SimSage Metadata Operations

SimSage has a series of powerful metadata manipulation operations. This section discussed the operation of the Metadata Enhancement Engine.

Navigate to any of your sources, and click the Metadata tab as shown. This will present you with three sub-tabs, click the Metadata Enhancements tab.

the SimSage metadata enhancement system

The enhancement system is created for matching parts of paths of documents, and adding any number of extra metadata values to a document or a record.

An example of this would be that all documents inside a path called `MIMR-2025-0125` need specific metadata added, because they are part of a particular project. However, this pattern is more general, in that anything that matches `MIMR-####-####` (the literal letters MIMR followed by a dash, four digits, another dash, and another four digits) requires a lookup.

The primary key pattern is just that. It can consist of

Pattern	Matches
#	any single digit 0 through to 9
.	any single letter (case insensitive) a..z
*	any single digit or single letter as above
Any other character	Must match that character itself, case sensitive

Once we've established this pattern, we can import a database of entries in CSV format.

we've set our primary-key pattern, and proceed to import CSV data

Click the "Import Csv Metadata" button.

Import Metadata Csv Data

Text

```
e.g., for a primary-key defined as MIMR-####-####:

key, mdAuthor, mdTitle, mdValue
MIMR-9023-1883, John Elderberries, A Tale of two Cities, 40
MIMR-0708-9015, John Smallberries, Constellation Drive Title, 25
```

Browse

- The first line of your Csv must define the column names
- first column must be named 'key'.
- One Csv record per line
- Csv records with commas inside must use double quotes

Cancel   Import

the Metadata importer

You can upload as many metadata items, as many times as you like. It is the primary-key (key) that makes a record unique. No extra values are created by importing the same items twice.

We can import such items either from a file, or by typing the entries into the window itself.

Import Metadata Csv Data

Text

```
key, mdauthor, mdtitle, mdvalue
MIMR-2020-0185, John Elderberries, Experiment 123, 50
MIMR-2020-0186, John Manyberries, Experiment 25, 5
MIMR-2020-0187, John Smallberries, Experiment 32, 25
MIMR-2020-0188, John Someberries, Experiment 0188, 28
```

- The first line of your Csv must define the column names
- first column must be named 'key'.
- One Csv record per line
- Csv records with commas inside must use double quotes

Cancel   Import

Some example values

The CSV file has a strict format:

- The CSV entries must Always have a header.
- Metadata column names are defined by this header.
- Metadata names themselves are restricted to 50 characters maximum, and are not allowed spaces, only hyphens and a..z (lowercase).
- The first column MUST BE NAMED 'key' (as shown)
- The other column names are up to you, and define what the user must enter to search for these values (e.g., `mdauthor` in the above example, combined with `John Elderberries` as a value, enables a search for `mdauthor: John` or `mdauthor: Elderberries` or `mdauthor: John Elderberries`)

Click "Import" to finish uploading your data.

Find Existing Records

Use the "Find Existing Records" to view what data is stored inside SimSage once you've uploaded your CSV metadata.

the results of a find

Or, use a partial pattern to return up to a maximum of five records to view.

Edit Source: rock-test

General   Web Crawler   **Metadata**   ACLs   MimeTypes   Rules   Check Document   Schedule

Metadata Mapper   Metadata Transform   Metadata Enhancements

Match part of a document's/record's path with a Primary Key Pattern, then add metadata to any document that matches the same key inside a set of records defined through a CSV database with name/value pairs.

The Primary Key pattern is case sensitive and uses
(hash) to match a single digit 0..9
. (dot / fullstop) to match single letters a..z or A..Z
* (asterisk / star) to match a single digit 0..9 or letter a..z, A..Z
e.g., MIMR-####-#### will match MIMR-2020-0128 and MIMR-2025-0125

Primary Key Pattern

MIMR-####-####   [Import Csv Metadata]

Find existing records by partially or fully matching the Primary Key Pattern. Will only show up to five records maximum.

Find Existing Records   MIMR-2020-01##   [Find]

(showing 4 of 4 records)

MIMR-2020-0185 : mdvalue=50, mdauthor=John Elderberries, mdtitle=Experiment 123
MIMR-2020-0186 : mdvalue=5, mdauthor=John Manyberries, mdtitle=Experiment 25
MIMR-2020-0187 : mdvalue=25, mdauthor=John Smallberries, mdtitle=Experiment 32
MIMR-2020-0188 : mdvalue=28, mdauthor=John Someberries, mdtitle=Experiment 0188

Cancel   [Test]   [Save]

Any Source that changes, or adds new data will need its files reprocessed using normal source operations. This is not necessary if you set up this database before starting processing any data.

Any change to this data, requires reprocessing!

The results will automatically come through and can be viewed in a document's metadata. Let's suppose we have a file called `https://dataset.simsage.co.uk/rock-test/MIMR-2020-0185/candy.odp` that matches one of these database entries through its folder/path structure (`MIMR-202-0185`). We can then take this URL and put it into the source's check document.

Check document, click Find

two of the new metadata values (the third `mdvalue` is off screen)

After indexing is complete, we can now look for this document and any documents that were inside the folder marked by `MIMR-2020-0185` by searching for any of the three metadata values automatically added.

mdauthor: John

one result

dataset.simsage.co.uk/rock-test/MIMR-2020-0185/...

Candy

https://dataset.simsage.co.uk/rock-test/MIMR-2020-0185/candy.odp

Last modified 10/6/2025

👍

+

a Search using the new metadata

dataset.simsage.co.uk/rock-test/MIMR-2020-0185/...