

Updating the Historical Images RDF Thesaurus file

1. The updating process begins when the Metadata/Copy Cataloging Unit Team Lead sends a list of terms (usually as a Word document) newly added to the local Historical Images thesaurus spreadsheet.
2. Start OpenRefine and navigate to **Create Project > Clipboard**. Paste in the new terms. Click **Next**.

Create a project by importing data. What kinds of data files can I import?
TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with Google Refine extensions.

Get data from

- This Computer
- Web Addresses (URLs)
- Clipboard**
- Google Data

Paste data from clipboard here:

Oslo (Norway)
Palaces-Norway-Bygdøy
Palaces-Sweden-Stockholm
Piers-Norway
Plazas-Sweden-Stockholm
Rivers-Norway
Stockholms slott (Stockholm, Sweden)
Stone bridges-Norway
Texas Christian University Horned Frogs (Football team)
University of Houston Cougars (Football team)
Urbanscreen (Firm)
Vigeland, Gustav, 1869-1943
Waterfronts-Sweden-Stockholm

Next »

3. Use the below parameters and click **Create Project**.

Parse data as

- Fixed-width field text files
- Line-based text files**
- CSV / TSV / separator-based files
- PC-Axis text files
- JSON files

Character encoding

Parse every lines into one row

- ☐ Store blank rows
- ☐ Store blank cells as nulls
- ☐ Store file source (file names, URLs) in each row

4. Microsoft Word routinely changes double-dashes (–) into the longer emdash (—). These need to be replaced with the correct double-dashes.

This change can be accomplished by navigating to **Edit Cells > Transform**:

27 rows

Show as: **rows** records Show: 5 10 25 50 rows

▼ All	▼ Column 1
★	1. Statues
★	2. Boathouses—Norway
★	3. Bygdøy (Norway)—Buildings, structures, etc.
★	4. Centennial (Rice University)
★	5. College football players—Louisiana
★	6. Colorado Buffaloes (Football team)
★	7. Emma Canyon, U.S.A.
★	8. Emma Canyon, U.S.A.
★	9. Emma Canyon, U.S.A.
★	10. Emma Canyon, U.S.A.
★	11. Emma Canyon, U.S.A.
★	12. Emma Canyon, U.S.A.
★	13. Emma Canyon, U.S.A.

Context menu options:

- Facet
- Text filter
- Edit cells
- Edit column
- Transpose
- Sort...
- View
- Reconcile
- Extract named entities...
- Transform...
- Common transforms
- Fill down
- Blank down
- Split multi-valued cells...
- Join multi-valued cells...
- Cluster and edit...

In the Transform window, using the following expression: `value.replace('—','--')`

Custom text transform on column Column 1

Expression: `value.replace('—','--')` Language: Google Refine Expression Language (GREL) ▼

No syntax error.

Preview

row	value	value.replace('—','--')
1.	Abel, Niels Henrik, 1802-1829—Statues	Abel, Niels Henrik, 1802-1829--Statues
2.	Boathouses—Norway	Boathouses--Norway
3.	Bygdøy (Norway)—Buildings, structures, etc.	Bygdøy (Norway)--Buildings, structures, etc.
4.	Centennial (Rice University)	Centennial (Rice University)
5.	College football players—Louisiana	College football players--Louisiana
6.	Colorado Buffaloes (Football team)	Colorado Buffaloes (Football team)
7.	Emma Canyon, U.S.A.	Emma Canyon, U.S.A.

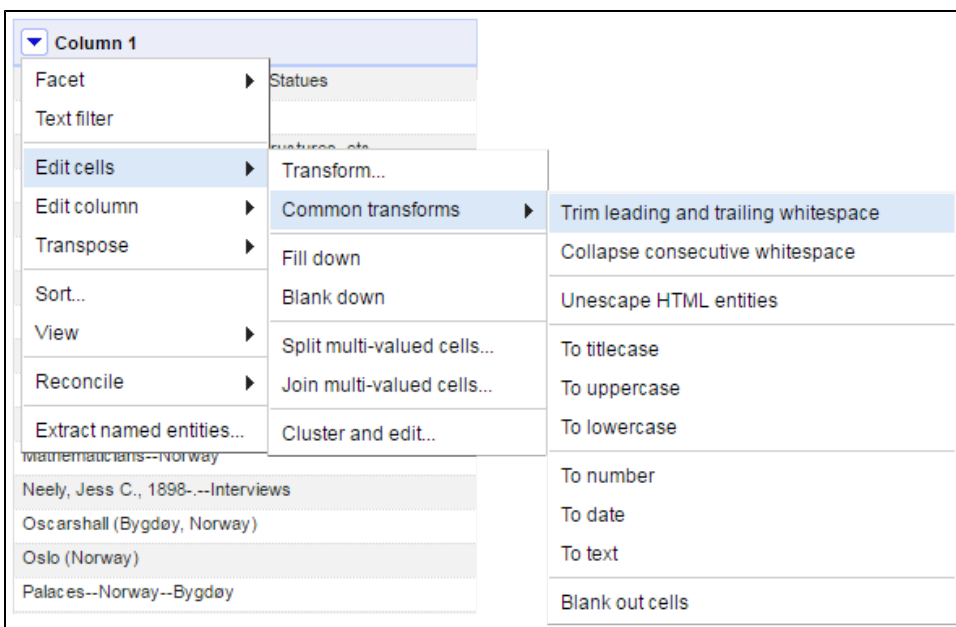
On error: ☒ keep original ☐ set to blank ☐ store error

☐ Re-transform up to 10 times until no change

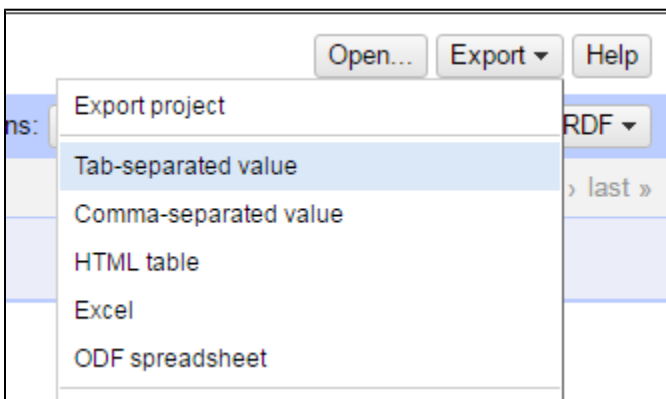
OK Cancel

Click okay. All emdashes should now be replaced.

- Return to **Edit Cells** to trim the whitespace in the new terms:



- Export your reconciled data to a new TSV file using the **Export** button at the top right hand part of the screen:



- Open the Excel spreadsheet that generates RDF entities – you will find it on the Projects server at: **\\Projects\\AuthorityWork\\RDF\\HHI-thesaurus-Updated.xlsx**
- Load your new terms from the TSV file into the spreadsheet in column A.
- The FORMULAS tab contains pre-written Excel FORMULAS to help you generate the RDF elements.

NOTE: this is not real-world linked data. The formulas concoct bunk URIs which are necessary for the RDF to function.

- Increment the URI number** (column B)
- Generate URI** (column C)
- Compile SKOS** (column G) – generates RDF XML elements from columns A & C.

- Next, you will add the SKOS RDF XML elements generated in Column G. To start, open the the RDF file – which can be found at **\\Projects\\AuthorityWork\\RDF\\HHI.rdf** – using Oxygen XML editor.

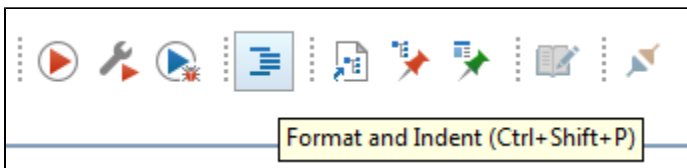
Copy the new terms and paste them in the RDF file just before the STATUS declaration at the end of the file. The pasted-in entities will look different than the ones that already exist:

```

24650 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5722" xml:lang="en">
24651   <skos:inScheme rdf:resource="http://sandbox.metadataregistry.org/uri/TV"/>
24652   <reg:status rdf:resource="http://metadataregistry.org/uri/RegStatus/1001"/>
24653   <skos:prefLabel xml:lang="en">University of Texas at Austin--Track and field</skos:prefLabel>
24654 </skos:Concept>
24655 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5723" xml:lang="en">
24656   <skos:inScheme rdf:resource="http://sandbox.metadataregistry.org/uri/TV"/>
24657   <reg:status rdf:resource="http://metadataregistry.org/uri/RegStatus/1001"/>
24658   <skos:prefLabel xml:lang="en">William M. Rice Institute--Caricatures and cartoons</skos:prefLabel>
24659 </skos:Concept>
24660 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5724" xml:lang="en"><skos:inScheme rdf:resource="http
24661 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5725" xml:lang="en"><skos:inScheme rdf:resource="http
24662 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5726" xml:lang="en"><skos:inScheme rdf:resource="http
24663 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5727" xml:lang="en"><skos:inScheme rdf:resource="http
24664 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5728" xml:lang="en"><skos:inScheme rdf:resource="http
24665 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5729" xml:lang="en"><skos:inScheme rdf:resource="http
24666 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5730" xml:lang="en"><skos:inScheme rdf:resource="http
24667 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5731" xml:lang="en"><skos:inScheme rdf:resource="http
24668 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5732" xml:lang="en"><skos:inScheme rdf:resource="http
24669 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5733" xml:lang="en"><skos:inScheme rdf:resource="http
24670 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5734" xml:lang="en"><skos:inScheme rdf:resource="http
24671 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5735" xml:lang="en"><skos:inScheme rdf:resource="http
24672 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5736" xml:lang="en"><skos:inScheme rdf:resource="http
24673 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5737" xml:lang="en"><skos:inScheme rdf:resource="http
24674 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5738" xml:lang="en"><skos:inScheme rdf:resource="http
24675 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5739" xml:lang="en"><skos:inScheme rdf:resource="http
24676 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5740" xml:lang="en"><skos:inScheme rdf:resource="http
24677 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5741" xml:lang="en"><skos:inScheme rdf:resource="http
24678 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5742" xml:lang="en"><skos:inScheme rdf:resource="http
24679 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5743" xml:lang="en"><skos:inScheme rdf:resource="http
24680 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5744" xml:lang="en"><skos:inScheme rdf:resource="http
24681 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5745" xml:lang="en"><skos:inScheme rdf:resource="http
24682 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5746" xml:lang="en"><skos:inScheme rdf:resource="http
24683 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5747" xml:lang="en"><skos:inScheme rdf:resource="http
24684 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5748" xml:lang="en"><skos:inScheme rdf:resource="http
24685 <skos:Concept rdf:about="http://sandbox.metadataregistry.org/uri/TV/5749" xml:lang="en"><skos:inScheme rdf:resource="http
24686 <!-- Status properties used in this document -->
24687 <skos:Concept rdf:about="http://metadataregistry.org/uri/RegStatus/1001">
24688   <skos:prefLabel xml:lang="en">Published</skos:prefLabel>
24689 </skos:Concept>
24690 </rdf:RDF>

```

To fix this, click on the **Format and Indent** button in the toolbar:



11. Your code should be indented and beautiful. Save the RDF file.
12. Finish by emailing the Metadata/Copy Cataloging Unit Team Lead to say the RDF file is updated.