

File Analysis Suite

Software Version 3.8.0

Release Notes



Document Release Date: November 2022
Software Release Date: November 2022

Legal notices

© Copyright 2019-2022 Micro Focus or one of its affiliates.

The only warranties for products and services of Micro Focus and its affiliates and licensors ("Micro Focus") are as may be set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Micro Focus shall not be liable for technical or editorial errors or omissions contained herein. The information contained herein is subject to change without notice.

Except as specifically indicated otherwise, this document contains confidential information and a valid license is required for possession, use or copying. If this work is provided to the U.S. Government, consistent with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under vendor's standard commercial license.

Contents

- File Analysis Suite Release Notes 1
 - About File Analysis Suite 1
- What's new in 3.8.0 2
 - General 2
 - Processing Agent 5
 - Connect 6
 - Analyze 8
 - Manage 8
 - Administration 9
 - API updates 9
- Resolved issues in 3.8.0 10
- Known issues in 3.8.0 11
- Documentation 12
- Send documentation feedback 13

File Analysis Suite Release Notes

Software version: 3.8.0

Publication date: November 2022

This document is an overview of the changes made to File Analysis Suite.

Support matrix

For information about the installation requirements and compatibility with other products, see the File Analysis Suite Support Matrix. The support matrix may be updated between releases so it is available only from the [Support portal](#).

About File Analysis Suite

Micro Focus File Analysis Suite lets you find, protect, and secure sensitive and high-value data within on-premises and cloud unstructured data repositories across your enterprise. Identify, collect, and organize content to ensure discovery of sensitive data. Configure how sources and repositories are processed and categorized with Connect. Analyze your data under management with Analyze. Organize, review, and take action with Manage.

What's new in 3.8.0

The following features and enhancements are included in this release.

General

The following features and enhancements apply to more than one component in File Analysis Suite.

- Added the ability to maintain a pared down File Analysis Suite index. You can now identify and classify files without moving or storing any content in File Analysis Suite. This reduces the risk and cost of discovering, protecting, and managing your data.
 - When creating a repository, the "Analyze" action no longer includes indexing the document content. When creating a repository, the "SmartScan (Analyzed)" and "Full Scan (Analyzed)" options now index the metadata and analyze the metadata and body content to identify tag values but do not store the body content.
 - When creating a repository, the new "Store content as text" option for Analyze actions lets you optionally store the document content in the index.
 - On the Data Volume dashboard in Connect, replaced the Documents on Hold information with Content Stored Documents information. This new card shows the number of documents for which the metadata and the content has been analyzed and stored in the index. Extracted attachments are not included in the count. Documents within email archives and compressed archives are included in the count only if they were processed as part of the repository's Primary Capture Rules.
 - Added Content Stored Documents count to the Metrics tab of the repository detail panel.
 - Added Content Stored Count and Content Stored File Size to the Usage report in Administration.
- Added support for protecting files with Microsoft Purview Information Protection. You can use your existing Microsoft Purview Information Protection labels to identify and protect sensitive data managed by file system repositories in File Analysis Suite.

This new file protection system type is referred to as "Microsoft data protection" in File Analysis Suite. For more information about protecting files with Microsoft Purview Information Protection, see "Microsoft Purview Information Protection connection" in the Connect Help Center.

- Removed support for protecting managed files with SmartCipher.
- Added the ability to use an external application to extract metadata from documents stored in your file systems. You can process the content of the resulting metadata files in conjunction with the original documents in file system repositories. You can process external metadata for documents regardless of whether you store the document content. During processing, File Analysis Suite reads metadata information from the external metadata files and applies the information according to the defined external metadata file processor. The external metadata files are not processed as part of the repository primary capture rules.

When creating a file system repository, you can specify whether to store additional metadata that is read from a defined file. The additional metadata is stored with the original document. When enabled (selected), you will define the file extensions of the files that contain the external metadata information and the external metadata file processor that will be used to read the metadata files and apply the information.

For more information about external metadata file capture, see "Process external metadata" in the Connect Help Center.

- Added the concept of grammar sets—groupings of grammar constructs used to identify potentially sensitive data in your environment. When you create repositories, you select the grammar sets to apply. This lets you apply consistent grammar rules with less effort.

File Analysis Suite includes built-in grammar sets that pertain to region-specific regulations that you can apply across data repositories. For example, the "EU General Protection Regulation (UK)" grammar set contains grammar rules that identify data pertinent to GDPR and are specific to data patterns for the United Kingdom (addresses, phone numbers, bank details, and so on). Built-in grammar sets cannot be edited or deleted.

You can also create custom grammar sets based on grammar classes, grammar types, and grammar rules to identify data that is important to your company. Custom grammar sets can include built-in and custom grammar rules.

Updated terminology related to grammars:

- "Grammars" are now referred to "grammar classes".
- "Grammar categories" are now referred to as "grammar types".
- "Entities" are now referred to as "grammar rules".
- "Enrich" is now referred to as "extract grammar values".
- The processing speed of documents with a large number of grammar value matches has been improved.

As an example, a processing job that identified 5000 grammar values, in a single document, completed approximately 40% quicker than prior to this improvement.

NOTE: This is an example of the improvement you may experience. The type of processing, document size, and number of grammar values identified will affect processing speeds.











- Updated grammar processing for content within tables and spreadsheets. Grammar processing now considers the content in the context of column heading content.




For example, a table in a document contains a column for date of birth. The column header is "DOB". The dates in this column are now identified specifically as dates of birth as opposed to simply *a date*.

- Added the ability to search by protection systems and protection rules when performing an advanced search in Analyze and Manage. In Connect, you can now also create tags that include criteria based on protection systems and protection rules.
- Added file type groupings to let you search for documents by more generic file types, such as "image" or "compressed file". Use the "File Types (Grouped)" option in the filter search panel

and search builder when refining a document list in Analyze or Manage and in the tag criteria when creating tags in Connect.

- Updated the information icons in the document list in Analyze and Manage.

Icon	Description
	<p>(circles in varying colors, solid) Is associated with a weighted label, but content has <i>not</i> been analyzed. The color reflects the color assigned to the weighted label in Connect.</p> <p>If more than one weighted label is associated with the document, the color for the label with the highest weight displays.</p>
	<p>(circles in varying colors with checkmark) Is associated with a weighted label and content has been analyzed. The color reflects the color assigned to the weighted label in Connect.</p> <p>If more than one weighted label is associated with the document, the color for the label with the highest weight displays.</p>
	<p>Only metadata has been processed; not indexed, content not analyzed, stored, or collected. Not associated with a weighted label.</p> <p>You cannot see the content of the file or any attachments or images. When you select a document for which only the metadata has been processed, the document view panel does not display.</p>
	<p>Has been analyzed (metadata indexed, tag values identified) but not associated with a weighted label.</p> <p>If the content has not been stored or collected, you cannot see the content of the file or any attachments or images and the document view panel does not display.</p>
	<p>Content has been stored; metadata and content has been indexed. If not also collected, you will see the content of the file in the document view panel in a simplified text-only view but cannot view any attachments or images.</p>
	<p>Has been collected. You can see the content of the file in the document view panel in a near-native view with basic formatting. You can view images, attachments, and attachment contents. If an email, you can view the message sender and recipients.</p>
	<p>Has had grammar values extracted. Documents include identified words, phrases, or blocks of information managed by grammars defined in Connect.</p> <p>From the data analysis tab () of the document detail panel, you can select a specific grammar to highlight the identified information in the content of the document. For more information, see View entities found in a document.</p>
	<p>Has been sent to a target. From the metadata tab () of the document detail panel, you can view the destination and target the document was sent to.</p>

Icon	Description
	Includes an attachment. If the file has been collected, you can click the attachment name from within the document view panel to view the contents of the attachment.
	Is on hold. Holds are managed at the workspace level.
	Is protected. Protected documents have had security rules applied, such as encryption.

- Added Content Manager Classification and CM Unique Identifier (CM ID) fields to the available options in the filter search panel and search builder when refining a document list in Analyze or Manage and in the tag criteria when creating tags in Connect.
- Added an indicator in the document view panel for documents whose content has been stored but the preview has been truncated because of its large size.

Processing Agent

The following features and enhancements apply specifically to the processing agent.

- Streamlined the processing agent installation and upgrade.
 - Removed reliance on SQL Server and SQL Server Express. This also removes the need to complete additional steps to install databases.
 - Single installation path instead of selecting a "Typical" or "Custom" installation, simplifying the process.
 - Single installation download from the Agent download page in Connect.
- File Analysis Suite now supports processing shared Google Drives within your organization.

When creating a Google Drive repository, the Primary Capture rules now include the following additional options.

- Select **User** to process items from a specific Google Drive account. This option represents the previously available processing in File Analysis Suite.
- Select **Shared Drive** to process items from a Google Drive shared within your organization. You will define a user account that has access to the shared drives. You can further specify whether to scan all drives the user is a member of or to scan only a specific shared drive.
- Added advanced configurations for Exchange to the Agent Admin UI to exclude hidden folders and specific folders from processing.
 - **ProcessHiddenFolders** specifies whether to process hidden Exchange folders. The default is **False** and hidden folders are excluded from processing.
 - **ExcludeFolderList** defines a comma delimited list of Exchange folders to exclude from processing. The default list of folders to exclude includes the following Exchange folders and any subfolders:

- Calendar
- Contacts
- Suggested Contacts
- Drafts
- Deleted Items
- Conflict
- Sync Issues
- Versions
- Purges
- MigratedMessages
- Conversation Action Settings
- Files
- Social Activity Notifications
- PersonMetadata
- Yammer Root
- ExternalContacts

The exclusion list also includes any folder name that starts with "Quick Step Settings".

- Improved efficiency of processing data sets that include large files.
- Added context to processing errors generated outside of File Analysis Suite, such as from SharePoint, to provide more detail.

Connect

The following features and enhancements apply specifically to Connect, and are in addition to changes already mentioned in relation to general enhancements.

- When creating a repository, added the option to extract grammar values as part of the primary capture rules.
- When editing a repository, added the ability to import the security options from a repository template, even if the repository is based on a different template. This action overrides any existing security options and can be refined further as needed.
 - To import the security options from a template, click **LOAD FROM TEMPLATE** on the Security page.

In the resulting dialog, select the desired repository template and then click **OK**. Only the security options from the selected template are loaded to the Security page and override any options previously selected.

- In the repository detail pane, added an indicator if the grammar sets have been updated since the repository was created. A warning icon (⚠) and a **Re-analyze** link display next to the Grammar Sets information. The name of the grammar set applied to the repository that has been updated (is "out of sync") displays in red text.

IMPORTANT: Re-analyzing the repository to update the grammar sets is optional. Changes to the grammar sets for this repository are not automatically updated for the repository. Re-analyzing a repository for which the content has not been stored requires reprocessing the full repository and may take additional time and impact server load.

- On the repositories list page, added a column to show the processing options (analysis) performed for the repository.

Analysis	Description
Metadata Scan	Indicates that only the document metadata has been analyzed and indexed for all documents in this repository
n% SmartScan	Indicates that document metadata has been indexed and document metadata and body content has been analyzed to identify tag values for the defined percentage of documents in this repository.
Full Scan	Indicates that document metadata has been indexed and document metadata and body content has been analyzed to identify tag values for all documents in this repository.
C	Indicates that content has been stored for documents in this repository. The original document itself was not originally stored.
E	Indicates that grammar values have been extracted for documents in this repository. Because the content has not been stored, you can see that grammar values exist, but you cannot see the document content in Analyze or Manage.
CE	Indicates that the content has been stored and grammar values have been extracted for documents in this repository.

- Updated the primary capture options for repositories.

For file system repositories:

- Updated the date range option to "Limit by modified date".

For Exchange repositories:

- When defining a default action for a group, you now select from a list of existing File Analysis Suite groups. Previously, you had to type the name of the desired group.
- Added a date range option, "Limit by sent date".
- Removed the option to force collection for all mail in mailboxes from a defined group.

For SharePoint, SharePoint Online, and Content Manager repositories:

- Added a date range option, "Limit by modified date".
- When viewing repositories organized by source and an existing repository row is selected and you then click **NEW REPOSITORY**, the source and options of the selected repository are pre-populated in the new repository wizard.
- Added the ability to create static tags based on reporting groups and weighted labels.
- Added the ability to cancel processing jobs that are in progress. Added an inline cancel icon (X) to an in-progress activity on the Agent Activity page to let you cancel a job before it completes.

To cancel an in-progress processing run

1. In the agent activity list, click or hover over the row for the desired agent activity and then click the associated cancel icon (X).
2. In the confirmation dialog, click **YES** to confirm the action.

The activity is canceled and the status updates when the cancellation is completed.

- Added an indicator to the Agent Clusters page to identify processing agents that are not up to date.
- Added an indicator to the Agent Clusters page to identify processing agents whose host machine is close to the low disk space threshold.
- Renamed the **UPDATE** button in the repository detail panel to **SCAN**. This change more accurately reflects the action being taken.
- Renamed the "Sensitive Data" grammar category to "Other Sensitive Data". This built-in grammar category includes the Racial Ethnic Origin grammar rule.

Analyze

The following features and enhancements apply specifically to Analyze, and are in addition to changes already mentioned in relation to general enhancements.

- Added file type groupings to the settings options of the **File Types** chart in the Data Volume Analysis dashboard. You can now show file types individually (such as jpeg, bmp, png) or grouped by general file type (such as image) when you click the settings icon.

Manage

The following features and enhancements apply specifically to Manage, and are in addition to changes already mentioned in relation to general enhancements.

- With the separation of file analysis and indexing (storing content), updated the workbook file analysis tasks that can be acted upon. For a given workbook, you can now analyze documents without having to store the content. Previously, you could analyze the documents in a workbook only in conjunction with indexing, performing OCR, or enriching. These actions are taken from the **ACTIVITY** tab of the workbook detail panel.

- Added an **Activity** view to the individual workspace Overview page that shows currently running activities related to the workspace and its related workbooks and documents. If there are currently running activities, the Activity view is the default view for the workspace summary chart; if no activities are currently running, the Timeline view displays.
- Replaced the card view of data subjects, data sources, exports, collection activity, and holds on the individual workspace Overview page with a tabbed view of data sources, data subjects, holds and exports. The updated layout lets you see all objects for each view, giving you an in-depth perspective.

Administration

The following features and enhancements apply specifically to Administration, and are in addition to changes already mentioned in relation to general enhancements.

- None

API updates

The following updates have been made to the File Analysis Suite APIs.

- The `/v1/auth/login-agent-user` method has been removed; use the `/v1/auth/login` method instead.
- The APIs for creating and editing repositories and repository templates no longer support individual grammar rules. The following APIs no longer support the `customGrammars` and `regionCountries` properties.
 - `POST /ajp/v1/job`
 - `PUT /ajp/v1/job/{jobId}`
 - `POST /cc/v1/repository-template`
 - `PUT /cc/v1/repository-template`

To associate a repository with grammar sets, use the `grammarSets` property.

Resolved issues in 3.8.0

The following issues have been resolved in this release of File Analysis Suite.

- Resolved an issue where un-escaped double quotes in the term list values caused errors during document processing.
- Resolved an issue where processing of large files used an increased amount of disk space and resulted in the processing agent stopping.
- When searching by record date, only top-level documents matching the date are now correctly returned.

Known issues in 3.8.0

The following issues are present in this release of File Analysis Suite.

- If upgrading from 3.7.x to 3.8.0, all users of the File Analysis Suite UIs must manually clear their browser cache the first time they access File Analysis Suite after the upgrade. Failure to do so may result in errors in the UIs. The errors are resolved by manually clearing the browser cache.
- When creating Google Drive repositories for Shared Drives, the member account that is used to access the drive must be unique for each repository within a given Google Drive source. This limitation will be addressed in a future release.
- Performing OCR from a workbook is not supported for items processed from an Exchange repository. This will be addressed in a future release.
- To allow OCR for a given repository, the agent cluster to which the repository's source is assigned must have OCR enabled and the repository must have **Store content as text** enabled.

Documentation

File Analysis Suite includes Help Centers that are incorporated into each User Interface and are updated with each software release as appropriate.

To view the File Analysis Suite documentation outside of the product, visit Support & Services on the Micro Focus web site, www.microfocus.com.

To navigate to the File Analysis Suite documentation

1. From the Micro Focus home page click **Support & Services** in the primary menu along the top of the page and then click **Support**.
2. In the Browse Resources section, click **Documentation**.
3. In the product selection list, begin typing **File Analysis Suite**. As you type, products matching what you type display; click **File Analysis Suite (FAS)**.

Select the desired release version.

You must have Adobe® Reader installed to view files in PDF format (*.pdf). To download Adobe Reader, go to the [Adobe](http://adobe.com) web site.

Send documentation feedback

If you have comments about this document, you can [contact the documentation team](#) by email. If an email client is configured on this system, click the link above and an email window opens with the following information in the subject line:

Feedback on Micro Focus File Analysis Suite 3.8.0 Release Notes

Add your feedback to the email and click **Send**.

If no email client is available, copy the information above to a new message in a web mail client, and send your feedback to FAS.docFeedback@microfocus.com.

We appreciate your feedback!