



Load File Specifications

v6.6

August 25, 2010



Contents

1	Overview.....	3
2	Metadata, Extracted Text, and Native Files.....	3
2.1	Encoding.....	3
2.2	Header Row.....	3
2.3	Field Order	3
2.4	Delimiters.....	3
2.4.1	Default delimiters	4
2.5	Required Fields.....	4
2.6	System Fields.....	5
2.7	Other fields.....	6
3	Images Files.....	6
3.1	Image File Formats	7
3.2	Load file Format.....	7
3.3	Importing Extracted Text during an Image Load	7
4	Disclaimer	8

1 Overview

This document details Relativity load files specifications. These load files can be used to load:

- Metadata, Extracted text, and Native files
- Images, Extracted text

2 Metadata, Extracted Text, and Native Files

Metadata, document level extracted text, and natives files are loaded into Relativity using a flat, document-level load file. Each line should represent one document.

2.1 Encoding

A variety of load file encoding options are supported, including:

- Western European (Windows)
- Unicode
- Unicode (Big-Endian)
- Unicode (UTF-7)
- Unicode (UTF-8)
- Other options, based on your SQL sever

2.2 Header Row

Relativity does not require load file header rows. However, they are strongly recommended to ensure accuracy.

The field names in your header do not need to match the field names in your case.

2.3 Field Order

Relativity does not require any specific load file field order. During the load process, you can match your load file fields to the fields in your case.

2.4 Delimiters

During import, Relativity allows you to designate which delimiters are used in your load file. You can select each delimiter from the ASCII characters, 001 – 255.

The delimiters are:

- **Column** – the character which separates load file columns.

- **Quote** – the character which marks the beginning and end of each load file field. Also known as a text qualifier.
- **Newline** – the character which marks the end of a line in any extracted or long text field.
- **Multi-value** – the character which separates distinct values in a column. Note that this delimiter is only used when importing into a Relativity multi-choice field.
- **Nested-values** – the character which denotes the hierarchy of a choice. Note that this delimiter is only used when importing into a Relativity multi-choice field.



For example, say a load file contained the following entry, and was being imported into a multi-choice field:

“Hot\Really Hot\Super Hot; Look at Later”

With the multi-value delimiter set as “;” and the nested value delimiter set as “\”, the choices would appear in Relativity as:

- Hot
 - Really Hot
 - Super Hot
 - en fuego
- 정
- 박사 (Med Expert)
- Look at Later

2.4.1 Default delimiters

If you generate your own load files, you may choose to use Relativity’s defaults:

- **Column** - ASCII 020
- **Quote** - ASCII 254
- **Newline** - ASCII 174
- **Multi-Value** - ASCII 059
- **Nested Values** - ASCII 092

2.5 Required Fields

There is only one required field for each load – the identifier.

When loading new records, this will be your case identifier.

When performing an overlay, you can use the case identifier or select another field as the identifier. This is useful when overlaying production data. For example, you could use the Bates number field rather than the document identifier in the case.

2.6 System Fields

All fields besides “Identifier” are optional. However, you may find some of the following system fields beneficial.

1. **Identifier** – the unique identifier of the record. Required.
2. **Group Identifier** – the identifier of a document’s family group.
 - The group identifier repeats for all records in the group.
 - Usually, this is the document identifier of the group’s parent document. For example:
 - If an email with the document identifier of AS00001 has several attachments, the email and its attachments will have a group identifier of AS00001.
 - If a group identifier for a record is not set, the document identifier is used to populate the group identifier field in the case. This effectively creates a “group” of one document.
3. **MD5 Hash** – the duplicate hash value of the record.
 - You can enter any type of hash value (and rename the field in your case).
 - If documents share the same hash value, they will be identified as a duplicate group.
 - If a hash field for a record is not set, the document identifier is used to populate the hash field in the case. This effectively creates a “group” of one document.
4. **Extracted Text** – The text of the document. Either the OCR or Full Text.
 - The extracted text will be displayed in the viewer, and added to search indexes. This field can contain either:
 - The actual OCR or Full Text
 - The path to a document level text file containing the OCR or Full Text.
5. **Native File Path** – the path to any native files you’d like to load.
 - Both relative and absolute paths are supported.
6. **Folder Info** – builds the folder browser structure for the documents.
 - This field is backslash “\” delimited.

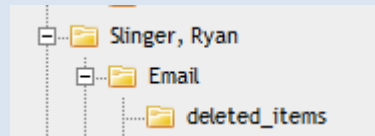
- If not set, the documents will be loaded to the root of the case.
- Each entry between backslashes will be a folder in the Relativity folder browser.
- Each backslash indicates a new subfolder in the browser.



For example, if the load file contained the following entry in the folder info field:

“Slinger, Ryan\Email\deleted_items”

Relativity would build the following folder structure:



Each document with the above entry would be stored in the “deleted_items” folder.

7. **Relativity Native Time Zone Offset** - Relativity’s native viewer technology displays all email header dates and times in GMT. This numeric field offsets how email header dates and times appear in the viewer.
 - If the value in this field is blank, or 0, for a document, then the email header date and time will be displayed in GMT.
 - You can enter a whole number in this field – positive or negative – to offset the time from GMT to the local time zone of the document.
 - For example, if the document was collected from US CDT time, you would enter “-5” in the field, because the CDT offset from GMT is -5.
 - Note this **ONLY** applies when viewing email header dates and times in the Relativity Native File Viewer. Your metadata fields will be displayed as imported.

2.7 Other fields

Relativity allows you to create any number of case fields to store metadata or coding. You can match any load file fields to your case fields during the import process.

3 Images Files

Images are loaded into Relativity using an Opticon load file. This process can additionally be used to import page level Extracted text.

3.1 Image File Formats

Relativity accepts only the following file types for image loads:

- Single page, Group IV TIFs (1 bit, B&W)
- Single page JPGs

Multi page TIFs and PDFs can be imported into the system, but they must be loaded as native files.

3.2 Load file Format

The Opticon load file is a page level load file, with each line representing one image.

Below is a sample:

```
REL00001,REL01,D:\IMAGES\001\REL00001.TIF,Y,,,3
REL00002,REL01,D:\IMAGES\001\REL00002.TIF,,,,
REL00003,REL01,D:\IMAGES\001\REL00003.TIF,,,,
REL00004,REL01,D:\IMAGES\001\REL00004.TIF,Y,,,2
REL00005,REL01,D:\IMAGES\001\REL00005.TIF,,,,
```

The fields are, from left to right:

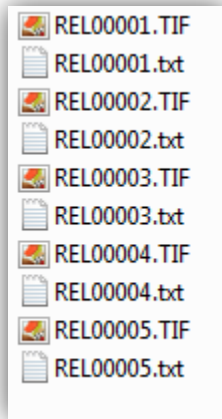
- Field One – (REL00001) – The page identifier
- Field Two – (REL01) – The volume identifier – not required
- Field Three – (D:\IMAGES\001\REL00001.TIF) – a path to the image to be loaded
- Field Four – (Y) – Document marker – a “Y” indicates the start of a unique document.
- Field Five – (blank) – Can be used to indicate box
- Field Six – (blank) – Can be used to indicate folder
- Field Seven – (3) – often used to store page count, but unused in Relativity

3.3 Importing Extracted Text during an Image Load

You can also import extracted text during the image import process.

This is accomplished by setting an option in the Relativity Desktop Client. For more information about importing extracted text during an image load, see the Relativity Administrative Manual.

No changes are needed in the Opticon load file. If the aforementioned setting is active, Relativity will look for page level txt files named identical to their corresponding TIFs. For example:



Page Level text files

4 Disclaimer

This documentation is proprietary information of kCura Corporation and may be modified, altered, or repurposed only in accordance with written consent from kCura.

© 2010. All rights reserved.