

# Summation Data Loading Guide





# AccessData Legal and Contact Information

Document date: January 13, 2015

## Legal Information

©2015 AccessData Group, Inc. All rights reserved. No part of this publication may be reproduced, photocopied, stored on a retrieval system, or transmitted without the express written consent of the publisher.

AccessData Group, Inc. makes no representations or warranties with respect to the contents or use of this documentation, and specifically disclaims any express or implied warranties of merchantability or fitness for any particular purpose. Further, AccessData Group, Inc. reserves the right to revise this publication and to make changes to its content, at any time, without obligation to notify any person or entity of such revisions or changes.

Further, AccessData Group, Inc. makes no representations or warranties with respect to any software, and specifically disclaims any express or implied warranties of merchantability or fitness for any particular purpose. Further, AccessData Group, Inc. reserves the right to make changes to any and all parts of AccessData software, at any time, without any obligation to notify any person or entity of such changes.

You may not export or re-export this product in violation of any applicable laws or regulations including, without limitation, U.S. export regulations or the laws of the country in which you reside.

AccessData Group, Inc.  
1100 Alma Street  
Menlo Park, California 94025  
USA

[www.accessdata.com](http://www.accessdata.com)

## AccessData Trademarks and Copyright Information

AccessData®	MPE+ Velocitor™
AccessData Certified Examiner® (ACE®)	Password Recovery Toolkit®
AD Summation®	PRTK®
Discovery Cracker®	Registry Viewer®
Distributed Network Attack®	Resolution1™
DNA®	SilentRunner®
Forensic Toolkit® (FTK®)	Summation®
Mobile Phone Examiner Plus®	ThreatBridge™

A trademark symbol (®, ™, etc.) denotes an AccessData Group, Inc. trademark. With few exceptions, and unless otherwise notated, all third-party product names are spelled and capitalized the same way the owner spells and and capitalizes its product name. Third-party trademarks and copyrights are the property of the trademark and copyright holders. AccessData claims no responsibility for the function or performance of third-party products.

Third party acknowledgements:

- FreeBSD ® Copyright 1992-2011. The FreeBSD Project .
- AFF® and AFFLIB® Copyright© 2005, 2006, 2007, 2008 Simson L. Garfinkel and Basis Technology Corp. All rights reserved.
- Copyright © 2005 - 2009 Ayende Rahien

BSD License: Copyright (c) 2009-2011, Andriy Syrov. All rights reserved. Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met: Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer; Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution; Neither the name of Andriy Syrov nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission. THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

WordNet License

This license is available as the file LICENSE in any downloaded version of WordNet.

WordNet 3.0 license: ([Download](#))

WordNet Release 3.0 This software and database is being provided to you, the LICENSEE, by Princeton University under the following license. By obtaining, using and/or copying this software and database, you agree that you have read, understood, and will comply with these terms and conditions.: Permission to use, copy, modify and distribute this software and database and its documentation for any purpose and without fee or royalty is hereby granted, provided that you agree to comply with the following copyright notice and statements, including the disclaimer, and that the same appear on ALL copies of the software, database and documentation, including modifications that you make for internal use or for distribution. WordNet 3.0 Copyright 2006 by Princeton University. All rights reserved. THIS SOFTWARE AND DATABASE IS PROVIDED "AS IS" AND PRINCETON UNIVERSITY MAKES NO REPRESENTATIONS OR WARRANTIES, EXPRESS OR IMPLIED. BY WAY OF EXAMPLE, BUT NOT LIMITATION, PRINCETON UNIVERSITY MAKES NO REPRESENTATIONS OR WARRANTIES OF MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OR THAT THE USE OF THE LICENSED SOFTWARE, DATABASE OR DOCUMENTATION WILL NOT INFRINGE ANY THIRD PARTY PATENTS, COPYRIGHTS, TRADEMARKS OR OTHER RIGHTS. The name of Princeton University or

Princeton may not be used in advertising or publicity pertaining to distribution of the software and/or database. Title to copyright in this software, database and any associated documentation shall at all times remain with Princeton University and LICENSEE agrees to preserve same.

## Documentation Conventions

In AccessData documentation, a number of text variations are used to indicate meanings or actions. For example, a greater-than symbol (>) is used to separate actions within a step. Where an entry must be typed in using the keyboard, the variable data is set apart using *[variable\_data]* format. Steps that require the user to click on a button or icon are indicated by **Bolded text**. This *Italic* font indicates a label or non-interactive item in the user interface.

A trademark symbol (®, ™, etc.) denotes an AccessData Group, Inc. trademark. Unless otherwise notated, all third-party product names are spelled and capitalized the same way the owner spells and capitalizes its product name. Third-party trademarks and copyrights are the property of the trademark and copyright holders. AccessData claims no responsibility for the function or performance of third-party products.

## Registration

The AccessData product registration is done at AccessData after a purchase is made, and before the product is shipped. The licenses are bound to either a USB security device, or a Virtual CmStick, according to your purchase.

## Subscriptions

AccessData provides a one-year licensing subscription with all new product purchases. The subscription allows you to access technical support, and to download and install the latest releases for your licensed products during the active license period.

Following the initial licensing period, a subscription renewal is required annually for continued support and for updating your products. You can renew your subscriptions through your AccessData Sales Representative.

Use License Manager to view your current registration information, to check for product updates and to download the latest product versions, where they are available for download. You can also visit our web site, [www.accessdata.com](http://www.accessdata.com) anytime to find the latest releases of our products.

For more information, see Managing Licenses in your product manual or on the AccessData website.

## AccessData Contact Information

Your AccessData Sales Representative is your main contact with AccessData. Also, listed below are the general AccessData telephone number and mailing address, and telephone numbers for contacting individual departments

## Mailing Address and General Phone Numbers

You can contact AccessData in the following ways:

### AccessData Mailing Address, Hours, and Department Phone Numbers

Corporate Headquarters:	AccessData Group, Inc. 1100 Alma Street Menlo Park, California 94025 USAU.S.A.  <i>Voice: 801.377.5410; Fax: 801.377.5426</i>
General Corporate Hours:	Monday through Friday, 8:00 AM – 5:00 PM (MST) AccessData is closed on US Federal Holidays
State and Local Law Enforcement Sales:	<i>Voice: 800.574.5199, option 1; Fax: 801.765.4370</i> <i>Email: Sales@AccessData.com</i>
Federal Sales:	<i>Voice: 800.574.5199, option 2; Fax: 801.765.4370</i> <i>Email: Sales@AccessData.com</i>
Corporate Sales:	<i>Voice: 801.377.5410, option 3; Fax: 801.765.4370</i> <i>Email: Sales@AccessData.com</i>
Training:	<i>Voice: 801.377.5410, option 6; Fax: 801.765.4370</i> <i>Email: Training@AccessData.com</i>
Accounting:	<i>Voice: 801.377.5410, option 4</i>

## Technical Support

Free technical support is available on all currently licensed AccessData solutions.

You can contact AccessData Customer and Technical Support in the following ways:

### AD Customer & Technical Support Contact Information

<b>AD SUMMATION and AD EDISCOVERY</b>	Americas/Asia-Pacific: 800.786.8369 (North America) 801.377.5410, option 5 Email: <a href="mailto:legalsupport@accessdata.com">legalsupport@accessdata.com</a>
<b>AD IBLAZE and ENTERPRISE:</b>	Americas/Asia-Pacific: 800.786.2778 (North America) 801.377.5410, option 5 Email: <a href="mailto:support@summation.com">support@summation.com</a>
<b>All other AD SOLUTIONS</b>	Americas/Asia-Pacific: 800.658.5199 (North America) 801.377.5410, option 5 Email: <a href="mailto:support@accessdata.com">support@accessdata.com</a>
<b>AD INTERNATIONAL SUPPORT</b>	Europe/Middle East/Africa: +44 (0) 207 010 7817 (United Kingdom) Email: <a href="mailto:emeasupport@accessdata.com">emeasupport@accessdata.com</a>

## AD Customer & Technical Support Contact Information (Continued)

<i>Hours of Support:</i>	Americas/Asia-Pacific: Monday through Friday, 6:00 AM– 6:00 PM (PST), except corporate holidays. Europe/Middle East/Africa: Monday through Friday, 8:00 AM– 5:00 PM (UK-London) except corporate holidays.
<i>Web Site:</i>	<a href="http://www.accessdata.com/support/technical-customer-support">http://www.accessdata.com/support/technical-customer-support</a>
	The Support website allows access to Discussion Forums, Downloads, Previous Releases, our Knowledge base, a way to submit and track your “trouble tickets”, and in-depth contact information.

## Documentation

Please email AccessData regarding any typos, inaccuracies, or other problems you find with the documentation: [documentation@accessdata.com](mailto:documentation@accessdata.com)

## Professional Services

The AccessData Professional Services staff comes with a varied and extensive background in digital investigations including law enforcement, counter-intelligence, and corporate security. Their collective experience in working with both government and commercial entities, as well as in providing expert testimony, enables them to provide a full range of computer forensic and eDiscovery services.

At this time, Professional Services provides support for sales, installation, training, and utilization of FTK, FTK Pro, Enterprise, eDiscovery, Lab and the entire Resolution One platform. They can help you resolve any questions or problems you may have regarding these solutions.

## Contact Information for Professional Services

Contact AccessData Professional Services in the following ways:

### AccessData Professional Services Contact Information

Contact Method	Number or Address
<i>Phone</i>	North America Toll Free: 800-489-5199, option 7
	International: +1.801.377.5410, option 7
<i>Email</i>	<a href="mailto:services@accessdata.com">services@accessdata.com</a>

# Contents

- AccessData Legal and Contact Information . . . . . 3**
- Contents . . . . . 8**
- Chapter 1: Introduction to Loading Data . . . . . 10**
  - Importing Data . . . . . 10
- Chapter 2: Using the Evidence Wizard . . . . . 11**
  - Using the Evidence Wizard . . . . . 11
    - About Associating People with Evidence . . . . . 13
    - Using the CSV Import Method for Importing Evidence . . . . . 13
    - Using the Immediate Children Method for Importing . . . . . 15
  - Adding Evidence to a Project Using the Evidence Wizard . . . . . 17
    - Evidence Time Zone Setting . . . . . 19
- Chapter 3: Importing Evidence . . . . . 20**
  - About Importing Evidence Using Import . . . . . 20
    - About Mapping Field Values . . . . . 20
  - Importing Evidence into a Project . . . . . 21
- Chapter 4: Data Loading Requirements . . . . . 23**
  - Document Groups . . . . . 23
    - Images . . . . . 23
    - Full-Text or OCR . . . . . 24
    - DII Load File Format for Image/OCR . . . . . 24
  - Email & eDocs . . . . . 26
  - Coding . . . . . 28
  - Related Documents . . . . . 31
  - Transcripts and Exhibits . . . . . 32
    - Transcripts . . . . . 32
    - Exhibits . . . . . 33
  - Work Product . . . . . 34
  - Sample DII Files . . . . . 35
    - eDoc DII Load Files . . . . . 35
    - eMail DII Load Files . . . . . 36



DII Tokens . . . . .	39
<b>Chapter 5: Analyzing Document Content . . . . .</b>	<b>43</b>
Using Cluster Analysis . . . . .	43
About Cluster Analysis . . . . .	43
Filtering Documents by Cluster Topic . . . . .	44
Using Entity Extraction . . . . .	46
About Entity Extraction . . . . .	46
Enabling Entity Extraction . . . . .	48
Viewing Entity Extraction Data . . . . .	48
<b>Chapter 6: Editing Evidence . . . . .</b>	<b>49</b>
Editing Evidence Items in the Evidence Tab . . . . .	49
Evidence Tab . . . . .	50

# Chapter 1

## Introduction to Loading Data

---

### Importing Data

This document will help you import data into your project. You create projects in order to organize data. Data can be added to projects in the forms of native files, such as DOC, PDF, XLS, PPT, and PST files, or as evidence images, such as AD1, E01, and OFF files.

To manage evidence, administrators, and users with the Create/Edit Projects permission, can do the following:

- Add evidence items to a project
- View properties about evidence items in a project
- Edit properties about evidence items in a project
- Associate people to evidence items in a project

---

**Note:** You will normally want to have people created and selected before you process evidence.

---

See [About Associating People with Evidence](#) on page 13.

See the following chapters for more information:

#### To import data

1. Log in as a project manager.
2. Click the **Add Data** button next to the project in the *Project List* panel.
3. In the *Add Data* dialog, select one of the methods by which you want to import data. The following methods are available:
  - Evidence (wizard): See [Using the Evidence Wizard](#) on page 11.
  - Job (Resolution1 applications): See [About Jobs](#) on page 377.
  - Import: See [Importing Evidence](#) on page 20.
  - Cluster Analysis: See [Using Cluster Analysis](#) on page 43.

# Chapter 2

## Using the Evidence Wizard

---

### Using the Evidence Wizard

When you add evidence to a project, you can use the *Add Evidence Wizard* to specify the data that you want to add. You specify to add either parent folders or individual files.

---

**Note:** If you activated Cluster Analysis as a processing option when you created the project, cluster analysis will automatically run after processing data.

---

You select sets of data that are called “evidence items.” It is useful to organize data into evidence items because each evidence item can be associated with a unique person.

For example, you could have a parent folder with a set of subfolders.

```
\\10.10.3.39\EvidenceSource\  
\\10.10.3.39\EvidenceSource\John Smith  
\\10.10.3.39\EvidenceSource\Bobby Jones  
\\10.10.3.39\EvidenceSource\Samuel Johnson  
\\10.10.3.39\EvidenceSource\Edward Peterson  
\\10.10.3.39\EvidenceSource\Jeremy Lane
```

You could import the parent `\\10.10.3.39\EvidenceSource\` as one evidence item. If you associated a person to it, all files under the parent would have the same person.

On the other hand, you could have each subfolder be its own evidence item, and then you could associate a unique person to each item.

An evidence item can either be a folder or a single file. If the item is a folder, it can have other subfolders, but they would be included in the item.

When you use the Evidence Wizard to import evidence, you have options that will determine how the evidence is organized in evidence items.

When you add evidence, you select from the following types of files.

### Evidence File Types

File Type	Description
Evidence Images	You can add AD1, E01, or AFF evidence image files.
Native Files	You can add native files, such as PDF, JPG, DOC PPT, PST, XLSX, and so on.

When you add evidence, you also select one of the following import methods.

### Import Methods

Method	Description
CSV Import	<p>This method lets you create and import a CSV file that lists multiple paths of evidence and optionally automatically creates people and associates each evidence item with a person.</p> <p>Like the other methods, you specify whether the parent folder contains native files or image files.</p> <p>See <a href="#">Using the CSV Import Method for Importing Evidence</a> on page 13.</p> <p>This is similar to adding people by importing a file.</p> <p>See the Project Manager Guide for more information on adding people by importing a file.</p>
Immediate Children	<p>This method takes the immediate subfolders of the specified path and imports each of those subfolders' content as a unique evidence item. You can automatically create a person based on the child folder's name (if the child folder has a first and last name separated by a space) and have it associated with the data in the subfolder.</p> <p>See <a href="#">Using the Immediate Children Method for Importing</a> on page 15.</p> <p>Like the other methods, you specify if the parent folder contains native files or image files.</p>
Folder Import	<p>This method lets you select a parent folder and all data in that folder will be imported. You specify that the folder contains either native files (JPG, PPT) or image files (AD1, E01, AFF).</p> <p>A parent folder can have both subfolders and files.</p> <p>Using this method, each parent folder that you import is its own evidence item and can be associated with one person.</p> <p>For example, if a parent folder had several AD1 files, all data from each AD1 file can have one associated person. Likewise, if a parent folder has several native files, all of the contents of that parent folder can have one associated person.</p>
Individual File(s)	<p>This method lets you select individual files to import. You specify that these individual files are either native files (JPG, PPT) or image files (AD1, E01, AFF).</p> <p>Using this method, each individual file that you import is its own evidence item and can be associated with a person.</p> <p>For example, all data from an AD1 file can have an associated person. Likewise, each PDF, or JPG can have its own associated person.</p>

---

**Note:** The source network share permissions are defined by the administrator credentials.

---

## About Associating People with Evidence

When you add evidence items to a project, you can specify people, or custodians, that are associated with the evidence. These custodians are listed as People on the *Data Sources* tab.

In the *Add Evidence Wizard*, after specifying the evidence that you want to add, you can then associate that evidence to a person. You can select an existing person or create a new person.

**Important:** If you want to select an existing Person, that person must already be associated to the project. You can either do that for the project on the *Home* page > *People* tab, or you can do it on the *Data Sources* page > *People* tab.

You can create people in the following ways:

- On the *Data Sources* tab before creating a project.  
See the *Data Sources* chapter.
- When adding evidence to a project within the *Add Evidence Wizard*.  
See [Adding Evidence to a Project Using the Evidence Wizard](#) on page 17.
- On the *People* tab on the *Home* page for a project that has already been created.

## About Creating People when Adding Evidence Items

In the *Add Evidence Wizard*, you can create people as you add evidence. There are three ways you can create people while adding evidence to a project:

- Using a CSV Evidence Import.  
See [Using the CSV Import Method for Importing Evidence](#) on page 13.
- Importing immediate children.  
See [Using the Immediate Children Method for Importing](#) on page 15.
- Adding a person in the *Add Evidence Wizard*.  
You can select a person from the drop-down in the wizard or enter a new person name.  
See the Project Manager Guide for more information on creating people.

## Using the CSV Import Method for Importing Evidence

When specifying evidence to import in the *Add Evidence Wizard*, you can use one of two general options:

- Manually browse to all evidence folders and files.
- Specify folders, files, and people in a CSV file.  
There are several benefits of using a CSV file:
  - You can more easily and accurately plan for all of the evidence items to be included in a project by including all sources of evidence in a single file.
  - You can more easily and accurately make sure that you add all of the evidence items to be included in a project.
  - If you have multiple folders or files, it is quicker to enter all of the paths in the CSV file than to browse to each one in the wizard.
  - If you are going to specify people, you can specify the person for each evidence item. This will automatically add those people to the system rather than having to manually add each person.

When using a CSV, each path or file that you specify will be its own evidence item. The benefit of having multiple items is that each item can have its own associated person. This is in contrast with the Folder Import method, where only one person can be associated with all data under that folder.

Specifying people is not required. However, if you do not specify people, when the data is imported, no people are created or associated with evidence items. Person data will not be usable in Project Review.

See the Project Manager Guide for information on associating a person to an evidence item.

If you do specify people in the CSV file, you use the first column to specify the person's name and the second column for the path.

If you do not specify people, you will only use one column for paths. When you load the CSV file in the *Add Evidence Wizard*, you will specify that the first column does not contain people's names. That way, the wizard imports the first column as paths and not people.

If you do specify people, they can be in one of two formats:

- A single name or text string with no spaces  
For example, JSmith or John\_Smith
- First and last name separated by a space  
For example, John Smith or Bill Jones

In the CSV file, you can optionally have column headers. You will specify in the wizard whether it should use the first row as data or ignore the first row as headers.

## CSV Example 1

This example includes headers and people.

In the wizard, you select both **First row contains headers** and **First column contains people names** check boxes.

When the data is imported, the people are created and associated to the project and the appropriate evidence item.

### People, Paths

JSmith,\\10.10.3.39\EvidenceSource\JSmith

JSmith,\\10.10.3.39\EvidenceSource\Sales\Projections.xlsx

Bill Jones,\\10.10.3.39\EvidenceSource\BJones

Sarah Johnson,\\10.10.3.39\EvidenceSource\SJohnson

Evan\_Peterson,\\10.10.3.39\EvidenceSource\EPeterson

Evan\_Peterson,\\10.10.3.39\EvidenceSource\HR

Jill Lane,\\10.10.3.39\EvidenceSource\JLane

Jill Lane,\\10.10.3.39\EvidenceSource\Marketing

This will import any individual files that are specified as well as all of the files (and additional subfolders) under a listed subfolder.

You may normally use the same naming convention for people. This example shows different conventions simply as examples.

## CSV Example 2

This example does not include headers or people.

In the wizard, you clear both **First row contains headers** and **First column contains people names** check boxes.

When the data is imported, no people are created or associated with evidence items.

```
\\10.10.3.39\EvidenceSource\JSmith
\\10.10.3.39\EvidenceSource\Sales\Projections.xlsx
\\10.10.3.39\EvidenceSource\BJones
\\10.10.3.39\EvidenceSource\SJohnson
\\10.10.3.39\EvidenceSource\EPeterson
\\10.10.3.39\EvidenceSource\HR
\\10.10.3.39\EvidenceSource\JLane
\\10.10.3.39\EvidenceSource\Marketing
```

## *Using the Immediate Children Method for Importing*

If you have a parent folder that has children subfolders, when importing it through the *Add Evidence Wizard*, you can use one of three methods:

- Folder Import
- Immediate Children
- CSV Import

See [Using the CSV Import Method for Importing Evidence](#) on page 13.

When using the Immediate Children method, each child subfolder of the parent folder will be its own evidence item. The benefit of having multiple evidence items is that each item can have its own associated person. This is in contrast with the Folder Import method, where all data under that folder is a single evidence item with only one possible person associated with it.

Specifying people is not required. However, if you do not specify people, when the data is imported, no people are created or associated with evidence items. Person data will not be usable in Project Review.

See the Project Manager Guide for more information on associating a person to evidence.

When you select a parent folder in the *Add Evidence Wizard*, you select whether or not to specify people.

If you do specify people, the names of people are based on the name of the child folders.

Imported names of people can be imported in one of two formats:

- A single name or text string with no spaces  
For example, JSmith or John\_Smith

- First and last name separated by a space

For example, John Smith or Bill Jones

For example, suppose a parent folder had four subfolders, each containing data from a different user. Using the Immediate Children method, each subfolder would be imported as a unique evidence item and the subfolder name could be the associated person.

\Userdata\ (parent folder that is selected)

\Userdata\INewstead (unique evidence item with INewstead as a person)

\Userdata\KHetfield (unique evidence item with KHetfield as a person)

\Userdata\James Ulrich (unique evidence item with James Ulrich as a person)

\Userdata\Jill\_Hammett (unique evidence item with Jill\_Hammett as a person)

---

**Note:** In the Add Evidence Wizard, you can manually rename the people if needed.

---

The child folder may be a parent folder itself, but anything under it would be one evidence item.

This method is similar to the CSV Import method in that it automatically creates people and associates them to evidence items. The difference is that when using this method, everything is configured in the wizard and not in an external CSV file.



# Adding Evidence to a Project Using the Evidence Wizard

You can import evidence for projects for which you have permissions.

When you add evidence, it is processed so that it can be reviewed in Project Review.

Some data cannot be changed after it has been processed. Before adding and processing evidence, do the following:



- Configure the Processing Options the way you want them.  
See the Admin Guide for more information on default processing options.
- Plan whether or not you want to specify people.  
See the Project Manager Guide for more information on associating a person to evidence.
- Unless you are importing people as part of the evidence, you must have people already associated with the project.  
See the Project Manager Guide for more information on creating people.

---

**Note:** Deduplication can only occur with evidence brought into the application using evidence processing. Deduplication cannot be used on data that is imported.

---

## To import evidence for a project

1. In the project list, click  (add evidence) in the project that you want to add evidence to.
2. Select **Evidence**.
3. In the *Add Evidence Wizard*, select the *Evidence Data Type* and the *Import Method*.  
See [Using the Evidence Wizard](#) on page 11.
4. Click **Next**.
5. Select the evidence folder or files that you want to import.  
This screen will differ depending on the *Import Method* that you selected.
  - 5a. If you are using the *CSV Import* method, do the following:
    - If the CSV file uses the first row as headers rather than folder paths, select the **First row contains headers** check box, otherwise, clear it.
    - If the CSV file uses the first column to specify people, select the **First column contains people's names** check box, otherwise, clear it.  
See [Using the CSV Import Method for Importing Evidence](#) on page 13.
      - Click **Browse**.
      - Browse to the CSV file and click **OK**.  
The CSV data is imported based on the check box settings.  
Confirm that the people and evidence paths are correct.  
You can edit any information in the list.  
If the wizard can't validate something in the CSV, it will highlight the item in red and place a red box around the problem value.  
If a new person will be created, it will be designated by .
  - 5b. If you are using the *Immediate Children* method, do the following:
    - If you want to automatically create people, select **Sub folders are people's names**, otherwise, clear it.  
See [Using the Immediate Children Method for Importing](#) on page 15.
    - Click **Browse**.
    - Enter the IP address of the server where the evidence files are located and click **Go**.

For example, 10.10.2.29

- Browse to the parent folder and click **Select**.


Each child folder is listed as a unique evidence item.

If you selected to create people, they are listed as well.

Confirm that the people and evidence paths are correct.

You can edit any information in the list.

If the wizard can't validate something, it will highlight the item in red and place a red box around the problem value.

If a new person will be created, it will be designated by .

5c. If you are using the Folder Input or Individual Files method, do the following:

- Click **Browse**.
- Enter the IP address of the server where the evidence files are located and click **Go**.

For example, 10.10.2.29


- Expand the folders in the left pane to browse the server.
- In the right pane highlight the parent folder or file and click **Select**.

If you are selecting files, you can use Ctrl-click or Shift-click to select multiple files in one folder.

The folder or file is listed as a unique evidence item.

6. If you want to specify a person to be associated with this evidence, select one from the *Person Name* drop-down list or type in a new person name to be added.

See [About Associating People with Evidence](#) on page 13.

If you enter a new person that will be created, it will be designated by .

You can also edit a person's name if it was imported.

7. Specify a Timezone.

From the Timezone drop-down list, select a time zone.


See [Evidence Time Zone Setting](#) on page 19.

8. (Optional) Enter a *Description*.

This is used as a short description that is displayed with each item in the *Evidence* tab.

For example, "Imported from Filename.csv" or "Children of *path*".

This can be added or edited later in the *Evidence* tab.


9. (Optional) If you need to delete an evidence item, click the  for the item.

10. Click **Next**.

11. In the *Evidence to be Added and Processed* screen, you can view the evidence that you selected so far. From this screen, you can perform one of the following actions:

- *Add More*: Click this button to return to the *Add Evidence* screen.
- *Add Evidence and Process*: Click this button to add and process the evidence listed.

When you are done, you are returned to the project list. After a few moments, the job will start and the project status should change to *Processing*.

12. If you need to manually update the list or status, click  **Refresh**.

13. When the evidence import is completed, you can view the evidence items in the *Evidence* and *People* labels.

## *Evidence Time Zone Setting*

Because of worldwide differences in the time zone implementation and Daylight Savings Time, you select a time zone when you add an evidence item to a project.

In a FAT volume, times are stored in a localized format according to the time zone information the operating system has at the time the entry is stored. For example, if the actual date is Jan 1, 2005, and the time is 1:00 p.m. on the East Coast, the time would be stored as 1:00 p.m. with no adjustment made for relevance to Greenwich Mean Time (GMT). Anytime this file time is displayed, it is not adjusted for time zone offset prior to being displayed.

If the same file is then stored on an NTFS volume, an adjustment is made to GMT according to the settings of the computer storing the file. For example, if the computer has a time zone setting of -5:00 from GMT, this file time is advanced 5 hours to 6:00 p.m. GMT and stored in this format. Anytime this file time is displayed, it is adjusted for time zone offset prior to being displayed.

For proper time analysis to occur, it is necessary to bring all times and their corresponding dates into a single format for comparison. When processing a FAT volume, you select a time zone and indicate whether or not Daylight Savings Time was being used. If the volume (such as removable media) does not contain time zone information, select a time zone based on other associated computers. If they do not exist, then select your local time zone settings.

With this information, the system creates the project database and converts all FAT times to GMT and stores them as such. Adjustments are made for each entry depending on historical use data and Daylight Savings Time. Every NTFS volume will have the times stored with no adjustment made.

With all times stored in a comparable manner, you need only set your local machine to the same time and date settings as the project evidence to correctly display all dates and times.

# Chapter 3

## Importing Evidence

---

### About Importing Evidence Using Import

As an Administrator or Project Manager with the Create/Edit Projects permissions, you can import evidence for a project.

You import evidence by using a load file, which allows you to import metadata and physical files, such as native, image, and/or text files that were obtained from another source, such as a scanning program or another processing program. You can import the following types of load files:

- Summation DII - A proprietary file type from Summation. See [Data Loading Requirements](#) on page 23.
- Generic - A delimited file type, such as a CSV file.
- Concordance/Relativity - A delimited DAT file type that has established guidelines as to what delimiter should be used in the fields. This file should have a corresponding LFP or OPT image file to import.

Transcripts and exhibits are uploaded from *Project Review* and not from the *Import* dialog. See the Project Manager Guide for more information on how to upload transcripts and exhibits.

### *About Mapping Field Values*

When importing you must specify which import file fields should be mapped to database fields. Mapping the fields will put the correct information about the document in the correct columns in the *Project Review*.

After clicking **Map Fields**, a process runs that checks the imported load file against existing project fields. Most of the import file fields will automatically be mapped for you. Any fields that could not be automatically mapped are flagged as needing to be mapped.

---

**Note:** If you need custom fields, you must create them in the *Custom Fields* tab on the *Home* page before you can map to those fields during the import. If the custom names are the same, they will be automatically mapped as well.

---

Any errors that have to be corrected before the file can be imported are reported at this time.

When importing a CSV or DAT load file that is missing the unique identifier used to map to the DocID file, an error message will be displayed.


Notes:

- If a record contains the same values for the DocID as the ParentID, an error is logged in the log file and the record is not imported. This allows you to correct the problem record and make sure all records in the family are included in the loadfile correctly.

- In review, the AttachmentCount value is displayed under the EmailDirectAttachCount column.
- The Importance value is not imported as a text string but is converted and stored in the database as an integer representing a value of either *Low*, *Normal*, *High*, or blank. These values are case sensitive and in the import file must be an exact match.
- The Sensitivity value is not imported as a text string but is converted and stored in the database as an integer representing a value of either *Confidential*, *Private*, *Personal*, or *Normal*. These values are case sensitive and in the import file must be an exact match.
- The Language value is not imported as a text string but is converted and stored in the database as an integer representing one of 67 languages.
- Body text that is mapped to the *Body* database field is imported as an email body stream and is viewable in the Natural viewer. When importing all file types, the import *Body* field is now automatically mapped to the *Body* database field.

## Importing Evidence into a Project

### To import evidence into a project

1. Log into the application as an Administrator or a user with Create/Edit Project rights.
2. In the *Project List* panel, click **Add Evidence**  next to the project.
3. Click **Import**.
4. In the *Import* dialog, select the file type (EDII, Concordance/Relativity, or Generic ).
  - 4a. Enter the location of the file or **Browse** to the file's location.
  - 4b. (optional - Available only for Concordance/Relativity) Select the *Image Type* and enter the location of the file, or **Browse** to the file's location. You can choose from the following file options:
    - OPT - Concordance file type that contains preferences and option settings associated with the files.
    - LFP - Ipro file type that contains load images and related information.
5. Perform field mapping.
 

Most fields will be automatically mapped. If some fields need to be manually mapped, you will see an orange triangle.

  - 5a. Click **Map Fields** to map the fields from the load file to the appropriate fields.  
See [About Mapping Field Values](#) on page 20.
  - 5b. To skip any items that do not map, select **Skip Unmapped**.
  - 5c. To return the fields back to their original state, click **Reset**.

---

**Note:** Every time you click the *Map Fields* button, the fields are reset to their original state.

---
6. Select the *Import Destination*.
  - 6a. Choose from one of the following:
    - **Existing Document Group:** This option adds the documents to an existing document group. Select the group from the drop-down menu.  
See the Project Manager Guide for more information on managing document groups.
    - **Create New Document Group:** This option adds the documents to a new document group. Enter the name of the group in the field next to this radio button.

7. Select the *Import Options* for the file. These options will differ depending on whether you select DII, Concordance/Relativity, or Generic.
  - DII Options:
    - **Page Count Follows Doc ID:** Select this option if your DII file has an @T value that contains both a Doc ID and a page count.
    - **Import OCR/Full Text:** Select this option to import OCR or Full Text documents for each record.
    - **Import Native Documents/Images:** Select this option to import Native Documents and Images for each record.
  - Concordance/Relativity, or Generic Options:
    - **First Row Contains Field Names:** Select this option if the file being imported contains a row header.
    - **Field, Quote, and Multi-Entry Separators:** From the pull-down menu, select the symbols for the different separators that the file being imported contains. Each separator value must match the imported file separators exactly or the field being imported for each record is not populated correctly.
    - **Return Placeholder:** From the pull-down menu, select the same value contained in the file being imported as a replacement value for carriage return and line feed characters. Each return placeholder value must match the imported file separators.
8. Configure the **Date Options**.
  - Select the date format from the **Date Format** drop-down menu.

This option allows you to configure what date format appears in the load file system, allowing the system to properly parse the date to store in the database. All dates are stored in the database in a yyyy-mm-dd hh:mm:ss format.
  - Select the *Load File Time Zone*.

Choose the time zone that the load file was created in so the date and time values can be converted to a normalized UTC value in the database.  
See [Normalized Time Zones](#) on page 157.
9. Select the Record Handling Options.
  - **New Record:**
    - **Add:** Select to add new records.
    - **Skip:** Select to ignore new records.
  - **Existing Record:**
    - **Update:** Select to update duplicate records with the record being imported.
    - **Overwrite:** Select to overwrite any duplicate records with the record being imported.
    - **Skip:** Select to skip any duplicate records.
10. **Validation:** This option verifies that:
  - The path information within the load file is correct
  - The records contain the correct fields. For example, the system verifies that the delimiters and fields in a Generic or Concordance/Relativity file are correct.
  - You have all of the physical files (that is, Native, Image, and Text) that are listed in the load file.
11. (optional) **Drop DB Indexes.** Database indexes improve performance, but slow processing when inserting data. If this option is checked, all of the data reindexes every time more data is loaded. Only select this option if you want to load a large amount of data quickly before data is reviewed.
12. Click **Start**.

# Chapter 4

## Data Loading Requirements

---

This chapter describes the data loading requirements of Resolution1 Platform and Summation and contains the following sections:

- [Document Groups](#) (page 23)
- [Email & eDocs](#) (page 26)
- [Coding](#) (page 28)
- [Related Documents](#) (page 31)
- [Transcripts and Exhibits](#) (page 32)
- [Work Product](#) (page 34)
- [Sample DII Files](#) (page 35)
- [DII Tokens](#) (page 39)

## Document Groups

---

**Note:** You can import and display Latin and non-Latin Unicode characters. While the application supports the display of fielded data in either Latin or non-Latin Unicode characters, the modification of fielded data is supported only in Latin Unicode characters.

---

**Note:** The display of non-Latin Unicode characters does not apply to transcript filenames, since transcript deponents are defined by project users, or work product filenames, which are not displayed in the application.

---

## *Images*

The following describes the required and recommended formats for images.

### Required

- A DII load file is required to load image documents. 0
- Group IV TIFFS: single or multi-page, black and white (or color), compressed images, no DPI minimum.
- Single page JPEGs for color images.

## Full-Text or OCR

The following describes the required and recommended formats for full-text or OCR.

### Required

- If submitting document level OCR, page breaks should be included between each page of text in the document text file.  
Failure to insert page breaks will result in a one page text file for a multi-page document. The ASCII character 12 (decimal) is used for the “Page Break” character. All instances of the character 12 as page breaks will be interpreted.
- Document level OCR or page level OCR.
- All OCR files should be in ANSI or Unicode text file format, with a \*.txt extension.
- A DII load file. Loading Control List (.LST) files are not supported.

### Recommended

- OCR text files should be stored in the same directories as image files.
- Page level OCR is recommended to ensure proper page breaks.

## DII Load File Format for Image/OCR

---

**Note:** When selecting the **Copy ESI** option, the DII and source files *must* reside in a location accessible by the IEP server; otherwise, import jobs will fail during the **Check File** process.

---

The following describes the required format for a DII load file to load images and OCR.

### Required

- A blank line after each document summary.
- **@T** to identify each document summary.
- **@T** should equal the beginning Bates number.
- If OCR is included, then use **@FULLTEXT** at the beginning of the DII file (**@FULLTEXT DOC** or **@FULLTEXT PAGE**).
- If **@FULLTEXT DOC** is included, OCR text files are assumed to be in the **Image** folder location with the same name as the first image (TIFF or JPG) file.
- If **@FULLTEXT PAGE** is included, OCR text files are assumed to be in the **Image** folder location with the same name as the image files (each page should have its own txt file).
- If **@O** token is used, **@FULLTEXT** token is not required.
- If Fulltext is located in another directory other than images, use **@FULLTEXTDIR** followed by the directory path.



- The page count identifier on the @T line can be interpreted ONLY if it is denoted with a space character.

For example:

```
@FULLTEXT PAGE
@T AAA0000001 2
@D @\IMAGES\01\
AAA0000001.TIF
AAA0000002.TIF
@T AAA0000003 1
@D @\IMAGES\02\
AAA0000003.TIF
```

Import controls the **Page Count Follows DocID** option. If this option is deselected, the page count identifier on the @T line would not be recognized.

## Recommended

- DII load file names should mirror that of the respective volume (for easy association and identification).
- @T values (that is, the BegBates) and EndBates should include no more than 50 characters. Non-alphabetical and non-numerical characters should be avoided.

# Email & eDocs

You can host email, email attachments, and eDocs (electronic documents in native format) for review and attorney coding, as well as associated full-text and metadata. It is also possible to include an imaged version (in TIFF format) of the file at loading. A DII load file is required in order to load e-mail and electronic documents.

---

**Note:** You can import and display of Latin and non-Latin Unicode characters. While the application supports the display of fielded data in either Latin or non-Latin Unicode characters, the modification of fielded data is supported only in Latin Unicode characters.

---

---

**Note:** The display of non-Latin Unicode characters does not apply to transcript filenames, since transcript deponents are defined by users, or work product filenames, which are not displayed.

---

## General Requirements

The following describes the required and recommended formats for DII files that are used to load email, email attachments, and eDocs.

A DII load file with a \*.dii file extension, using only the tokens, is listed in [DII Tokens](#) (page 39).

- **@T** to identify each email, email attachment, or eDoc record.
- **@T** is the first line for each summary.
- **@T** equals the unique **Docid** for each email, email attachment, or eDoc record. There should be only one **@T** per record.
- A blank line between document records.
- **@EATTACH** token is required for email attachments and **@EDOC** for eDocs. These tokens contain a relative path to the native file.
- **@MEDIA** is required for email data with a value of **eMail** or **Attachment**. For eDocs, the **@MEDIA** value must be **eDoc**.
- **@EATTACH** is required when **@MEDIA** has a value of **Attachment** and is not required when **@MEDIA** has a value of **eMail**.
- To maintain the parent/child relationship between an e-mail and its attachments (family relationships for eDocs), the **@PARENTID** and **@ATTACH** tokens are used.
- To include images along with the native file delivery, use the **@D @I** tokens at the end of the record.
- **@O** token is extended to support loading FullText into eDoc and eMails also.  
If record has both **@O** and **@EDOC/@EATTACH** tokens, FullText is loaded from the file specified by the **@O** token. If **@O** token does NOT exist for the record, FullText is extracted from the file specified by the **@EDOC/@EATTACH** token.
- **@AUTHOR** and **@ITEMTYPE** tokens are NOT supported.

## Recommended

- **@T** values (Begbates/Docid) should include no more than 50 characters. Non-alphabetical and non-numerical characters should be avoided.
- Specify parent-child relationship in the DII file based on the following rule:

- In the DII file, email attachments should immediately follow the parent record, that is:

@T ABC000123

@MEDIA eMail

@EMAIL-BODY

Please reply with a copy of the completed report.

Thanks for your input.

Beth

@EMAIL-END

@ATTACH ABC000124; ABC000125

@T ABC000124

@MEDIA Attachment

@EATTACH \Native\ABC000124.doc

@PARENTID ABC000123

@T ABC000125

@MEDIA Attachment

@EATTACH \Native\ABC000125.doc

@PARENTID ABC000123

# Coding

The following describes the required and recommended formats for coded data.

## Recommended

- Coded data should be submitted in a delimited text file, with a \*.txt extension.
- Use the following default delimiter characters:

Field Separator	
Multi-entry Separator	;
Return Placeholder	~
Quote Separator	^

Users can, however, specify any custom character in the Import user interface for any of the separators above.

- The standard comma and quote characters (',' ""') are accepted. When these characters are present within coded data, different characters must be used as separators.

For instance,

DOCID|SUMMARY|AUTHOR

^DOJ000001^|^Test "Summary1"^^Smith, John^

In the above file,

Field Separator |

Quote Separator ^

- Date field values should have any of the following formats. The date 16<sup>th</sup> August 2009 can be represented in the load file as:
  - 08/16/2009
  - 16/08/2009
  - 20090816

In addition, fuzzy dates are also supported. Currently only **DOCDATE** field supports fuzzy dates.

- If a day is fuzzy, then replace dd with 00.
- If a month is fuzzy, then replace mm with 00.
- If a year is fuzzy, replace yyyy with 0000.

<b>Format</b>	<b>Example</b>
mm/dd/yyyy	00/16/2009 (month fuzzy)
	08/00/2009 (day fuzzy)
	08/16/0000 (year fuzzy)
	00/16/0000 (month and year fuzzy)
	08/00/0000 (day and year fuzzy)
	00/00/2009 (month and day fuzzy)
	00/00/0000 (all fuzzy)
	08/16/2009 (no fuzzy)
yyyymmdd	00000816 (year fuzzy)
	20090016 (month fuzzy)
	20090800 (day fuzzy)
	00000016 (year and month fuzzy)
	00000800 (year and day fuzzy)
	20090000 (month and day fuzzy)
	00000000 (all fuzzy)
	20090816 (no fuzzy)
dd/mm/yyyy	00/08/2009 (day fuzzy)
	16/00/2009 (month fuzzy)
	16/08/0000 (year fuzzy)
	16/00/0000 (month and year fuzzy)
	00/08/0000 (day and year fuzzy)
	00/00/2009 (day and month fuzzy)
	00/00/0000 (all fuzzy)
	16/08/2009 – no fuzzy

- Time values should have any of the following formats. The time 1:27 PM can be represented in the load file as:
  - 1:27 PM
  - 01:27 PM
  - 1:27:00 PM
  - 01:27:00 PM
  - 13:27
  - 13:27:00

Time values for standard tokens @TIMESENT/@TIMERCVD/@TIMESAVED/TIMECREATED will not be loaded for a document unless accompanied by a corresponding DATE token DATESENT/ @DATERCVD/ @DATESAVED/@DATECREATED.

## Recommended

- You can use Field Mapping where the user can select different fields to be populated from the DII/CSV files. Fields would be automatically mapped during Import if the name of the database field matches the name of the field within the DII/CSV file.
- Field names within the header row will appear exactly as they appear within the delimited text file. Use consistent field naming for subsequent data deliveries.
- DocID/BegBates/EndBates values should include no more than 50 characters. Non-alphabetical and non-numerical characters should be avoided.
- Coding file names should mirror that of the respective volume (for easy association and identification). For example:

DOCID|TITLE|AUTHOR

^AAA-000001|^Report to XYZ Corp|^Jillson, Deborah;Ward, Simon;LaBelle, Paige^

^AAA-000005|^Financial Statement|^Mubark, Byju;Aminov, Marina^

^AAA-000008|^Memo|^McMahon, Brian^

## Related Documents

You can review related documents the **@ATTACHRANGE** token or the **@PARENTID** and **@ATTACH** tokens. .

The related documents must be coded in sequential order by their DOCID. The sequence determines the first document and the last document in the related document set.

---

**Note:** Bates number of the first document in **@ATTACHRANGE** populates the ParentDoc column.

---

---

**Note:** **@ParentID** populates the ParentDoc field and **@ATTACH** populates the AttachIDs.

---

Either **@Attachrange** or **@ParentID** can be used at a time.

For example:

**@ATTACHRANGE** ABC001-ABC005

OR

**@PARENTID** ABC001

OR

**@ATTACH** ABC001;ABC002;ABC003;ABC004;ABC005

# Transcripts and Exhibits

---

**Note:** You can import and display of Latin and non-Latin Unicode characters. While the application supports the display of fielded data in either Latin or non-Latin Unicode characters, the modification of fielded data is supported only in Latin Unicode characters.

---

**Note:** The display of non-Latin Unicode characters does not apply to transcript filenames, since transcript deponents are defined by users, or work product filenames, which are not displayed.

---

From **Menu > Transcript > Manage**, you can upload new transcripts to any transcript collection to which they have access. All transcripts are displayed individually, and each has its own menu that controls various transcript management functions.

## *Transcripts*

The following describes the required and recommended formats for transcripts.

### Required

- ASCII or Unicode files (\*.txt) in AMICUS format.

### Recommended

- Transcript size is less than one megabyte.
- Page number specifications:
  - All transcript pages are numbered.
  - Page numbers are up against the left margin. The first digit of the page number should appear in Column 1. See the figure below.
  - Page numbers appear at the top of each page.
  - Page numbers contain no more than six digits, including zeros, if necessary. For example, Page 34 would be shown as **0034**, **00034**, or **000034**.
  - The first line of the transcript (Line 1 of the title page) contains the starting page number of that volume. For example, if the volume starts on Page 1, either **0001** or **00001** are correct. If the volume starts on Page 123, either **0123** or **00123** are correct.
  - Line numbers appear in Columns 2 and 3.
  - Text starts at least one space after the line number. It is recommended to start text in Column 7.
  - No lines are longer than 78 characters (including letters and spaces).
  - No page breaks, if possible. If page breaks are necessary, they should be on the line preceding the page number.
  - Consistent numbers of lines per page, if neither page breaks nor page number formats are used.
  - No headers or footers.
  - All transcript lines are numbered.



Column numbers:		1234567	
No page breaks,	22	Q	Okay. Will you produce that in 14 days,
headers or footers	23		please?
	24	A	Okay.
Zero-filled →	00028		
page numbers	1	Q	Off the top of your head, how many appraisals
start in Column 1	2		do you have pending?
	3	A	Nine, I believe.
Line numbers →	4	Q	Okay. How many properties do you have listed
start in Column 2	5		for sale?
	6	A	I think there's only one that's currently
Text starts	7		listed.
in Column 7	8	Q	Okay. Are you sharing any listings or
	9		appraisals with any other brokers or appraisers?

**Preferred Transcript Format**

*Exhibits*

The following describes the required format for Exhibits.

**Required**

- Exhibits that will be loaded must be in PDF format.
- If an Exhibit has multiple pages, all pages must be contained in one file instead of a file per page.

# Work Product

---

**Note:** You can import and display of Latin and non-Latin Unicode characters. While the application supports the display of fielded data in either Latin or non-Latin Unicode characters, the modification of fielded data is supported only in Latin Unicode characters.

---

**Note:** The display of non-Latin Unicode characters does not apply to transcript filenames, since transcript deponents are defined by users, or work product filenames, which are not displayed.

---

From **Menu > Work Product > Manage** you can upload, view, and review Work Product files. Work Product can be any type of file: text, word processing, PDF, or even MP3. (MP3 files are useful when you wish to send an audio transcript or message to the members of the group who have access to Work Product). The application does not maintain edits or keep version control information for the documents stored. Users working with Work Product documents must have the appropriate native application, such as Microsoft Word or Adobe Acrobat, to open them.

# Sample DII Files

---

**Note:** You can import and display of Latin and non-Latin Unicode characters. While the application supports the display of fielded data in either Latin or non-Latin Unicode characters, the modification of fielded data is supported only in Latin Unicode characters.

---

**Note:** The display of non-Latin Unicode characters does not apply to transcript filenames, since transcript deponents are defined by users, or work product filenames, which are not displayed.

---

**Note:** When selecting the **Copy ESI** option, the DII source files *must* reside in a location accessible by the IEP server; otherwise, import jobs will fail during the **Check File** process.

---

## *eDoc DII Load Files*

### Required DII Format (eDocs)

@T SSS00000007  
@MEDIA eDoc  
@EDOC \folder\SSS00000007.xls

@T SSS00000008  
@MEDIA eDoc  
@EDOC \Native\SSS00000008.doc

### Recommended DII format (eDocs)

@T ABC00000123  
@MEDIA eDoc  
@EDOC \Natives\ABC00000123.xls  
@APPLICATION Microsoft Excel  
@DATECREATED 05/25/2002  
@DATESAVED 06/05/2002  
@SOURCE Dee Vader

## *eMail DII Load Files*

### Required DII File Format for Parent Email (Emails)

@T ABC000123

@MEDIA eMail

@EMAIL-BODY

Please reply with a copy of the completed report.

Thanks for your input.

Beth

@EMAIL-END

@ATTACH ABC000124;ABC000125

### Required DII File Format for Related Email Attachment (Emails)

@T ABC000124

@MEDIA Attachment

@EATTACH \Native\ABC000124.doc

@PARENTID ABC000123

## Recommended DII Format for Parent Email (Emails)

@T ABC000123  
@MEDIA eMail  
@ATTACH ABC000124; ABC000125  
@EMAIL-BODY  
Please reply with a copy of the completed report.

Thanks for your input.

Beth

@EMAIL-END  
@FROM Abe Normal (anormal@ctsummation.com)  
@TO abcody@ctsummation.com; rob.hood@wolterskluwer.com  
@CC Willie Jo  
@BCC Jopp@ctsummation.com  
@SUBJECT Please reply  
@APPLICATION Microsoft Outlook  
@DATECREATED 06/16/2006  
@DATERCVD 06/16/2006  
@DATESENT 06/16/2006  
@FOLDERNAME \Normal\Sent Items  
@READ Y  
@SOURCE Abe Normal  
@TIMERCVD 1:36 PM  
@TIMESENT 1:35 PM

## Recommended DII Format for Related Email Attachments (Emails)

@T ABC000124  
@MEDIA Attachment  
@EATTACH \Native\ABC000124.doc  
@PARENTID ABC000123  
@APPLICATION Microsoft Word  
@DATECREATED 05/25/2005  
@DATESAVED 06/05/2005  
@SOURCE Abe Normal  
@AUTHOR Abe Normal  
@DOCTITLE Sales Report June 2005

## Recommended DII Format for Native Plus Images Deliveries (Email and eDocs)

(Append to the previous recommended DII formats for eDocs or email.)

@D @|\Images\  
ABC000124-001.tif

ABC000124-002.tif

## DII Tokens

Data for all tokens must be in a single line except the @OCR...@OCR-END, @EMAIL-BODY ... @EMAIL-END and @HEADER ... @HEADER-END.

TOKEN	FIELD POPULATED	DESCRIPTION OF USAGE
@T	DOCID & BEGBATES	This token is required for each DII record. This must be the first token listed for the document. This must be unique in the case. The @BEGBATES or @DOCID should not be used. @T ABC000123
@APPLICATION	Application	The application used to view the electronic document. For example: @APPLICATION Microsoft Word
@ATTACH	AttachDocs	IDs of attached documents. For example: @ATTACH ABC000124;ABC000125
@ATTACHRANGE	ParentDoc	The document number range of all attachments if more than one attachment exists. The beginning number in the range populates the PARENTDOC. For example: @ATTACHRANGE WGH000008 – WGH0000010
@ATTMSG	Media & Native file is copied into the filesystem using the path provided	The file name of the e-mail attachment (that is an e-mail message itself) including the relative or absolute path to the document. The relative path is evaluated using the path to the DII file as the root path. The native file is then loaded. The Media field is populated with the value eMail.
@BATESBEG	Begbates	Beginning Bates number, used with @BATESEND. For example: @BATESBEG SGD00001
@BATESEND	EndBates	Ending Bates number. For example: @BATESEND SGD00055
@BCC	EmailBCC	Anyone sent a blind copy on an e-mail message. For example: @BCC Nick Thomas
@C	Custom Field	Code used to load a custom field in the database. The syntax for the @C token is: @C <FIELDNAME> <DATA> The FIELDNAME value cannot contain spaces. For example, to fill in the DEPARTMENT field of the database with the value Accounting, the line would read: @C DEPARTMENT Accounting
@CC	EmailCC	Anyone copied on an e-mail message. For example: @CC John Ace

@D @I	Link to images	<p>Required token for each DII record that has an image associated with it. This designates the directory location of the image file(s). Note that only the “@D @I” sequence is allowed. The “@D @V” sequence is not recognized.</p> <p>The following 2 examples are equivalent:</p> <p>--Example 1  @D @I\Images\001\  ABC00123.tif  ABC00124.tif</p> <p>--Example 2  @D @I\Images\  001\ABC00123.tif  001\ABC00124.tif. Note the directory should be relative to the load file. If this token is in the record, it must be the last token in the record.</p> <p>Also UNC paths in the Image Directory field (For example @D <a href="#">\\Server\PFranc\Images</a>) are recognized but no hard coded drive letters.</p>
@DATECREATED	CreationDateFT	The date that the file was created. For example: @DATECREATED 01/04/2003
@DATERCVD	DeliveryTimeFT	Date that the e-mail message was received.
@DATESAVED	ModificationDateFT	Date that the file was saved.
@DATESENT	SubmitTimeFT	Date that the e-mail message was sent.
@EATTACH	Native file is copied into the filesystem using the path provided	Relative path (from the load file location) of the native file to be loaded. Valid for Attachments.
@EDOC	Native file is copied into the filesystem using the path provided	Same as @EATTACH except for eDocs. For example @EDOC \Attachments\ABC000123.xls Valid for edocs only.
@EMAIL-BODY @EMAIL-END	Email body is copied into a file in the file system.	Body of an e-mail message. Must be a string of text contained between @EMAIL-BODY and @EMAIL-END. The @EMAIL-END token must be on its own line. For example: @EMAIL-BODY Bill, This looks excellent. Ted @EMAIL-END
@FILENAME	Filename of the native	Original Filename of the native file (Edoc/Email/Attachment) For example @FILENAME AnnualReport.xls
@FOLDERNAME	FolderNameID	The name of the folder that the e-mail message came from. For example: @FOLDERNAME \Inbox\Projects\ARProject
@FROM	EmailFrom	From field in an e-mail message. For example: @FROM Kelly Morris



@FULLTEXT	N/A (text processing directive)	Determines how OCR is associated with the document. This token should be placed at the top of the file, before any @T tokens. The OCR files must have the same names as the images (not including the extension), and they must be located in the same directory. Variations: @FULLTEXT DOC - One text file exists for each database record. The name of the file must be the same name as the first image file. @FULLTEXT PAGE - One text file exists for each page.
@FULLTEXTDIR	Link to Full text Directory	The @FULLTEXTDIR token is a partner to the @FULLTEXT token. @FULLTEXTDIR allows specifying a directory from which the full-text will be copied during the import. Therefore, the full-text files do not have to be located in the same directory as the images at the time of import. The @FULLTEXTDIR token gives you the flexibility to import the DII file and full-text files without requiring you to copy the full-text files to the network first. For example: @FULLTEXTDIR Vol001\Box001\ocrFiles The above example shows a relative path. The application searches for the full-text files in the same location as the DII file that is imported and follows any subdirectories listed after the @FULLTEXTDIR token. The @FULLTEXTDIR token applies to all subsequent records in the DII file until it is changed or turned off.
@HEADER @HEADER-END	EmailHeader	E-mail header content. The @HEADER-END token must be on its own line. For example: @HEADER <Header Text> @HEADER-END
@INTMSGID	InternetMessageID	Internet message ID. For example: @INTMSGID <00180c34fe5\$bf2d5\$050@SKEETER>
@MEDIA	Media	Indicates the type of document. This must be populated with one of the following values: {email, attachment, and eDoc} This value is REQUIRED. This value is used by the application to determine how to display the document. For example : @MEDIA eDoc
@MSGID	EntryID	E-mail message ID generated by Microsoft Outlook or Lotus Notes. For example: @MSGID 00E8324B3A0A800F4E954B8AB427196A1304012000
@MULTILINE	Any custom field with multiple lines	Allows carriage returns and multiple lines of text to populate a specified text field. Text must be between @MULTILINE and @MULTILINE-END. The @MULTILINE-END token must be on its own line. For example: @MULTILINE FIELDNAME Here is the first line. Here is the second line. Here is the third line. Here is the last line. @MULTILINE-END
@O	OCRTEXT / FULLTEXT is copied into a file in the file system	This token is used to load full-text documents. The text files can be located someplace other than the image location as specified by the @D line of the DII file. There can only be one text file for the record. The value following the @O should contain the relative path (from the load file location) of the .txt file. @O \Text\ABC000123.txt

@OCR @OCR-END	OCRTEXT is copied into a file in the file system	The @OCR and @OCR-END tokens offer the flexibility to include the full-text (including carriage returns) in the DII file. The @OCR-END token must appear on a separate line. For example: @OCR <full-text extracted from the electronic document, which can span multiple lines> @OCR-END
@PARENTID	ParentDoc	Parent document ID of an attachment. For example: @PARENTID ABC000123
@PSTFILE0	PSTFilePath and PSTStoreNameID	<p>The original PST File name and ID</p> <ol style="list-style-type: none"> <li>1) The name and/or location of the .PST file.</li> <li>2) The unique ID of the .PST file.</li> </ol> <p>The two values are separated by a comma. The unique ID can be any unique value that identifies the .PST file. For example: @PSTFILE EMAIL001\PFranc.pst, PFranc_14April_07</p> <p>The .PST file's unique ID (the second value) is populated into the PST ID field designated in eMail Defaults.</p> <p>The PST ID value specified by the @PSTFILE token is assigned to the record it appears in and will apply to all subsequent e-mail records. The value is applied until either the @PSTFILE token is turned off by setting the token to a blank value or the value changes. The @PSTFILE token can occur multiple times in a single DII file and assign a different value each time. This allows processing multiple .PST files and presenting the data for all .PST files in a single DII file.</p> <p>As a best practice, the @PSTFILE token should be placed above the @T token.</p>
@READ	IsUnread (stores 0 if Y and 1 if N)	Notes whether the e-mail message was read. For example: @READ Y
@RELATED	LinkedDocs	The document IDs of related documents. For example: @RELATED WGH000006
@SOURCE	Source	Custodian of the data. You can quickly filter documents by this field. @SOURCE Joe Custodian
@SUBJECT	Subject	The subject of an e-mail message. For example: @SUBJECT RE: Town Issues
@TIMECREATED	CreationDateFT	Time the file/e-mail/edoc was created
@TIMERCVD	DeliveryTimeFT	Time that the e-mail message was received.
@TIMESAVED	ModificationDateFT	Time that the file/e-mail/edoc was last saved
@TIMESENT	SubmitTimeFT	Time that the e-mail message was sent.
@TO	EmailTo	To field in an e-mail message. For example: @TO Conner Stevens
@UUID	UUID	Customer-specific and unique identifier for a record (not used internally by the application) For example : @UUID AE01R95

# Chapter 5

## Analyzing Document Content

---

### Using Cluster Analysis

#### *About Cluster Analysis*

You can use Cluster Analysis to group Email Threaded data and Near Duplicate data together for quicker review.

---

**Note:** If you activated Cluster Analysis as a processing option when you created the project, cluster analysis will automatically run after processing data and will not need to be run manually.

---

Cluster Analysis is performed on the following file types:

- Documents (including PDFs)
- Spreadsheets
- Presentations
- Emails

Cluster Analysis is also performed on text extracted from OCR if the OCR text comes from a PDF. Cluster Analysis cannot be performed on OCR text extracted from a graphic.

#### **To perform cluster analysis**

1. Load the email thread or near duplicate data using Evidence Processing or Import.
2. On the *Home* page, in the *Project List* panel, click the **Add Evidence** button next to the project.
3. In the *Add Data* dialog, click **Cluster Analysis**.
4. Select a threshold to group the documents based on similarity. The default value is 80%.
5. Click **Start**.

The data for the email thread appears in the *Conversation* tab in *Project Review*. The data for Near Duplicate appears in the *Related* tab in *Project Review*.

An entry for cluster analysis will appear in the *Work List*.

#### Words Excluded from Cluster Analysis Processing

Noise words, such as “if,” “and,” “or,” are excluded from Cluster Analysis processing. The following words are excluded in the processing:

a, able, about, across, after, ain't, all, almost, also, am, among, an, and, any, are, aren't, as, at, be, because, been, but, by, can, can't, cannot, could, could've, couldn't, dear, did, didn't, do, does, doesn't, don't, either, else,

ever, every, for, from, get, got, had, hadn't, has, hasn't, have, haven't, he, her, hers, him, his, how, however, i, if, in, into, is, isn't, it, it's, its, just, least, let, like, likely, may, me, might, most, must, my, neither, no, nor, not, of, off, often, on, only, or, other, our, own, rather, said, say, says, she, should, shouldn't, since, so, some, than, that, the, their, them, then, there, these, they, they're, this, tis, to, too, twas, us, wants, was, wasn't, we, we're, we've, were, weren't, what, when, where, which, while, who, whom, why, will, with, would, would've, wouldn't, yet, you, you'd, you'll, you're, you've, your

## *Filtering Documents by Cluster Topic*

Documents processed with Cluster Analysis can be filtered by the content of the documents in the evidence. The Cluster Topic filter is created in Review under the Document Contents filter from data processed with Cluster Analysis. Data included in the Cluster Topic is taken from the following types of documents: Word documents and other text documents, spreadsheets, emails, and presentations.

In order for the application to filter the data with the Cluster Topic filter, the following must occur:

- [Prerequisites for Cluster Topic](#) (page 44)
- [How Cluster Topic Works](#) (page 44)
- [Filtering with Cluster Topic](#) (page 45)
- [Considerations of Cluster Topic](#) (page 45)

## Prerequisites for Cluster Topic

Before Cluster Topic filter facets can be created, the data in the project must be processed by Cluster Analysis. The data can be processed automatically when Cluster Analysis is selected in the Processing options or you can process the data manually by performing **Cluster Analysis** in the *Add Evidence* dialog.

[Evidence Processing and Deduplication Options](#) (page 159)

## How Cluster Topic Works

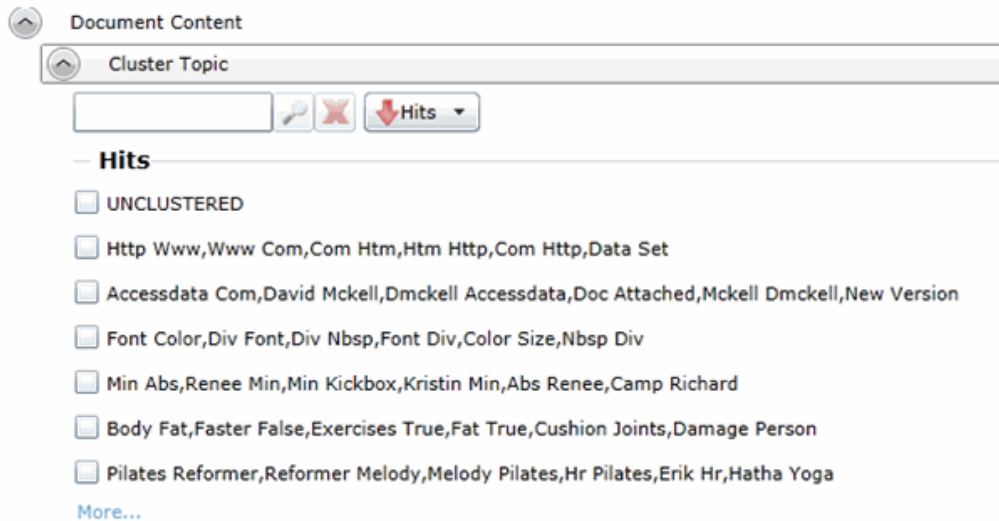
The application uses an algorithm to cluster the data. The algorithm accomplishes this by creating an initial set of cluster centers called pivots. The pivots are created by sampling documents that are dissimilar in content. For example, a pivot may be created by sampling one document that may contain information about children's books and sampling another document that may contain information about an oil drilling operation in the Arctic. Once this initial set of pivots is created, the algorithm examines the entire data set to locate documents that contain content that might match the pivot's perimeters. The algorithm continues to create pivots and clusters documents around the pivots. As more data is added to the project and processed, the algorithm uses the additional data to create more clusters.

Word frequency or occurrence count is used by the algorithm to determine the importance of content within the data set. Noise words that are excluded from Cluster Analysis processing are also not included in the Cluster Topic pivots or clusters.

## Filtering with Cluster Topic

Once data has been processed by Cluster Analysis and facets created under the Cluster Topic filter, you can filter the data by these facets.

### Cluster Topic Filters



The topics of the facets available are cluster terms created. Documents containing these terms are included in the cluster and are displayed when the filter is applied. Topics are comprised of two word phrases that occur in the documents. This is to make the topic more legible.

The UNCLUSTERED facet contains any documents that are not included under a Cluster Topic filter.

For more information, see *Filtering Data in Case Review* in the *Reviewer Guide*.

## Considerations of Cluster Topic

You need to aware the following considerations when examining the Cluster Topic filters:

- Not all data will be grouped into clusters at once. The application creates clusters in an incremental fashion in order to return results as quickly as possible. Since the application is continually creating clusters, the Cluster Topic facets are continually updated.
- Duplicate documents are clustered together as they match a specific cluster. However, if a project is particularly large, duplicate documents may not be included as part of any cluster. This is to avoid performance issues. You can examine any duplicate documents or any documents not included in a cluster by applying the UNCLUSTERED facet of the Cluster Topic filter.

# Using Entity Extraction

## About Entity Extraction

You can extract entity data from the content of files in your evidence and then view those entities.

You can extract the following types of entity data:

- Credit Card Numbers
- Email Addresses
- People
- Phone Numbers
- Social Security Numbers

The data that is extracted is from the body of documents, not the meta data.

For example, email addresses that are in the *To:* or *From:* fields in emails are already extracted as meta data and available for filtering. This option will extract email addresses that are contained in the body text of an email.

Using entity extraction is a two-step process:

1. Process the data with the *Entity Extraction* processing options enabled.  
You can select which types of data to extract.
2. View the extracted entities in *Review*.

The following tables provides details about the type of data that is identified and extracted:

Type	Examples
<b>Credit Card Numbers</b>	Numbers in the following formats will be extracted as credit card numbers:
16-digit numbers used by VISA, MasterCard, and Discover in the following formats.	For example, <ul style="list-style-type: none"><li>• 1234-5678-9012-3456 (segmented by dashes)</li><li>• 1234 5678 9012 3456 (segmented by spaces)</li></ul> Not: <ul style="list-style-type: none"><li>• 1234567890123456 (no segments)</li><li>• 12345678-90123456 (other segments)</li></ul>
15-digit numbers used by American Express in the following formats.	For example, <ul style="list-style-type: none"><li>• 1234-5678-9012-345 (segmented by dashes)</li><li>• 1234 5678 9012 345 (segmented by spaces)</li></ul>
	Notes: Other formats, such as 14-digit Diners Club numbers, will not be extracted as credit card numbers

Type	Examples
<b>Email Addresses</b>	Text in standard email format, such as jsmith@yahoo.com will be extracted.
	Note: Email addresses that are in the <i>To:</i> or <i>From:</i> fields in emails are already extracted as meta data and available for filtering. This option will extract email addresses that are contained in the body text of an email.
<b>People</b>	Text that is in the form of proper names will be extracted as people.
	Proper names in the content are compared against personal names from 1880 - 2013 U.S. census data in order to validate names.

Type	Examples
<b>Phone Numbers</b>	Numbers in the following formats will be extracted as phone numbers:
Standard 7-digit	For example: <ul style="list-style-type: none"> <li>• 123-4567</li> <li>• 123.4567</li> <li>• 123 4567</li> </ul> Not: 1234567 (not segmented)
Standard 10-digit	For example: <ul style="list-style-type: none"> <li>• (123)456-7890</li> <li>• (123)456 7890</li> <li>• (123) 456-7809</li> <li>• (123) 456.7809</li> <li>• +1 (123) 456.7809</li> <li>• 123 456 7809</li> </ul> Not 1234567890 (not segmented) <p>Note: A leading 1, for long-distance or 001 for international, is not included in the extraction, however, a +1 is.</p>

Type	Examples
International	<p>Some international formats are extracted, for example,</p> <ul style="list-style-type: none"> <li>• +12-34-567-8901</li> <li>• +12 34 567 8901</li> <li>• +12-34-5678-9012</li> <li>• +12 34 5678 9012</li> </ul> <p>Not 12345678901 (not segmented)</p> <p>Other international formats are not extracted, for example,</p> <ul style="list-style-type: none"> <li>• 123-45678</li> <li>• (10) 69445464</li> <li>• 07700 954 321</li> <li>• (0295) 416,72,16</li> </ul> <p>Notes: Be aware that you may get some false positives. For example, a credit number 5105-1051-051-5100 may also be extracted as the phone number 510-5100.</p>

Type	Examples
<b>Social Security Numbers</b>	<p>Numbers in the following formats will be extracted as Social Security Numbers:</p> <ul style="list-style-type: none"> <li>• 123-45-6789 (segmented by dashes)</li> <li>• 123 45 6789 (segmented by spaces)</li> </ul> <p>The following will not be extracted as Social Security Numbers:</p> <ul style="list-style-type: none"> <li>• 123456789 (not segmented)</li> <li>• 12345-6789 (other segments)</li> </ul>

## Enabling Entity Extraction

### To enable entity extracting processing options:

1. You enable *Entity Extraction* when creating a project and configuring processing options. See [Evidence Processing and Deduplication Options](#) on page 159.

## Viewing Entity Extraction Data

### To view extracted entity data

1. For the project, open *Review*.
2. In the *Facet* pane, expand the *Document Content* node.
3. Expand the *Document Content* category.
4. Expand a sub-category, such as *Credit Card Numbers* or *Phone Numbers*.
5. Apply one or more facets to show the files in the *Item List* that contain the extracted data.



# Chapter 6

## Editing Evidence

---

### Editing Evidence Items in the Evidence Tab

Users with Create/Edit project admin permissions can view and edit evidence for a project using the Evidence tab on the Home page.

#### To edit evidence in the Evidence tab

1. Log in as a user with *Create/Edit* project admin permissions.
2. Select a project from the *Project List* panel.
3. Click on the **Evidence** tab.
4. Select the evidence item you want to edit and click the **Edit** button.
5. In the *External Evidence Details* form, edit the desired information.

# Evidence Tab

Users with permissions can view information about the evidence that has been added to a project. To view the *Evidence* tab, users need one of the following permissions: Administrator, Create/Edit Project, or Manage Evidence.


## Evidence Tab

The screenshot displays the Evidence Tab interface. At the top, there is a toolbar with various icons. Below it is a 'Filter Options' section. The main area contains a table with the following data:


Path	Description	Evidence Type
\\cvg-dcstorage\cases\Agi\doc		Native
\\cvg-dcstorage\cases\Agi\doc		Native
\\cvg-dcstorage\cases\TestData_Load\DII\CVG7_002\CVG7_002_img01.tif		Native
\\cvg-dcstorage\cases\TestData_Load\Sally\Small Control Set\gmail.pst		Native

Below the table, there are pagination controls: Page Size: 15, Total: 4, Page 1 of 1. To the right of the table is the 'External Evidence Details' panel, which includes fields for Path, Description, Evidence Type (Native), Associated Person Name, Created By (Administrator), and Created Date (12/8/2011 9:01:42 PM). At the bottom of the interface is a 'Processing Status' section with 'General' and 'Progress' tabs, and areas for 'Error Messages' and 'Messages'.

## Elements of the Evidence Tab

Element	Description
Filter Options	Allows the user to filter the list.
Evidence Path List	Displays the paths of evidence in the project. Click the column headers to sort by the column.
Refresh 	Refreshes the Evidence Path List.

## Elements of the Evidence Tab (Continued)

Element	Description
Columns 	Click to adjust what columns display in the Evidence Path List.
External Evidence Details	Includes editable information about imported evidence. Information includes: <ul style="list-style-type: none"><li>• That path from which the evidence was imported</li><li>• A description of the project, if you entered one</li><li>• The evidence file type</li><li>• What people were associated with the evidence</li><li>• Who added the evidence</li><li>• When the evidence was added</li></ul>
Processing Status	Lists any messages that occurred during processing.