the dominant language.

Before you can perform language detections, you must first index the import job.

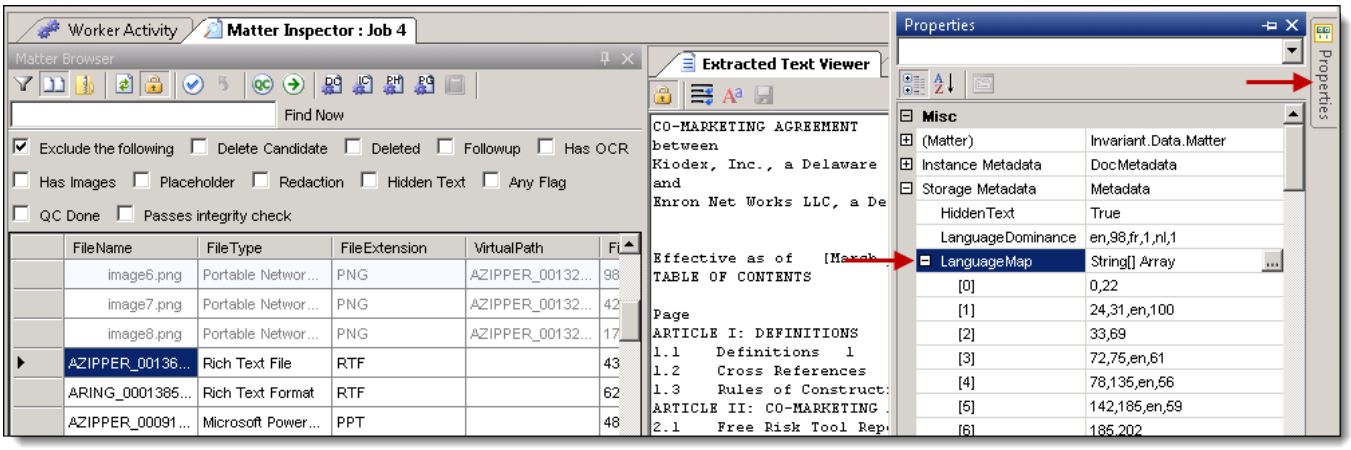
1. In the Data Stores window, drill down to an import job.
2. Right-click on the import job. Point to **Indexing**, and click **Detect Languages**.



3. Select the **View** menu, and click **Properties** window.



4. Right-click on the import job, and click **Inspect** to open the Matter Inspector.
5. To view the language map, highlight the document in the Matter Inspector, and click **Properties**.
6. In the **Properties** window, expand **Storage Metadata**, and then expand **Language Map** to display a list of language ranges.



7. The range for the language lists the starting character position, the character length, the language, and the percentage of the text in that language. For entry [3], the starting character position is 89, character length is 93, the language is zh (that is Chinese), and the text is 100% Chinese: [3] 89,93,zh,100. When the Language Dominance is below 50, this value usually indicates unreliable detection.

Note: For a complete list the two-digit codes that accompany each full language name, visit the ISO standards site.

5.3 Merging subindexes

The RPC performs parallel indexing by distributing jobs across multiple machines, which create subindexes. It then attempts to merge all the subindexes after the indexing completes. When a very large import job indexes, this process can over run dtSearch capabilities, which has a limit of about 4 billion words.

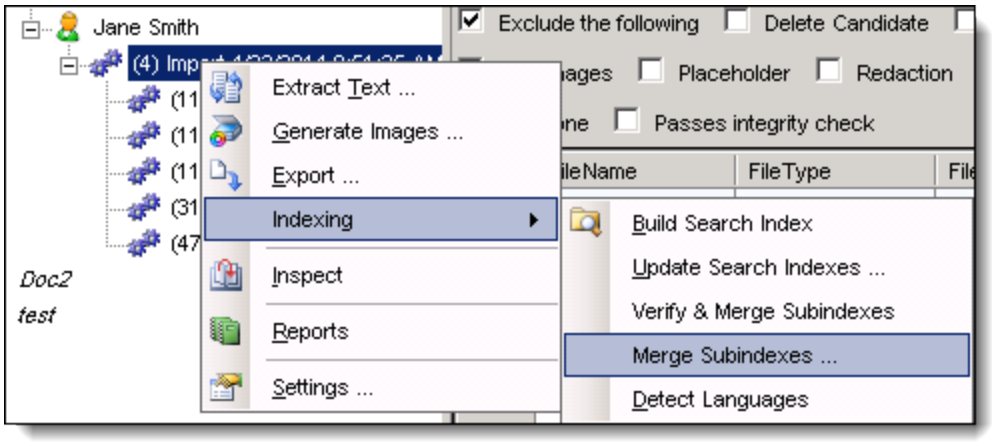
Errors can occur when the RPC automatically attempts to merge multiple subindexes into a single index. In this case, you can manually attempt to merge a few subindexes at a time, gradually building a large index.

Note: The RPC can search across multiple indexes. It automatically detects them and spans the search across them, but this process is slower than using a single index.

1. In the Data Stores window, drill down to an import job.

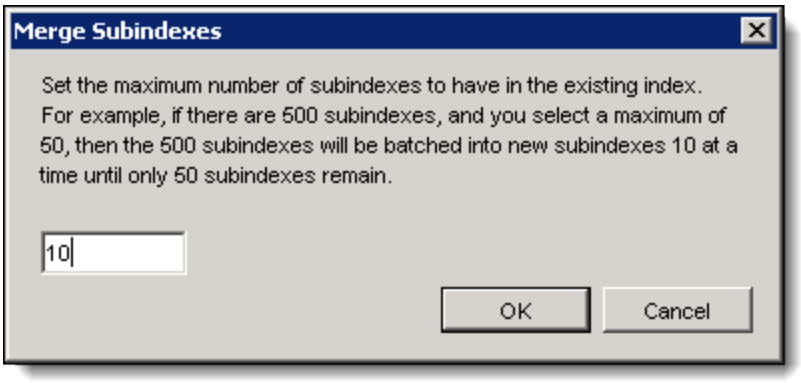
   Note: The option **Verify and Merge Subindexes** verifies that all expected documents are indexed, and then attempts to merge the subindexes.

2. Right-click on the import job. Point to **Indexing**, and click **Merge Subindexes**



.

3. Enter a value and click **OK**.



6 Generating images

The RPC references the options set on the Document and OCR/Image tab to determine how the images are generated, and OCR runs. During OCR, the RPC converts the native files to intermediate PDFs, and the text is then extracted from these files. For intermediates derived from natives with text, the text layer of the created PDF is extracted. For intermediates derived from images the text is created by the OCR engine. Natives with both text and images will have both of these methods performed on the intermediate file.

Note: You can index data at any point after text extraction, and you can image documents immediately after import. You can also index after image generation. The system automatically combines page-level extracted text into a document level extracted text and index that text.

The intermediate files are stored in the Intermediate Folder created for a job. This folder may contain intermediate files using the following naming conventions:

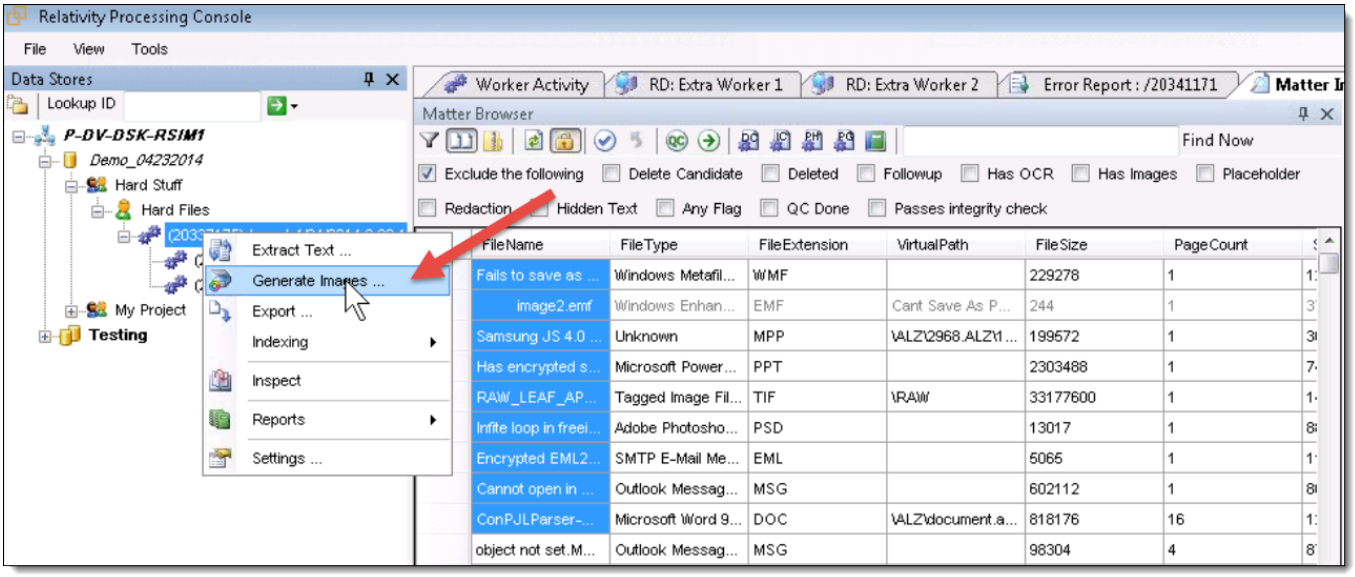
- storageID_PH.PDF: Indicates a placeholder file.
- storageID_R.PDF: Indicates a file containing redactions.
- storageID.PDF: Indicates a PDF of a page.

When exporting, the RPC looks in the Intermediate Folder first for placeholders, files with redactions, and then PDFs of pages. It uses the appropriate PDF to generate TIFs for export as determined by settings selected on OCR/Image tab for a specific job.

Note: To locate an Intermediate folder for a job, right-click a job in the Data Stores window > **Inspect**. Right-click a file in the **Matter Inspector** > **Open** > **Explore Intermediate Folder**. Use the same procedure to select **Explore Native Folder**.

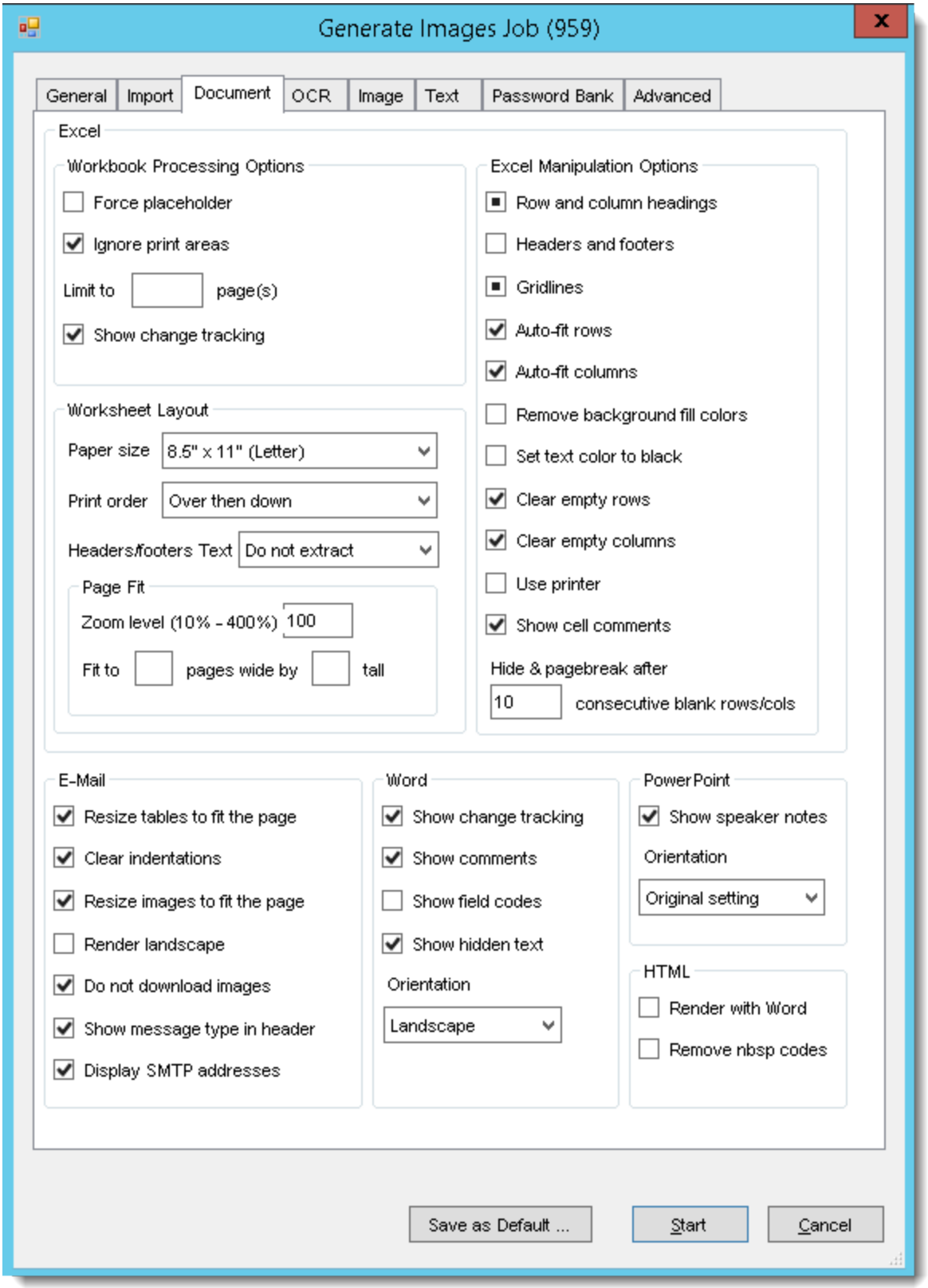
6.1 Generating images

1. In the Data Stores window, drill down to a completed import job. An import job is directly under a custodian.
2. Right-click the import job, and select **Generate Images**.



3. From the Generate Images Job window on the Documents tab, select options for the following fields.

Note: On the Document tab, the checkboxes have on, off, and indeterminate states. When the checkbox is in an indeterminate state, the document's internal setting for an option is used by default. When the option is turned on or off in the RPC, this new value overrides the default setting in the document.



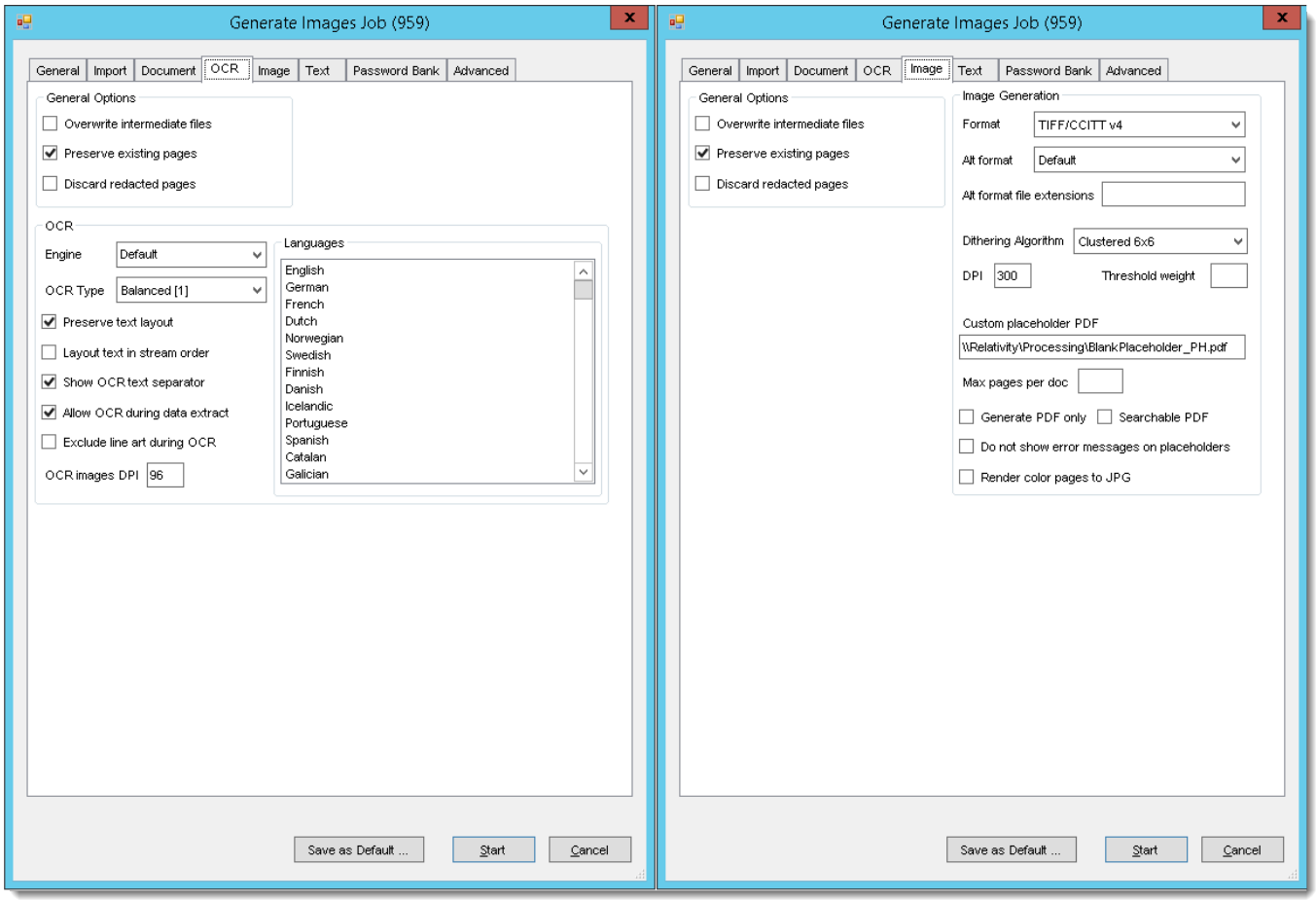
The following table lists the options available on the Document tab.

Option	Description
Excel - Workbook Processing Options	
Force placeholder	Use this option to generate a placeholder for each Excel file instead of an image
Ignore print areas	Use this option to override any pre-defined print areas existing in an Excel file. When this option is selected, the entire document is imaged.
Limit to X pages box	Use this option to image only the number of pages entered in this box.
Show change tracking	Use this option to display any modifications made to the document through the Track Changes option in Excel. By default, this option is selected.
Excel - Worksheet Layout	
Paper size	Choose an available paper size. The default value is the size internally stored in the file.
Print Order	Use this option to determine the order used to render the worksheets. (Print down and then over, or vice versa.)
Page Fit - Zoom Level	Reset the zoom level for a spreadsheet. The value in this field will override the default setting internally stored in the file.
Fit to X pages wide by X tall	Reset the page fit for a spreadsheet. The values in these fields will override the default settings internally stored in the file.
Excel - Manipulation Options	
Row and Column Headings	Select this option to render row and column headings (that is the number assigned to rows, and the letters assigned to columns).
Headers and footers	Select this option to render headers and footers in a spreadsheet. **Note:** Leave this option disabled. The printing of these items interferes with the ability of the RPC to detect blank pages.
Gridlines	Select this option to render the gridlines between columns and rows in a spreadsheet.
Auto-fit rows and Auto-fit columns	Use these options to expand the dimensions of rows and columns to accommodate their content. When these options are selected, any hidden columns or rows are displayed. They are selected by default. **Note:** You can prevent the RPC from displaying hidden columns and rows by setting the **DontUnhide** option on the Advanced tab to True.
Remove all background fill colors	Select this option to remove background color. It ensures that any hidden text or rows formatted to match the background color are displayed.
Set text color to black	Select this option to display the font color of text as black. It ensures that any hidden text is displayed, such as text with a white font on a white background.
Clear empty rows and Clear empty columns	Use these options to remove empty rows and columns from a spreadsheet, and render as few pages as possible. By default, these options are selected.
Use printer	Use a printer driver to render the PDF rather than internal processing provided by the RPC. By default, this option is not selected.

Option	Description
	Note: Select this option to speed up processing when you are imaging spreadsheets created in Excel 2007, especially if they contain charts or graphs.
Show cell comments	This determines whether Relativity renders comments.
Hide and page break after X consecutive blank rows box	Set a maximum value for consecutive blank rows or columns. By default, the RPC checks for a run of 10 blank rows or columns, and compresses them to limit the number of pages rendered for a spreadsheet.
E-Mail	
Resize tables to fit page	Modifies the table sizes to make them compatible with the page size. By default, this option is selected.
Clear indentations	Removes the tabs from e-mail threads to improve their display. (Select this option to prevent the entries in the thread from aligning too far to the right.)
Resize images to fit page	Modifies the image sizes to make them compatible with the page size. By default, this option is selected.
Render landscape	Changes the orientation of the message to landscape.
Do not download images	Select this option to prevent the downloading of images from the internet.
Show message type in header	Adds text in the header identifying the image as a message, appointment, distribution list, and so on. By default, this option is selected.
Display SMTP address	Select this option to render SMTP addresses for To, Cc, and Bcc fields instead of Microsoft Exchange information.
Word	
Show change tracking	Select this option to render any markup added to the document through the Track Changes feature in Word.
Show comments	Select this option to render any comments added to the document through the Review feature in Word.
Show field codes	Select this option to render the field code rather than the value in the document. **Note:** In general, this option is not enabled, because the RPC can preserve the actual value for a field code by preventing modifications to it when the document is opened. For example, field codes for file path and date will not be automatically updated with the current user settings, but will retain the original settings.
Show hidden text	Select this option to render any text added to the document with Hidden feature in Word.
PowerPoint	
Show speaker notes	Renders the slide at the top of the page, and speaker's notes at the bottom. By default, this option is selected, and the image orientation is portrait even when the document does not contain speaker notes. When this checkbox is not selected, the image orientation for the PowerPoint slides is landscape.

Option	Description
HTML	
Render with Word	Select this option to render HTML in Word. By default, the RPC will render HTML documents in Internet Explorer, and then generate a PDF. For some problematic HTML documents, a better image will generated if the file is rendered in Word.
Remove nbsp codes	Select this option to remove long rows of non-breaking spaces (nbsp) codes, which prevent the text from wrapping properly. When it renders HTML as a PDF, the RPC will automatically format page breaks without cutting text or margins.

4. On the OCR tab, select options for the following fields.



Option	Description
General Options	
Overwrite intermediate files	Use this option to generate a new intermediate file. The RPC does not store multiple copies of an intermediate file for a document. When an intermediate file already exists, the RPC will use it to generate images unless this option is selected. By default, this option is not selected.
Preserve existing pages	Use this option to prevent existing PDFs from being overwritten. The RPC will skip existing pages when they do not need to be regenerated. By default, this option is selected.
Discard redacted pages	Use this option to ignore redacted pages during processing, and select intermediate PDFs (named as storageID.PDF). By default, this option is not selected. You can use this option when redacted PDFs (storageID_R.PDF) have been added to the

Option	Description
	Intermediate Folder for a job. Instead of preferring these files, the RPC will process unredacted files (storageID.PDF).
Image Generation	
Format	Use this option to change the image format. For most processing, use **Default**, which is CCITT v4 for generating black and white TIFs. This setting is the setting at installation but it can be modified at the project or data store level as well. Select another format if you want to force a specific format for images. Options include **Color JPG**, and for TIFs, **CCITT v3** and **CCITT v4**.
Alt format	See Format options above.
Alt format file extensions	Use this option to force a file type to be imaged in an alternate format. For example, you could set the **Alt Format** as Color JPG, and then enter **+XLS** in the file extensions box. The RPC will then render Excel files as Color JPGs. You can also exclude file types from rendering as an Alt Format by prefixing the file extension with a minus sign. This is a regular expression filter, which means that it will return partial matches to the extension you enter. For example, if you enter +XLS here, the RPC will also image Excel files with .xlsx and .xlt extensions.
Dithering Algorithm	Select one of the following algorithms for converting images from color to black and white: • Default (Uses Clustered 6x6) • Floyd & Steinberg • Dispersed 4x4 • Dispersed 8x8 • Clustered 6x6 • Clustered 8x8 • Clustered 16x16 • Dispersed 16x16 • No Dithering (Uses setting for Threshold Weight option listed below.)
Threshold Weight	Enter a value for this option when you set the Dithering Algorithm to **No Dithering**. The default threshold value is 128, while the threshold range is between 0 - 255. When the pixel RBG in an image is greater than the threshold, it is converted to black, and when less, it is converted to white.
DPI	Use this option to control the density of image (dots per inch). The default value is 300. A typical range used for setting the DPI is 200 - 600.
Custom placeholder PDF	Use this option to enter a path to a PDF file that you want used as a custom placeholder. • The file chosen will be used to create placeholders when imaging. It does not set what placeholder to use when exporting. The distinction is that the exporter looks for the imaged placeholder, not what it's in this field. If you have done imaging with the default placeholder and later enter a value for a custom placeholder you will need to re-image the placeholder documents otherwise the originally imaged default placeholders will be exported. • Use UNC pathing to the custom placeholder and be sure to save the placeholder in a

Option	Description
	location the Invariant workers can access. Also, make sure the name of your custom placeholder ends with '_PH' (Sample_PH.pdf). • You're unable to enter a dynamic file path in the middle of a placeholder name that changes for each file. • The RPC doesn't support adding slipsheets in front of each exported file with dynamic metadata fields on them. However, it is possible to use blank placeholders and apply an endorsement on them of that file's metadata through the use of switch statements. For more information, see Using a switch statement for custom logic on page 76.
Max pages per doc	Use this option to specify the maximum number of pages imaged for a document. (Set this value if you do not want the entire document imaged.)
Generate PDF only	Select this option to not create and page level text when imaging.
Searchable PDF	Select this option to perform in-place OCR on the PDF page image elements.
Do not show error messages on placeholders	Select this option to generate blank placeholders. By default, placeholders display an error message.
Render color pages to JPG	Select this option to substitute JPGs for TIFFs when the pages are in color. (TIFFS are generated only in black and white.)
OCR	
Engine	Use this option to choose the Nuance or no OCR engine. Select NoOCR when you want extracted text but don't want to OCR pages with images. The default engine is Nuance.
OCR Type	Use this option to control the performance of OCR job. Select Accurate for more precise OCR, Fast for improved performance, and Balanced to equalize precision and performance.
Preserve text layout	Use this option to maintain the current layout of the text when extracting text from PDF for OCR. By default, this option is selected.
Layout text in stream order	Use this option to maintain the order of text as in the PDF layout when extracting text for OCR.
Show OCR text separator	Use this option to display a separator between extracted text at the top of a page and text derived from OCR at the bottom of the page. The separator reads as, "--- OCR From Images ---". With the separator set to off, the OCR 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 selected.
Allow OCR during data extraction	Clear this option if you do not want documents that do not have extractable electronic text to undergo OCR. By default, this option is selected.
Exclude line art during OCR	Select this option to remove line art and annotations added to the original page. (For example, it will remove any text boxes drawn over the original content of the page.)
OCR images DPI box	Use this option to set the DPI level the Nuance engine will use when performing OCR. Changing this can often allow the OCR engine to successfully OCR if it failed using the default setting.
Languages	Use this option to make multiple language selections.

5. Click **Start**. The job is moved into the queue, which displays in the Job Activity window.
6. Monitor the progress of the job in the Job Activity window. See Managing workers and jobs on page 80.

Note: If your workgroup is set to Group 0, highlight the job and select another group to begin processing. If you have workers assigned to Group 0, this is not necessary, and the first available worker will begin the job.

Once you've generated images, you can QC images. See Performing Quality Control tasks on page 90.

7 Filtering data

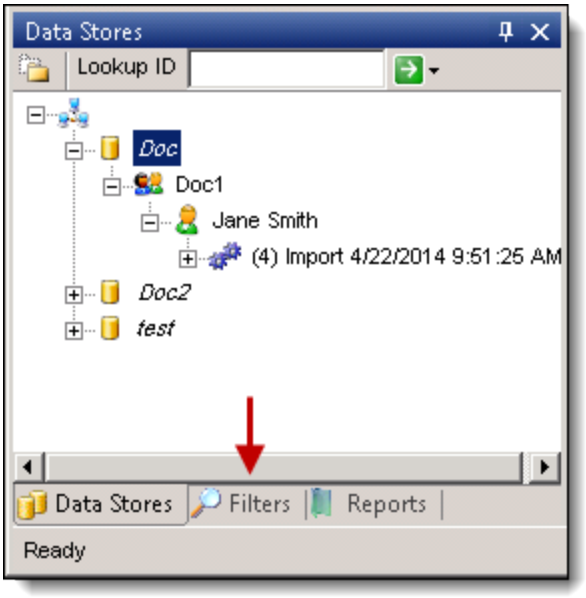
After you've imported data you can use filters to isolate a set of data. Use filtering to run quality control tasks to check or fix errors, extract text, as well as produce or export documents. Filters are similar to creating saved searches in Relativity and are especially useful when you're performing tasks that require you to include or exclude a specific set of documents.

Note: For more information on how to execute filters in the RPC, see the Filtering scenarios in the RPC video webinar on the Relativity Training site.

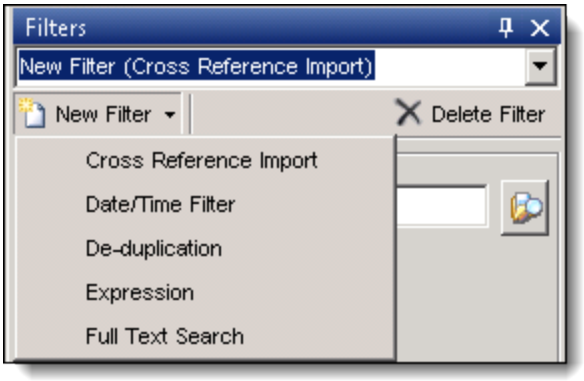
7.1 Creating a filter

This section includes generic steps for creating a filter. For information about specific filter type options, see Filter types on page 45.

1. In the Data Stores window, highlight a data store that you want to filter, and click the **Filters** tab at the bottom of the store list. If the Filters tab isn't visible here, right-click on the data store name and select **Filter(s)...** The Filters tab won't be visible if you previously had it open but then closed out of it.

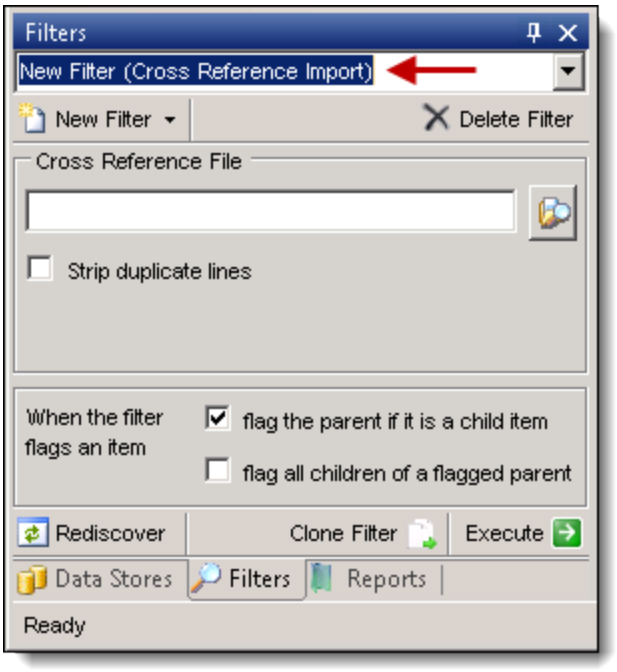


2. Click **New Filter** to display a list of filter types.



3. Select a filter type, and enter a name in the **New Filter** box and press **Tab**. We recommend clicking Tab after entering the filter name, as this will ensure that the name is saved. Under some circumstances, not clicking Tab

lists the filter merely by its default name of "New Filter (<Type>)."



4. Enter the required information for the filter type that you selected. You will see specific fields for each filter type displayed in the center of the tab. See Filter types on page 45.
5. Select one or both of the following options to determine additional actions when the filter flags an item:
    - **Flag the parent if it is a child item** - if the filter flags a child item, then the parent will also be flagged.
    - **Flag all children of flagged parent** - if the filter flags the parent, then the children will be flagged. Select this option if you only want to search for parent documents, but you also want to include the children of the documents that had a hit.
        - If you simply click Execute without flagging anything, the filter you've set up will only flag documents that actually meet the search criteria. This could be problematic if you want to then export the filtered data, especially in situations in which you've hit on a keyword that's a child document but you didn't hit on the parent document. This is because you can't export a child document without also exporting its parent. Thus, it's recommended that you check either the **flag the parent if it is a child item** or the **flag all children of a flagged parent** option. Checking both of these brings in the entire family, which you can then export.
        - Flagging all children of a flagged parent ensures that you capture all children of a parent that contained a keyword hit.
        - If you have a parent item and it has, for example, ten children and one of the children hits on the keywords you entered and you check the flag for parent and then execute the filter and then export the data, it will display as one parent and one child.
6. Click one of these buttons:
    - **Rediscover** - queues the files in the filter for re-import. The filter must be executed first for this button to function.
    - **Clone Filter** - clones the filter.
        - You can use the clone option to create a filter using identical date settings or keyword terms and then adjust the job list it's filtering. Alternatively, you can keep the jobs being filtered on and just change, for example, the end date on a date filter.

- The newly created filter clone won't have any items in the filter until you click **Execute**. To clone a filter, click **Clone Filter**, enter a name for the new filter, press the **Tab** key, modify the filter according to your needs, and click **Execute**.
- You can't modify filters used in an export that uses a setting of New, Replacement or Supplemental. If you need to alter a locked filter, you must first clone the filter and then edit the clone.

- **Execute** - runs the filter and populates the data in the RPC. Until you click Execute in the Filters window, you haven't yet set the filters you've created or added.

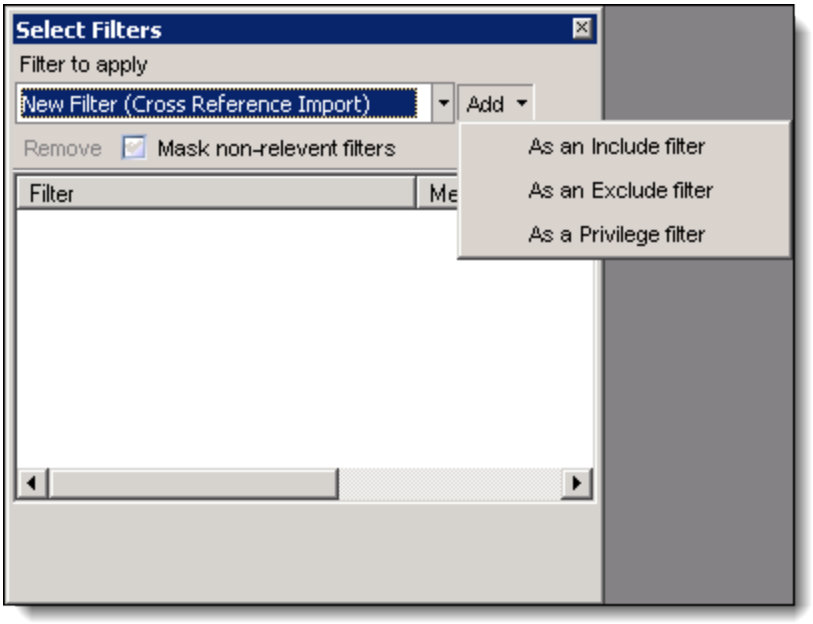
Note: When you see a list of import jobs in the filter window after selecting a filter type, those jobs are sorted by Job ID by default, but you can sort on any column (Custodian, Project, Job ID, CreateOn, etc.) you want to.

Once you've filtered data, you can create export files for review or production. See Exporting data on page 59.

7.2 Displaying filter results

To display the filter results, complete the following steps:

1. In the Data Stores window, right-click on an import job, and select **Inspect**.
2. From the Matter Inspector, click , and select a filter in the **Filter to apply** box.



3. Click the **Add** drop-down arrow to select the filter type.
   - **As an Include filter** - includes any document that matches the criteria set in this filter. For example, you could add a keyword search as an include filter, so that all documents with this keyword are exported.
   - **As an Exclude filter** - excludes any document that matches the criteria set in this filter.
   - **As a Privilege filter** - flags documents that are potentially privileged based on their inclusion of privilege-relevant terms found in the full text search.
     - In the matter inspector, adding a privilege filter has no impact on the documents displayed.
     - Privilege filters are limited to exporting; however, a Boolean search is most often used for a privilege filter.

4. Close the window after you have selected all the required filters.
5. Click **Refresh Needed** to display a list of documents match your filter settings. For more information about the Matter Inspector, see Performing Quality Control Tasks.

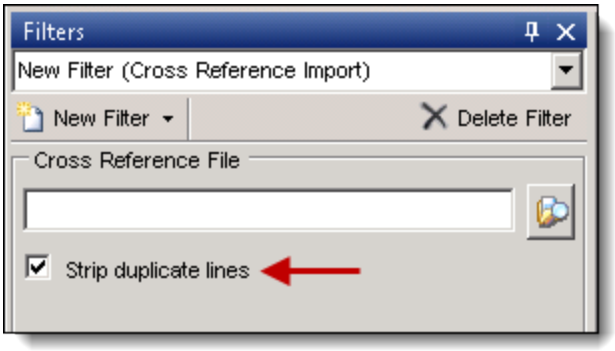
7.3 Filter types

This section describes how to use the various filter types supported by the RPC.

7.3.1 Cross reference import filter

When you are importing data, this filter uses the file ID as a cross reference. You can have multiple cross reference import filters for a single data store. You can define five metadata fields in the cross reference text file used by this filter. See Formatting the cross reference file below.

1. Select **Cross Reference Import** in the **New Filter** menu.
2. Enter a name in the **New Filter** box and press **Tab**.
3. To configure the filter, browse or enter the path to your file.
4. Select **Strip duplicate lines** if you want any duplicate file IDs in the cross reference filter removed.



5. To run the filter, continue with Step 5 in Creating a filter on page 42.

7.3.1.1 Formatting the cross reference file

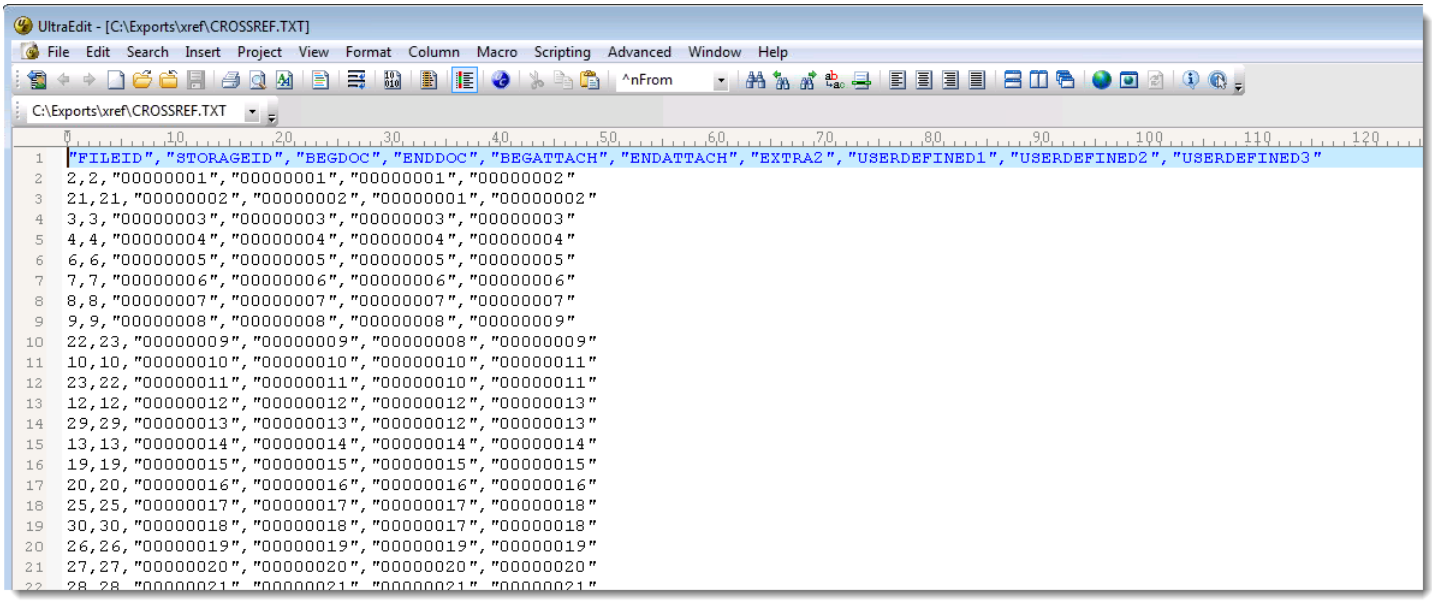
When you import data using the Cross Reference Import filter, you need to supply a text file containing additional metadata fields. These user-defined fields can be used for external data, such as branding, tagging, reference, and export information, as well as external control numbers.

Note: The FileID is the only required field. Other fields are optional.

You must follow the required format for the cross reference file. The first row in the file must be a header row, which is not imported when you run the filter. You can also define up to five metadata fields in the cross reference text file:

- EXTRA1
- EXTRA2
- USERDEFINED1
- USERDEFINED2
- USERDEFINED3

Use the format in the sample file illustrated below:



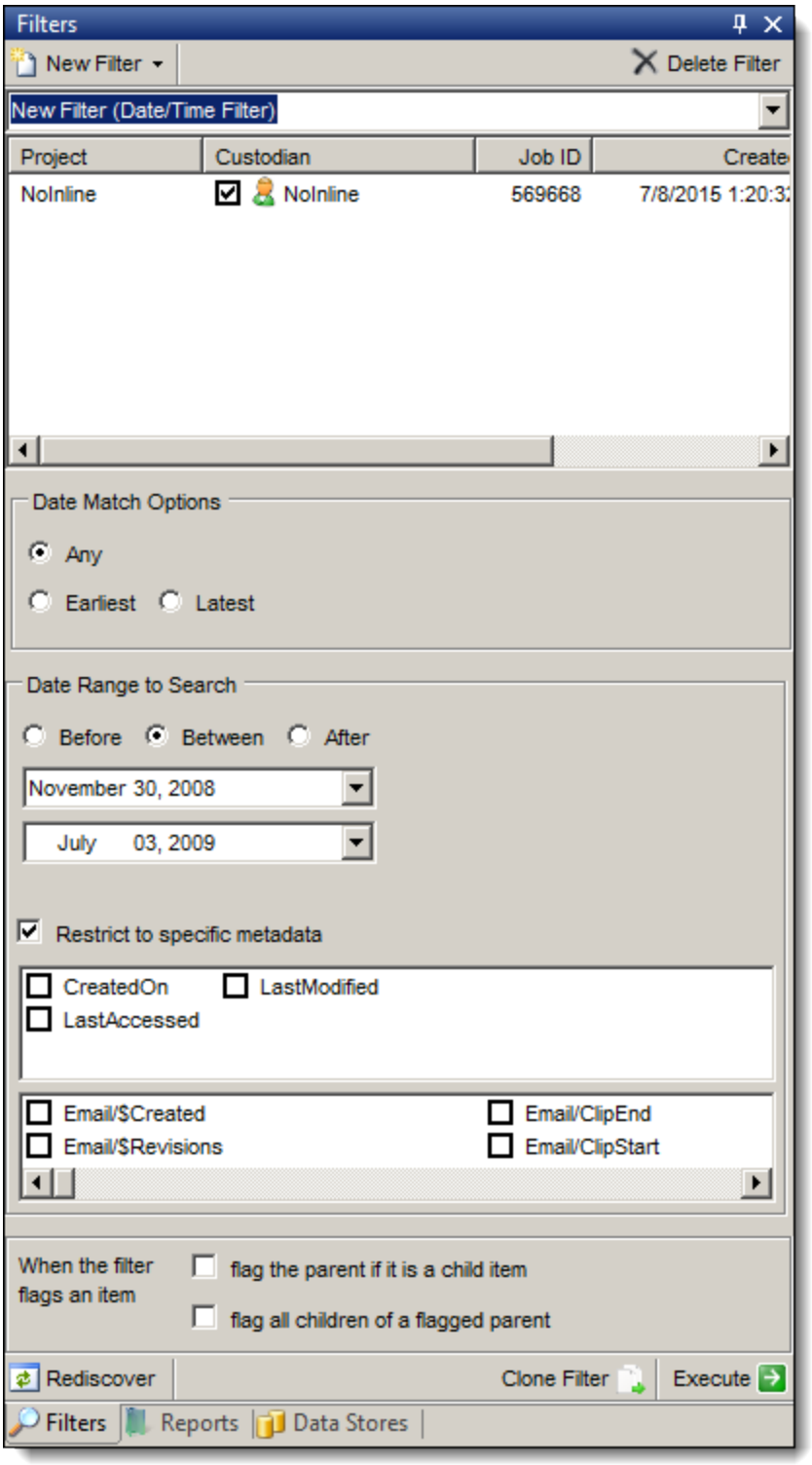
Note: Use two sets of double quotation marks to represent a blank.

7.3.2 Date/Time filter

For this filter, you can set specific date ranges and metadata after ingestion. This filter uses a metadata search so it does required extracted text for the selected custodian.

1. Select **Date/Time Filter** in the **New Filter** menu.
2. Enter a name in the **New Filter** box and press **Tab**.

3. Click the **Custodian** checkbox for one or more jobs that you want to filter for date and time.



4. Select a Date Match Option:
   - **Any** - filters on any date available for a document. Each document is associated with multiple dates, such as document sent date, delivered date, date last printed, and so on. When you select this option, the document will be returned if any of the dates match.
   - **Earliest** - filters on the earliest date in the metadata list. Usually filters on the create date.
   - **Latest** - filters on the latest date in the metadata list. Usually filters on the last modified date.

5. To specify a data range, select **Before**, **Between**, or **After**. Use the date drop-down option to select a date. (The beginning date starts at midnight, while the end date terminates at 11:59.)
6. To control the metadata used by the filter, select **Restrict to specific metadata**, and then choose the metadata fields. These fields are based on the metadata discovered in the document collection during import. They are pulled dynamically from the database.

   Note: When running a date filter, make sure to restrict that filter to specific metadata or else it will return hits on any date that meets the parameters you entered. This may include parameters that you don't care about when executing a date filter, such as Last Printed On. In this way, not restricting to specific metadata often results in an over-inclusive document list and costs you time during your QC process.

7. To run the filter, click **Execute**.

7.3.3 Deduplication filter

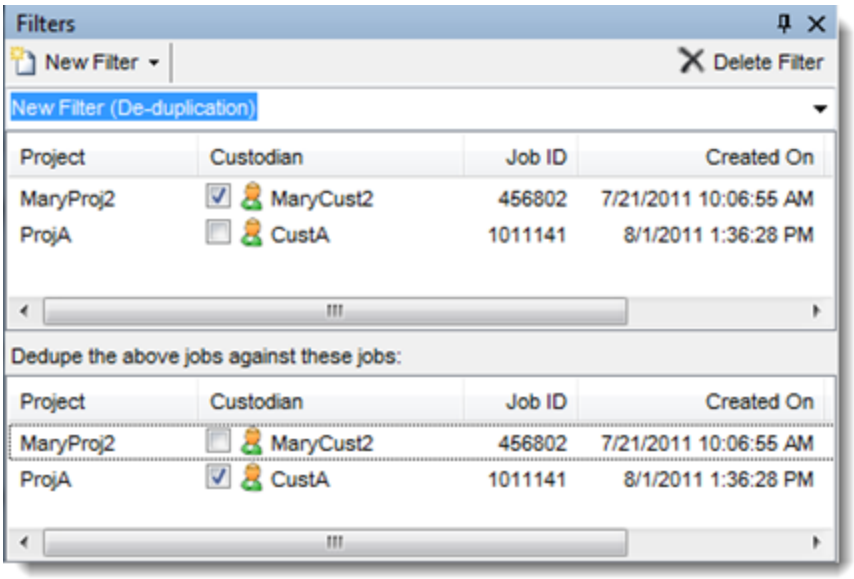
Use the Deduplication filter to remove multiple copies of a document from a data set by selecting two groups of custodians for this process. When you run a deduplication filter for the first time, the RPC generates the hash tables used for it, and adds them to the database. The performance time is improved when you re-run the filter because these tables are already built.

You can use this filter in the following ways:

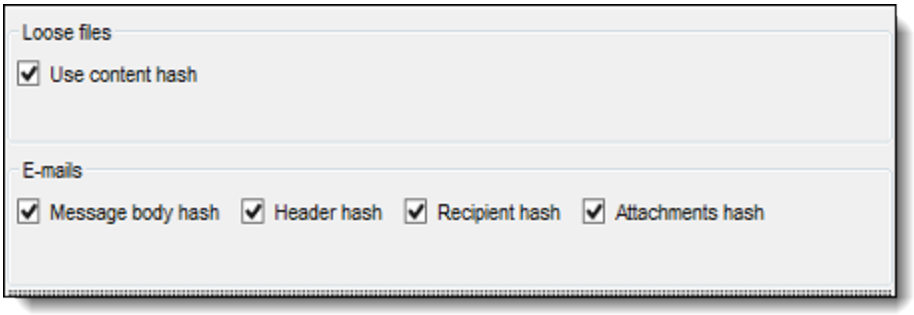
- **Dedupe the custodian against self** - you can de-dupe the data set of a custodian against itself. For example, you could de-dupe the laptop against the desktop of a custodian.
- **Dedupe one custodian against another** - you can select one custodian whose document set may contain duplicates, and then another custodian whose data set determines the documents to be removed. For example, you may want to dedupe the messages in John's mailbox against those in Jane's. See Targeted deduplication filter scenario on page 50.
- **Dedupe against multiple custodians** - you can dedupe one custodian against multiple custodians. The RPC uses a weighting algorithm to determine which documents to dedupe, when the same custodian is in both groups.

Note: The Deduplication filter does not delete any documents. It simply does not return them in the list of documents.

1. Select **Deduplication** in the **New Filter** menu.
2. Enter a name in the **New Filter** box and press **Tab**.
3. Click the **Custodian** checkbox for one or more jobs in the top window that you want de-duplicated against the jobs selected in the bottom window.



4. Ensure that the **Use content hash** option under the **Loose Files section** is checked. This option should always be enabled. Note that this option will be removed from the GUI in an upcoming release.



5. Select filtering options for E-mails. You can select a combination of different hashes that the RPC will use to identify duplicate files.
    - Processing generates four different hashes for emails and keeps each hash value separate, which allows users to de-duplicate in processing based on individual hashes and not an all-inclusive hash string. For example, if you're using processing, you have the ability to de-duplicate one custodian's files against those of another custodian based only on the body hash and not the attachment or recipient hashes.
        - **Body hash** - takes the text of the body of the e-mail and generates a hash.
        - **Header hash** - takes the message time, subject, author's name and e-mail, and generates a hash.
        - **Recipient hash** - takes the recipient's name and emails and generates a hash.
        - **Attachment hash** - takes each SHA256 hash of each attachment and hashes the SHA256 hashes together.
6. Select **Filtered Dedupe** if you want to de-duplicate the data returned from another filter. Next, select the filter in the activated list box. You will be performing the deduplication on a smaller data set, which speeds up the processing. For example, you may want to use a keyword filter on a document.
7. To run the filter, click **Execute**.

7.3.3.1 Targeted deduplication filter scenario

Let's say that you want to dedupe John's documents against Jane's documents. Before deduplicating, you need to ensure that any documents in John's data set containing the term "Fruit Loops" are relevant and therefore should be included in the results. Likewise, any documents in Jane's data set containing the term "Cheerios" should also be included in the results.

To perform this targeted deduplication filter:

1. Create Full text search filters on:
    a. John's data set for the word "Fruit Loops". See Full text search filter on page 53.
    b. Jane's data set for the word "Cheerios". See Full text search filter on page 53.
2. Create a new Deduplication filter and ensure the following are set:
    - **Source filter** - Jane's Cheerios filter. Add as an Exclude filter.
    - **Target filter** - John's Fruit Loops filter. Add as an Exclude filter.
3. Execute the filter to dedupe against both custodians. Documents returned will still contain documents with terms "Fruit Loops" and "Cheerios" specified in John's and Jane's Exclude filters.

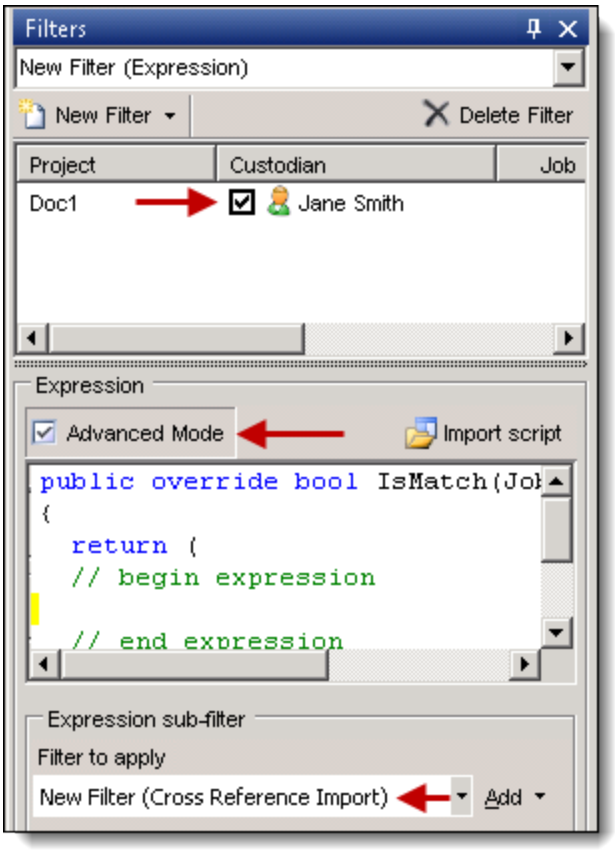
7.3.4 Expression filter

With an Expression Filter, you can use C# syntax to write custom code to manipulate the data set. The code runs on the desktop, and it is dynamically compiled and executed on the database.

This filter type is frequently used for checking errors that occurred during text extraction.

1. Select Expression in the **New Filter** menu.
2. Enter a name in the **New Filter** box and press **Tab**.

3. Click the **Custodian** checkbox for one or more jobs that you want to filter with your custom code.



4. Enter a code for a simple filter in the **Expression** box. If you want to use C# syntax, click **Advanced Expression** to auto-generate an outline of a method, and then enter your code.
5. To run the expression on a limited set of data, select a filter in the **Sub-Filter** box.
6. To run the filter, continue with Step 5 in Creating a filter on page 42.

The following sections provide examples of expression filters. Note that none of these examples requires advanced mode.

7.3.4.1 Example: Find documents with no error message and aren't flagged as unprocessable

Use the following expression to isolate documents that have an error message but are not flagged as unprocessable. This filter is useful for situations in which you've run text extraction and want to rerun anything that threw an error.

```
public override bool IsMatch(Job job, Matter matter)
{
return (
// begin expression
IsIn((matter.Message != "") && (matter.Unprocessable == false))
// end expression
);
}
```

7.3.4.2 Example: Find password-protected documents

Use the following expression to isolate password-protected documents.

```
public override bool IsMatch(Job job, Matter matter)
{
return (
// begin expression
IsIn(matter.Message.Contains("Password"))||(matter.Message.Contains("password"))
// end expression
);
}
```

7.3.4.3 Example: Find text within the VirtualPath

Use the following expression to isolate a string of text within the VirtualPath.

```
public override bool IsMatch(Job job, Matter matter)
{
return (
// begin expression
IsIn(matter.VirtualPath.Contains("Sent Items")
// end expression
);
}
```

7.3.4.4 Example: Find specific file extensions

Use the following expression to isolate specific file extensions. You must enter all file extensions in CAPS, not in lower case; if you enter the extensions in lower case, the filter will provide no results.

```
public override bool IsMatch(Job job, Matter matter)
{
return (
// begin expression
IsIn(matter.FileExtension, "XLS", "XLSX", "XLSM", "XLSB", "XLA", "XLB")
// end expression
);
}
```

7.3.4.5 Example: Find documents whose names start with a particular string

Use the following expression to isolate documents with names that start with a particular string.

```
public override bool IsMatch(Job job, Matter matter)
{
return (
```

```
// begin expression
IsIn(matter.FileName.StartsWith("image")
// end expression
);
}
```

7.3.4.6 Example: Find documents whose names end with a particular string

Use the following expression to isolate documents with names that end with a particular string. This is useful for finding files with a specific literal extension instead of the identified extension.

```
public override bool IsMatch(Job job, Matter matter)
{
return (
// begin expression
IsIn(matter.FileName.EndsWith("emf")
// end expression
);
}
```

7.3.5 Full text search filter

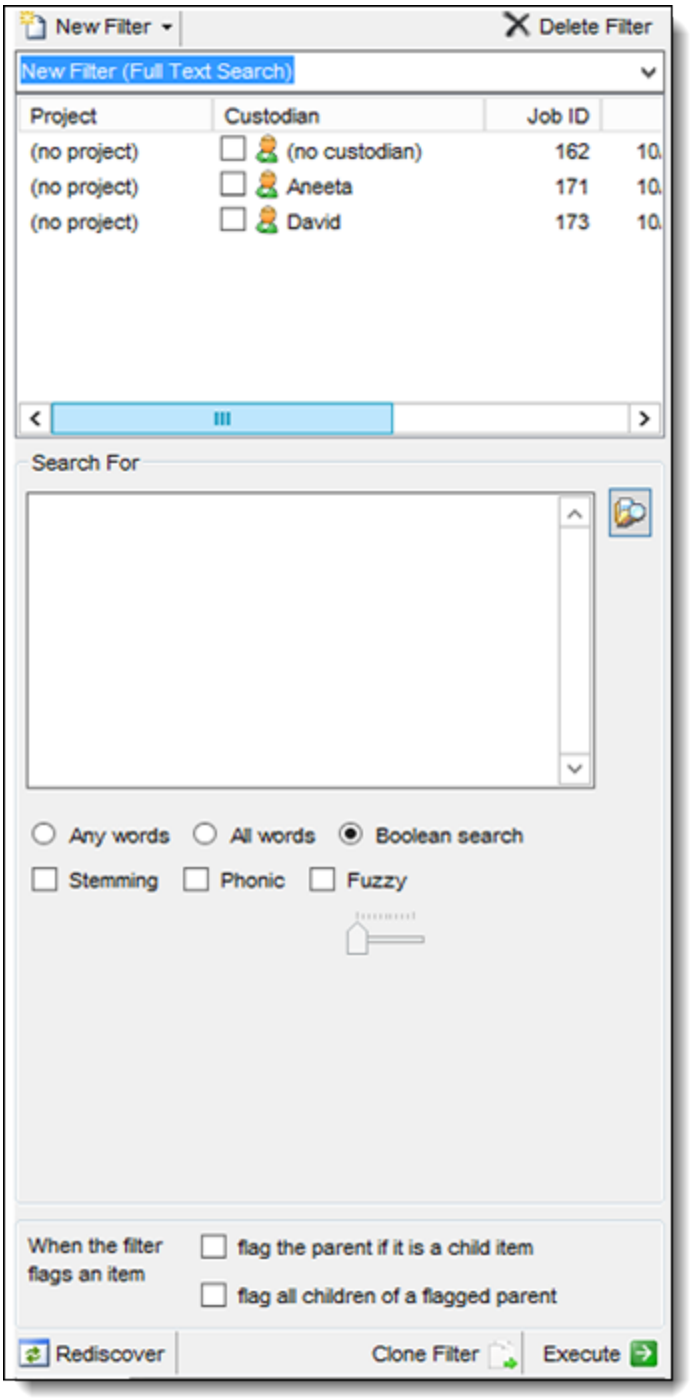
You can use the Full Text Search filter to run keyword searches on data indexes built for specific import jobs. The RPC uses the dtSearch engine to run these filters. However, it executes the search for each term in parallel using a separate thread, and then combines the results of these searches.

A common use for a full text search filter is the removal of extraneous documents from a data set. For example, you could run the filter on a group of documents before importing them into Relativity for review. You could also use this filter to tag documents that are potentially privileged. See Exporting data on page 59.

Note: You can run the Comprehensive Hits Report to display a list of statistics about the number of documents that match a search term and other data. See Running standard reports on page 99.

1. Select **Full Text Search** in the **New Filter** menu.
2. Enter a name in the **New Filter** box and press **Tab**.

3. Click the **Custodian** checkbox for one or more jobs that you want included in the full text search.



4. Either click  next to the Search For window to browse for a file containing your search terms, or enter then in the following format: "relativity","processing"
    - When you import a list of terms through the browse option, the RPC automatically sets the search to use the Boolean option.
    - If you manually enter your own terms, the search defaults to Any words, and you must remember to switch to Boolean if this is the option you want.

- You can also add operators used by the dtSearch engine, such as Boolean operators AND, OR, NOT.
- You can manually resize the Search For window.

5. Select any of the searching options that you want to use. These options follow the searching conventions used by the dtSearch engine.
6. To run the filter, continue with Step 5 in Creating a filter on page 42.

7.3.5.1 Boolean filter

To create and execute a boolean filter:

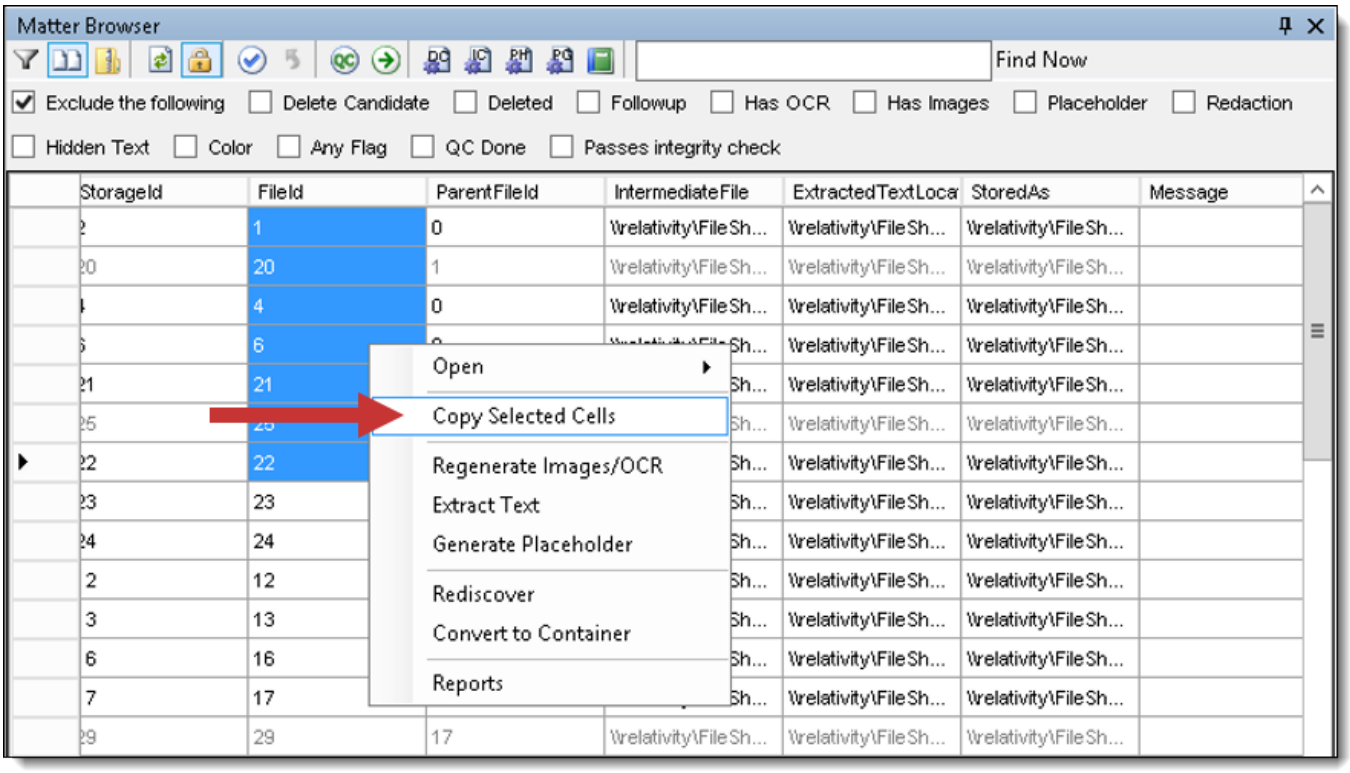
1. Click **New Filter** and select **Full Text Filter**.
2. Name the filter at the top of the custodian window.
3. Select the data sets to which you want to apply the filter. Note that each data set has its own index and each index records those noise words that were used when that index was built. Note also that noise words aren't configurable in the RPC.
4. Enter the terms you'd like to include in the filter. Each line is its own search, and each line will be executed in its own thread. These will be joined together by an OR statement, in that the entire search string is an implicit OR since you've selected the Boolean. The fact that they're on separate lines allows you to search for them in parallel across multiple threads.
5. Select **Boolean search**.
6. Click **Execute**. Note that the RPC now displays which threads it's currently searching, as well as the fact that it's currently populating the search results.

7.3.6 Retrying imaging or text extraction errors through a filter

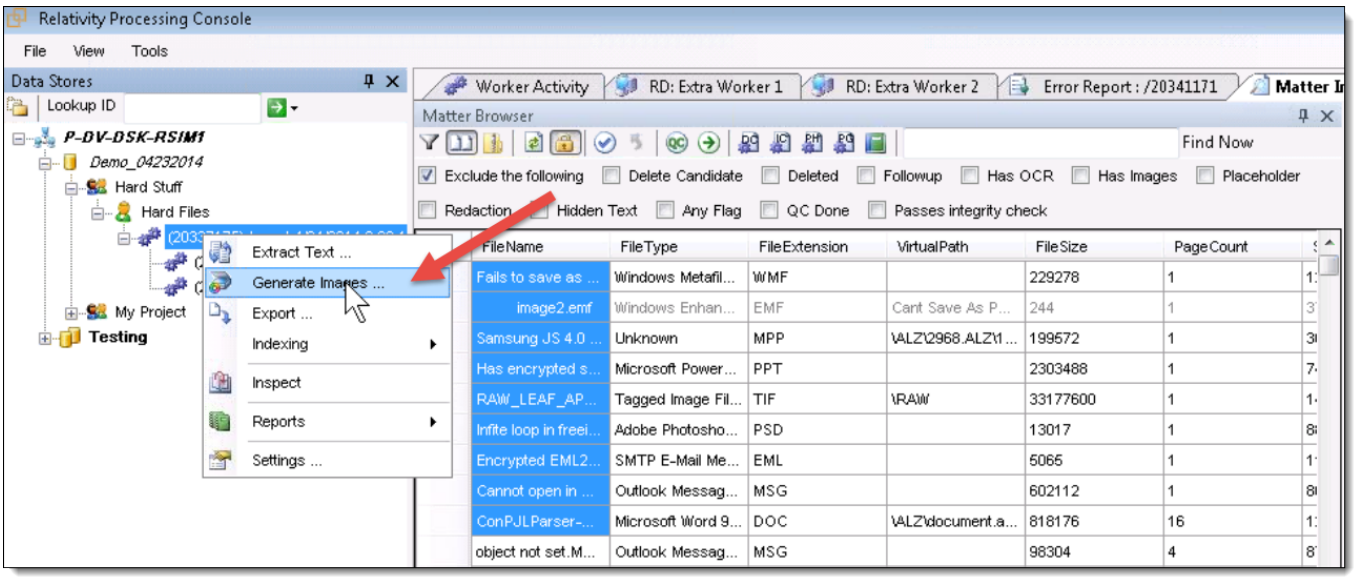
There are many ways to isolate documents that you may want the system to retry, depending on your needs. For example, in the matter inspector you may want to sort on the Message column to group all items with an error message together. You can also de-select the Exclude the following checkbox and check the Placeholder checkbox and then refresh the screen to display only items that received a placeholder image. Alternatively, you could sort on the Page Count column to sort anything with a page count of -1 to the top.

To retry a group of errors through a filter in the RPC, perform the following steps:

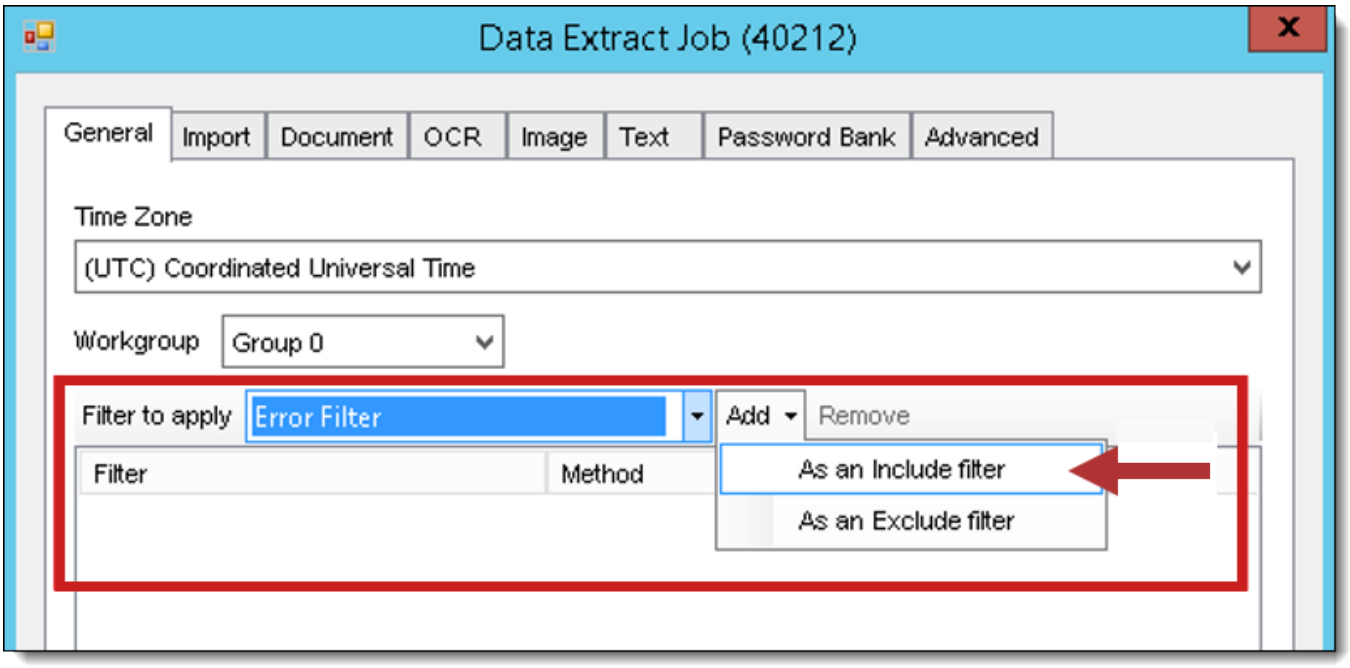
1. Open the matter inspector for the job that contains the errors.
2. Select the File Ids of all the files that you want to retry.
3. Right-click and select **Copy Selected Cells**.



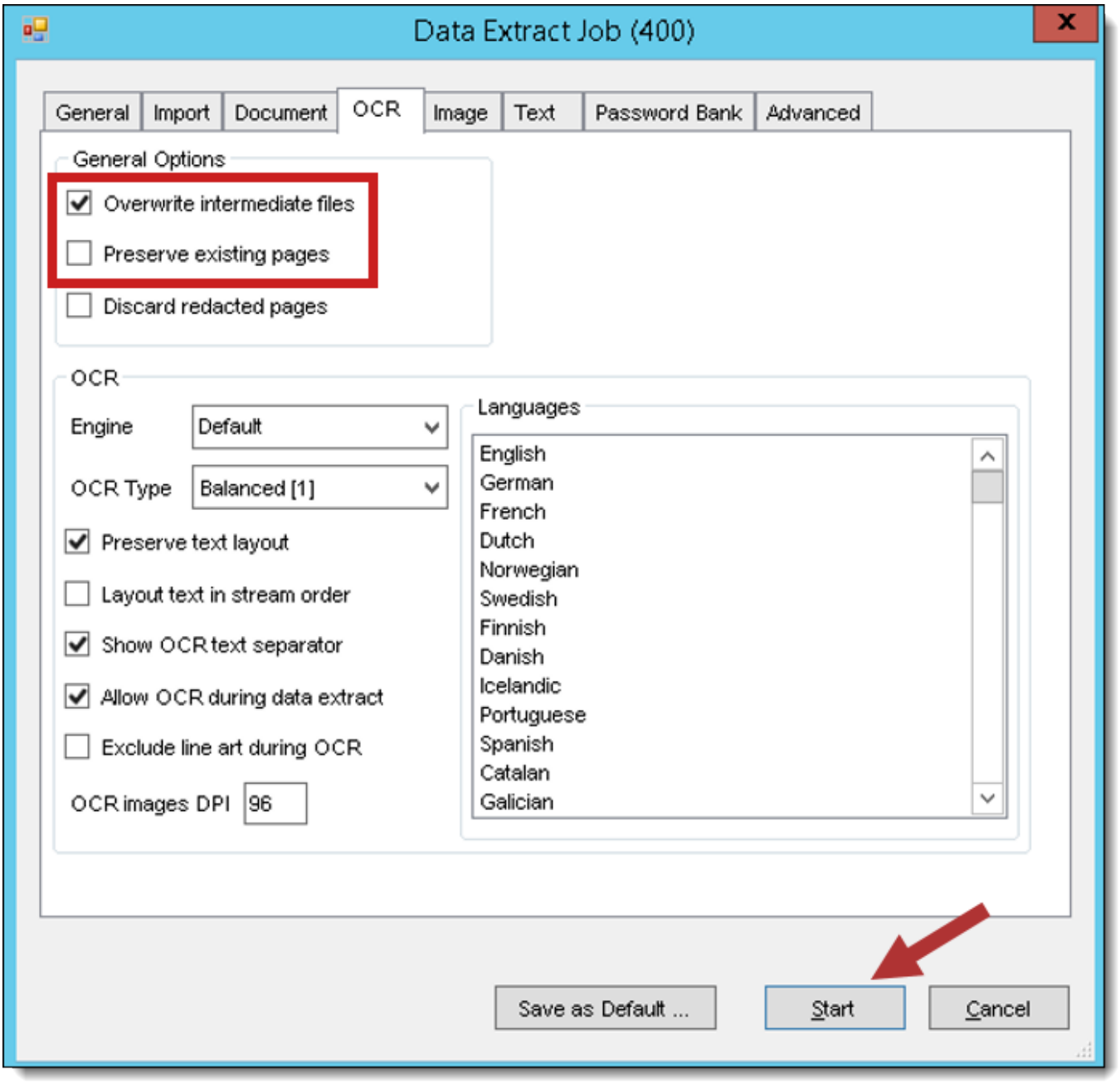
4. Use the File IDs that you copied to the clipboard to create a cross reference import filter. See Cross reference import filter on page 45 for more information.
5. Right-click on the import job and select **Generate Images** or **Extract Text**, as appropriate.



6. In the Generate Images or Data Extract Job window, select the filter you just created from the Filter to apply drop-down list and select **As an Include filter** from the **Add** drop-down list.



7. Select the **OCR** tab, check the box next to the **Overwrite intermediate files** option, and un-check the **Preserve existing pages** box. Then click **Start** to retry all of the errors you filtered for.



Note: Sometimes documents that error on a data extract job are handled without error on an image generation job. You can try generating images for data extract errors prior to indexing. When you index the job, the software uses page-level text derived from the image generation job if the doc level extracted text doesn't exist. Additionally, during the indexing process the system copies the page-level text files into a single document-level text file and saves it where the document-level text file would have been created had the file not thrown an error during data extraction.

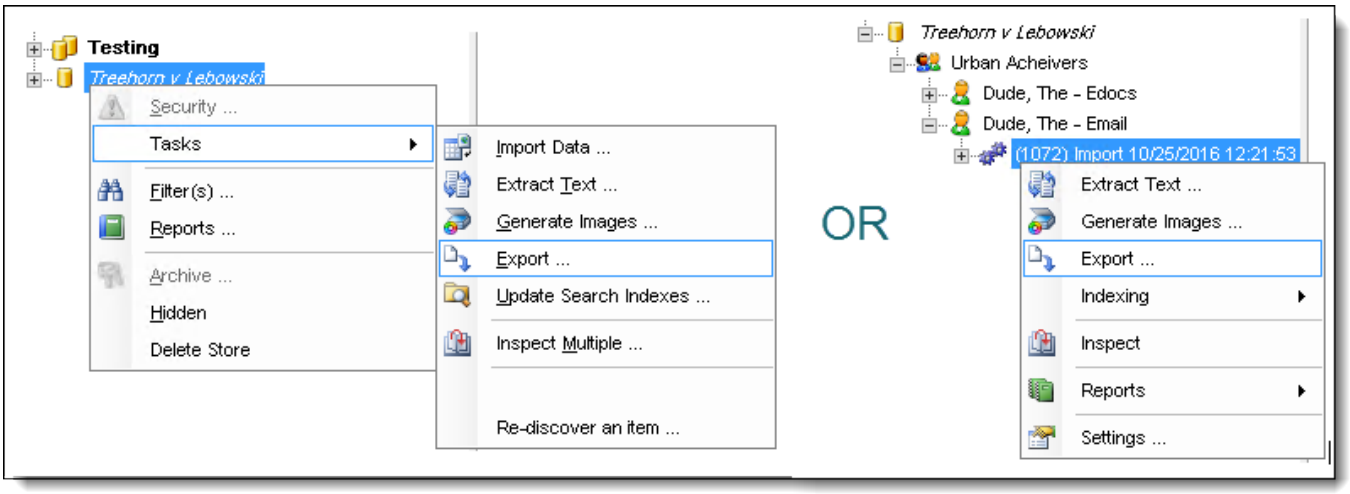
8 Exporting data

When you are ready, you can export files for review, production or any other purpose you may have. Use the steps provided in the following sections to run your exports.

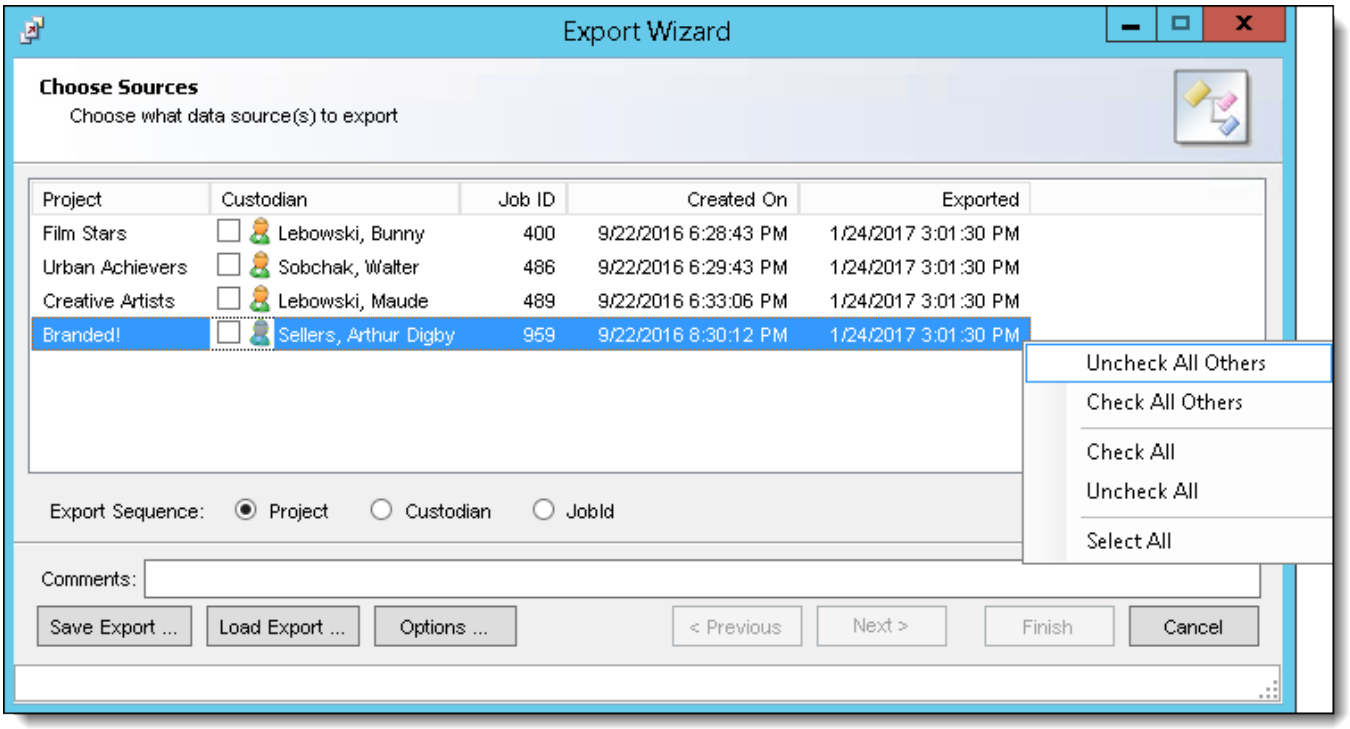
Note: For more information on how to export through the RPC, see the Using the RPC Exporter video webinar on the Relativity Training site.

8.1 Creating an export job

1. In the Data Stores window either right click on a data store or drill down to a specific import job, right click on either, and click **Export** to open the Export Wizard.



2. When the Export Wizard is opened either no import jobs are checked (if opened by clicking on the data store) or a single import job is checked (if opened by clicking on a specific import. Regardless of which option was chosen, you can check or un-check import jobs as needed for your export. If you want to use a saved export file (.EXF) from a previous job, click **Load Export** to select it before you pick which jobs to export. There are no adverse effects of selecting the jobs first, but loading an export un-checks all import jobs automatically and you'll need to reselect the jobs again.

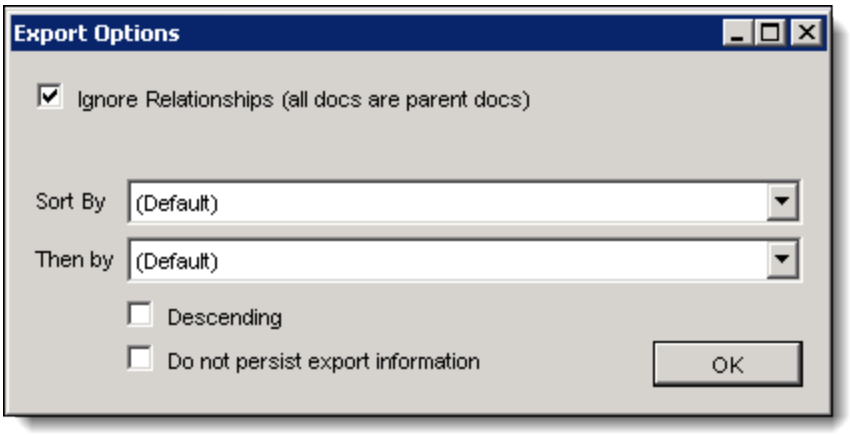


- When choosing which import jobs to export you can right click on any of them and make use of the options displayed in the image above. They are all self-explanatory except for 'Select All' which highlights all the import jobs but doesn't check or un-check any of them. You can also left-click and drag to highlight a number of jobs then press the space bar to select or deselect them.
- The default export sequence is by Project. If one or more import jobs have no value entered for the Project, as soon as an import is checked a pop-up window will inform you of that the export sequence is automatically being changed to JobId.

3. If you are selecting multiple imports you can choose which export sequence to be used. The sequence setting dictates the order in which the exporter exports the selected items.
    - The Export Sequence is relevant when you export multiple custodians in a single export job. The sequence setting dictates the order in which the exporter exports those sources.
        - **Project** - exports alphabetically by project name.
        - **Custodian** - exports alphabetically by custodian name.
        - **JobId** - exports by the JobId value in ascending order.

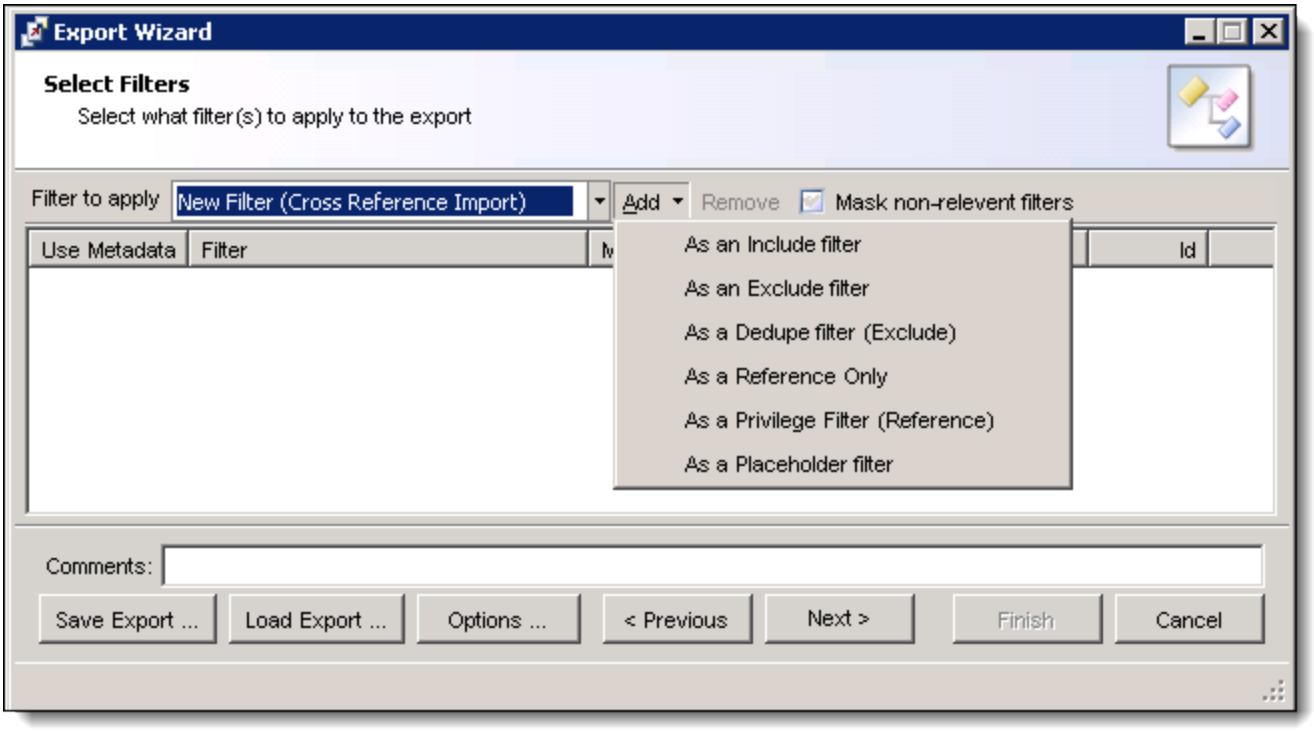
    Note: The Comments window serves no purpose at this time.

4. If needed, click **Options** to display the Export Options dialog box. None of these are required to run an export



- **Ignore Relationships** - select this option to treat all documents like parents. For example, you might want to export only the documents with keyword hits, so you may need to export a child without a parent. You can use this option to flatten all family relationships.
- **Sort By** - performs primary sort based on FileId, FileName, ImportSource, ParentFileId, RootFileId, or Virtual Path.
- **Then By** - performs secondary sort using selected metadata field.
- **Descending** - select to perform a descending sort. For example, you may want to use a descending sort on the creation date of e-mail messages, so that you list the messages from newest to oldest.
- **Do not persist export information** - enabling this prevents any export information such as completion time and assigned Bates numbers from being tracked in an export table in the Store database. This is useful for cases in which you want to run a test export and you don't want that job's data permanently tracked.

5. Click **OK** to redisplay the Choose Sources window of the Export Wizard.
6. Click **Next** to display the Select Filters window.



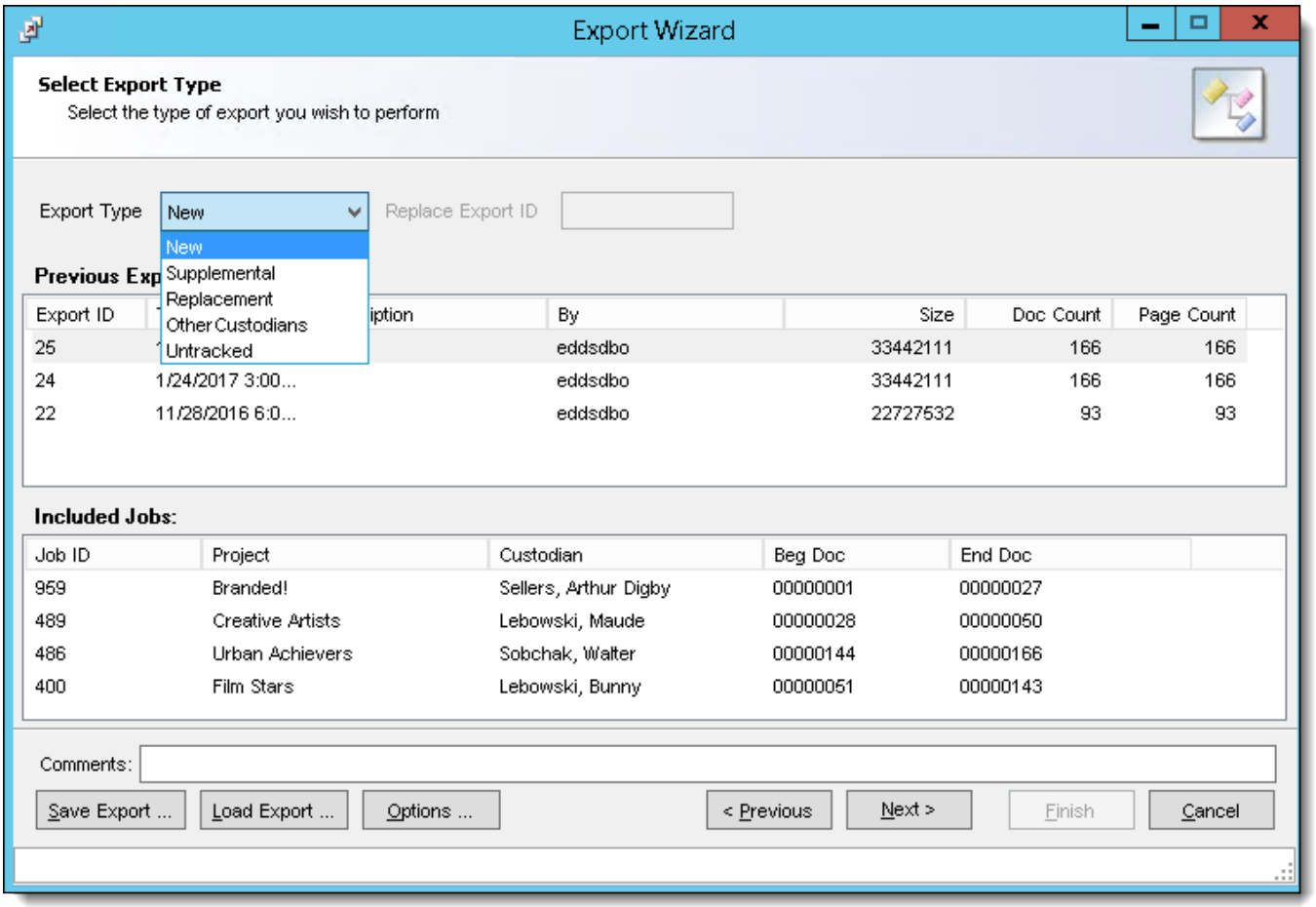
7. (Optional) Select a filter from the **Filter to apply** drop-down list.
8. Click **Add**, and select one of following options to control the filter behavior.

Option	Description
As an Include filter	Includes any document that matches the criteria set in this filter. For example, you could add a keyword search as an include filter, so that all documents with this keyword are exported. **Note:** Select the checkbox under **Use Metadata** to export keyword, and its frequency in a document. This functionality is available because metadata is actually stored in filters. New columns with the metadata are added.
As an Exclude filter	Excludes any document that matches the criteria set in this filter.
As a Dedupe filter (Exclude)	Same as an Exclude filter, except that the RPC uses a different logging code for deduplication. You can run a report on this code to identify files excluded due to deduplication.
As a Reference Only	References the criteria in the filter and may perform an action based on a match. For example, tagging information obtained from Relativity could be used to apply tags to documents that match the metadata in the filter criteria. All documents are exported and tagged as required.
As a Privilege Filter (Reference)	Similar to a Reference Only filter. The Potentially Privileged column is updated to Y for any document that matches the filter criteria. All documents are exported.
As a Placeholder filter	Inserts a placeholder for any document that meets the criteria set in this filter. For example, you could use this filter to substitute a placeholder for all Excel/Power Point or large text files in a production. When you select this filter, the RPC runs the export job and substitutes the placeholder for these files without modifying the PDFs already generated. This eliminates the

Option	Description
	need for you to image these files, since you can leave them as-is.

- When multiple include/exclude filters are used, the logic used to combine them is an AND statement only. Reference, Privilege and Placeholder filters do nothing to limit any documents from being exported.
- If any filters were added incorrectly, highlight them and click Remove. The Mask non-relevant filters checkbox will hide any filters that do not apply to the import jobs selected for export. This button also can be used to refresh the filters in the drop-down list if you created new ones while the export wizard was open.

9. Select **Next** to display the Export Type window.



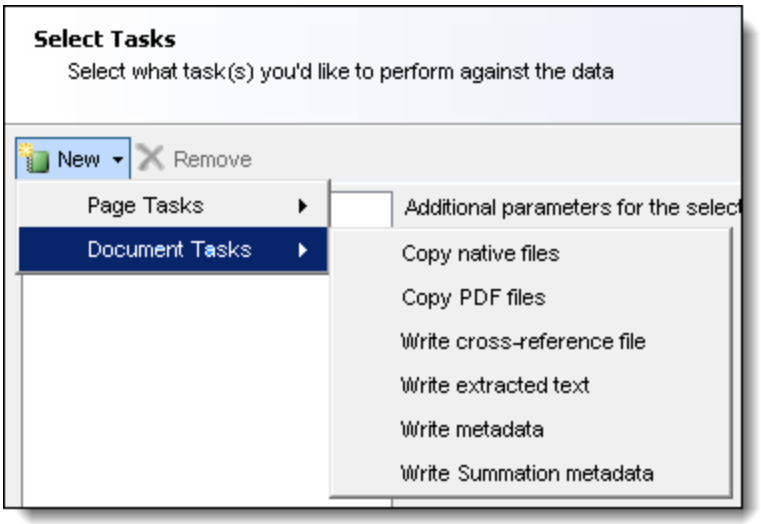
10. Select an **Export Type**:
    - **New** - this is typically used for a first-time export. The tracking status for the export is updated and tracked in the database. The RPC tracks bates numbers, page counts, and other information so that you know where to begin numbering for your next export with the same custodian. This information is displayed in the Previous Exports and Included Jobs panes of the export wizard on future exports.
    - **Supplemental** - is intended for partial exports for the same data set in sequence. In other words, use this option if you want to add documents to an export job that already exists. Note that you still need to specify a starting number for the supplemental documents, as the RPC won't automatically number them.
    - **Replacement** - replace an entire export job with a completely different export job. When you select this option, you must specify a valid export ID value in the Replace Export ID field to the right, which is only

enabled when you select Replacement as the export type. Simply highlight the job you want to replace in the Previous Exports pane to populate the field.

Note: You can't modify filters used in an export that uses a setting of New, Replacement or Supplemental. If you need to alter a locked filter, you must first clone the filter and then edit the clone.

- **OtherCustodians** - export the DeDuped Custodian and Deduped Path information via Relativity (the front end). You can't export either of these fields if you don't choose this option. As a result, standard practice is to export all other metadata using one of the other dropdown settings and to export these two fields to a separate overlay metadata file using this option.
- **Untracked** - no tracking information for the export is added to the database. You may want to select this option if you are doing some experimental export jobs that you do not want tracked in the database. If you are building a new export file definition, you may run several test exports to see how the data is displayed with the current settings.

11. Click **Next** to display Select Tasks window.
12. Click **New**. Point to **Page Tasks** or **Document Tasks**, and click on one or more export tasks. Each of these options will have a variety of task parameters that can be set. For details, see Updating task parameters on page 66.
    - **Page Tasks** - available options are Copy Images and Copy Text.
        - **Copy Images** - writes TIF and/or JPG image files to a desired location. An image generation job must have been performed first. The exporter will throw an error for each document exported that was not previously imaged.
        - **Copy Text** - writes page level text files to a desired location. An image generation job must have been performed first. The exporter will throw an error for each document exported that was not previously imaged.
    - **Document Tasks** - available options are Copy Native files, Copy PDF Files, Write cross-reference file, Write extracted text, Write metadata, and Write Summation metadata.

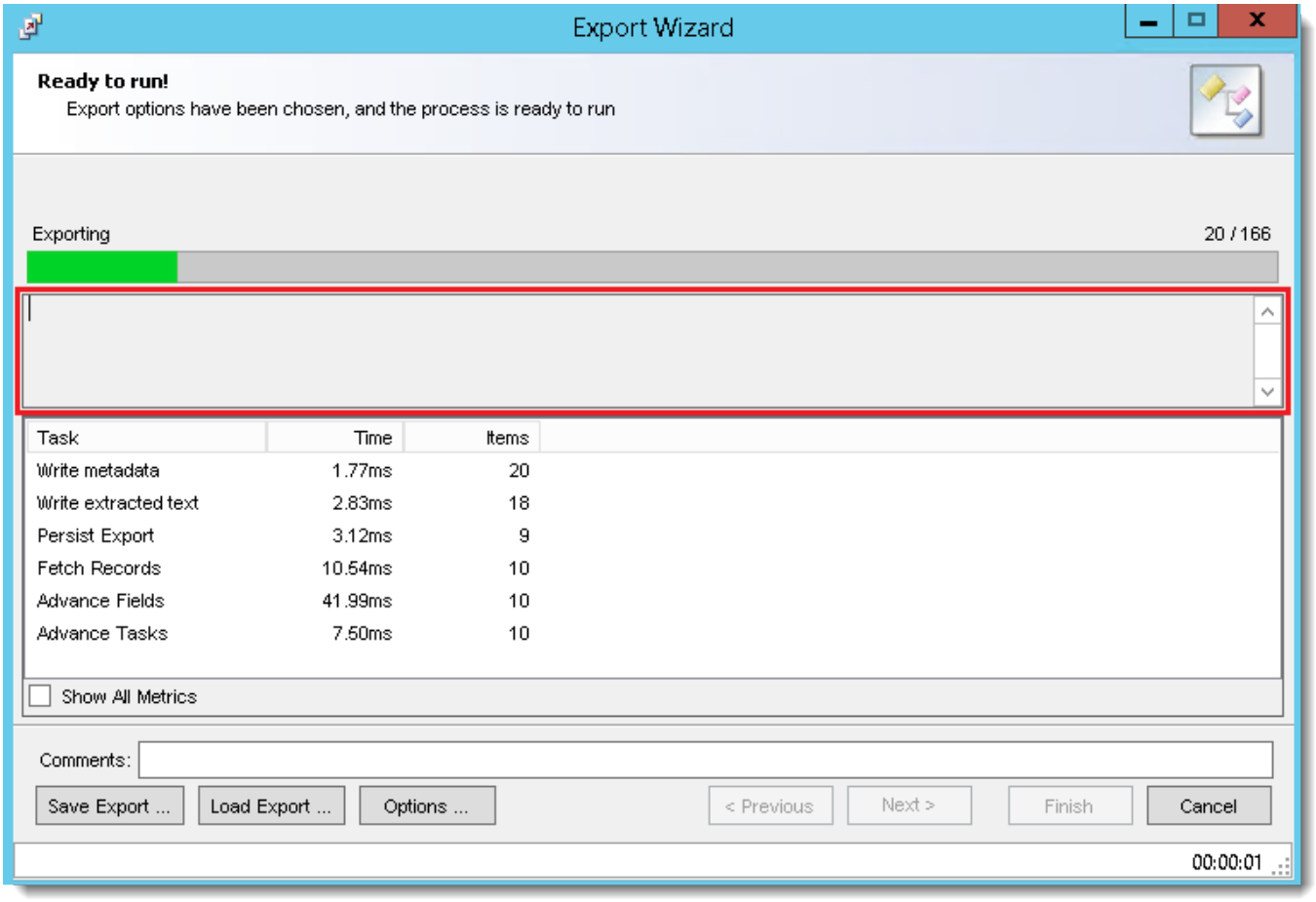


        - **Copy native files** - copies the native file for an exported document to a desired location.
        - **Copy PDF files** - copies a PDF file for an exported document to a desired location. An image generation job must have been performed first. The exporter will throw an error for each document exported that was not previously imaged.

- **Write cross-reference file** - writes a stock-formatted cross-reference file. It is recommended to include this task on ALL exports. It frequently is a very handy file to have.
- **Write extracted text** - writes document level extracted text to a desired location. A text extraction or image generation job must have been performed first. If an image generation job was run but not a text extraction job, the page level text files will be copied together to form a doc level text file automatically during the export. The exporter will throw an error for each document exported that was not previously imaged or text extracted.
- **Write metadata** - writes custom metadata to a desired location. Alternatively, this task can be used to write a file that contains no metadata. For example, you could write a readme.txt file to include reminders about a customer's preferences for their deliverables.
- **Write Summation metadata** - currently inoperative.

Note: You can add the same task multiple times to an export job. Depending on the task, you can assign a different file name or folder location to each copy of the task. For example, you could export a set of metadata for a client and another for opposing counsel by adding the Write metadata task twice. You would then modify the name of the task added under Document Tasks, and the name of file output for each task. If the export job needs to be repeated, you do not need to redo each of the document or page tasks. You can clear the checkboxes for the tasks that do not need to be redone, and then perform the export. For example, you may want to add new fields to the metadata file, so you can select only Write metadata under Document Tasks.

13. Click **Next** and then **Finish** to run the export job. The Finished box contains any errors that occurred along with the File ID for troubleshooting purposes. The window outlined in red below will contain any errors that occurred along with the File ID for troubleshooting purposes.



14. (Optional) Click **Save Export** to save your current settings as a reusable export file (.EXF). You can select this load file for use with another export job by clicking Load Export. You can save your export at any point during the export process, you do not have to actually export anything before saving. If you are creating a complicated export, you may want to consider saving your export periodically as you add tasks.

8.2 Updating task parameters

In the export wizard, each document and page task has multiple parameters that you can update to meet the requirements of a specific export job. Some of these parameters are universal to all export tasks while others are specific to that task. The table below outlines the various parameters, what tasks utilize them and if they are universal or specific.

Parameter	Associated Tasks	Universal or Specific	Notes
Bates Number	All	Universal	Also serves as control number for document level exports. Should be left as {BatesBeginDoc}, expand this section to format the bates/control number.
DocLevel	All	Universal	If True, numbers are incremented at the doc level.
Ignore Number	All	Universal	If True, the value in the StartAt field will be disregarded and only the prefix and suffix will be used.
Prefix	All	Universal	The prefix to prepend each bates/control number.
Reset Condition	All	Universal	When the value entered here changes, the control/bates number will reset to the StartAt number. Typically used with {Custodian}

Parameter	Associated Tasks	Universal or Specific	Notes
Reset Counters	All	Universal	When set to true, any generic or file counters will reset when the reset condition is met.
StartAt	All	Universal	The starting number to assign.
Suffix	All	Universal	The suffix to append to each bates/control number.
CopyMode	Copy Images	Specific	Normal = export images and create image load files, BuildNoCopy = create image load files but don't actually export the images. This setting is rarely changed from Normal.
Custom Foldering	Copy Images & Copy Text (page level)	Specific	When set to False the system will automatically the standard 00\00 folder structure after the Image Sub-Folder. When True the system will use the value in the Image Sub-Folder field only, allowing the user to enter custom counters if desired.
Distributed	Copy Images, Copy Text, Copy Natives & Copy PDFs	Universal	True = all copying and branding is performed by the workers. False = these tasks are performed by the machine running the export.
Endorsement	Copy Images, Copy Text & Copy PDFs	Semi-Universal*	Settings for what to brand on images, page level text and PDFs.
Bottom Left	Copy Images, Copy Text & Copy PDFs	Semi-Universal*	What to endorse in the bottom left corner.
Bottom Right	Copy Images, Copy Text & Copy PDFs	Semi-Universal*	What to endorse in the bottom right corner.
Gap	Copy Images, Copy Text & Copy PDFs	Semi-Universal*	Space between endorsement and image in inches.
Margin	Copy Images, Copy Text & Copy PDFs	Semi-Universal*	The distance between the edge of the page and the endorsement.
Top Left	Copy Images, Copy Text & Copy PDFs	Semi-Universal*	What to endorse in the top left corner.
Top Right	Copy Images, Copy Text & Copy PDFs	Semi-Universal*	What to endorse in the top right corner.
Image / Text / PDF / Extracted Sub-Folder	Copy Images, Copy Text, Write extracted text & Copy PDFs	Specific	The subfolder to write the files to. Will be a subfolder of the value entered in the Working Folder field.
Jpeg Natives	Copy Images	Universal	True = use the native file if the native is a jpeg. False = use the imaged version of the native. Setting is ignored if endorsing is enabled.
LFP Filename	Copy Images, Copy PDFs	Specific	LFP file name to use. An OPT is also automatically created and will have the same name.

Parameter	Associated Tasks	Universal or Specific	Notes
Overwrite Images / Files	All	Specific	Overwrite existing files in the destination folder.
Production Name	Copy Images, Copy PDFs	Universal	The export volume name. This is the default value used for the LFP name. This value is used to populate the volume column in the LFP & OPT.
Working Folder	All	Univeral	The base folder to write all output to. If the export work is to be distributed to the workers it needs to be a location they have write access to. UNC pathing is strongly recommended.
Compress Text	Copy Text, Write extracted text	Specific	When set to true, all blank lines and leading or trailing spaces are removed from the text.
Default Text	Copy Text, Write extracted text	Specific	If a document has no extractable text, this text is used instead of nothing.
Multipage	Copy Text	Specific	Specifies whether to combine single page text files into a doc level text file or not.
Page Separator	Copy Text	Specific	Specifies what separator to use between each page of text if Multipage is set to True.
Strip Non-ASCII	Copy Text, Write extracted text, Write cross-reference file, Write metadata	Specific	Strips non-ASCII characters from the data or text before writing.
Text Encoding	Copy Text, Write extracted text, Write cross-reference file, Write metadata	Universal	Specifies the type of encoding to use when writing to the file.
Wrap Lines	Copy Text, Write extracted text	Specific	Word-wrap lines so they do not exceed the specified length. Use 0 to disable.
Copy Condition	Copy native file, Copy PDFs	Specific	The condition to evaluate to determine whether or not to copy the file. 0 or False means do not copy the file, 1 or True means copy the file.
Default Extension	Copy native file, Copy PDFs	Specific	The extension to use if the file doesn't have an extension.
Filename	All document level tasks	Specific	The file name to use.
MaxExtLength	Copy native file, Copy PDFs	Specific	The maximum number of characters in the file extension, 0 for no limit.
MaxNameLength	Copy native file, Copy PDFs	Specific	The maximum number of characters in the file name, 0 for no limit.
Placeholders Only	Copy native file, Copy PDFs	Specific	True = File will only be copied if the image for the document is a placeholder.
Source File	Copy native file, Copy PDFs	Specific	You won't need to change this. Ever. We promise.
UnicodeFilenames	Copy native file, Copy PDFs	Specific	If set to False, filenames are converted to ANSI and unconvertable characters are replaced with an

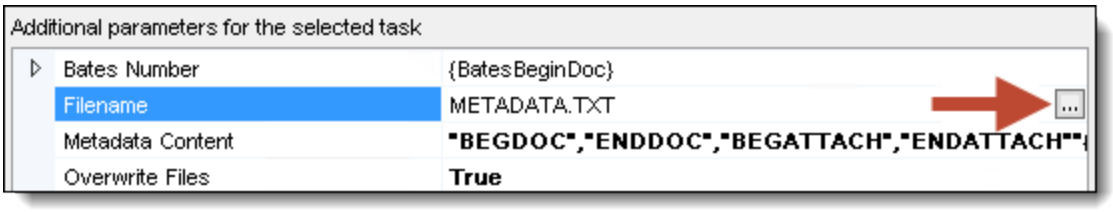
Parameter	Associated Tasks	Universal or Specific	Notes
			underscore. If True, the original Unicode file name is used.
Highlight	Write extracted text	Specific	Highlight search hits from the filters applied to the export. Requires Text Format to be set to Rich Text
Text Format	Write extracted text	Specific	Specifies to use Text or Rich Text as the format.
Metadata Content	Write Metadata	Specific	The metadata to write to the file. Can include any number of metadata fields in just about any format. Can also have no metadata if desired.

* Semi-universal items only apply to other tasks of the same type. For example, if you made endorsement settings to one image export task, those settings would also apply to any other image export task but they won't apply to endorsements on a Copy PDF task.

Note: You can rename any of the tasks if desired, just left click once on the name to highlight and a second time to edit it. For instance, if you are exporting two different metadata files, you may want to name one 'Opposing Metadata' and the other 'Our Metadata.'

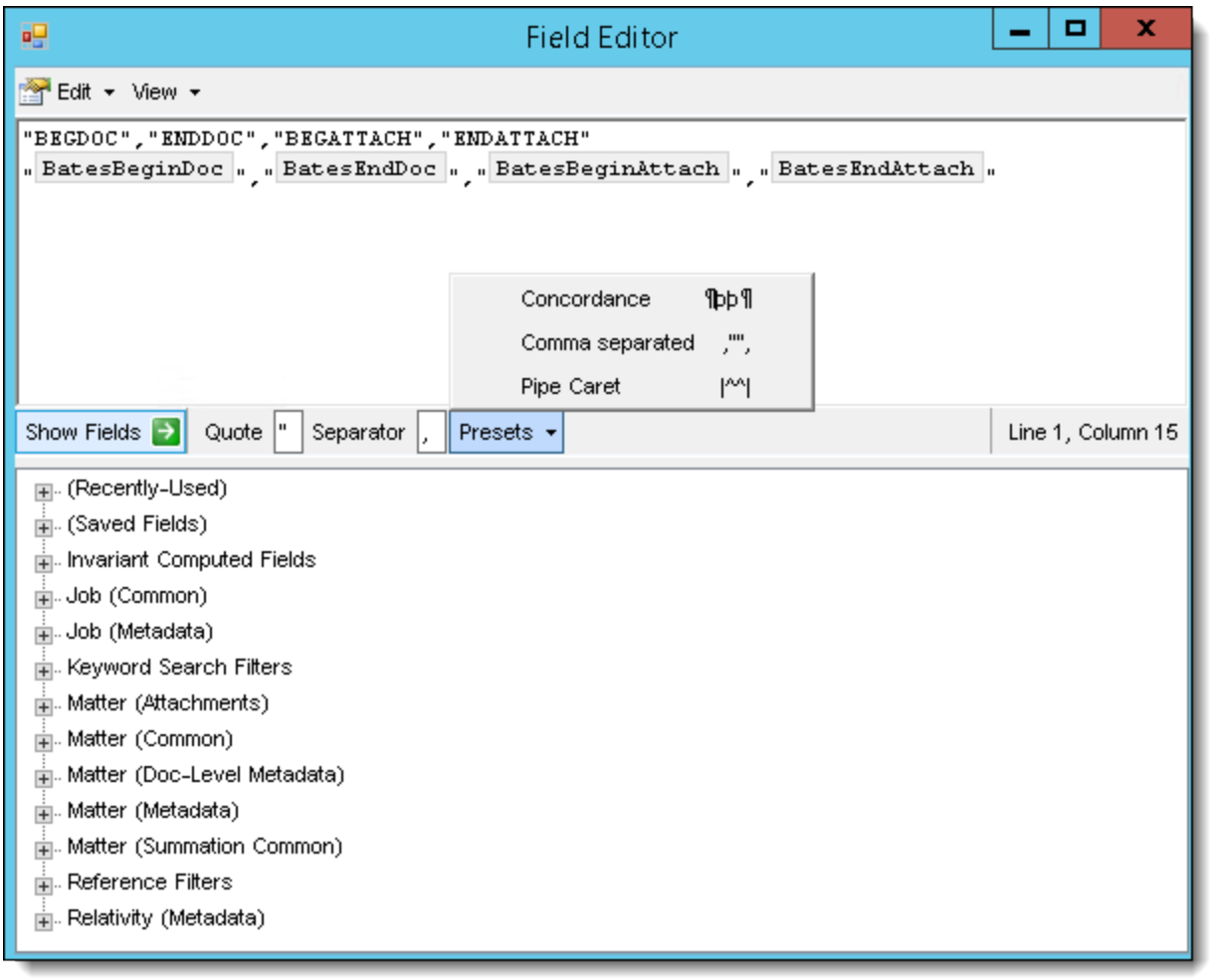
8.2.1 Using the field editor

Any parameters that aren't accessible via a drop-down option, with the exception of Working Folder, have an ellipsis button providing access to a Field Editor for that setting.



Note: You aren't required to use the field editor; you can enter simple data directly in the window.

The field editor is a powerful tool that provides access to all metadata captured so far in the instance, as well as numerous ways to manipulate that data.



The upper pane displays what will be returned for a given parameter. In the view above you see the default fields for Metadata Content. The top line represents the header and the second line the fields of metadata.

The Show Fields button opens the lower pane which provides access to the various fields of metadata. The presets option lets you quickly change between three common delimiter settings but you aren't limited to these.

You can enter any values you want in the Quote and Separator fields to the left. To add a field of metadata, find what metadata you want to include in your export in the lower pane and do one of three things:

- Ctrl-Left Click will add both the field of metadata to the second line as well as adding the field name to the header.
- Left Click only will only add the metadata without adding anything to the header.
- Shift-Click will add the name of the field but not the actual metadata.

Once a field is added, if it's an editable field, you can edit it by double clicking on the field in the upper pane. Fields that are editable include dates and multi-value fields.

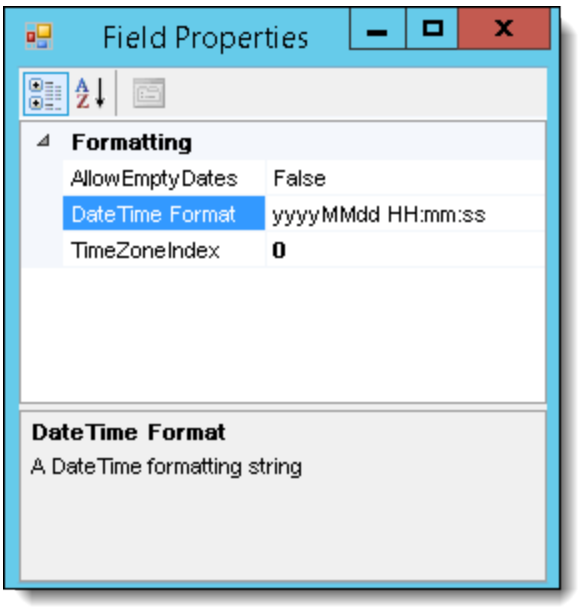
There are a multitude of advanced functions available under the Edit dropdown. Some are only available if you have a field of metadata highlighted. The more common features are detailed below but all of them provide a brief description if you mouse-over them. If you have questions about any of the ones not listed (or listed for that matter), contact support.

- **Save this field** - used to save a customized field of metadata or switch statement for use in other exports. Saving the field also allows the field to be mapped through the Relativity front end.

- **Apply Formatting > Child Values** - the exporter will return the corresponding metadata values of a given document's children instead of its own. For child documents it will return its own value only.
- **Apply Formatting > Parent Values** - the exporter will return the corresponding metadata value of a given document's parent instead of its own. For parent documents it will return its own value only.
- **Remove Formatting** - removes any applied formatting.
- **File Counter** - enter a customized counter that increments with every document. Double clicking on this allows you to customize the format of the number returned. This is typically used for creating subfolder names for native file and extracted text exports.
- **Generic Counter** - similar to the File Counter but can used to increment on any field of metadata rather than just with every document.
- **Control Flow > Switch Statement** - see below for information on creating a Switch Statement.

8.2.2 Editing dates

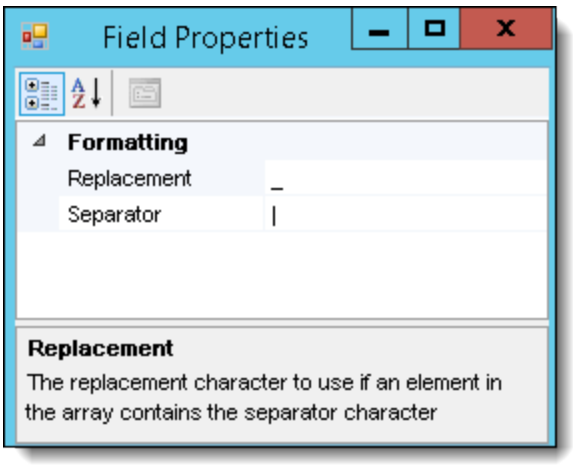
Use the following settings to edit dates in the field editor:



- **Allow Empty Dates** - True = Invalid or missing dates will return null (empty). False = Invalid or missing dates will return 1/1/1900.
- **Date Time Format** - you have the full Microsoft custom date and time formatting options available.
- **Time Zone Index** - 0 uses the time zone entered in the Job Settings. To use a different time zone on the export, enter the appropriate index number (GMT = 831169467). It may be easier to keep this at 0 and change the Job Settings instead.

8.2.3 Editing multi-value fields

Use the following settings to edit multi-value fields in the field editor:



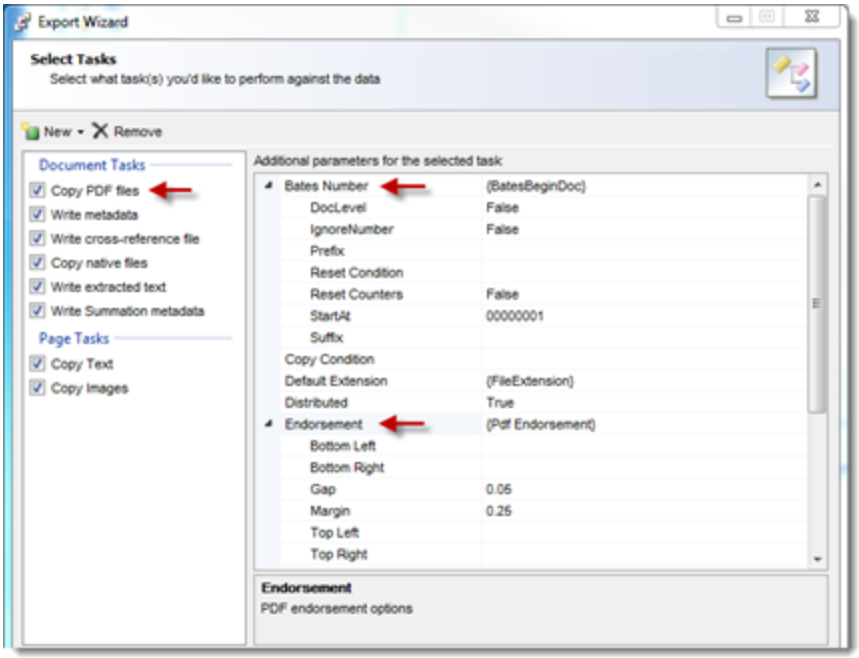
- **Replacement** - if the designated separator character exists in the metadata, it will be replaced with this character.
- **Separator** - the character to be used between each value being exported.

Note: Unlike the preset field separator values, there is no function to change all multi-value separators. If you are changing them from the default | character, be sure to change them on all multi-value fields.

8.2.4 Endorsing PDFs

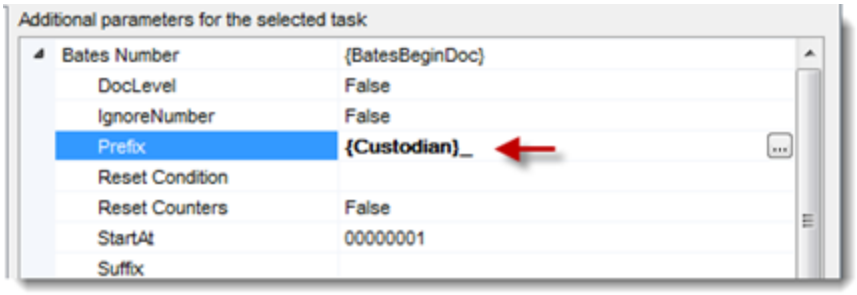
This example illustrates how to endorse a PDF by adding a Bates number to it.

1. Under Document Tasks, highlight **Copy PDF files** to display a list of parameters that you can set for this task.
2. Expand **Bates Number** and **Endorsement** in the Parameters box.

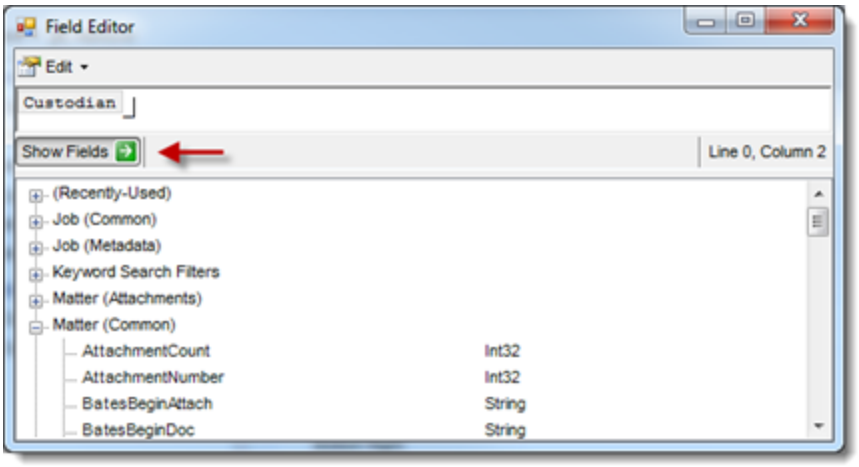


Note: The Distributed parameter is set to True by default in the previous illustration. This export job will queue up a task to the workers. The workers can then do the branding, imaging, generating PDFs and other tasks. You do need to ensure that the destination location is visible to these worker machines. When this parameter is set to False, the job will run only on the machine that you are using, but it will be multithreaded. You might disable this parameter if you were exporting to a USB drive that other machines cannot see, or if you were debugging.

3. Under **Bates Number**, highlight **Prefix**. Enter text or a metadata field from the document, job settings, custodian information, or other source. In the following illustration, a variable substitution (requiring curly braces) for custodian metadata has been added to the Prefix field.

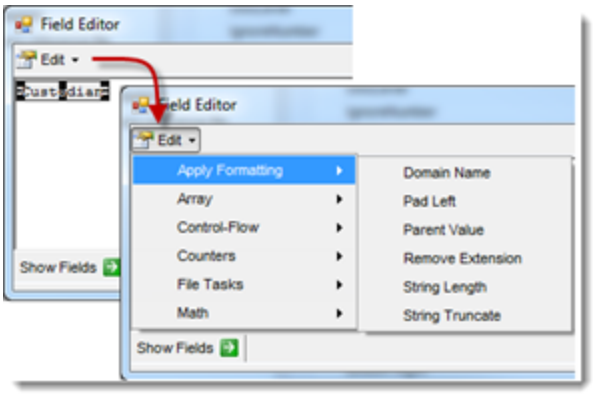


You can also edit this field by clicking the Browse button. In the Field Editor dialog box, click Show Fields to display a list of available metadata. Expand a metadata group, and double-click on a field to add it to list box.

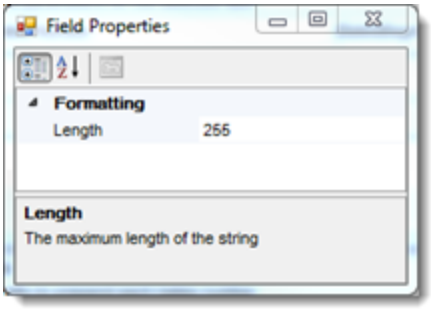


Some fields also include statistics about the document, such as the ClientSubmitTime under Matter (Metadata). The occurrence value for this field lists how frequently it appears in the collection. You can use this occurrence information to determine if a metadata field is relevant to the current data set.

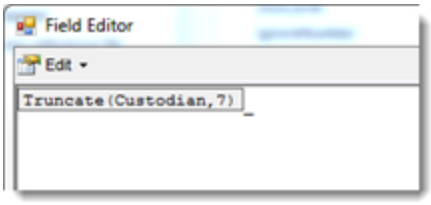
4. To apply formatting, select the metadata (such as Custodian), and click **Edit** to display a list of formatting options.



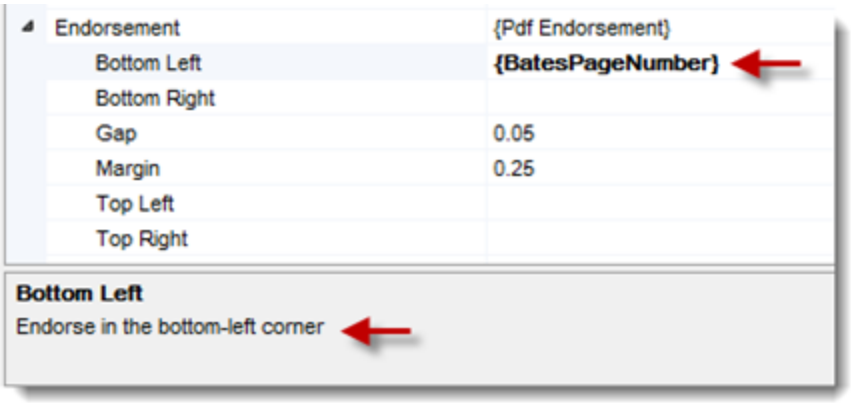
If you want to truncate the string, click String Truncate to display the Field Properties pop-up where you can define the maximum length.



Change the default value and close the pop-up to display the Field Editor with the updated formatting information.



5. Close the Field Properties and Field Editor pop-ups.
6. In the **Parameters** box, locate **Endorsement**. Use the **Browse** button to select the metadata **BatesPageNumber** in **Bottom Left** field.



8.2.5 Switch statements

A switch statement is an IF/THEN/ELSE statement. A simple IF/THEN/ELSE statement might look like the following:

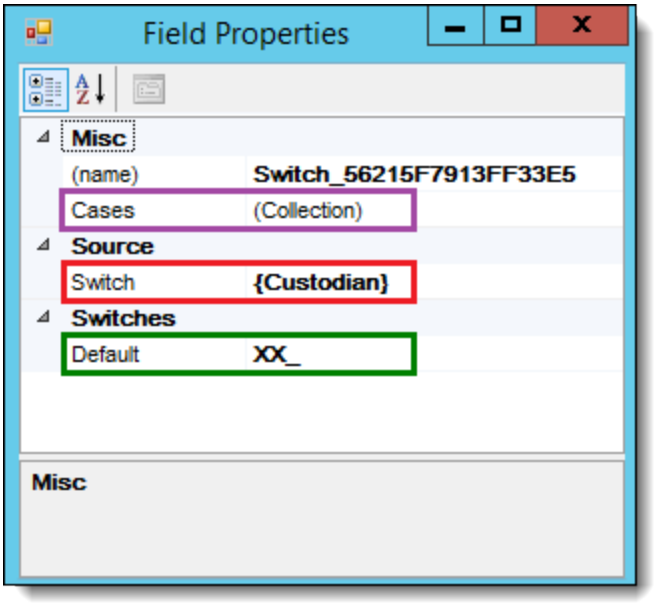
if (X = 1)

print “Hello”

else

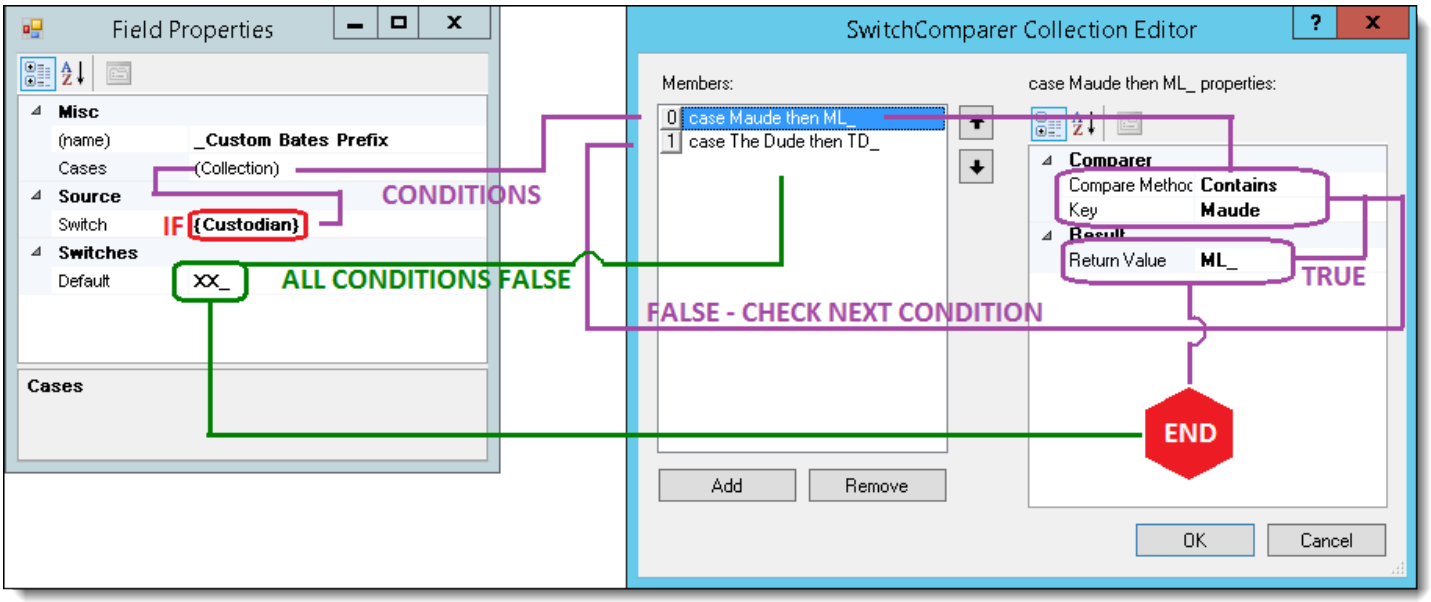
print “Good Bye”

The colored text above and throughout this section corresponds to these parts of the switch statement:



A switch statement can make multiple checks against the source and return different values depending on the condition met.

The creation of a switch statement involves using two windows, the field properties window above and the Switch Comparer Collection Editor.



- If the Custodian contains Maude return ML_ and close the switch statement.
- If the Custodian contains The Dude return TD_ and close the switch statement.
- Else return XX_ and close the switch statement.
- You can have as many conditions as you want in your switch statement but remember that they are checked in order and the switch statement terminates on the first true condition. Because of this you must make certain that your logic for the conditions accurately covers all possibilities. Let's say we have two custodians named John Smith and Sara Smith. Our first condition checks if the custodian name contained Smith then returns JS_ and the second condition looks for Sara and returns SS_. In this case, none of Sara's documents would get a return value of SS_ because they all met condition 1 first, returned JS_, and the switch statement ended looking at no further logic.
- You can use switch statements inside of switch statements. It is perfectly acceptable to use another switch statement for the Key value, Return value, Default value or even the Switch value itself – or any combination of those – or use more switch statements inside those switches. This of course adds a great deal of complexity but in return the user can perform very in-depth logic calculations on the data to return exactly what is needed for the situation.

The field properties contains the following settings:

- **Name** - the system will automatically give a new switch statement a name consisting of a random string of letters and numbers. You can change this to fit your needs however the name of the switch must not match the name of an existing field of metadata. A simple way to avoid accidentally doing this is to precede the name of your switch with an underscore (_).
- **Cases** - the various checks that will be made against the entry in the switch field.
- **Switch** - the field of metadata or fixed value that the collection of cases is compared to.
- **Default** - the value to return if none of the collection of cases is true.

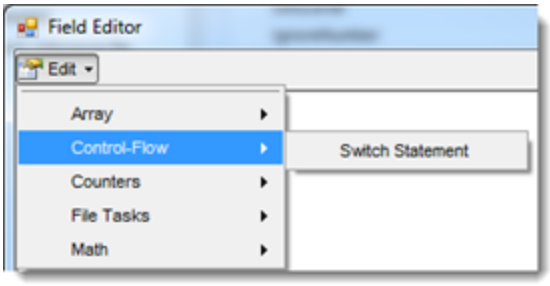
The Switch Comparer Collection Editor contains the following settings:

- **Members pane** - list of cases that will be checked
- **Properties Pane** - displays the definition of the highlighted case in the Members pane
- **Compare Method** - how the Switch value will be compared the Key. The available options are:
  - Equal To
  - Less Than
  - Less or Equal
  - Not Equal
  - Contains
  - Not Contains
  - Greater Than
  - Greater or Equal
- **Key** -the value the Switch is compared against.
- **Return Value** - the value returned if the comparison is true.

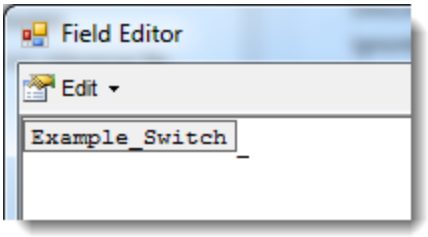
8.2.5.1 Using a switch statement for custom logic

This example illustrates how to define a switch statement used to determine the prefix applied to a Bates number.

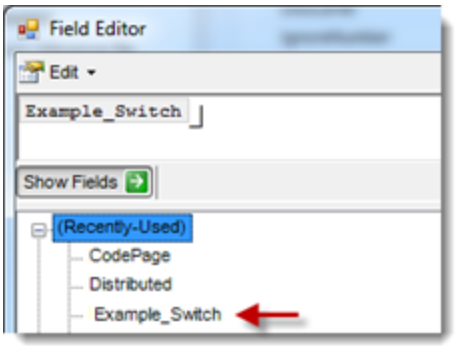
1. Under Document Tasks, highlight **Copy PDF files** to display a list of parameters that you can set for this task.
2. Expand **Bates Number**, and highlight **Prefix**. Click the Browse button to display the Field Editor.
3. Click **Edit** to display a list of options.



4. Point to **Control-Flow**, and click **Switch Statement**.
5. Enter a **name** for the switch and a metadata field used in the **Switch** box. In this example, the name is Example_Switch, and a variable {Custodian} is used for the switch.
6. Highlight **Cases** and click the **Browse** button.
7. In the SwitchComparer Collection Editor, click **Add**.
8. In **Compare Method**, select an option. This example uses **Contains**.
9. Enter a value in the **Key** and **Return Value** fields. Click **Add**. As illustrated below, the first key-value pair are Doe and DOE_, while the second is Smith and SMITH_. When the custodian contains one of these names, the associated prefix will be added to the Bates number.
10. Click **OK** to display the Field Properties dialog box. For this example, the default switch is UNK_.
11. Close the dialog box to display the Field Editor with the new switch.



12. Highlight the switch, and click **Edit**. Select **Save this Field** from the menu.
13. In the Save Field As dialog box, enter a name for the switch. This field is now added to the Recently_ Used list under Show Fields, and be reused as necessary.

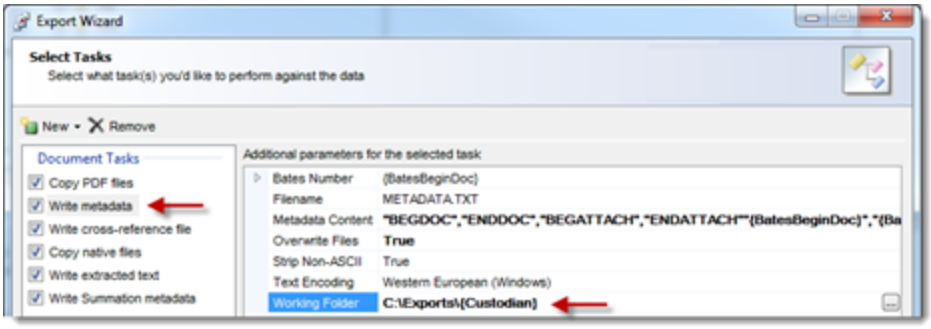


14. Close the dialog box.

8.2.6 Setting metadata content and folder path

This example illustrates how you can set the metadata fields written to a file, and how to set a Working Folder path.

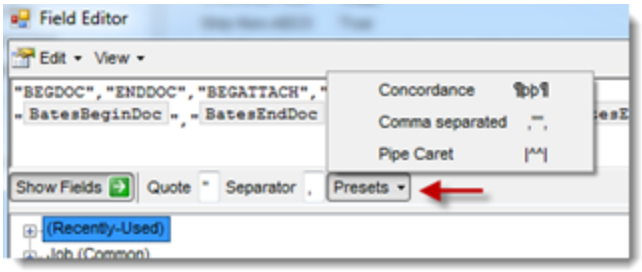
1. Under Document Tasks, highlight **Write metadata** to display a list of parameters that you can set for this task.



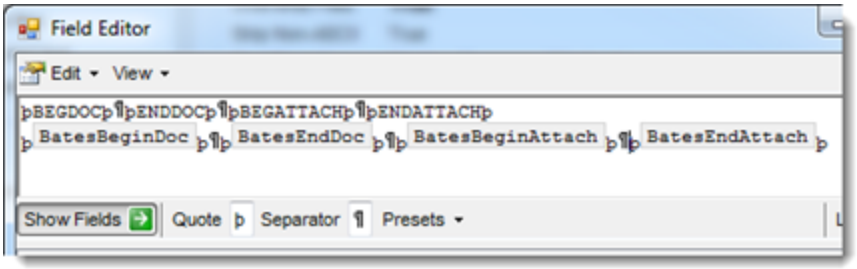
2. Highlight **Working Folder**. Enter a folder path or click the **Browse** button to select one.
Consider the following when defining a Working Folder path:

   - This is a shared field among the Document and Page tasks. For example, when you use this field for one task, it populates for all tasks with the Working Folder field specified.
   - You can't specify a separate folder for separate tasks.
   - You can also use variables in the folder path.
   - This example uses metadata as a variable, so that each custodian has an individual folder: c:\Exports\{Custodian}

3. Highlight **Metadata Content**, and click the **Browse** button to select the metadata fields to include in the file as described in the previous section.
4. Click **Presets** to select a delimiter for use in the file.



For example, select **Concordance** to separate the metadata fields as illustrated below:

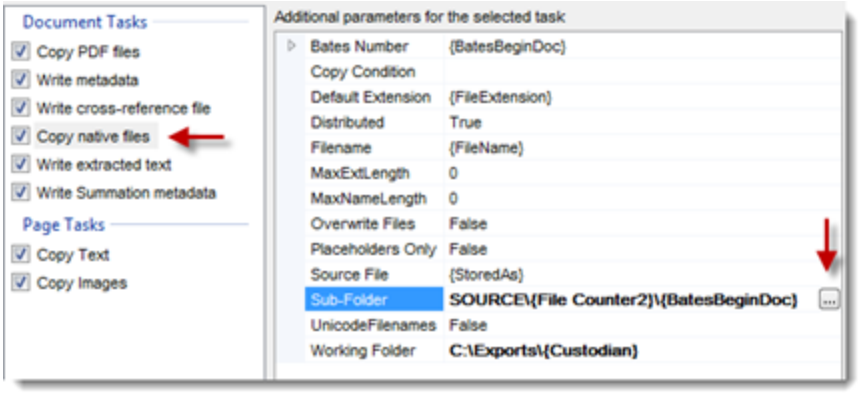


5. Close the dialog boxes.

8.2.7 Defining a copy condition

You can set up a copy condition that permits the file to be copied only when the given expression returns true. In this example, a copy condition is defined for native files.

1. Under Document Tasks, highlight **Copy native files** to display a list of parameters that you can set for this task.



2. Click **Copy Condition** to display the Field Editor.
3. Enter your own custom fields or click **Show Fields** to select them.
4. Close the dialog boxes.

9 Managing workers and jobs

The RPC uses a queue manager to control the activities of worker machines. All jobs executed by the RPC are designed to be distributed across these machines to improve performance. To centrally manage these workers, the RPC includes the Worker Activity window, which displays the workers and their current statuses. This central management console also provides you with the ability to manage the activities of these machines.

Read more about worker machines in the RPC
In the RPC, you can coordinate the activities of worker machines and the processing of jobs by assigning them both to workgroups. A worker machine processes only the jobs that are assigned to its workgroup. You can assign multiple workers to a workgroup, and they will all work on the same job. The RPC breaks ups large jobs into smaller subtasks to leverage parallelism and multithreading. For example, PST and NSF files are divided into jobs of a 100 documents, while ZIP and RAR files are chunked into jobs of 25 documents. The size of these jobs is determined by document type, and it is hard-coded in the system.

For example, you might create an Import job and add it to Workgroup 1. Any worker machines assigned to that workgroup will immediately begin processing the job. If you have a large job, you may want to add more workers to the workgroup. The use of multiple workers splits a job across numerous machines leveraging parallel processing, which increases the performance of the RPC.

Note: If you're starting an import job of a large PST or similar container you may note that only a single worker in the assigned group is doing any work for an extended period of time. This is because the file must be copied to the repository and have hash values calculated for it. This is not a distributable task and can take considerable time on very large files which is normal. Once these tasks are complete the other workers will be able to work on importing messages from the PST simultaneously.

Since you can constantly monitor the queue, you can also dynamically reallocate workers from one workgroup to another as the job load requires. You must stop a worker in order to reallocate it to another workgroup. The worker will complete its current job, and then display its status as Stopped.

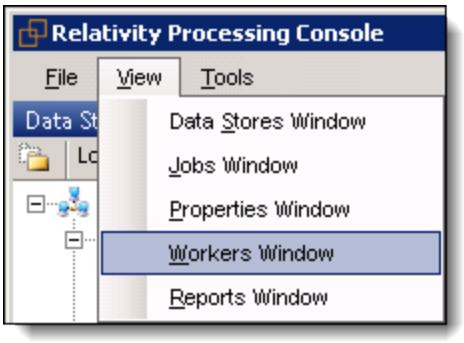
You also have the option to bring the worker offline immediately. In this case, the worker doesn't complete the current job. When you restart the worker, it automatically performs a series of cleanup tasks based on stored procedures in the database. It then cleans up any entries added to the Matter table and other places in the system, as well as removes any sub-jobs that it added to the queue. The restarted worker resets and re-executes any of its open jobs from the beginning.

While you can queue multiple jobs to the same workgroup, we don't recommend this method because of the first-in-first-out design of the queue. This method can create contention for shared resources that may degrade performance if small and large jobs intermix. We instead recommend utilizing multiple workgroups with their own workers processing a single job.

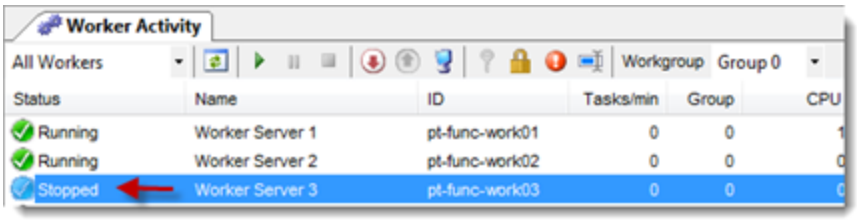
9.1 Starting worker machines

Use these instructions to start a worker machine:

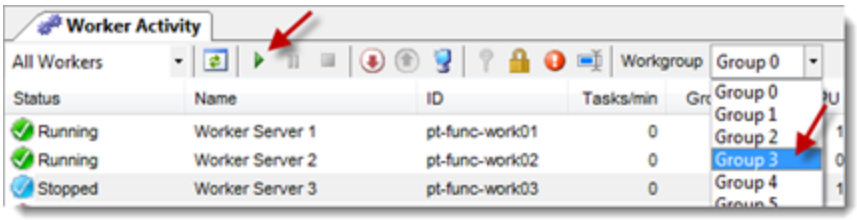
1. In the **View** menu, click **Workers Window**. The Workers Activity window displays a list of available workers and their current statuses.



2. From the Workers Activity window, highlight a worker machine with the Status of **Stopped**.



3. In the **Workgroup** box, select a group. The worker is now assigned to this group, and will only process the jobs that are added to it.



4. Click **Run** to start the worker. Its status will be updated to **Running**.

9.1.1 Stopping versus taking worker machines offline

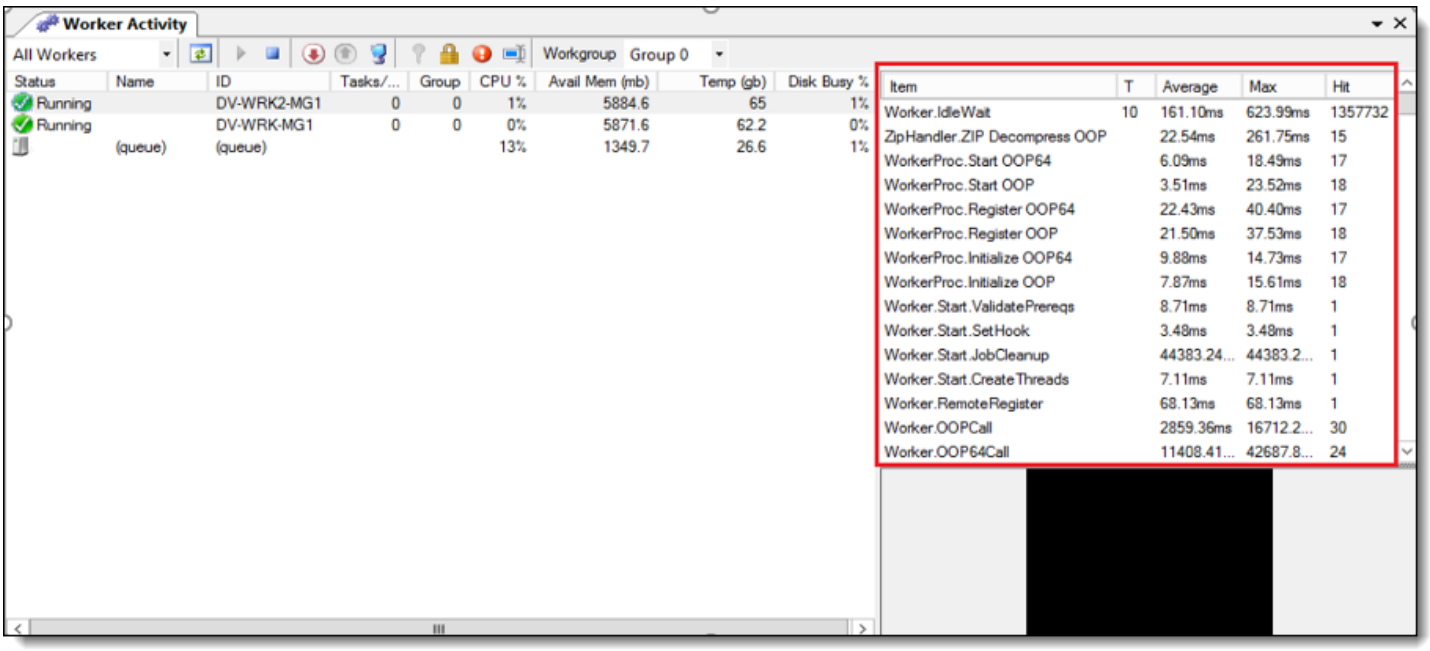
Note the following difference between stopping a worker versus taking it offline:

- **Stopping a worker** - the worker stops receiving work from the queue, and it finishes only the tasks it's working on, not the entire root job. The remainder of that job is then available to be picked up by another worker.
- **Taking a worker offline** - the worker immediately stops and is brought offline, regardless of whether or not it's working on any tasks. After a short period of time the worker should bring itself back online automatically. If you are concerned that a worker is hung up on a task it is recommended to try stopping the worker first to allow as many of the tasks the worker has picked up to end gracefully. If the worker remains in a "stopping" state, then it may be necessary to bring the worker offline.

Note: In order to stop a worker and take a worker offline, the account running the RPC must be a local administrator on the worker servers.

9.2 Viewing worker activities

You can monitor the activities of a worker machine or the queue on the side panel in the Worker Activity window. As illustrated below, you can highlight a worker and expand this panel to display statistics about the number of microseconds (ms) taken to perform various tasks.



The panel to the right of the Worker Activity window provides the following information:

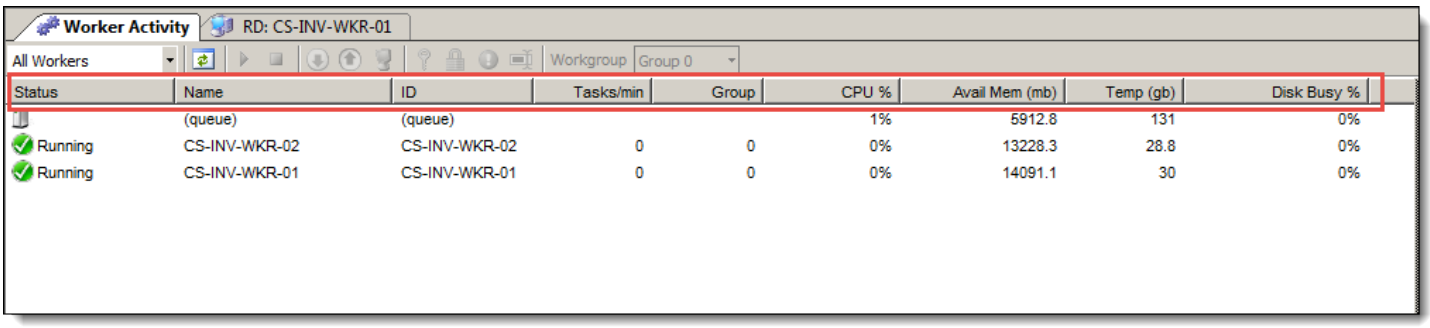
- **Item** - a list of tasks that the worker has performed since it was brought online.
- **T (Threads)** - a count of how many threads are currently working on the listed Item. It is recommended to sort on this column so that items that are currently being worked on are always displayed at the top of the list.
- **Average** - the average amount of time the worker has spent executing the items in that row since it was last brought online (displayed in milliseconds).
- **Max** - the maximum amount of time the worker has spent executing an item in that row since it was last brought online (displayed in milliseconds).
- **Hit** - the total number of times the worker has executed the item in that row since the worker was last brought online.

Note the following details about the behavior of the Worker Activity window:

- If no worker is highlighted, the Tasks window displays the last worker selected.
- You aren't able to display tasks for all workers simultaneously. You can only display tasks for a single worker.

9.3 Viewing worker status

The following columns/fields are available in the Worker Activity window for you to monitor the current state of your workers. Note that these also correspond to the data displayed in the Worker status tab in Relativity. You can sort on any of the columns in this window.



- **Status** - reflects the current status of the worker. This displays one of the following values:
  - **Running** - the worker thread is performing one of its designated jobs.
  - **Stopping** - the worker has been instructed to stop and is attempting to complete and tasks that it had already picked up. Once the worker has completed all tasks that it had picked up it will change to a status of stopped. If the worker is unable to complete the tasks it has picked up the worker will remain in a stopping state indefinitely and will need to be brought offline.
  - **Offline** - the worker has been brought offline by the user or the InvariantWorker.exe program was somehow forced to close. As long as the queue manager is running, the worker will automatically try to bring itself back online even if it was deliberately taken offline by the user. While the worker is offline only the Status, Name, ID, and Group columns will be populated.
  - **Stopped** - the worker has been manually stopped. If the worker is either stopped of logged off, the remaining columns will contain no data. A worker being brought online will briefly go into a stopped state before going online.
  - **Logged Off** - the worker is either powered off or the relativity service account is not currently logged in as the user. While the worker is logged off only the Status, Name, ID, and Group columns will be populated.

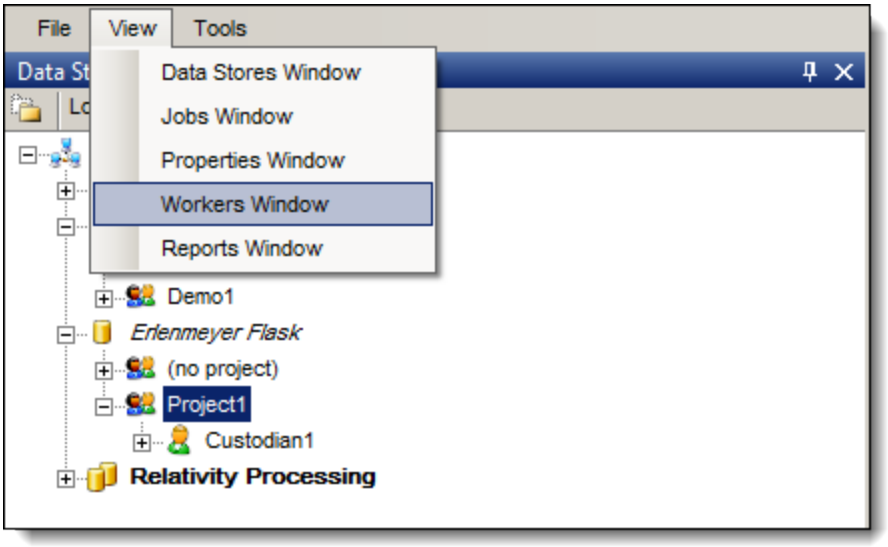
  Note: The network on the Utility Server isn't set up to view the status of your workers; therefore, you'll see all workers logged off in the Worker Activity window in the RPC, and you'll need to refer to the All Workers view of the Processing Administration tab tab in Relativity to see the status of your workers. For details on the Processing Administration tab, see the Processing User Guide.
- **Name**- the name of the worker.
- **ID** - the identifier of the worker.
- **Tasks/min** - how many singular units of designated work the worker is performing per minute.
  - Examples of tasks are discovering a document and converting a native.
- **Group** - the workgroup the worker belongs to.
- **CPU %** - 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.
- **Avail Mem (mb)** - how many MB of RAM are still available after counting what the worker is currently using.

- **Temp (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.
- **Disk Busy %** - 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.

9.4 Naming a worker

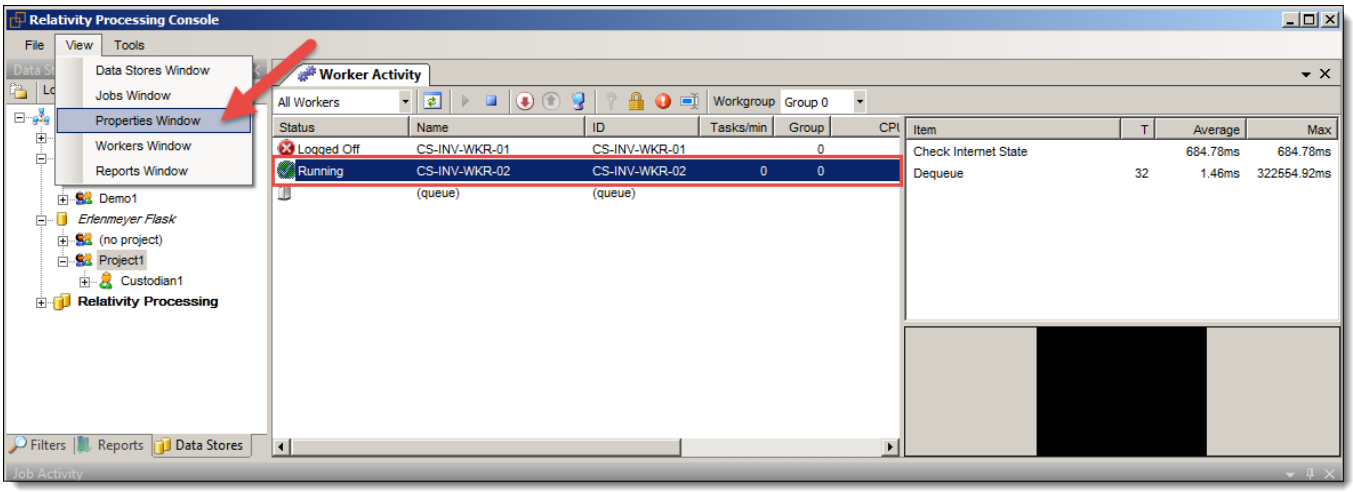
To provide or edit a worker name, perform the following steps:

1. If the Worker Activity window isn't already open, navigate to the **View** menu and select **Workers Window**.

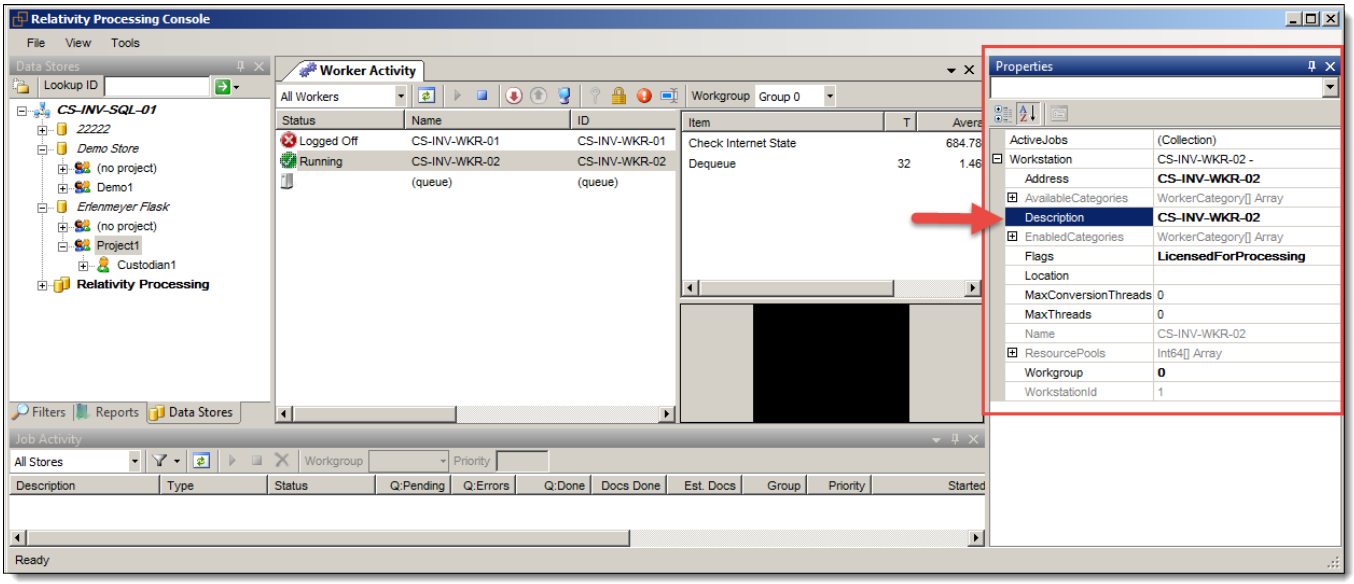


2. In the Worker Activity pane, select the worker you want to name.

3. If the Properties window isn't already open on the right side, navigate to the **View** menu and select **Properties Window**.



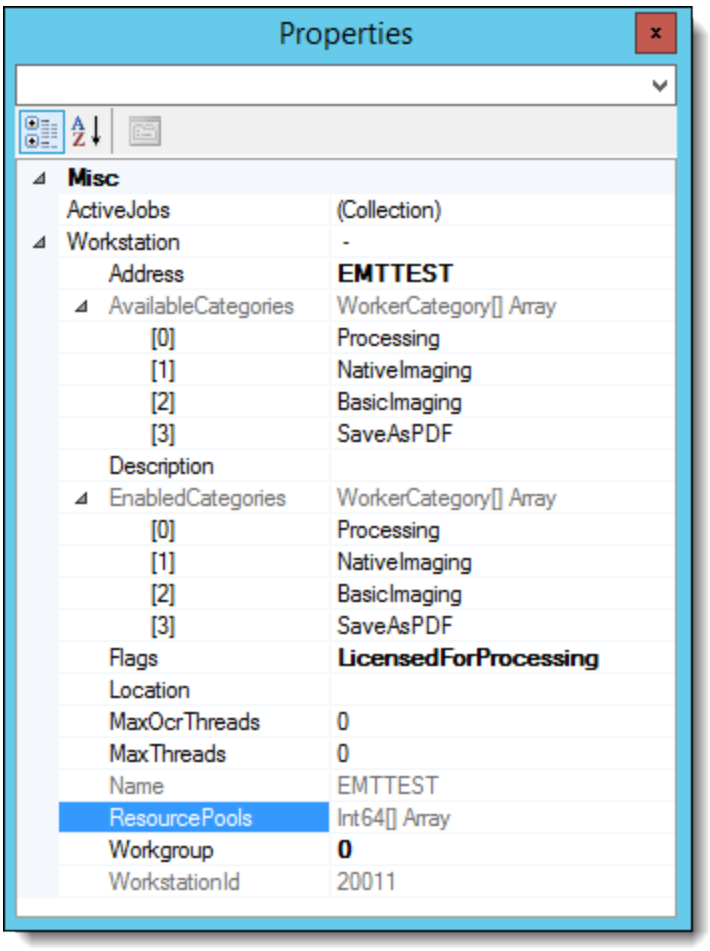
4. In the Properties Window on the right, expand the **Workstation** row, and scroll down the **Description** row. This is the name of your worker. Edit the value directly to the right in the Description row by clicking once inside the current name and modifying it.



5. Click the refresh button in the Worker Activity pane to display the changed values.

9.4.1 Viewing basic worker properties

The Properties window provides information about the worker that you've selected in the Worker Activity pane. To access this information, expand the **Workstation** column under the **ActiveJobs** column in the Properties window. Note that the Workstation column represents the EDDS.Workstation table in the database and that each value below it corresponds one-to-one with a column in the Workstation table.



- **Address** - the address of the worker server, as specified in the Address column of the workstation table.
- **Available Categories** - the categories available for selection when designating a worker for different kinds of work in Relativity.
- **Description** - a description of the worker, as specified in the Description column of the workstation table. Note that this value can be identical to the address and the workstation Name below.
- **Enabled Categories** - the categories of work for which the worker has been designated in Relativity. For example, if the worker has been designated to perform native imaging and processing, those will appear here.
- **Flags** - any database flags that have been applied to the worker.
- **Location** - the location of the worker server, as specified in the Location column of the workstation table.
- **Max Threads** - the maximum number of threads the worker is allowed to make use of. If this is set to 0, the worker is permitted to make use of all available threads.
    - The main reason to set this to something other than 0 is for troubleshooting purposes.
    - On rare occasions, if you are aware that the worker machine needs to process difficult files (large CAD drawings, blueprints, complicated excels, etc.), lowering the max threads can keep the worker from opening too many of these simultaneously, which may result in out-of-memory errors or similar issues.
    - If you do set this to something other than 0 for one of these purposes make sure to reset it back to 0 when done.
- **Name** - the name of the worker server, as specified in the Name column of the workstation table. This can be identical to the Description and the Address columns.

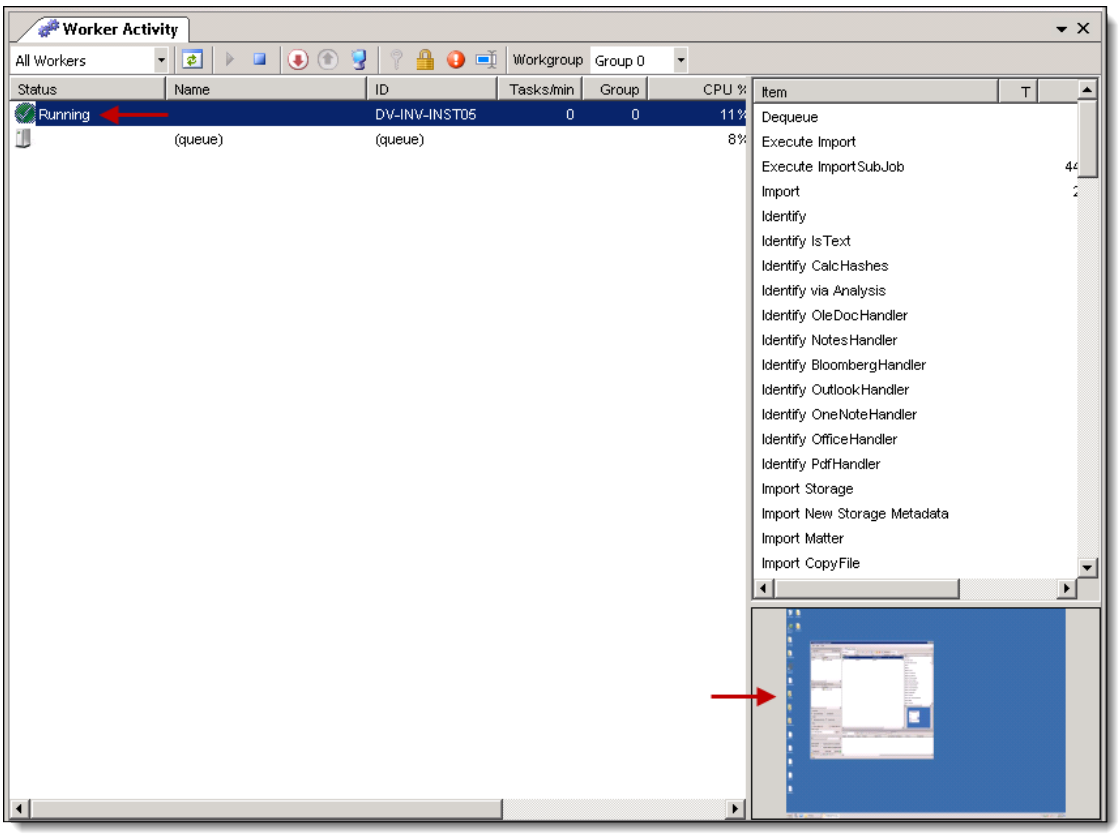
- **Resource Pools** - the artifact IDs of the resource pools to which the worker belongs in Relativity.
- **Workgroup** - the number identifier of the workgroup to which the worker belongs.
- **Workstation ID** - the ID of the workstation as specified in the WorkstationID column of the workstation table.

9.5 Other control options

The following table breaks down the other control options available in the worker activity window.

Menu Options	Description
Refresh 	Displays newly added workers.
Online 	Brings a worker back online after being taken offline if you don't want to wait for the worker to come back online on its own in a few seconds.
Remote Desktop 	Opens a remote desktop connection to the worker, in which you can bring a worker back online and add to your worker machines in order to troubleshoot issues.
Remote Logon 	Logs in to the worker as the RCA if the worker has been logged off.
Remote Logoff 	Logs the worker off, which closes any tasks that were opened while logged in. This is sometimes a useful last resort before rebooting a worker if it has been working on many problem files that have been causing the worker to hang.
Reboot Worker 	Reboots the worker if it needs it. If a worker is unresponsive this may work to reboot it but usually if the worker is in such state it won't respond to an instruction to reboot either. If the worker is up and running fine though and you need to cycle the worker, this button will work.
Remote command 	Brings up a text box in which you can enter a command as though you had clicked Start then Run on the worker machine. This feature should not be needed for typical day to day work.

Selected workers appear in a screen shot in the user interface bottom right corner, and updates every five seconds.



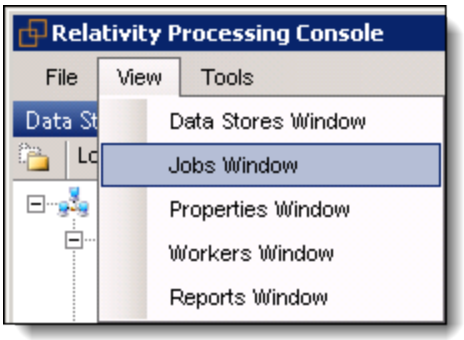
9.6 Running a job

This section includes the steps for manually starting a job. You need to complete these steps when a job has been added to workgroup that is not assigned to a started worker machine. You may need to complete these steps for Group 0, the default workgroup in the RPC.

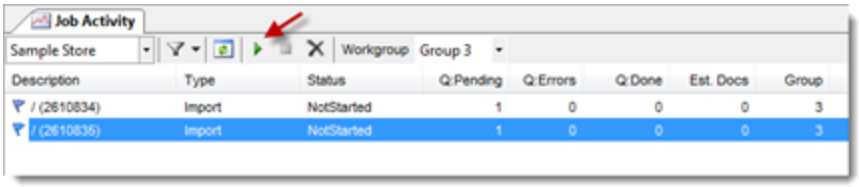
However, you can create jobs that will start automatically if you select workgroup that also has workers with the status of Running. (For most job types, you can select the Workgroup on the **General** tab of the associated job settings window.)

Note: Making changes to the workgroup or priority, or starting, stopping or deleting a job can have serious consequences, especially to jobs that originated from Processing.

1. In the **View** menu, click **Jobs Window.** The Jobs window displays a list of available workers and their current statuses.



2. In the Job Activity window, highlight your job. It will have the status of **NotStarted**. See the Job Activity section in the RPC user interface on page 104 topic.



3. Click to start the job. The job's status is updated to Running. See Viewing worker activities on page 81

10 Performing Quality Control tasks

You can view extracted text and imaged documents, as well as perform other quality control tasks in the Matter Inspector.

Note: For more information on how to use the matter inspector, see the Using the RPC Matter Inspector video webinar on the Relativity Training site.

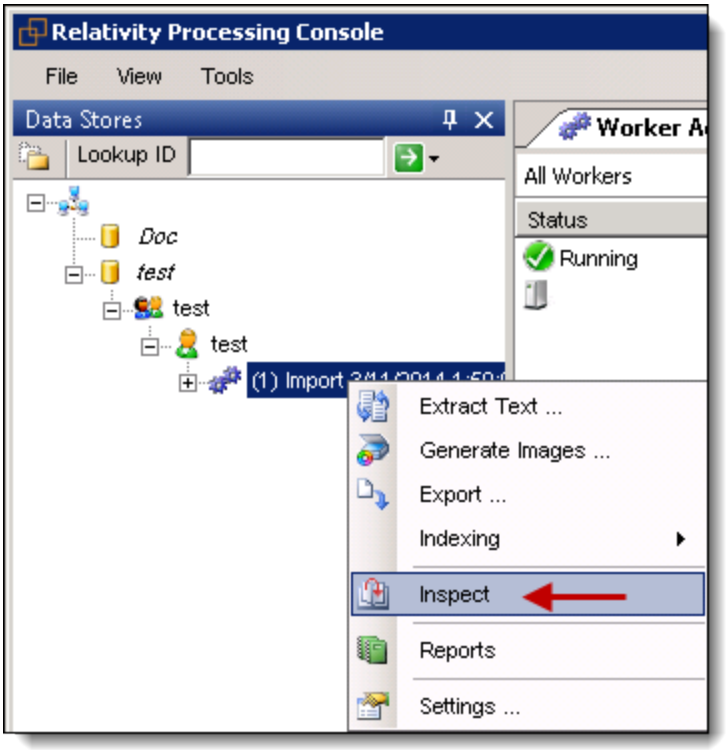
In the Matter Inspector you can:

- View all metadata for a selected file.
- Inspect multiple jobs.
- Apply filters to inspect specific files.
- View native files.

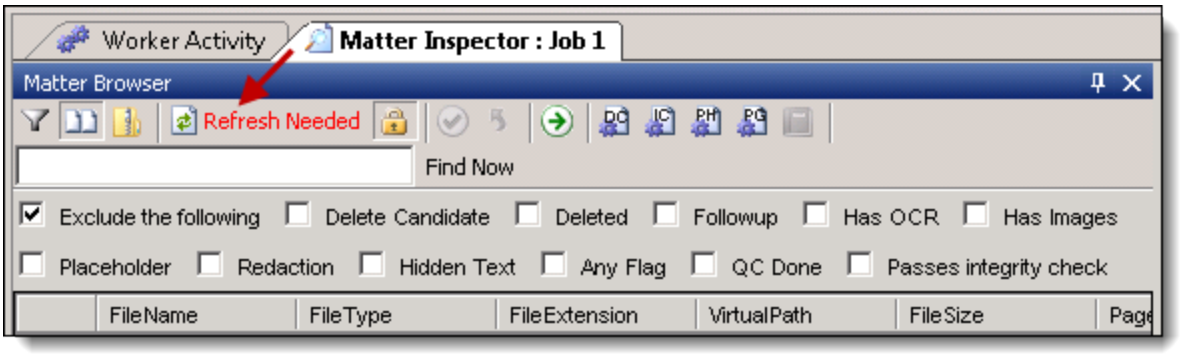
Note: There is no option for batching out documents for QC.

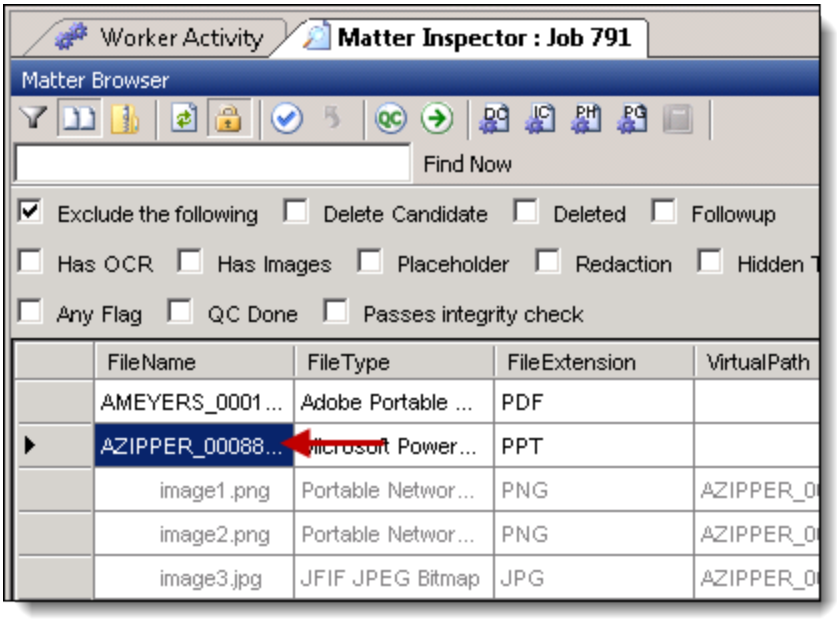
To QC documents:

1. In the Data Stores window, drill down to the import job that you want to review.
2. Right-click the import job, and click **Inspect**.



3. From the Matter Inspector window, click **Refresh Needed** to display a list of documents in the grid box.





Note: In the Matter Browser, click the column heading to sort the list of documents. An indented file is child of the file above it. Only one level of the parent-child hierarchy is displayed in the grid, but the database records the full hierarchy.

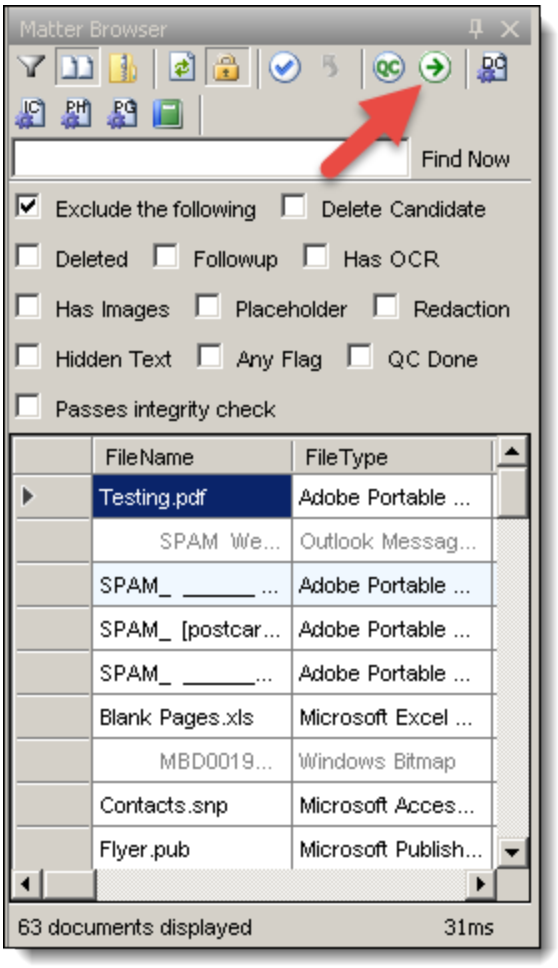
Note: You're unable to add more columns to the Matter Browser.

10.1 Performing automatic QC

The RPC's QC process uses a check-in/check-out system, multiple users can work with the same document at the same time. Automatic QC automatically checks out a document when navigating to it, and checks it back in when it moves on the next one. Automatic QC is especially useful when you have to perform QC on a large import job, as it presents you with a steady, controllable stream of images to inspect.

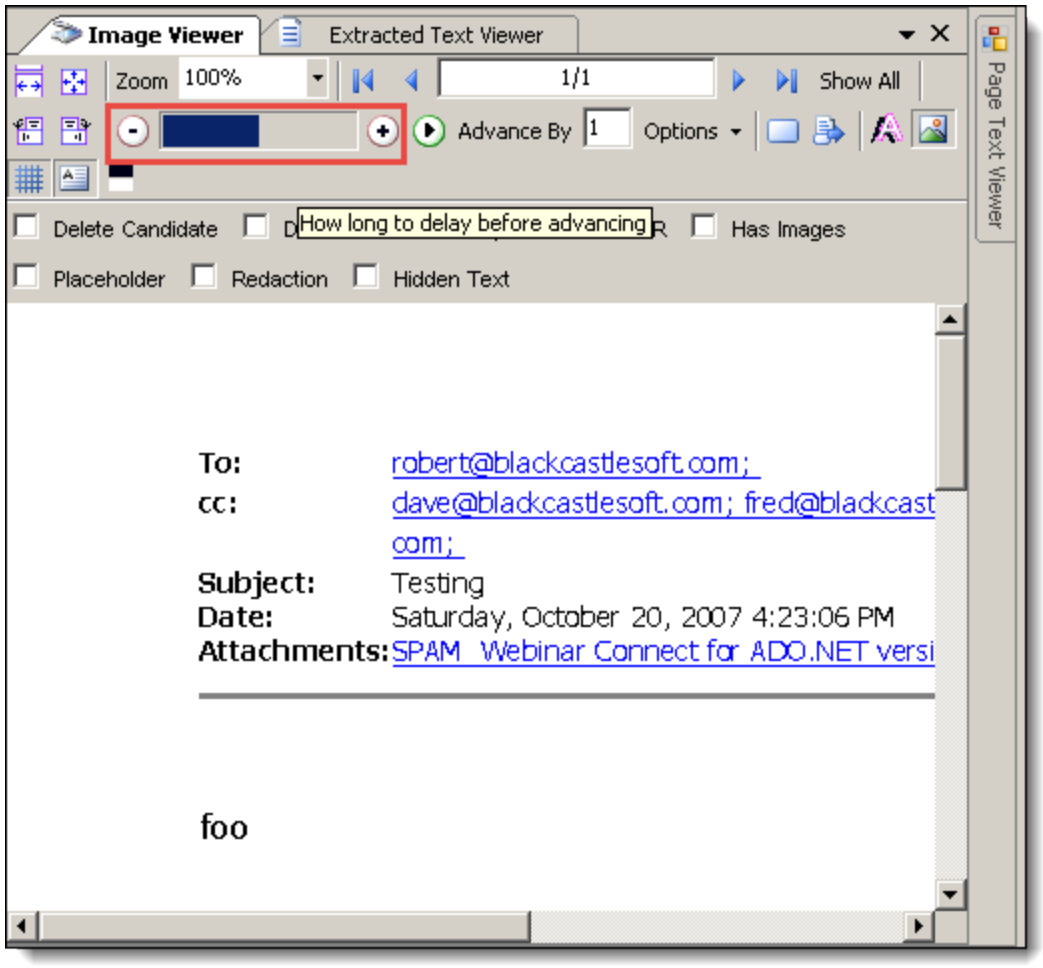
To perform automatic QC on a data set's documents:

1. In the Matter Browser, click the Automatic QC icon.

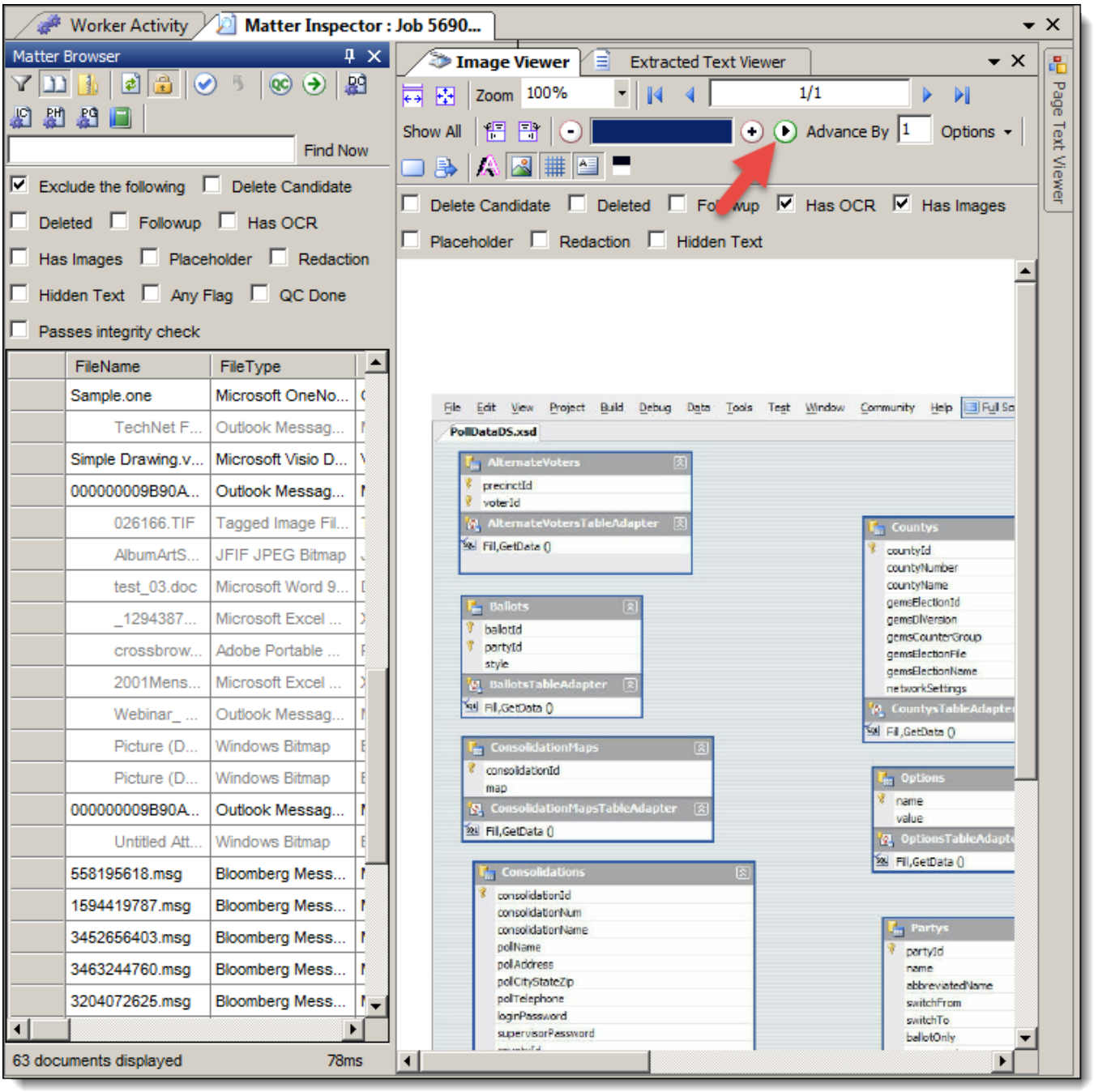


2. If needed, in the Image Viewer, increase or decrease the length of the delay before the RPC moves to the next image. Clicking the plus sign increases the delay and offers you more time to view the image and clicking the

minus sign shortens the delay, thus making the QC process move faster.



3. In the Image Viewer, click the play button to start the automatic advancement of images.

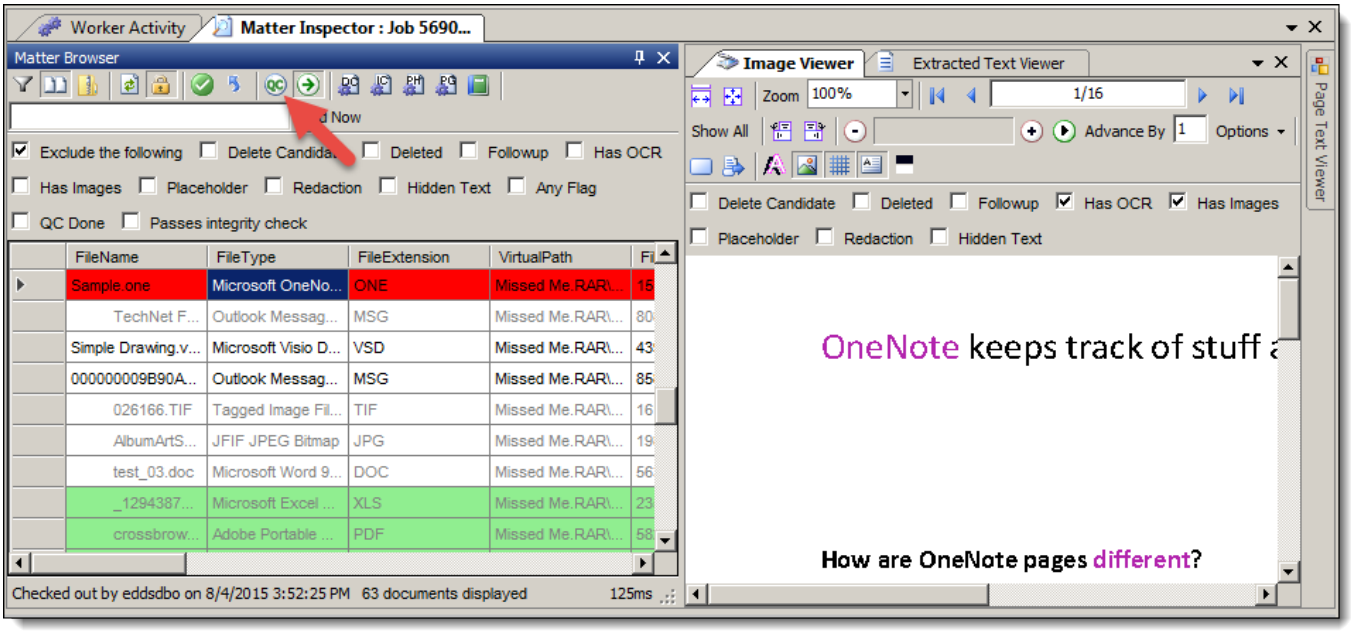


4. Perform QC on the images as they automatically appear, one at a time, in the Image Viewer.
   - The RPC only flips through each image and presents them to you for your manual QC, it doesn't perform the QC for you.
   - A Placeholder is a document that the RPC couldn't image and thus automatically made a placeholder for.

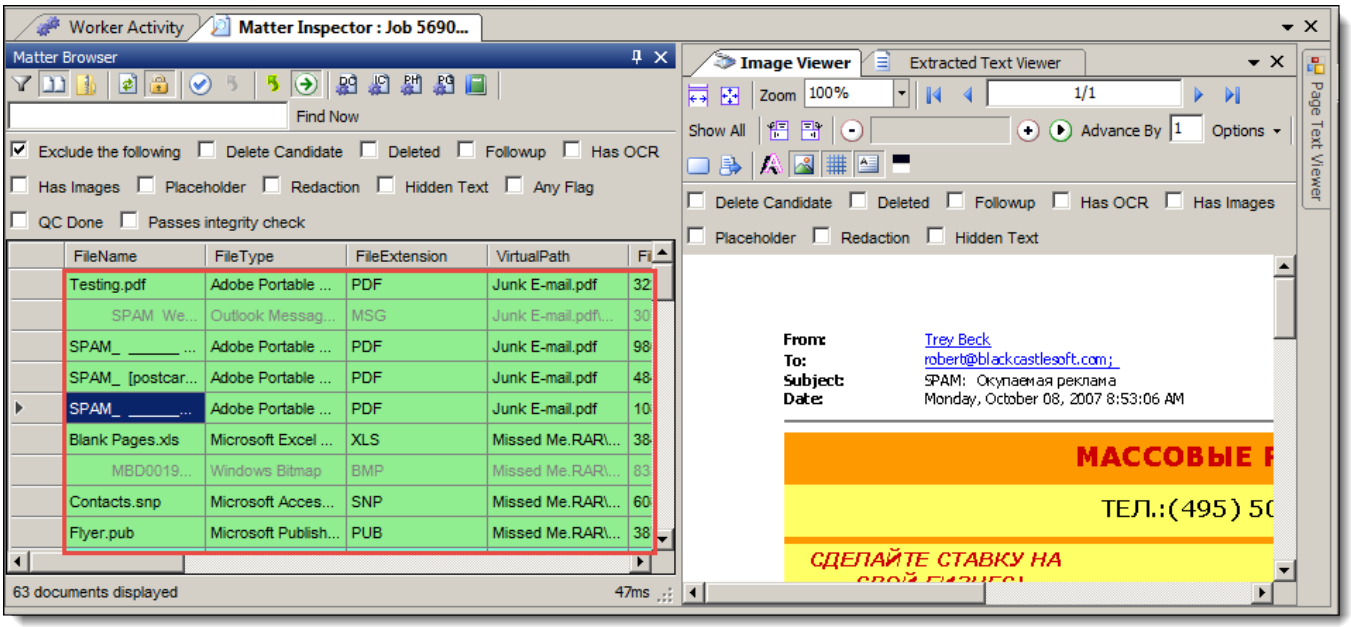
   Note: You can't actually configure a placeholder in the RPC; you can only generate a blank placeholder through the Generate Placeholder right-click option in the Matter Browser, or you can upload a PDF file created outside of the RPC for the Custom placeholder PDF option in the OCR/Image tab of the Job Settings window.

- If you need more time on a single image, click  to pause on that image.

5. When you're finished QC-ing, select the documents you want to mark as finalized and click the Finalize QC button.

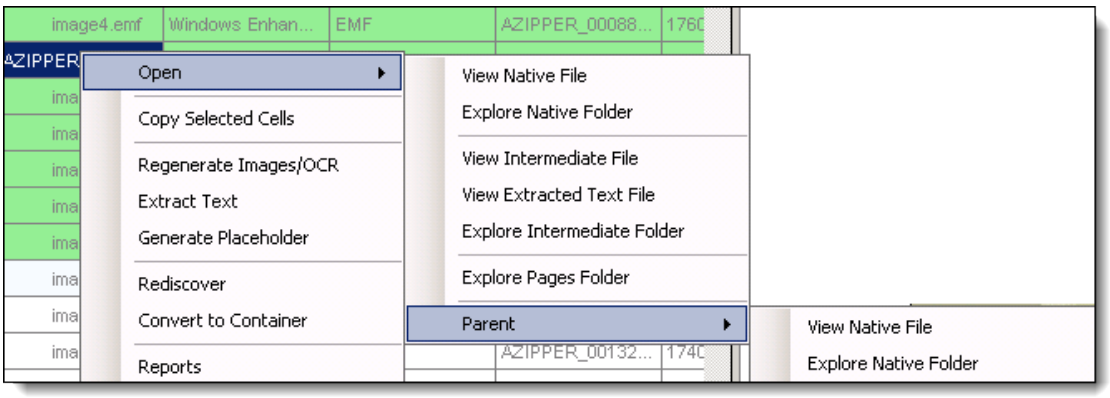


6. Note that the documents you marked as QC'd are green in the Matter Browser.



See Matter Inspector window on page 110 and Image Viewer window on page 113 for more QC options.

Alternatively, you can right-click on a file in the Matter Browser to display a pop-up menu. This menu includes options for generating placeholders, displaying intermediate folders, and others.

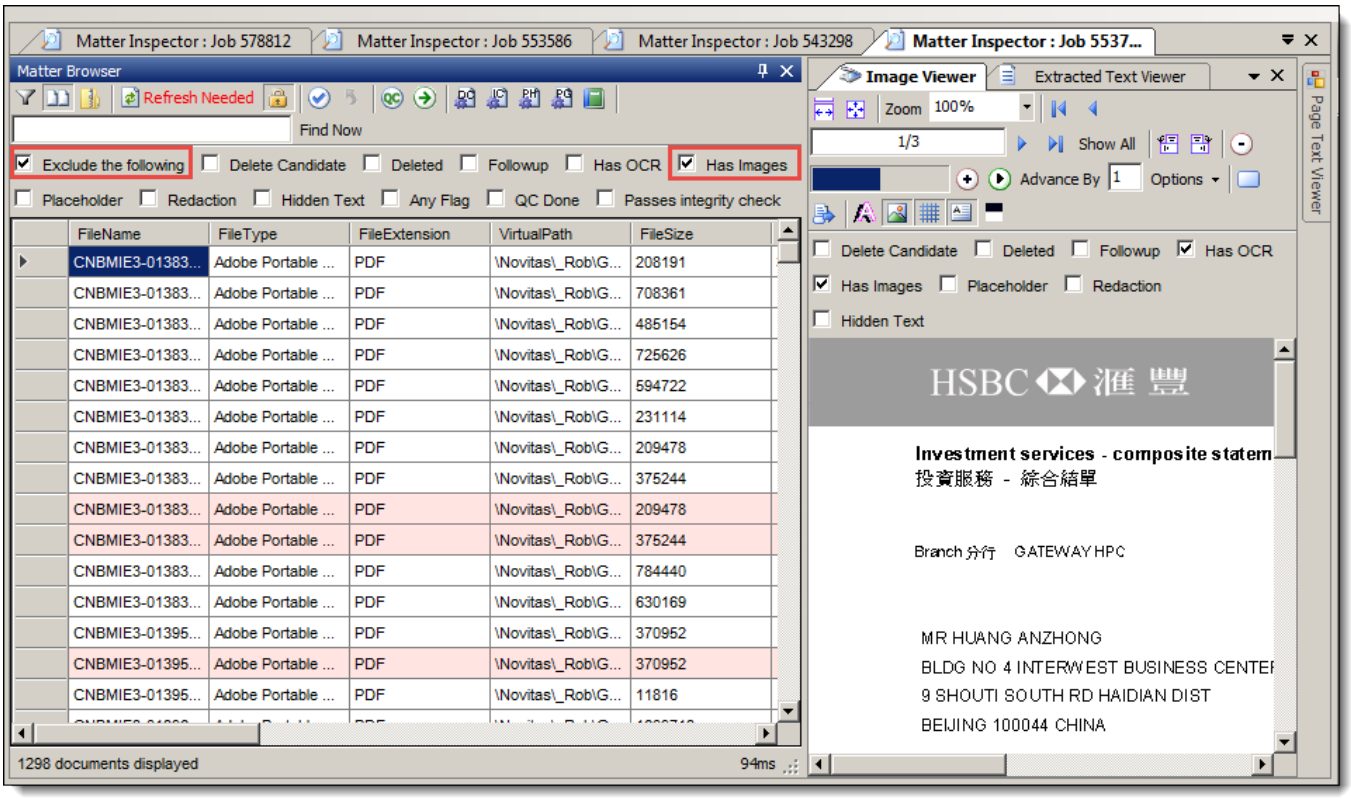


10.2 Non-image QC

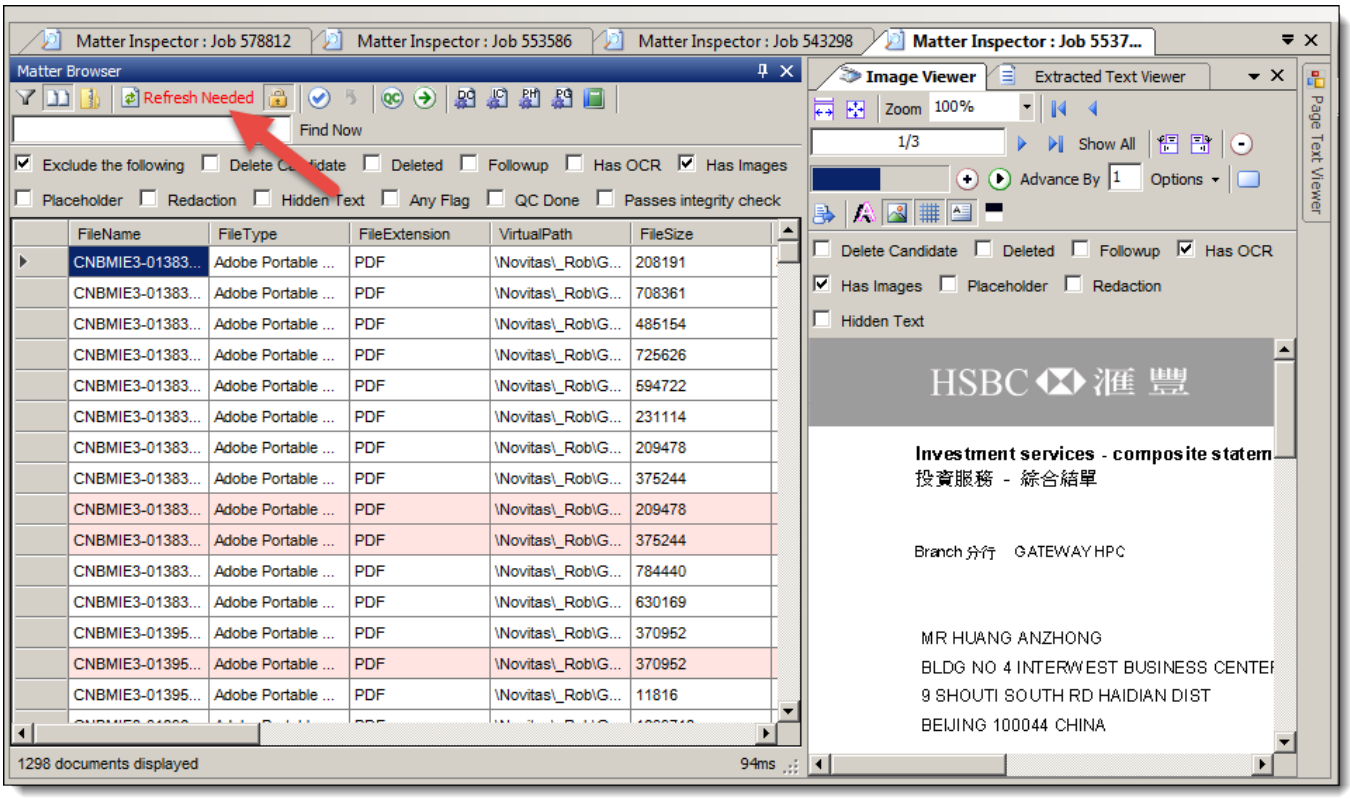
The RPC gives you the option of performing QC on non-image files. You can do this in either the Extracted Text or the Page Text viewers in the Image Viewer pane.

To perform non-image QC, follow these steps:

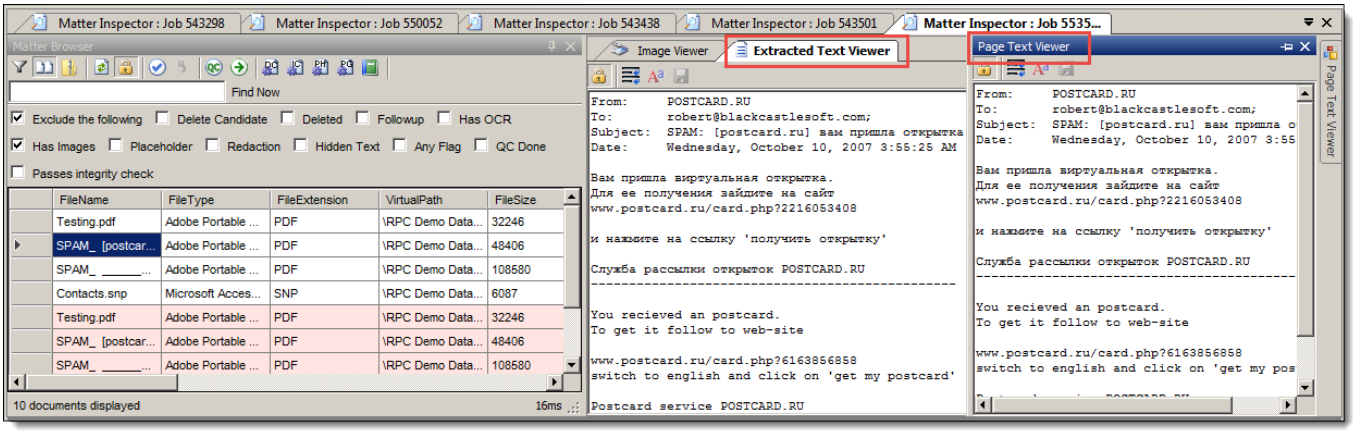
1. In the Matter Browser, select **Exclude the following** and select **Has Images**. Doing this will include only non-image files from the import in the viewer.



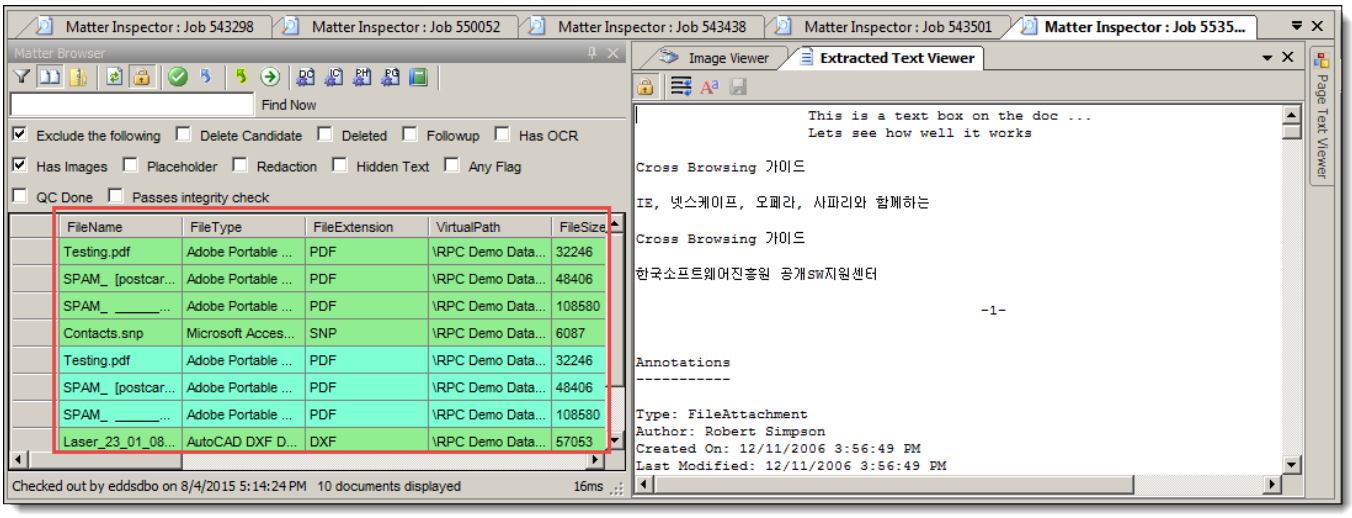
2. Click the **Refresh Needed** text when it appears. This will update the file list in the Matter Browser to meet the criteria you've just set.



3. Open the Extracted Text or Page Text Viewer to perform QC on each file you select.



4. Once you've finished QCing a document, click the Finalize QC icon. Files for which you've finished QC appear green in the Matter Browser.

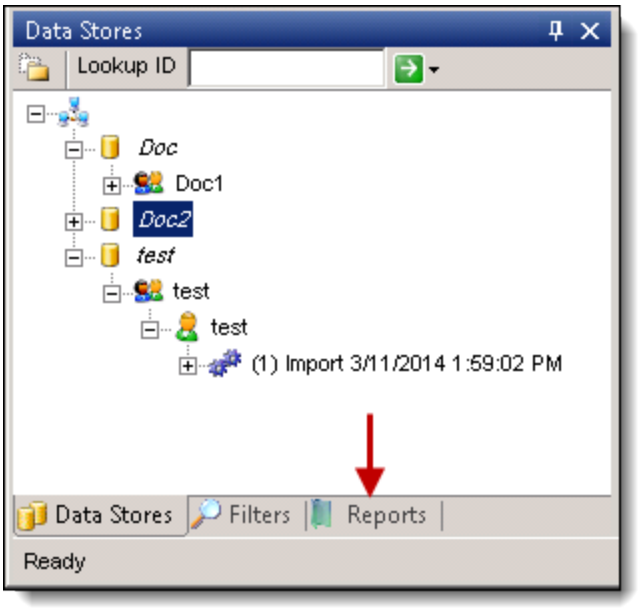


Once you've QC'd images, you can filter the data. See Filtering data on page 42.

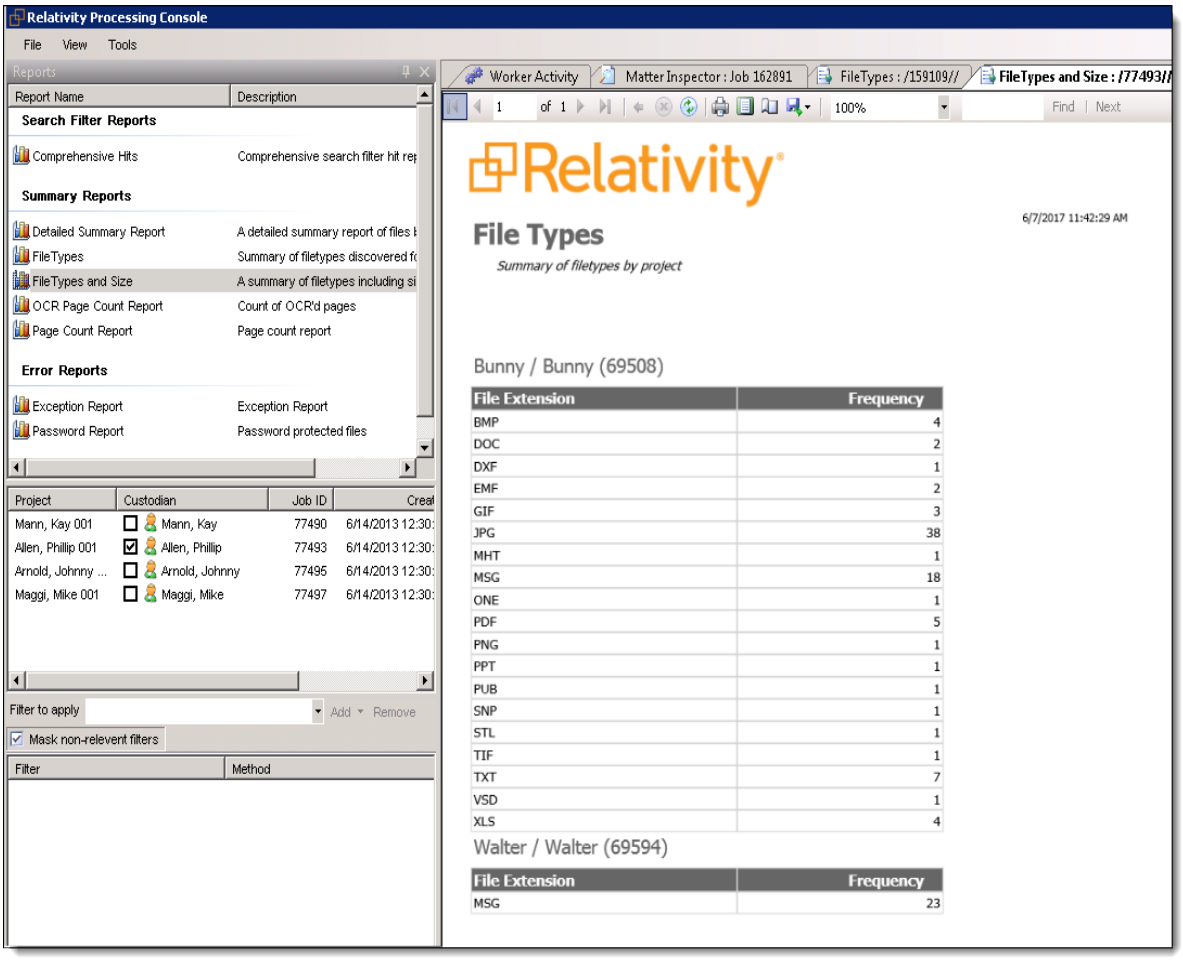
11 Running standard reports

This section includes generic steps for generating a standard reports listed on the Reports window. In the Data Store window, you can also generate reports for a specific job by using the right-click menu.

1. In the Data Store window, highlight a data store that you want to report on, and click the **Reports** tab. (You can also right-click on a data store, and select **Reports** from the menu.)



2. Highlight one of the following reports in the list box. See Reports window on page 108.



3. Select the jobs that you want to include in the report.

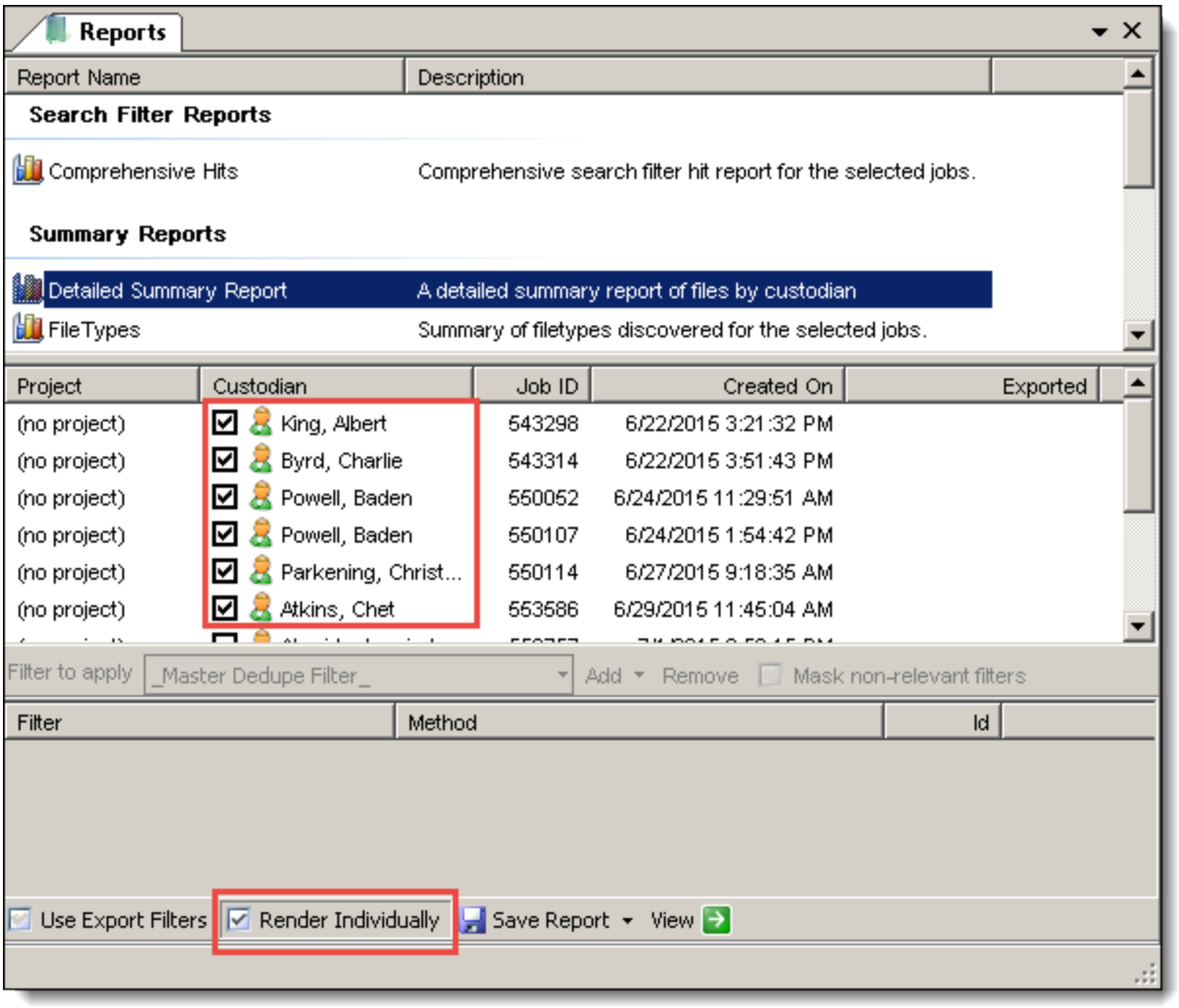
4. To run the report on a subset of documents, select filters in the **Filter to apply drop-down**. Click **Add**, and select the option to include or exclude the data returned by the filter.
    - **As an Include filter** - Includes any document that matches the criteria set in this filter. For example, you could add a keyword search as an include filter, so that all documents with this keyword are exported.
    - **As an Exclude filter** - Excludes any document that matches the criteria set in this filter.
    - **As a Privilege filter** - This has no function in running reports.
5. (Optional) Select **Mask non-relevant filters** to hide any non-relevant filters from the list. If you only want to see the filters that apply to one particular project out of several that exist in the RPC, you can select only that project in the Reports tab and check this option. Doing this blocks the filters that apply to the other projects from displaying in the Filters to apply drop-down.
6. (Optional) Click **Use Export Filters** to use the Export Filters on the report.
7. (Optional) Click **Render Individually** to view the report for each import job in a separate tab. When you save the report with this option checked, individual files generate for each import job. The report data for the selected

import jobs is combined into one tab when you don't select this option. To render reports individually, perform the following steps:

a. Select the report you'd like to run.

b. Select each import job you'd like to run an individual report for.

c. Check the box next to **Render Individually**.



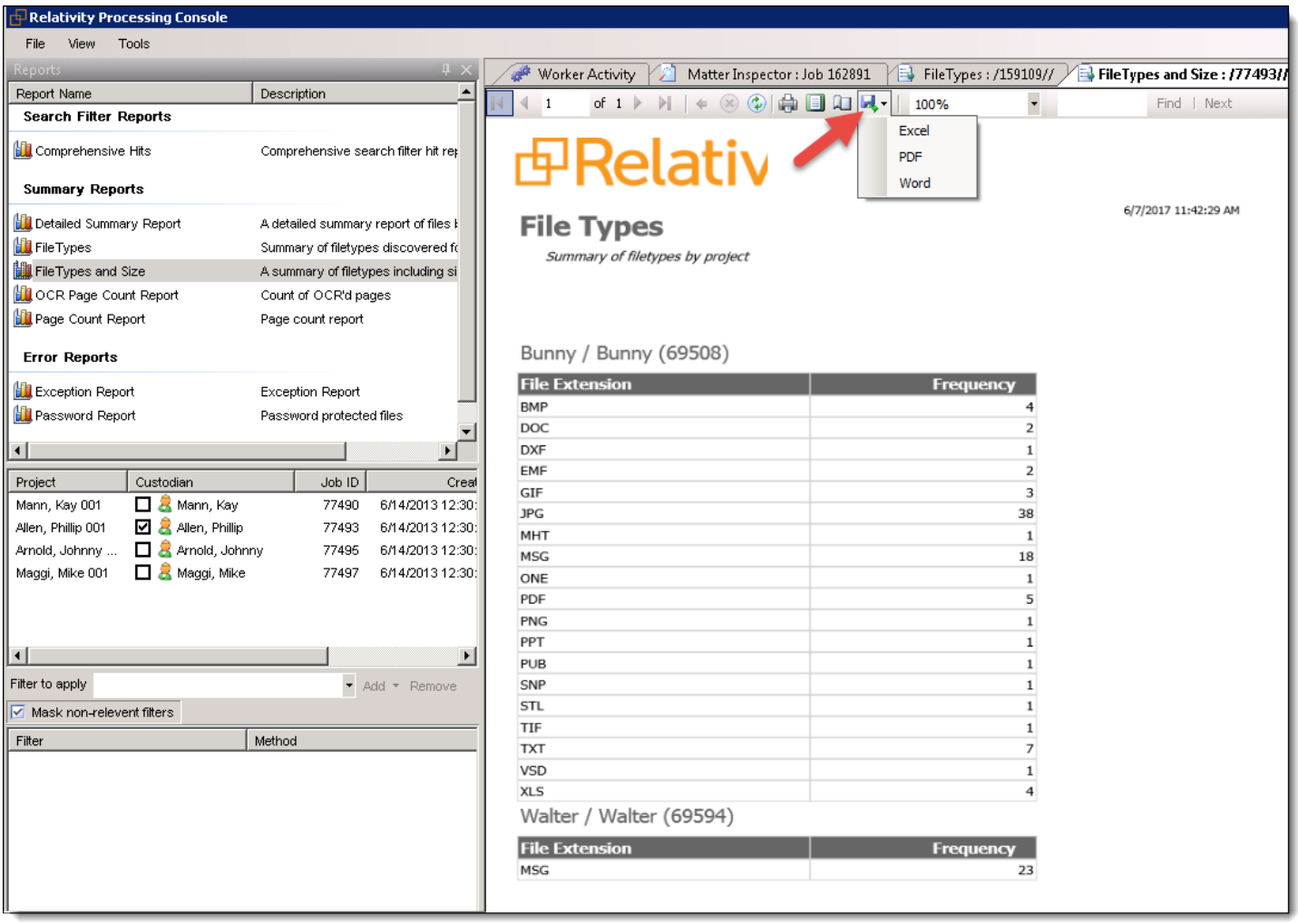
d. Click **View Report**. Notice that a tab opens up for each import job you selected.

8. To create a report, perform one of the following tasks:
   - Click **Save Report** to generate an Excel or Adobe PDF version of the report, and save it to directory of your choice.
   - Click **View** to display the report in the RPC.

There are additional reports available that you can run from the Data Stores window. See Right-click Data Store options on page 105 for more information about these types of reports.

11.0.1 Exporting a report

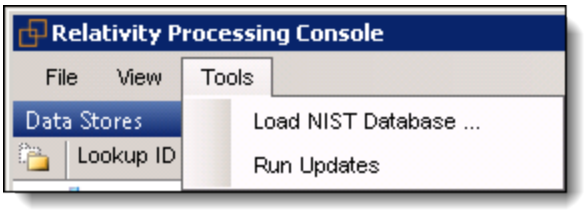
When you run a report on the data in your stores, you have the option of exporting it. To do this, click the export icon in the tool bar at the top of the report window. You have the option of exporting the report as an Excel, PDF, or Word file.


Relativity Processing Console
File
View
Tools
Reports
Report Name
Description
Search Filter Reports
Comprehensive Hits
Comprehensive search filter hit rep
Summary Reports
Detailed Summary Report
A detailed summary report of files b
FileTypes
Summary of filetypes discovered fo
FileTypes and Size
A summary of filetypes including si
OCR Page Count Report
Count of OCR'd pages
Page Count Report
Page count report
Error Reports
Exception Report
Exception Report
Password Report
Password protected files
Project
Custodian
Job ID
Mann, Kay 001
Mann, Kay
77490
Allen, Phillip 001
Allen, Phillip
77493
Arnold, Johnny ...
Arnold, Johnny
77495
Maggi, Mike 001
Maggi, Mike
77497
Filter to apply
Add
Remove
Mask non-relevent filters
Filter
Method
Worker Activity
Matter Inspector : Job 162891
FileTypes : /159109//
FileTypes and Size : /77493//
1
of 1
100%
Find
Next
Excel
PDF
Word
Relativ
6/7/2017 11:42:29 AM
File Types
Summary of filetypes by project
Bunny / Bunny (69508)
File Extension
Frequency
BMP 4
DOC 2
DXF 1
EMF 2
GIF 3
JPG 38
MHT 1
MSG 18
ONE 1
PDF 5
PNG 1
PPT 1
PUB 1
SNP 1
STL 1
TIF 1
TXT 7
VSD 1
XLS 4
Walter / Walter (69594)
File Extension
Frequency
MSG 23

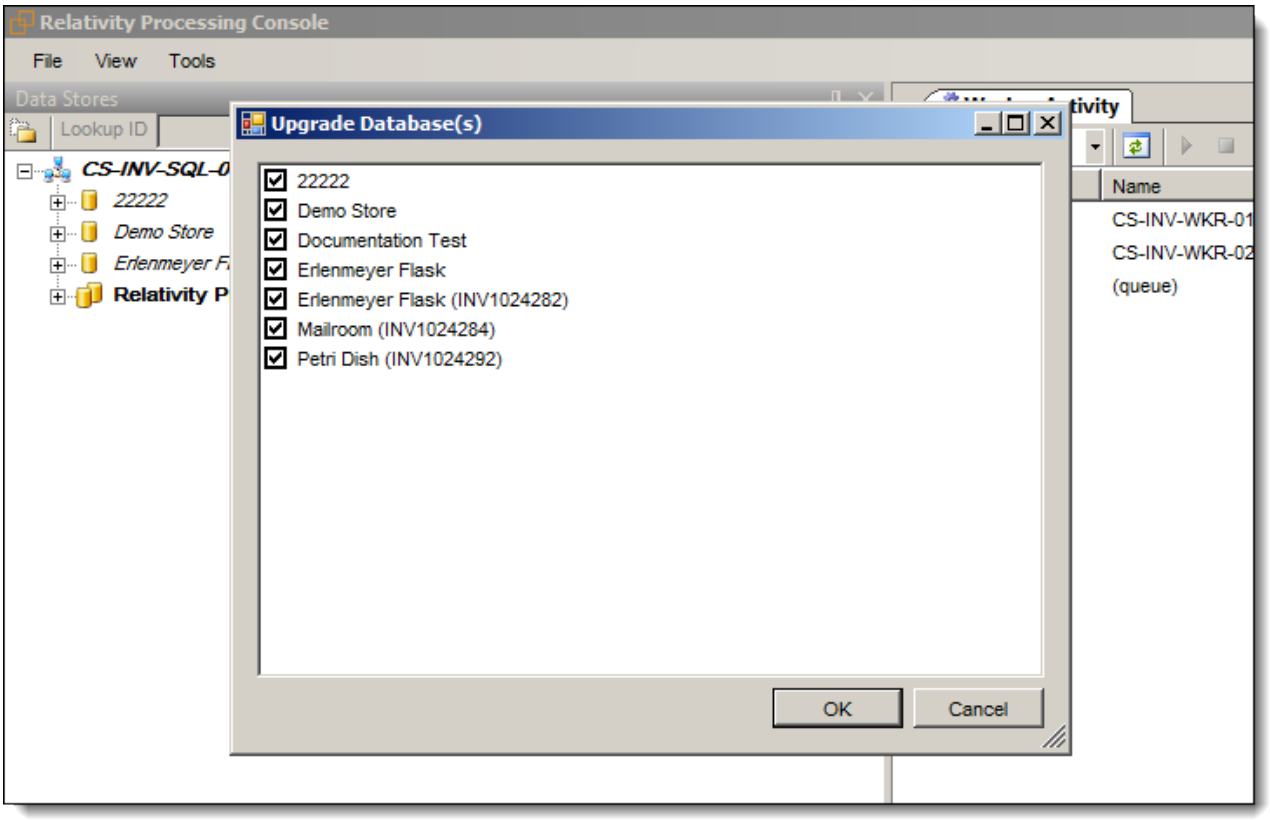
12 Maintenance tasks

In the RPC, the Tools menu provides options for performing maintenance tasks such as updates to databases updates, and XML file definitions.

1. Click the **Tools** menu to display a list of administrative options.



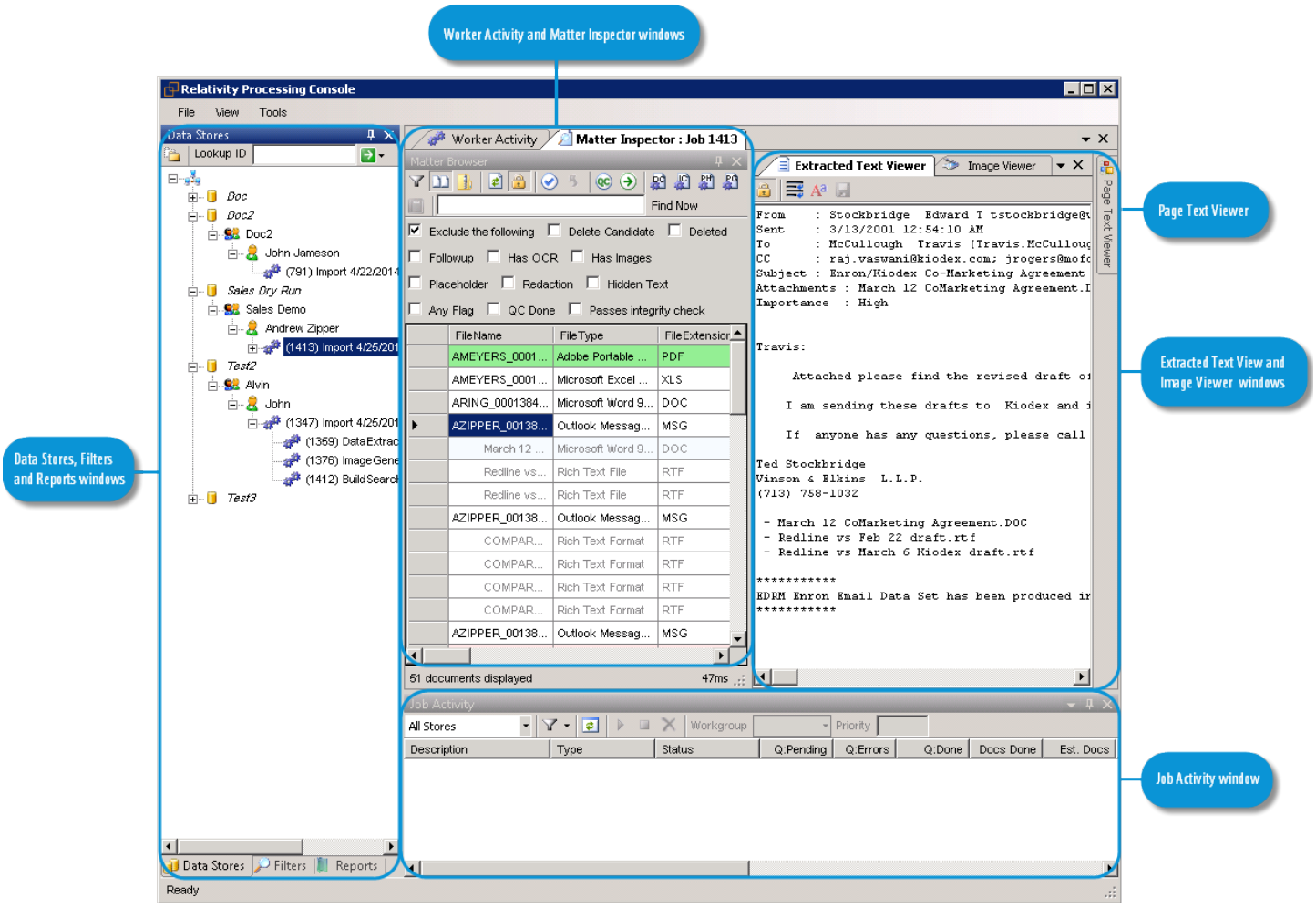
2. Select one of the following menu options:
   - **Load NIST Database** - download the NIST database. Clicking this takes you to a folder browser, from which you can select the folder where you've stored the NIST database. You can procure the most recent version of the NIST database by contacting Support.
   - **Run Updates** - updates the Invariant database schema to the most current version. This includes all stores. Running updates on these databases isn't necessary if you've already upgraded to the most recent version of Invariant. When you click this option, you bring up a window containing all databases available for upgrading.



13 RPC user interface

This referential page describes the main RPC user interface windows and buttons. Find any specific RPC task instructions in their respective topics.

The RPC user interface is comprised of a number of movable windows and panes. Click any of the windows to un-dock, move, and re-dock panes across your desktop.



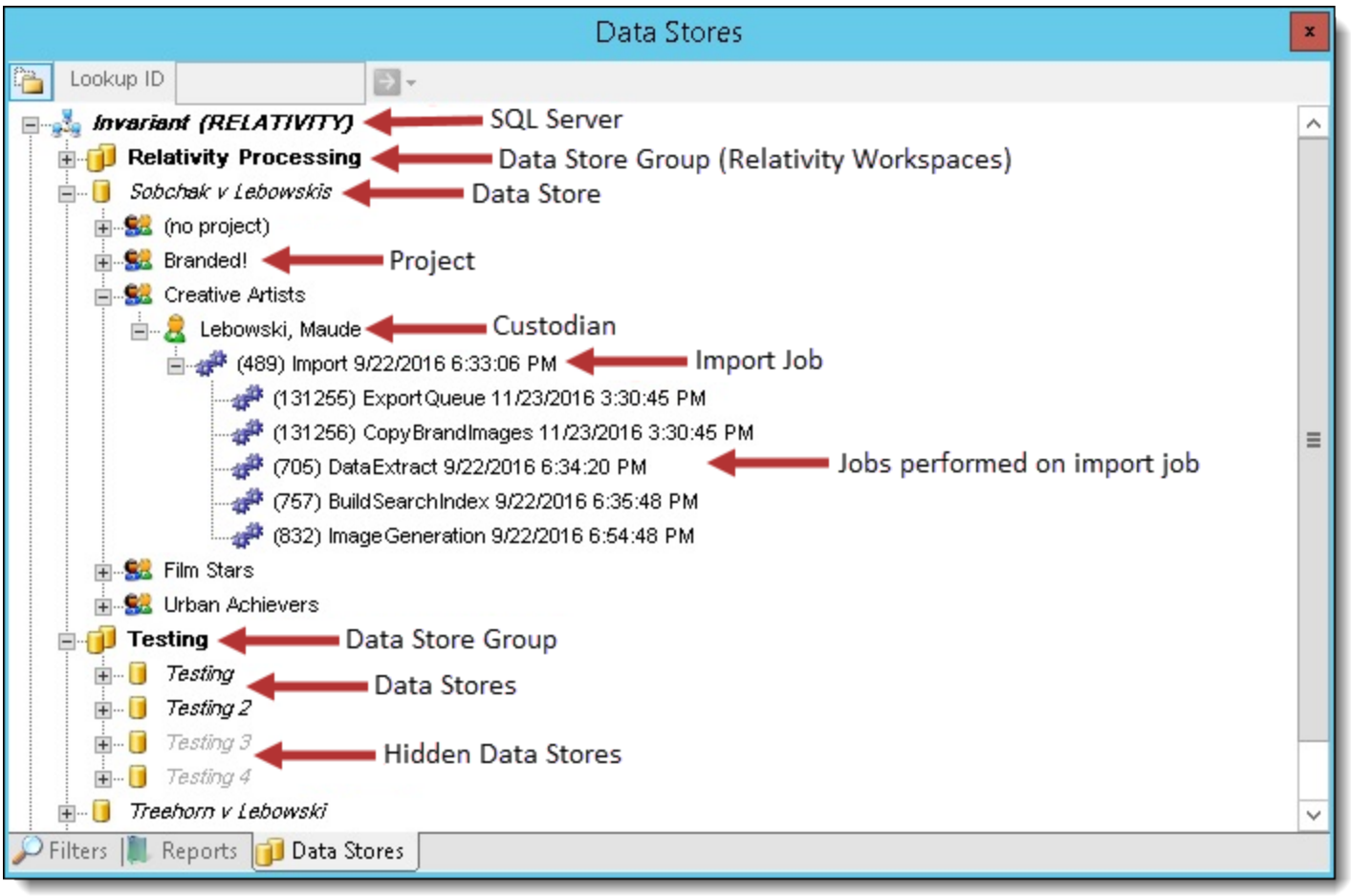
By default, only the Worker Activity, Job Activity, and Data Stores panes are open when you first access the RPC.

Note: Open the various windows and position them to where you will always want them. Then close the program without closing the individual windows first. If you do this, the desired window positions will be stored to a configuration file so that each time you open the RPC, the windows will be where you want them.

13.1 Data Stores window

The Data Stores window contains all the SQL Servers, data stores (workspaces), and imports done in those data stores, along with a listing of all work performed on the imports.

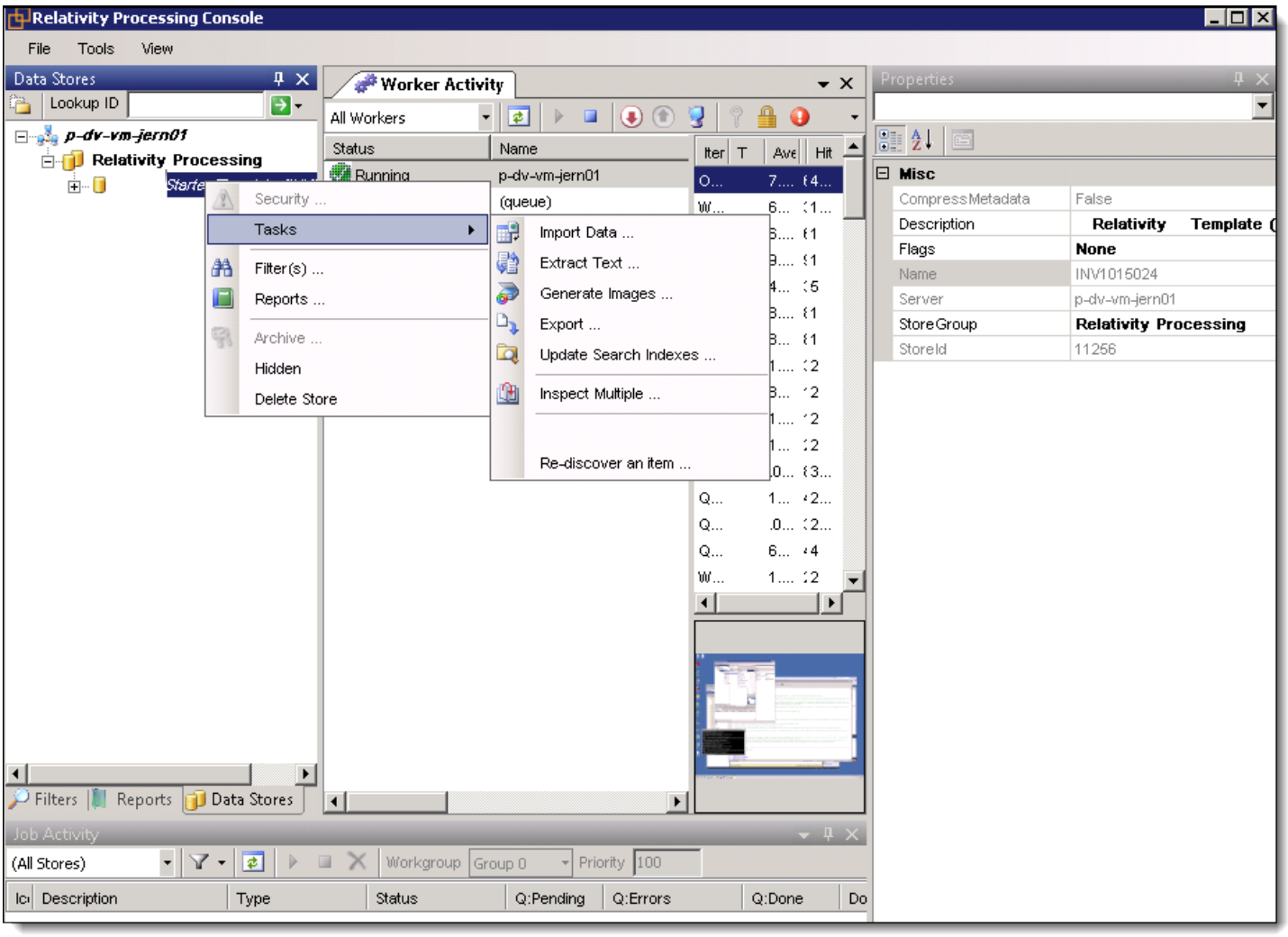
The top-down hierarchy of the display is SQL server > Data store (workspace) / Groups > Project > Custodian > Import > Jobs performed on an import.



Note: When you delete a workspace that corresponds to a data store in the RPC, that data store is not automatically removed from the RPC.

13.1.1 Right-click Data Store options

The following actions are available when you right-click a data store and select the **Tasks** option:



- **Import Data** - import a data set to the selected data store.

The following options are only available after you've imported data and when you perform them, they're applied to all imports in the data store. If you wish to perform any of these tasks on an individual import only, right-click on the import job instead of on the data store:

- **Extract Text** - perform text extraction on a data set.
- **Generate Images** - generate images from a data set.
- **Export** - Export files for review or production.
- **Update Search Indexes** - update the search indexes of a data set.
- **Inspect Multiple** - inspect multiple jobs in a data store.
- **Re-discover an item** - re-discover an item by file ID.

You can right-click on a data store and select **Filter(s)** or **Reports** to access those windows for that data store. You can also right-click on import and run one of these reports:

- **Error report** - gives a detailed list of errors on import job or any job performed on an import.
- **Summary report** - lists the frequency of each file type in the job, the total number of discovered documents and counts of de-NISTed files. It also includes file sizes of all files, de-NISTed files and the size of files remaining

after de-NISTIng not counting containers. Additionally, there is a brief entry describing each error encountered during import.

- **Bad Containers report** - lists all container files that Invariant was unable to pull any documents from. If a container throws an error on import but Invariant was able to get a single file from it, the container will not appear on this report however it will appear on the error report.
- **QC Report** - provides information on who performed various image QC tasks and when.

The Lookup ID section, look at stored files or folders by entering the File ID or Storage ID of a document. These are the choices available after entering an ID number:

- View Native File
- Explore Native Folder
- View Intermediate File
- Explore Intermediate Folder
- Explore Pages Folder

13.1.2 Hiding a data store

To hide a data store from the list, simply right-click on the store and select **Hidden** from the menu. The store will no longer be visible in the list.

To un-hide the store so that it displays in the list, click the **Show All** option in the upper left of the window to display all active and hidden stores, locate the hidden, grayed-out data store name, right-click on it and select **Hidden** again. The store will now be visible in the list.

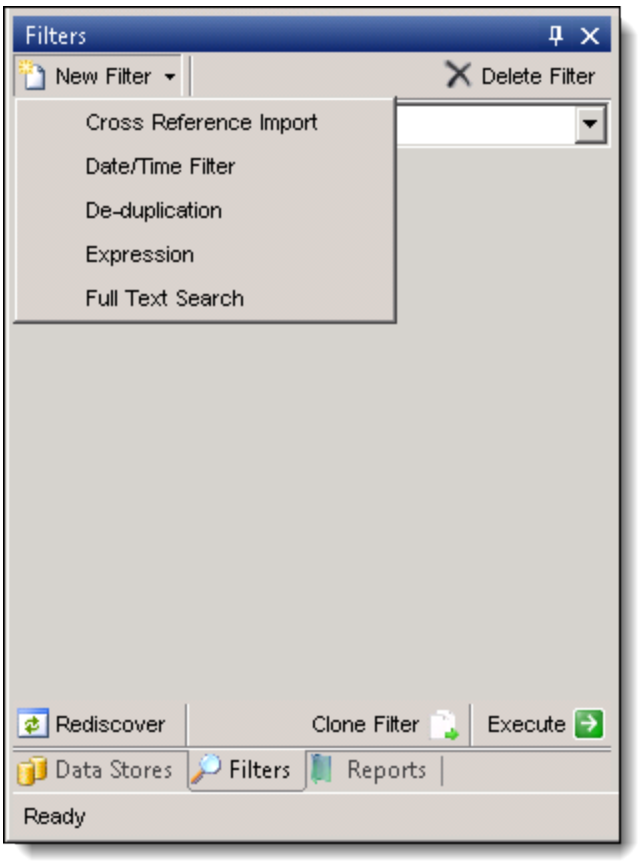
13.1.3 Deleting a data store

To delete a data store from the list, right-click on the store and select **Delete Store** from the menu. Note the following details about store deletion:

- You can't delete stores associated with active Relativity workspaces, meaning workspaces that haven't been deleted.
- You can manually delete a store through the right-click menu as soon as the workspace is marked for deletion by the Case Manager.
- If you don't manually delete a store and the workspace associate with it is deleted from Relativity, that store will automatically be removed from the store list in the RPC.

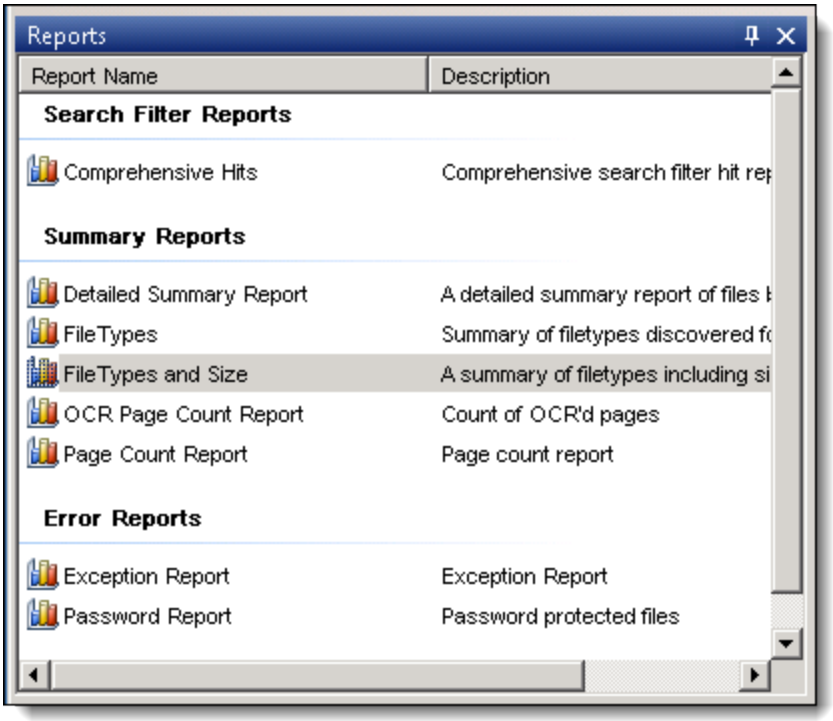
13.2 Filters window

The Filters window contains RPC Filters for use. Right-click a data store > Filters to access the Filters window or click the Filters tab at the bottom of the window. See Filtering data on page 42 for more information.



13.3 Reports window

The Reports window contains RPC reports to run on data sets in data stores. Select a data store and click the Reports tab at the bottom of the window. See Running standard reports on page 99 for more information.



Search Filter Reports

- **Comprehensive Hits** - displays a list of search terms and the frequency that they occur in the document set. You need to apply a search filter when you run this report.

Summary Reports

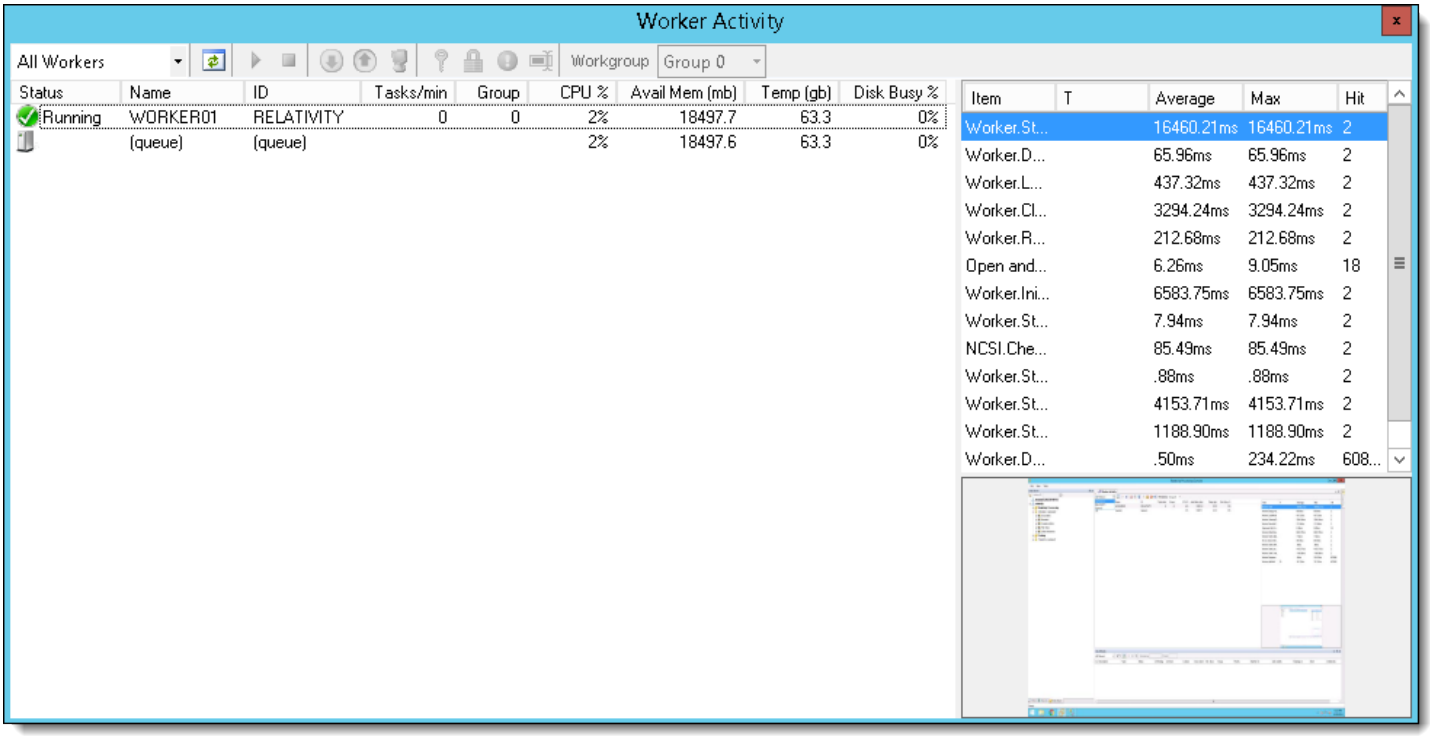
- **Detailed Summary Report** - contains data about the number and size of raw files, number of files before and after De-NISTing, number of e-mail messages and their attachments, and other information.
- **File Types** - lists the extensions of all the files in the data set, and the number of occurrences for each one. It also includes summary totals.
- **FileTypes and Size** - lists the extensions of all the files in the data set, the number of occurrences for each one, and file sizes. It also includes summary totals.
- **OCR Page Count** - lists the number of OCR'd pages.
- **Page Count**- lists the total page count, in addition to pages marked as deleted and pages marked as delete candidates.

Error Reports

- **Exception Report** - displays the exception totals for Unprocessable files, Password protected files, and Errored Files.
- **Password Report** - displays the exception totals for Password protected files and total file size count.

13.4 Worker Activity window

The Worker Activity window contains a list of all workers along with information on their current tasks per minute, group assignment, and CPU usage. Click View > Workers window to access the Workers window. See Managing workers and jobs on page 80 for more information.

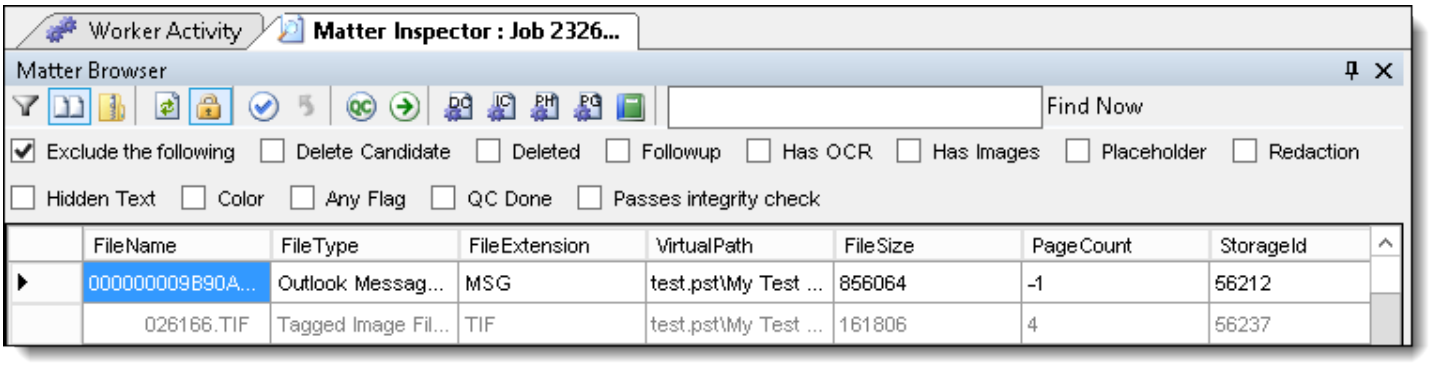


13.4.1 Worker Activity menu options

Menu Options	Description
All Workers drop-down list	Displays list of all workers.
Refresh	Refreshes the Worker Activity window.
Start	Starts the selected worker.
Stop	Stops the selected worker.
Offline	Takes the selected worker offline.
Online	Brings the selected worker online.
Remote Desktop	Open a remote desktop connection to the worker.
Remote Logon	Log in to a remote machine.
Remote Logoff	Log off of a remote machine.
Reboot Worker	Reboot a worker.
Remote command	Execute a remote command.
Workgroup drop-down list	Displays the selected Worker's workgroup.

13.5 Matter Inspector window

The Matter Inspector window displays a list of imported data files and a fixed set of metadata associated with them. See Performing Quality Control tasks on page 90 for more information.



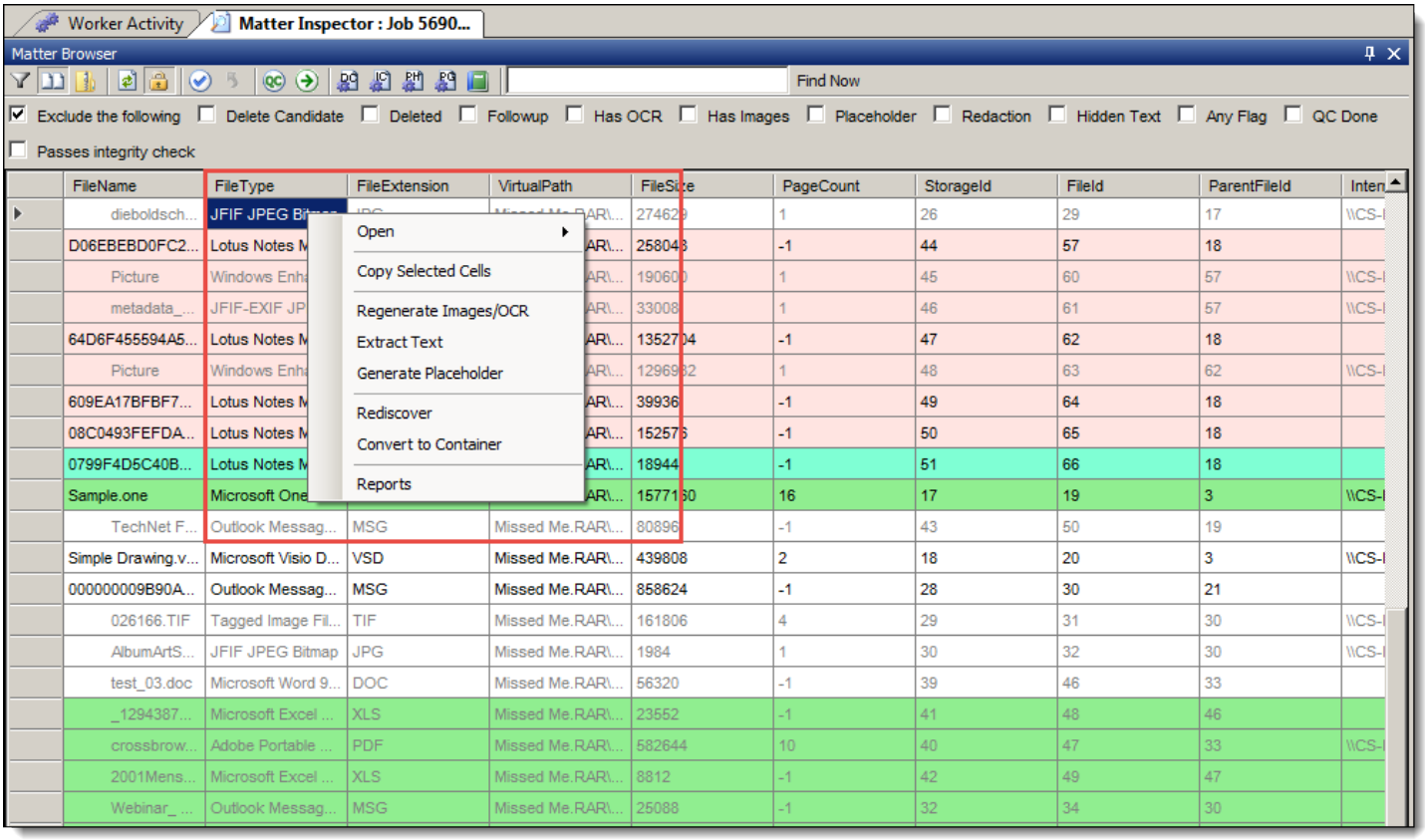
13.5.1 Matter Browser menu options

Menu options	Description
Filter	Click this icon to display the Select Filters dialog box. See Filtering data on page 42.
Show Duplicates	Use this option to show or hide duplicate documents. Duplicate documents are shaded in pink.
Show Containers	Use this option to show or hide containers, including ZIP, RAR, and PDF portfolios. (PDF portfolios contain other PDFs, but do not have content of their own.)
Check out Check in	Use these options when you want to check out a document so no other users can perform QC tasks on it, and then check the document in when the QC is completed. (Invariant will skip documents that are checked out when you browse through the list.)
Finalize QC	Highlight documents that you want marked as QC completed, and click this icon. The Matter Browser then displays these documents with green shading.
Automatic QC	Click this option to automatically flag documents as being QC'd, and then click the Play button in the Image Viewer to page through the documents.
Promote Delete Candidates	Click this option to change pages flagged as Delete Candidate to Delete. Note that this action will be applied to all documents currently displayed in the Matter inspector. Pages with a delete flag will not be exported.
Mark Integrity Check Completed	Click this option to mark this QC task as being completed. This verification step is listed in the QC Report.
Mark Placeholder Check Completed	Click this option to mark this QC task as being completed. This verification step is listed in the QC Report.
Mark Page by Page QC Completed	Click this option to mark this QC task as being completed. This verification step is listed in the QC Report.
Generate QC Report	Generates a report that displays who completed the above QC steps and when they were completed.
Exclude the Following	Use this option to exclude documents that have been flagged with any of the options listed below.
Deleted, Followup, Has OCR, Has Images, Placeholder, Redaction, Hidden Text, Any Flag, QC Done, Passes Integrity Check	Use these checkboxes to filter on documents with these flags that you do or do not want displayed in the grid based on the setting selected for 'Exclude the Following'. For example, you can select Delete Candidate if you want to display documents that have pages with the Delete Candidate flag so that you can QC them before promoting them to Deleted. • The Redaction flag identifies files that have been named with a naming convention of "storageID_R.PDF" and copied into the intermediate folder. The RPC supports using these redacted PDF's, but has no redaction capability itself.

Note: The Matter Inspector is not necessarily reflective of the complete document list that you would be exporting. This is because the inspector isn't concerned with document relationships, as it is merely a place for you to inspect documents that hit on, for example, any keywords you specified in your report settings. This means that if you don't flag parents or children and you review the resulting documents in the Matter Inspector, you may see a different number of results than you'd see in the comprehensive hit report. Likewise, the number of documents that you export might be different than what you viewed in the Matter Inspector because the exporter has to honor document relationships, and thus it will not export child documents without their parents.

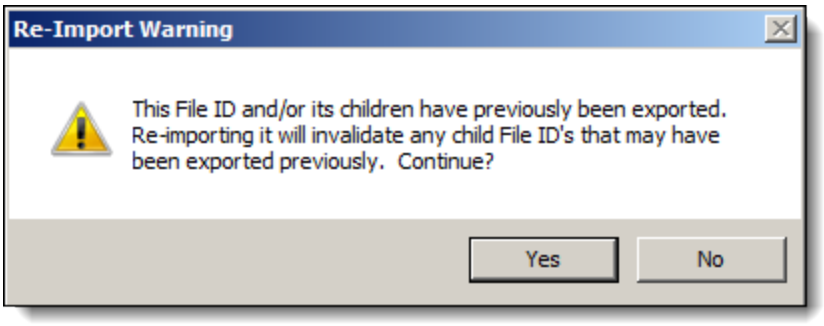
13.5.2 Right-click Matter Browser options

The following options are available to you when you right-click on a record in the Matter Browser:



- **Copy Selected Cells** - copies the cells you've selected to the clipboard. There is no subsequent window or explicit indication that you've taken action on the cells you've selected, but they have been copied once you select this option. This is the only right click option that can be used on more than a single record. Everything below will only work on the document where the mouse is when the right click was executed regardless of how many documents are highlighted.
- **Regenerate Images/OCR** - opens the Job Settings window, where you can switch to the Image or OCR tab to reconfigure your OCR and image settings and perform an imaging job on the selected document. Once you have made the desired settings click start and the document will be placed in the job queue.
- **Extract Text** - queues the selected document for text extraction. It does not first bring up the Job Settings window as the Regenerate Images/OCR option does. If you wish to make changes to the settings before executing extracting text you will need to right click on the appropriate import job in the Data Stores window and select 'Settings'. If the selected document has already been text extracted nothing will happen unless the 'Overwrite intermediate files' setting on the Text tab is checked.

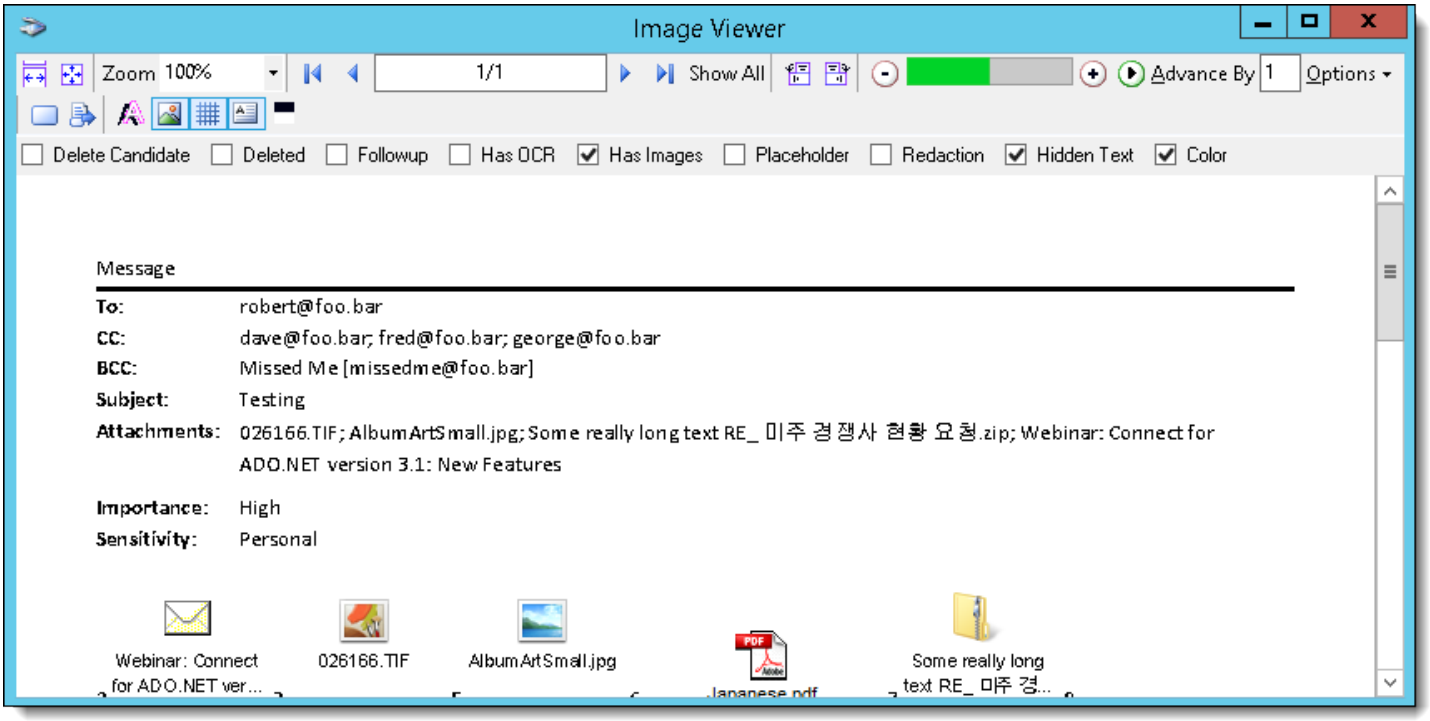
- **Generate Placeholder** - creates a generate placeholder job in the queue for the selected document. If you've already specified a file for the Custom placeholder PDF setting in the OCR/Image tab of the Job Settings window, the placeholder that you generate through this right-click option will display as that file. Thus, there is no option for you to override a custom placeholder PDF in the Matter Browser.
- **Rediscover** - starts a rediscover job for that file. This job appears in the Job Activity window as a ReImport type. Note that if you've already exported this file, you will receive the following confirmation message. Click **Yes** to continue.



- **Convert to Container** - takes what is currently a document in the job and converts it to a container file.
  - Containers are not exported or published, and their children are promoted up a level. For example, if a parent Word document with two child attachments is converted to a container file, then the two child attachments would become parents instead of children. They would then display as coming from a Word document parent container.
  - This option is useful when the RPC misidentifies a container file as a document, for example if it misidentifies a ZIP file and treats it like a document. In this case, the option to convert the document to a container allows you to correct the relationship.
  - When you select this option, the record is no longer available for individual inspection, as you aren't able to inspect containers.
  - You also have the option to convert a container to a document, which does the reverse of container conversion.
- **Reports** - this option currently provides no functionality.

13.6 Image Viewer window

The Image viewer window contains the selected document's images. See Performing Quality Control tasks on page 90 for more information.



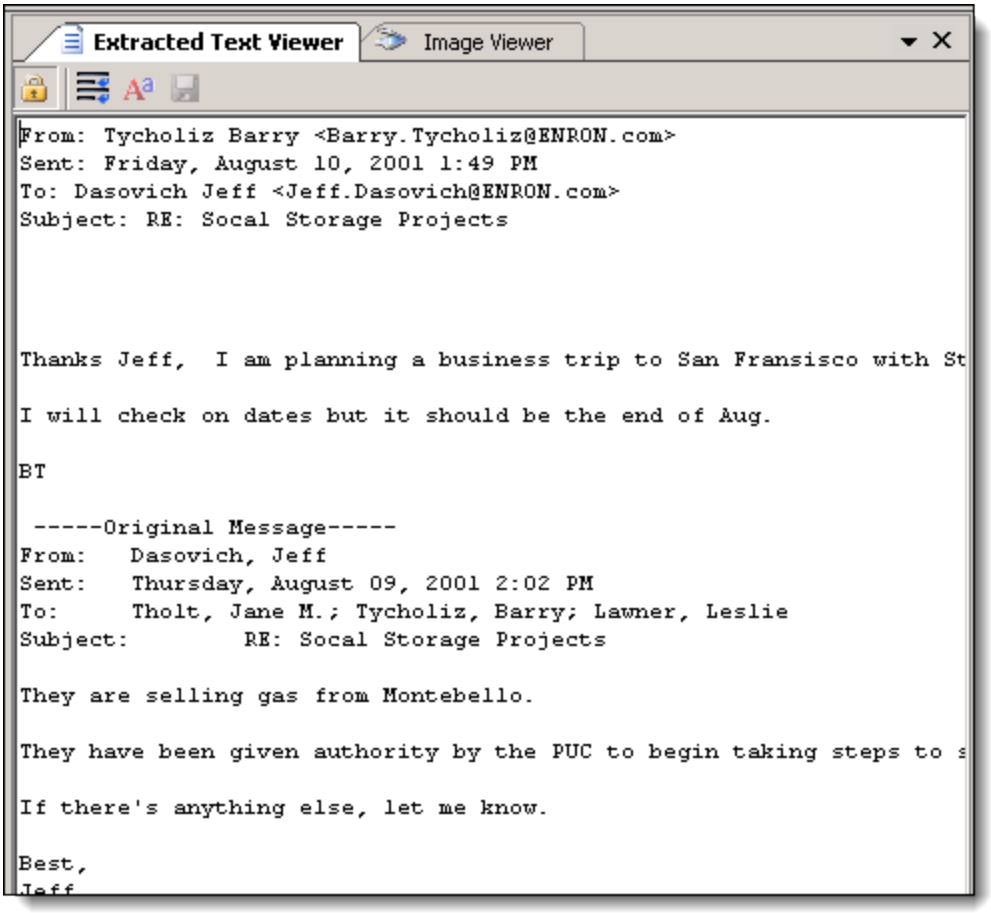
13.6.1 Image viewer menu options

Menu options	Description
Play	Use this option to perform a visual inspection of the images by paging through them. (You can use this option in combination with the **Automatic QC** button in the Matter Browser to confirm that the images are displayed correctly.)
Show All	Shows all pages of the current document regardless of what page level filters may be applied (e.g. Deleted, Delete Candidates).
Advance By 1	Set the number of images that you want to advance by. Regardless of what this number is set to, the image viewer will always show at least the first page of each document.
Options	These options are available when using the Image Viewer window: • Skip QC'd items during auto-advance. This option is on by default. • Apply current flags to the remaining pages in this doc. Useful most often when you want to add or remove the Deleted or Delete Candidate flag from the remaining pages in the document. • Apply current flags until stopped. Useful when needing to apply or remove a flag on many consecutive pages but not necessarily to the remainder of the document. • Apply rotation settings when applying flags.
Anti-Alias	Toggles on or off anti-aliasing to the displayed document to sharpen the image. Having this turned on will slow down the auto-advance of the images.
Show Images	Use this option to show or hide the image layer of the intermediate PDF currently being displayed in the image viewer.

Menu options	Description
Show Line Art	Use this option to show or hide the line art layer of the intermediate PDF currently being displayed in the image viewer. (Line art is most commonly associated with CAD drawings.)
Show Text	Use this option to show or hide the text layer of the intermediate PDF currently being displayed in the image viewer.
Black and White	Use this option to display an image in black and white.
Delete Candidate	Use this option to flag pages for deletion from the export. This is especially useful for when you encounter blank pages among a set of images and you want to exclude them from the export file.
Deleted, Followup, Has OCR, Has Images, Placeholder, Redaction, Hidden Text	Use these checkboxes to flag images. The setting for the flag is stored in the database. **Note:** Images marked as delete are not exported, but they remain in the database.

13.7 Extracted Text Viewer window

The Extracted Text Viewer window contains a selected item's extracted text.

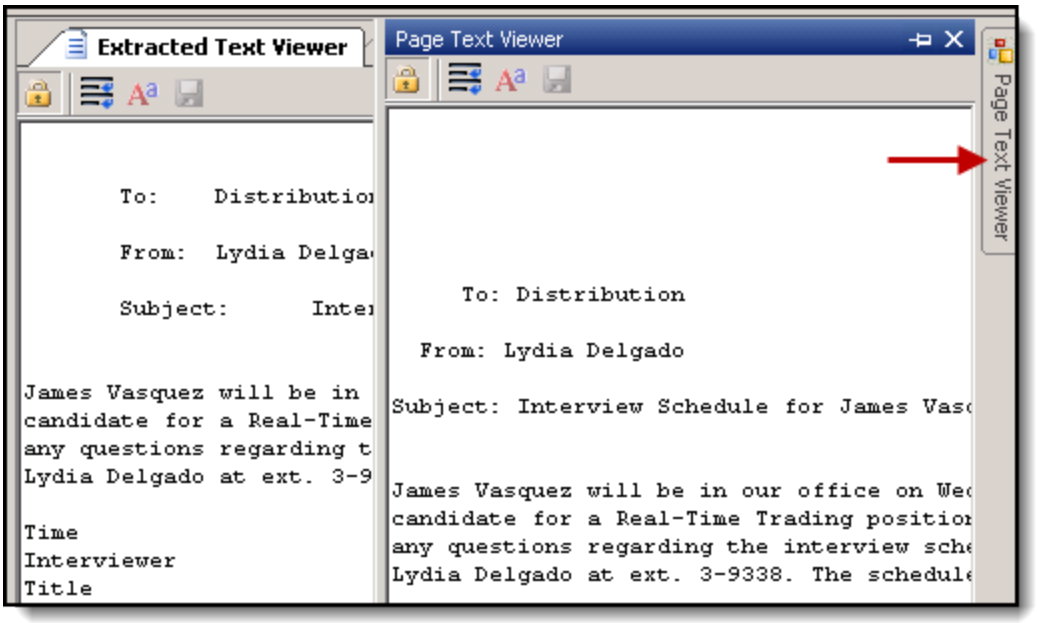


13.7.1 Extracted Text Viewer menu options

Menu Options	Description
Lock/Unlock Grid	Locks/unlocks the grid for editing.
Toggle Word Wrap	Toggle the word wrapping on and off.
Select Font	Select a different extracted text font to use.
Save Changes	Save your changes.

13.8 Page Text Viewer window

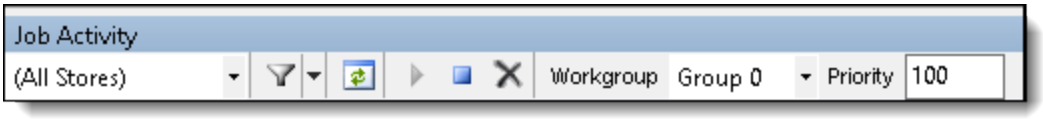
The Page Text Viewer window contains a selected item's page text. Click the horizontal Page Text Viewer button on the right to open the Page Text Viewer window.



The Page Text Viewer window menu options are the same as the Extracted Text Viewer menu options above.

13.9 Job Activity window

The Job Activity window contains all RPC active jobs and corresponding statuses.



13.9.1 Job Activity menu options

Menu Options	Description
All Stores drop-down list	Allows the user to select a specific data store to display jobs only from the selected store.
Filter Icon 	Displays a list of job statuses to be displayed in the Job Activity window with these options: • Not Started • Picked Up • Pending Worker • Started • Finished • Error • Interrupted • Stopped • Paused • Waiting • Canceling • Canceled **Note:** The 'Finished' status is unchecked by default. It is strongly recommended to change the All Stores drop-down list to a specific data store before checking the 'Finished' option. Not doing so will result in every job from every data store being displayed which can significantly slow the RPC or cause the program to lock up.
Refresh 	Refreshes the Job Activity window.
Start 	Starts the selected job.
Stop 	Stops the selected job.
Delete	Deletes the selected job.
Workgroup	Displays the selected job workgroup.
Priority	Displays the selected job priority.

Note: Making changes to the workgroup or priority, or starting, stopping or deleting a job can have serious consequences, especially to jobs that originated from Relativity processing. For more information on what will occur when using these features see Managing workers and jobs on page 80.

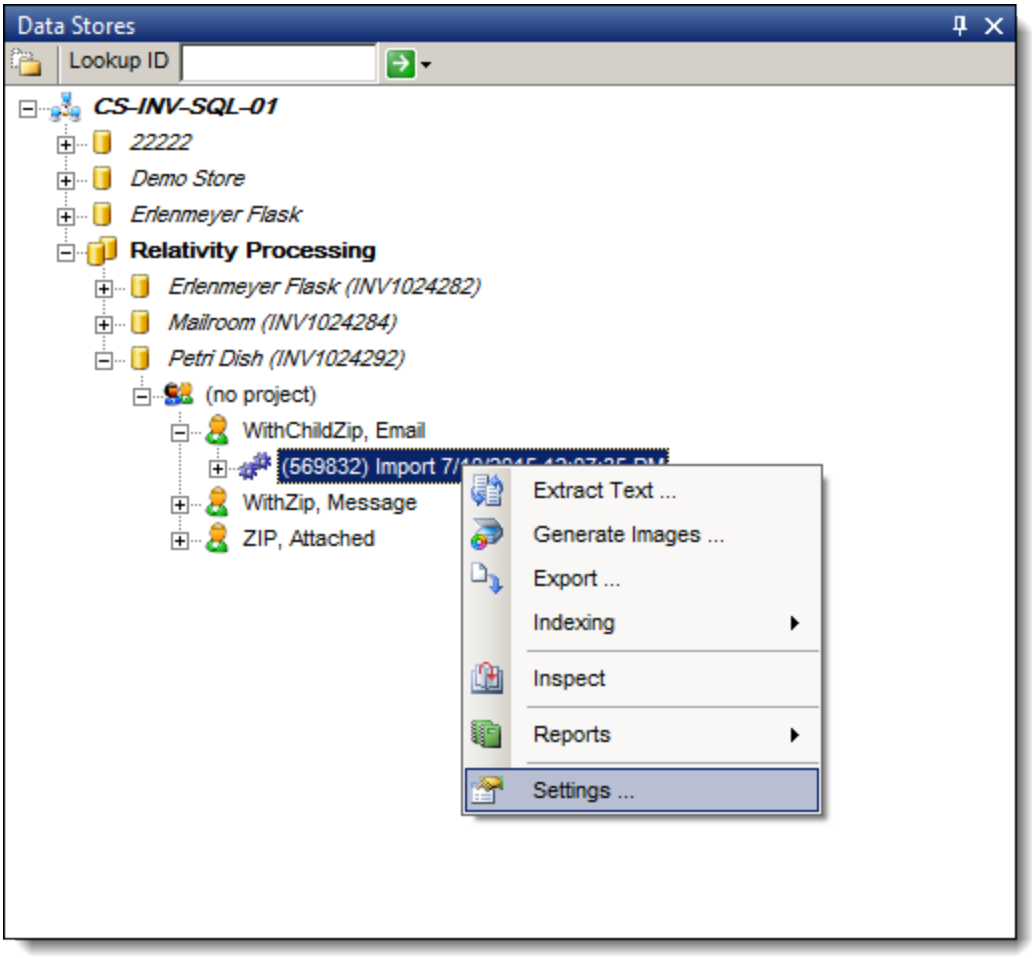
13.10 Configuring job settings

All of the tabs in the Settings pane are available at all times while you are in the RPC, regardless of what operation you're currently configuring. For example, the import settings are available even when you're preparing a data extraction job, even though import settings aren't logically applicable to extraction. Some of the Settings overlap, in that they are applicable to more than once type of job.

If you anticipate having to repeatedly perform jobs with identical settings, you have the option of saving a group of settings as the default for those future jobs.

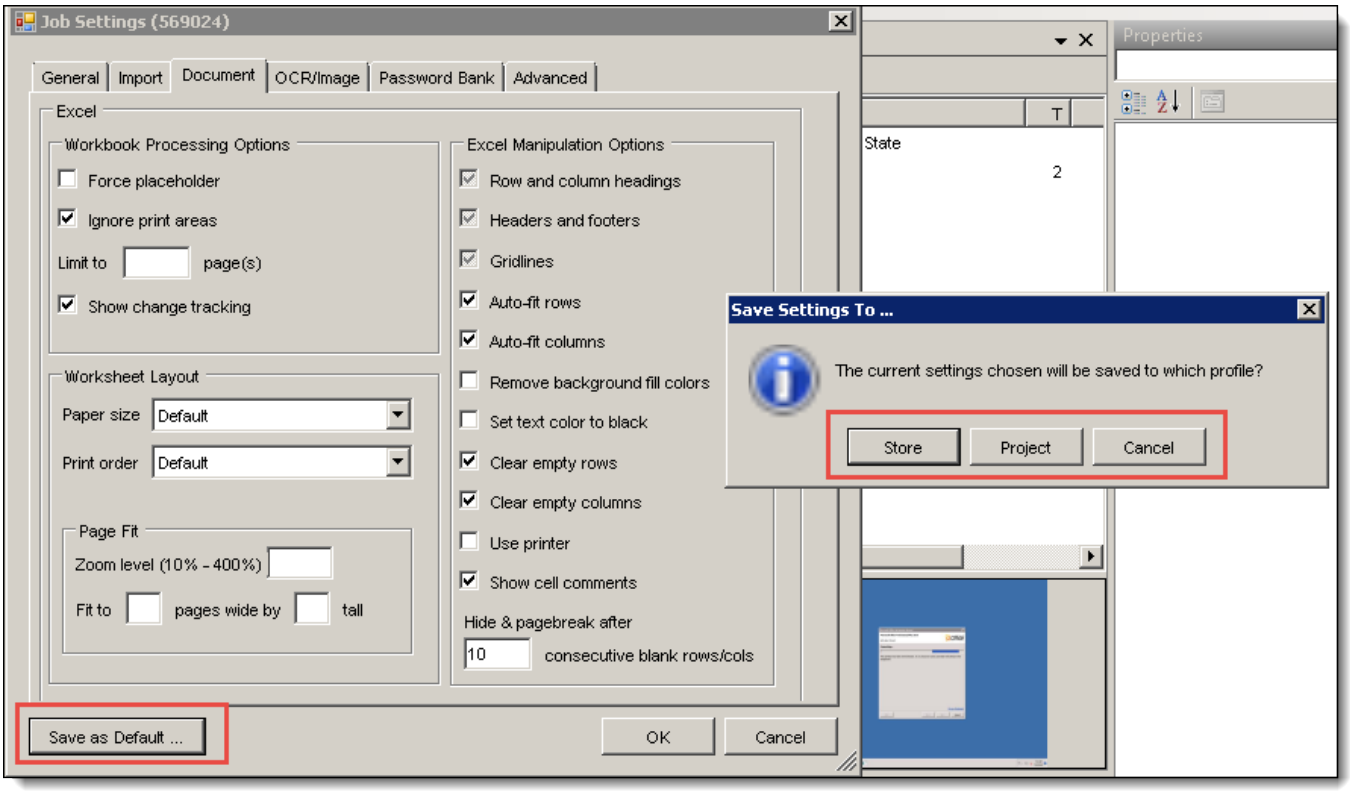
To save your settings as the default:

1. Select the job from the Data Stores pane, right-click, and select **Settings…**

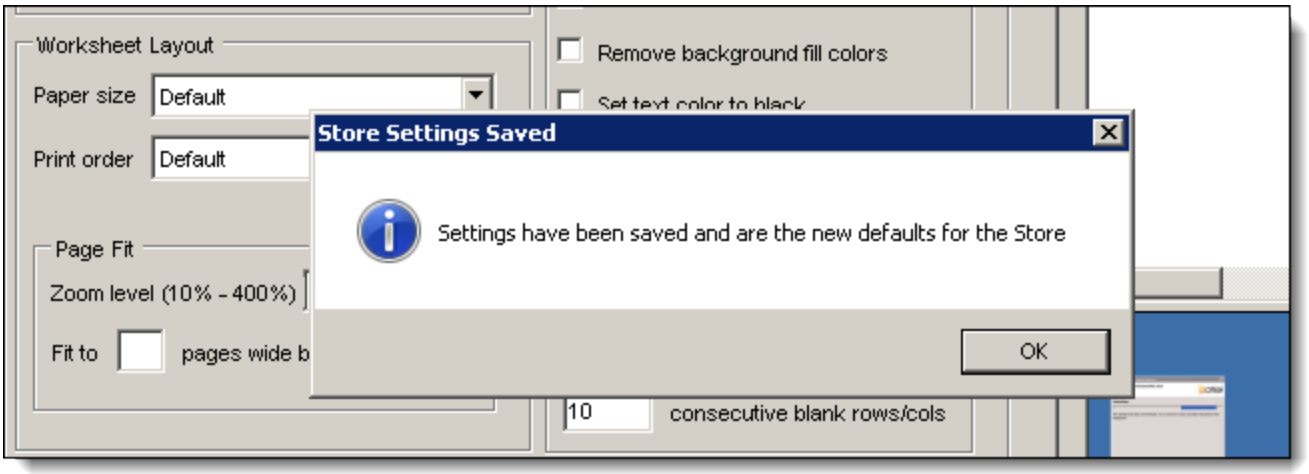


2. Specify settings for all applicable tabs in the Job Settings window.

3. Click **Save as Default** at the bottom left of the window.



4. Select whether to save your settings to a (Data) Store or a Project profile by clicking **Store** or **Project**. If you haven't specified a project name, you'll only have the option to save the defaults to a store. You aren't able to save the settings of the following areas as defaults:
    - The General tab
    - The File Handling section of the Import tab
    - The E-Mail | Ignore PST/OST CRC errors in the Import tab
5. Note the confirmation screen, which states that your settings have been saved and are now the defaults for the store or project.



13.11 Properties window

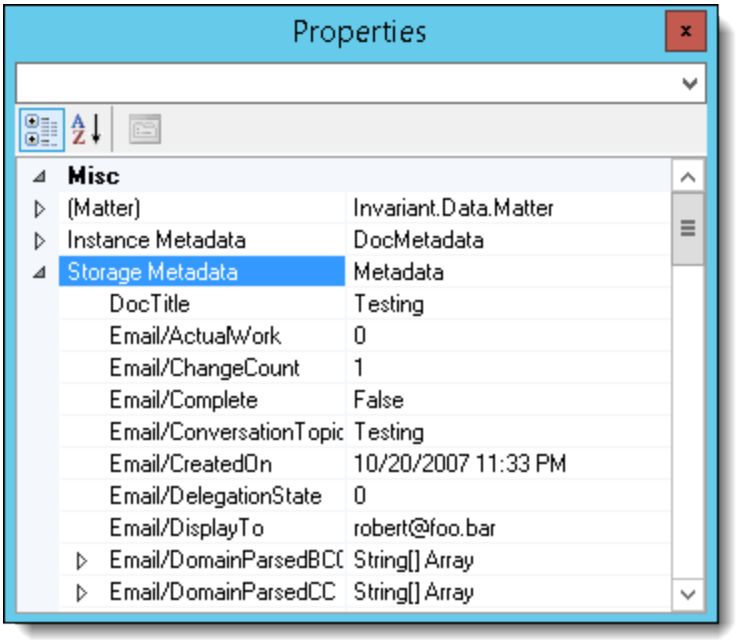
The properties window is a dynamic display of information related to whatever item is currently highlighted in the RPC. It displays information about a specific worker, job, data store or document, as well as any information stored in the underlying SQL tables on the given item being examined. You can also edit some information here.

- For workers, information about what types of work it the worker is set up to do is displayed, and you can edit the name of the worker , along with setting the max threads the worker can utilize. For a breakdown of the worker properties, see Viewing basic worker properties on page 85.
- For jobs, information about doc counts, error counts, Start/Finished/Last Progress times are displayed.
- The properties of a given data store are also available.
- The properties window can be especially useful when used in conjunction with the matter inspector. Here, you can click a specific document and find all available metadata on that document. You can also edit metadata here, even though this isn't something that would be done in a typical workflow.

You can click and highlight the following items to display their properties:

- Data Stores in the Data Stores window
- Jobs in either the Data Stores or Job Activity windows
- Workers or the Queue Manager in the Worker Activity window
- Individual documents in the Matter Inspector

Clicking on the **Project** or **Custodian** level item in the Data Stores window or any item in the **Reports** or **Filters** windows results in nothing being displayed in the Properties window.



- The Categorized or Alphabetical buttons will change the sort order of the various properties displayed. The default is Categorized.
- The Property Pages and blank drop-down above these buttons are currently non-functional.

14 RPC known issues

The following columns are included in the table:

- **Item ID** - the identifying number applied to an issue when it's reported.
- **Affected feature** - the RPC feature to which the defect applies.
- **Description** - a description of the issue.
    - **Resolution** - a resolution for the defect, if known.

Click the Item ID or Affected feature columns to sort defects.

Item ID	Affected feature	Description
REL-17034	Data Stores	The Data Stores tree doesn't automatically update after you run an import job. Resolution Manually collapse and expand the tree to update the Data Stores tree.
REL-17036	Configuration	The RPC doesn't display workstation information (such as the Data Store tree or jobs) when the AppSettings table isn't populated. **Resolution** Perform the required post-upgrade step in the Worker manager server installation guide.
REL-17037	Exporting	The Job Settings options accepts bogus data. For example, Advanced tab.DestinationPath. Resolution There's currently no resolution for this issue.
REL-17039	Importing	The Additional Settings job dialog appears first time with "null _applyTo" object and displays a System Exception for NULL object reference. Resolution There's currently no resolution for this issue.
REL-17049	UI behavior	The comment box clears out when you paste in long text paragraphs exceeding the maximum character limit. This occurs in the Export Wizard, Import Job Project, and Custodian dialog comment boxes. Resolution There's currently no resolution for this issue.
REL-17054	Exporting	The Task metadata in the Task window clears out when you click Show All Metrics. If you click Show All Metrics again, the metadata remains cleared out. Resolution There's currently no resolution for this issue.
REL-17082	Importing	When creating a new Data Store, the Additional Settings button on the Import dialog window is disabled. Resolution There's currently no resolution for this issue.
REL-17161	UI behavior	The Next button on the Task window is enabled without any task selected. Resolution There's currently no resolution for this issue.
REL-17201	UI behavior	The Context menu choice View Tracking Log is disabled. Resolution

Item ID	Affected feature	Description
		There's currently no resolution for this issue.
REL-17266	Exporting	The Want Containers check box option has no effect when selected in Export options wizard. Resolution There's currently no resolution for this issue.
REL-17322	Exporting	The ZIP file isn't included in .LFP output file when selected in the Export options wizard. Resolution There's currently no resolution for this issue.
REL-17346	Exporting	In the Export Wizard, the dialog boxes allow invalid inputs such as negative numbers, decimals, and letters. Resolution There's currently no resolution for this issue.
REL-17577	Exporting	In the Export Field Editor, the Array Merge functionality Write Metatdata and Copy PDF tasks export results are unexpected. Resolution There's currently no resolution for this issue.
REL-17633	Exporting	The Descending setting has no effect when selected in the Export options wizard. Resolution There's currently no resolution for this issue.
REL-17762	UI behavior	The Field and Text delimiters don't auto-fill in the Field Editor. Resolution There's currently no resolution for this issue.
REL-17820	UI behavior	The Any Flag button doesn't display items with a flag, even when items are flagged. Resolution There's currently no resolution for this issue.
REL-17821	UI behavior	The QC Finalize button doesn't display items that have been QC'd. Resolution There's currently no resolution for this issue.
REL-17822	UI behavior	The Promote Delete Candidates button doesn't promote items to be deleted. Resolution There's currently no resolution for this issue.
REL-17853	UI behavior	The Delete button isn't enabled for jobs with a status of Waiting. Resolution There's currently no resolution for this issue.
REL-17942	Filtering	The Remove Filter button doesn't always work. This usually occurs when you create two identical filters and then select the "wrong one" to delete. Resolution There's currently no resolution for this issue.
REL-18199	UI behavior	The RPC crashes when you attempt to rediscover an invalid item. Resolution There's currently no resolution for this issue.
REL-	UI behavior	The RPC crashes when you attempt to rediscover an invalid item with an input string.

Item ID	Affected feature	Description
18200		Resolution There's currently no resolution for this issue.
REL-18210	UI behavior	The Rediscover matter doesn't update file document messages with errors. Resolution There's currently no resolution for this issue.
REL-18416	Filtering	Extra fields aren't populated with input values when you create a method in the Expression filter. Resolution There's currently no resolution for this issue.
REL-18596	UI behavior	The RPC crashes when you cancel an Extract text job on an import job that already has text extraction job. Resolution There's currently no resolution for this issue.
REL-18625	UI behavior	The ActivityTimer window sorts numbers in the Average and Max columns as if they were strings instead of numbers. Resolution There's currently no resolution for this issue.
REL-18666	UI behavior	The RPC crashes when you submit a job without an ActivationKey. Resolution Submit one processing or inventory job from Relativity processing before using the RPC.
REL-19049	Filtering	An error occurs when you enter a file location in the Cross reference file window. Resolution There's currently no resolution for this issue.
REL-19052	Filtering	An error doesn't occur when you enter an invalid file location in the Cross reference file window. Resolution There's currently no resolution for this issue.
REL-19194	Reporting	The Job Error Report incorrectly states that no errors are found, even when errors exist. Resolution There's currently no resolution for this issue.
REL-19216	Importing	When you create an import job with the Extract Text option set to true, The export job doesn't create. Resolution There's currently no resolution for this issue.
REL-19221	UI behavior	The Matter Browser opens incorrect or unexpected selected files. For example, the Matter Browser opens native files when you select intermediate files and vice versa. Resolution There's currently no resolution for this issue.
REL-19222	Filtering	Enter the exact Filter name doesn't submit the Update Index job. Resolution There's currently no resolution for this issue.
REL-19255	Filtering	When using the Cross Reference Import filter, the importer will ignore the 1st line (Header row) and import the rest of the content.

Item ID	Affected feature	Description
		Resolution There's currently no resolution for this issue.
REL-19257	Filtering	The Cross References Import filter isn't able to parse filter results properly, when there are spaces in front of "". Resolution There's currently no resolution for this issue.
REL-19261	Filtering	Unicode characters display as number or letters (incorrectly) in the Filter Results table when you run a Full Text search on the characters . Resolution There's currently no resolution for this issue.
REL-19403	Filtering	The Regular Expression option in the Full Text filter doesn't display accurate results in the Filter Results table. Resolution There's currently no resolution for this issue.
REL-19404	Filtering	The Phonic option in Full Text filter doesn't display correct results for similar sounds words. Resolution There's currently no resolution for this issue.
REL-19609	Data Stores	An Invalid path in the Import Data Source displays files from an unknown location. Resolution There's currently no resolution for this issue.

15 Field mappings

The following table provides a breakdown of which processing fields match up with the most common fields in the RPC.

Processing Relativity field/source name	RPC field name	Field type	Is Unicode?	Description
Attachment Document IDs	ChildRelativityControlNumbers	Long Text	Yes	Attachment document IDs of all child items in family group, delimited by semicolon, only present on parent items.
Attachment List	ChildFileNames	Long Text	Yes	Attachment file names of all child items in a family group, delimited by semicolon, only present on parent items.
Author	Author	Fixed-Length Text (50)	Yes	Original composer of document or sender of email message.
BCC Address	EmailBCCSmtp	Long Text	Yes	The full SMTP value for the email address entered as a recipient of the Blind Carbon Copy of an email message.
CC Address	EmailCCSmtp	Long Text	Yes	The full SMTP value for the email address entered as a recipient of the Carbon Copy of an email message.
Child MD5 Hash Values	ChildMD5Hashes	Long Text	Yes	Attachment MD5 Hash values of all child items in a family group, delimited by semicolon, only present on parent items. The RPC can't calculate this value if you have FIPS (Federal Information Processing Standards cryptography) enabled for the worker manager server.
Child SHA1 Hash Values	ChildSHA1Hashes	Long Text	Yes	Attachment SHA1 Hash values of all child items in a family group, delimited by semicolon, only present on parent items.
Child SHA256 Hash Values	ChildSHA256Hashes	Long Text	Yes	Attachment SHA256 Hash values of all child items in a family group, delimited by semicolon, only present on parent items.
Comments	Comments	Long Text	Yes	Comments extracted from the metadata of the native file.
Company	Company	Fixed-	Yes	The internal value entered for the

Processing Relativity field/source name	RPC field name	Field type	Is Unicode?	Description
		Length Text (255)		company associated with a Microsoft Office document.
Contains Embedded Files	Office/EmbeddedItems	Yes/No	N/A	The yes/no indicator of whether a file such as a Microsoft Word document has additional files embedded in it.
Control Number Beg Attach	BatesBeginAttach	Fixed-Length Text (50)	Yes	The identifier of the first page of the first document in a family group. This is used for page-level numbering schemes.
Control Number End Attach	BatesEndAttach	Fixed-Length Text (50)	Yes	The identifier of the last page of the first document in a family group. This is used for page-level numbering schemes.
Conversation	EmailConversation	Long Text	Yes	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.
Conversation Family	ConversationFamily	Fixed-Length Text (44)	Yes	Relational field for conversation threads. This is a 44-character string of numbers and letters that is created in the initial email.
Conversation Index	EmailConversationIndex	Long Text	Yes	Email thread created by the email system. This is a 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.
Date Created	CreatedOn	Date	N/A	Date and time from the Date Created property extracted from the original file or email message.
Date Last Modified	LastModified	Date	N/A	Date and time from the Modified property of a document, representing the date and time that changes to the document were last saved.
Date Last Printed	LastPrinted	Date	N/A	Date and time that the document was last printed.
Date Received	EmailReceivedOn	Date	N/A	Date and time that the email message was received (according to ori-

Processing Relativity field/source name	RPC field name	Field type	Is Unicode?	Description
				ginal time zones).
Date Sent	EmailSentOn	Date	N/A	Date and time that the email message was sent (according to original time zones).
DeDuped Custodians	DeDupedCustodians	Multiple Object	N/A	Custodians associated with the deduped records of this document.
DeDuped Paths	DeDupedPaths	Long Text	Yes	Folder structure and paths to this document?s duplicates. Each path will contain the associated custodian.
Delivery Receipt	EmailDeliveryReceiptRequested	Yes/No	N/A	The yes/no indicator of whether a delivery receipt was requested for an e-mail.
Document Class	RelativityDocumentClass	Single Choice	N/A	A single choice field that can be one of: Email, Edoc, or Attach.
Document Extension	FileExtension	Fixed-Length Text (25)	Yes	Three (or more) character extension of the document that represents the file type to the Windows Operating System. Examples are PDF, DOC, or DOCX.
Document Subject	DocumentSubject	Long Text	Yes	Subject of the document extracted from the properties of the native file.
Domains (Email BCC)	Email/DomainParsedBCC	Multiple Object	N/A	Domains of 'Blind Carbon Copy' recipients of the email message.
Domains (Email CC)	Email/DomainParsedCC	Multiple Object	N/A	Domains of 'Carbon Copy' recipients of the email message.
Domains (Email From)	Email/DomainParsedFrom	Multiple Object	N/A	Domains of Originator of the email message.
Domains (Email To)	Email/DomainParsedTo	Multiple Object	N/A	Domains of 'To' recipients of the email message.
Email BCC	EmailDisplayBCC	Long Text	Yes	Recipients of 'Blind Carbon Copies' of the email message.
Email Categories	EmailKeywords	Long Text	Yes	Category assigned to an email message.
Email CC	EmailDisplayCC	Long Text	Yes	Recipients of 'Carbon Copies' of the email message.
Email From	EmailDisplaySender	Fixed-Length Text (255)	Yes	Originator of the email message.

Processing Relativity field/source name	RPC field name	Field type	Is Unicode?	Description
Email In Reply To ID	Email/In_Reply_To	Long Text	Yes	The internal metadata value within an email for the reply to ID.
Email Store Name	EmailContainer	Fixed-Length Text (255)	Yes	The identifier of the top-level container of an email message. For example, 'jdoe.nsf.' If a document comes from a rar/zip file attached to the email, the container is referred to in that file.
Email Subject	EmailSubject	Long Text	Yes	Subject of the email message.
Email To	EmailDisplayTo	Long Text	Yes	List of recipients or addressees of the email message.
File Name	FileName	Fixed-Length Text (25)	Yes	Original name of the file.
File Size	FileSize	Decimal	N/A	Generally a decimal number indicating the size in bytes of a file.
File Type	FileType	Fixed-Length Text (255)	Yes	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.
From Address	EmailSenderSmtp	Long Text	Yes	The full SMTP value for the sender of an email message.
Group Identifier	RelativityGroupId	Fixed-Length Text (40)	Yes	Group the file belongs to (used to identify the group if attachment fields are not used).
Has Hidden Data	HiddenText	Yes/No	N/A	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.
Importance	EmailImportance	Single Choice	N/A	Notation created for email messages to note a higher level of importance than other email messages added by the email originator.
Keywords	Office/Keywords	Long Text		The internal value entered for keywords associated with a Microsoft Office document.
Last Accessed	LastAccessed	Date	N/A	The date and time at which the loose

Processing Relativity field/source name	RPC field name	Field type	Is Unicode?	Description
Date/Time				file was last accessed.
Last Saved By	Office/LastAuthor	Fixed-Length Text (255)	Yes	The internal value indicating the last user to save a document.
Last Saved Date/Time	LastSaved	Date	N/A	The internal value entered for the date and time at which a document was last saved.
Lotus Notes Other Folders	Lotus/OtherFolders	Long Text	Yes	A semi-colon delimited listing of all non-primary folders that a Lotus Notes message or document was included.
MD5 Hash	MD5Hash	Fixed-Length Text (40)	Yes	Identifying value of an electronic record that can be used for deduplication and authentication generated using the MD5 hash algorithm.
Meeting End Date/Time	Email/EndDate	Date	N/A	The date and time at which a meeting item in Outlook or Lotus Notes ended.
Meeting Start Date/Time	Email/StartDate	Date	N/A	The date and time at which a meeting item in Outlook or Lotus Notes began.
Message Header	Email/MessageHeader	Long Text	Yes	The full string of values contained in an email message header.
Message ID	EmailMessageID	Fixed-Length Text (255)		The message number created by an email application and extracted from the email's metadata.
Message Type	Email/MessageType	Single Choice	N/A	An indication of 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.
Native File	StoredAs	Long Text	Yes	The path to a copy of a file for loading into Relativity.
Number of Attachments	AttachmentCount	Whole Number	N/A	Number of files attached to a parent document.
OCR Text	HasOcrText	Yes/No	N/A	The yes/no indicator of whether the

Processing Relativity field/source name	RPC field name	Field type	Is Unicode?	Description
				extracted text field contains OCR text.
Office Document Manager	Office/Manager	Fixed-Length Text (255)	Yes	The internal value entered for the manager of a document.
Office Revision Number	Office/Revision	Fixed-Length Text (255)	Yes	The internal value for the revision number within a Microsoft Office document.
Other Props	OtherProps	Long Text	Yes	Metadata extracted during processing for additional fields beyond the list of processing fields available for mapping. Field names and their corresponding values are delimited by semicolon.
Parent Document ID	ParentRelativityControlNumber	Fixed-Length Text (50)	Yes	Document ID of the parent document. This field will only be available on child items.
Password Protected	PasswordProtected	Single Choice	N/A	An indication of documents that were password protected. It will contain 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.
Primary Date	DocDate	Date	N/A	Date taken from Email Sent Date, Email Received Date, or Last Modified Date (in order of precedence).
Read Receipt	EmailReadReceiptRequested	Yes/No	N/A	The yes/no indicator of whether a read receipt was requested for an e-mail.
Sensitivity	EmailSensitivity	Single Choice	N/A	The indicator set on an email to denote the email's level of privacy.
SHA1 Hash	SHA1Hash	Fixed-Length Text (50)	Yes	Identifying value of an electronic record that can be used for deduplication and authentication generated using the SHA1 hash algorithm.
SHA256 Hash	SHA256Hash	Fixed-Length Text (70)	Yes	Identifying value of an electronic record that can be used for deduplication and authentication gen-

Processing Relativity field/source name	RPC field name	Field type	Is Unicode?	Description
				erated using the SHA256 hash algorithm.
Sort Date	SortDate	Date	N/A	Date taken from Email Sent Date, Email Received Date, or Last Modified Date (in order of precedence) repeated for parent document and all children items to allow for date sorting.
Speaker Notes	PowerPoint/HasSpeakerNotes	Yes/No	N/A	The yes/no indicator of whether a powerpoint file has speaker notes associated with its slides.
To Address	EmailToSmtp	Long Text	Yes	The full SMTP value for the recipient of an email message. For example, 'bob@example.com'
Track Changes	TrackChanges	Yes/No	N/A	The yes/no indicator of whether tracked changes exist in the document.
Unified Title	UnifiedSubject	Long Text	Yes	Subject of the document. If the document is an email, this field contains the email subject. If the document is not an email, this field contains the document's file name.
Unprocessable	Unprocessable	Yes/No	N/A	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.
Unread Flag	EmailIsUnread	Yes/No	N/A	The yes/no indicator of whether an e-mail was not read.

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.