

Getty Vocabularies OpenRefine Tutorial and Tips for Advanced Users

How to use the Getty Vocabularies Reconciliation Service for OpenRefine:
Step by Step Process from Program Installation to Implementation

released May 2019

latest revision: May 2023

Service by Gregg Garcia, Software Architect, Getty Digital

Tutorial revised by Getty Vocabulary Program and Gregg Garcia
Original documentation by Lindsey Gant

The Getty Vocabularies Reconciliation URL is <http://services.getty.edu/vocab/reconcile/>

Detailed instructions for first-time users are below.

Also below, FAQ, tips, and advanced techniques for using the service.

Document is subject to frequent additions and revisions

Latest revisions: addition of sections 2.7 and 2.8 by Jon Ward

Getty

Table of Contents

[1. About OpenRefine](#)

[2. Begin the Tutorial](#)

[2.1.-2. Steps 1-2: Download OpenRefine, etc.](#)

[2.3. Step 3: Connect to Getty Vocabularies Reconciliation Service](#)

[2.4. Step 4: Select the Vocabulary and Properties](#)

[2.5. Step 5: Begin OpenRefine Reconciliation](#)

[2.6. Step 6: Get Vocabulary identifiers](#)

[2.7. Step 7: Add Additional Fields Based on Reconciled Values](#)

[2.8. Step 8: Reconciling by Using Values \(IDs\) as Identifiers](#)

[3. Frequently Asked Questions and Tips](#)

[3.1. List of FAQs](#)

[Q1: Basics of Open Refine](#)

[Q1a: What are the basic functions in Open Refine?](#)

[Q1b. What does the scoring mean in Reconciliation results?](#)

[Q2: Creating concatenated name in ULAN](#)

[Q2a How to add a comma if name is already concatenated](#)

[Q3: How to separate broader contexts to reconcile TGN](#)

[Q4: Normalizing all caps in inconsistent ULAN data](#)

[Q5: Getting adjectival form of Geographic place from noun](#)

[Q6: Getting ULAN Nationalities](#)

[Q7: Getting full records XML](#)

[Q8: Fetching Birth Year based on ULAN reconciliation](#)

[Q9: Getting Facet Code for AAT](#)

[Q10: Getting Coordinates for TGN](#)

[Q11: Getting Arabic terms in TGN](#)

[Q12: Qualifying or constraining reconciliation criteria](#)

[3.2 Practical tips for using OR and the Reconciliation Service](#)

[3.2.1. Reconcile data sets of manageable size and content](#)

[3.2.2. Use auto-match with care](#)

[3.2.3. Normalize your data prior to reconciliation](#)

[3.2.4. Inserting IDs: Use Best Match with caution](#)

[3.2.5. When to format cells in Excel prior to loading in OpenRefine](#)

[3.2.6. Workflow for adding and reconciling new terms](#)

[4. Advanced OpenRefine Techniques using the Getty Vocabularies](#)

[4.1. OpenRefine ID Sync Example Using XML Web Services](#)

[4.2. OpenRefine Data Enrichment Example Using Full XML from Web Service](#)

[4.3. OpenRefine Example: Parse All Variant Terms from Full XML](#)

[4.4. OpenRefine Example: Get TGN Place Coordinates from SPARQL Endpoint](#)

[4.5. OpenRefine Example: Query Reconciliation Service Using HTTP](#)

[Online Resources](#)

[5. Library of GREL, SPARQL, SQL, and other expressions](#)

[5.1 GREL](#)

[5.2. SPARQL](#)

[5.3. SQL](#)

1. About OpenRefine

Quick Reference: If you are already an OpenRefine user, you can use the Getty Vocabulary Reconciliation Service by entering this URL from within OpenRefine:

<http://services.getty.edu/vocab/reconcile/>

For novice users: *OpenRefine*, formerly known as *Google Refine*, is a standalone open-source desktop application that may be used for data cleanup and transformation to other formats. OpenRefine allows data wrangling using methods similar to spreadsheet applications, with rows of data organized in columns. You may correct inconsistencies, devise global edits, and do many other functions on your data set with OpenRefine.

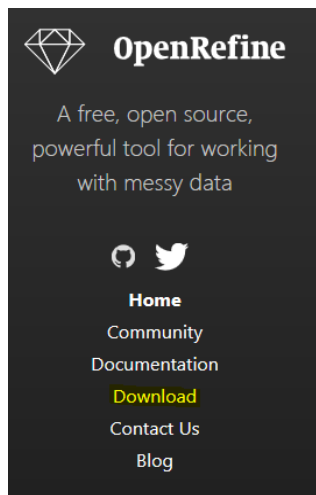
For a general tutorial on how to use OpenRefine, please see openrefine.org and various other sites devoted to this topic.

This document focuses on the Getty Vocabulary reconciliation service that is designed for use with documents that are opened in OpenRefine. Reconciliation is a semi-automated process by which you can match your data records, such as for artists, places, or work types, to the Getty Vocabularies: ULAN, TGN, or AAT. Through the OpenRefine Reconciliation Service using the Getty Vocabularies, matches are suggested. However, note that reconciliation is not a fully automated process, because this would result in inaccurate matches. Human oversight and judgment are essential. Through the reconciliation service, the user has the option to decide which data are modified by selecting from a list of results.

2. Begin the Tutorial

2.1. Step 1: Download OpenRefine

You must be in OpenRefine to use the Getty Vocabulary Reconciliation Service. If you have not already done so, download OpenRefine from this site: openrefine.org



Welcome!

OpenRefine (formerly Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

OpenRefine is available in English, Chinese, Spanish, French, Russian, Portuguese (Brazil), German, Japanese, Italian, Hungarian, Hebrew, Filipino, Cebuano, Tagalog

OpenRefine is supported by:



2.1.1. Download

Select the 'Download' option from the left column, as highlighted above. Choose a version for the Windows kit, Mac kit, or Linux kit, depending upon your preferred operating system. For this demonstration, we are using the Windows OS.

OpenRefine 3.2

The final release of OpenRefine 3.2. Please BACKUP your workspace directory before installing and report any problems that you encounter.

The final release of 3.2 was released on July 26, 2019. A change log is provided on [the release page](#).

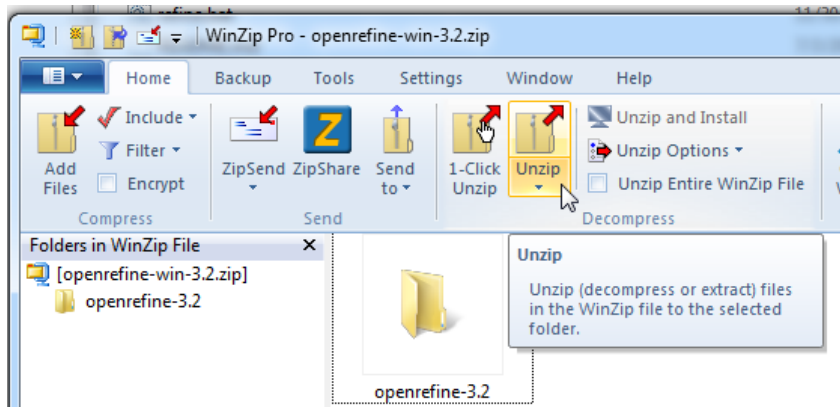
- **Windows kit**, Download, unzip, and double-click on *openrefine.exe*. If you're having issues with the above, try double-clicking on *refine.bat* instead.
- **Mac kit**, Download, open, drag icon into the Applications folder and double click on it.
- **Linux kit**, Download, extract, then type `./refine` to start.

Which version should you download? We recommend that you use the penultimate version (the next to last version) or the latest version. As of this writing, the most recent version is 3.7.2.

NOTE: For each project, all users should always use the same version of OpenRefine. For example, if two people are editing a project using OpenRefine, one person should not use Version 3.6 if the other person is using Version 3.7.

2.1.2. Unzip and save

After installation is complete, unzip the file and save to the desktop or a folder on your computer.



2.1.3. Open the OpenRefine program

To open the program, select (click) the application file **openrefine.exe** (highlighted below).

Name	Date modified	Type	Size
licenses	10/23/2018 11:19 ...	File folder	
server	10/23/2018 11:19 ...	File folder	
webapp	10/23/2018 11:19 ...	File folder	
LICENSE.txt	9/16/2018 10:20 AM	Text Document	2 KB
openrefine.exe	9/16/2018 12:00 PM	Application	39 KB
openrefine.l4j.ini	9/16/2018 10:20 AM	Configuration sett...	1 KB
refine.bat	9/16/2018 10:20 AM	Windows Batch File	7 KB
refine.ini	9/16/2018 10:20 AM	Configuration sett...	1 KB

Once the application is selected, a command line box will open (the black window below), initiating the web application. You will keep this box open in the background. OpenRefine is a web-based application that works in your browser.

(When you finish your session and have closed OpenRefine, close this box using Control-C rather than by simply clicking the 'x' at the upper right.)

```
openrefine
14:10:22.639 [ refine_server] Starting Server bound to '127.0.0.1:3333' (0ms)
14:10:22.733 [ refine_server] Initializing context: '/' from 'C:\Users\LGant\Downloads\openrefine-win-3.0\openrefine-3.0\webapp' (94ms)
14:10:32.730 [ refine] Starting OpenRefine 3.0 [TRUNK]... (9997ms)
14:10:32.730 [ refine] initializing FileProjectManager with dir (0ms)
14:10:32.746 [ refine] C:\Users\LGant\AppData\Local\OpenRefine (16ms)
14:10:38.546 [ database-extension] Initializing OpenRefine Database... (5800ms)
14:10:38.546 [ database-extension] Database Extension Mount point /extension/database/ [*] (0ms)
14:10:38.546 [ database-extension] Registering Database Extension Commands ..... (0ms)
14:10:38.842 [ database-extension] Database Extension Command Registration done!! (296ms)
14:10:38.842 [ database-extension] Database Operations Registered successfully... (0ms)
14:10:38.842 [ database-extension] Database Functions Registered successfully... (0ms)
14:10:38.920 [ DatabaseModuleImpl] *** Database Extension Module Initialization Completed!! *** (78ms)
14:10:51.481 [ org.morthbay.log] /favicon.ico (12561ms)
```

Once you click *openrefine.exe*, it may take a few moments for the application to open. Once the app opens, it should look like this:



2.2. Step 2: Import spreadsheet

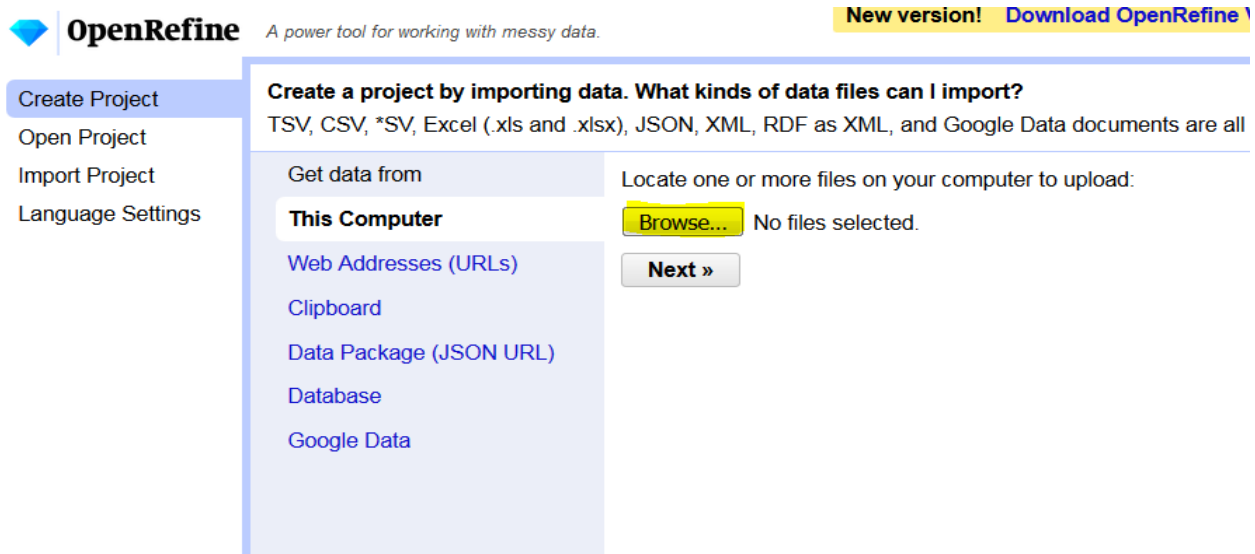
2.2.1. Import your file in OpenRefine

With 'Create Project' highlighted in the left column, click the 'Browse' button to find your file. Import the spreadsheet or other file that you wish to edit.

Alternatively, you may choose 'Open Project' or 'Import Project' to continue work on an in-process project

Formats may include Excel, CSV, XML, JSON, or others

NOTE: *Your spreadsheet must have values in column A for every cell down the column in order for OpenRefine to read each row as a record. If Column A is missing values for some rows, create a column A and number the cells in Excel before importing the column. To number the cells, the formula in Excel is **=row(A1)**, then pull the autofill handle down the column to apply this formula to cells.*



2.2.2. Create a project

Once the file is uploaded, you will see a Preview window, illustrated below. Give your project a name in the 'Project Name' field, or use the default name.

Select 'Create Project' at the upper right of the page.

You will be taken to another page from which you can edit the cells and columns on the spreadsheet and use reconciliation services.

« Start Over		Configure Parsing Options		Project name		XYZ.xls		Tags		Create Project »	
artistName	artistNonPrefName	artistBirth	artistDeath	artistNationality	artistBirthPlace 1	artistBirthPlace 2	artistBirthPlace 3	artistBirthPlace 4	artistBirthPlaceNation	artistRole	artist
1. Aaronson, Yakov	אהרונסון, יעקב	1924	2100	Israeli					Israel	31407/photographer	
2. Abadi, Abed	עבד, עבד	1942	2100	Israeli					Israel	31261/painter	
3. Abadi, Shay	עבאדי, שי	1965	2100	Israeli	Jerusalem	Israel				31100/artist	
4. Abakanowicz, Magdalena	אבאקונוביץ' מגדאלנה	1930	2017	Polish					Poland	34222/weaver	3126
5. Abba, Irit	אבא, עירית	1953	2100	Israeli					Israel	31557/designer	3322
6. Abbasi, Riza (attributed to)	אבאסי, ריזה, (מיוחס ל)	1200	2100	Iranian					Iran	31261/painter	
7. Abbasi, Riza (school of)	אבאסי, ריזה (אסכולה)	1200	2100	Iranian					Iran	Miniatures painter	
8. Abbasi, Riza	אבאסי, ריזה	1565	1635	Iranian					Iran	31261/painter	
9. Abbott, Berenice	אבוט, ברניס	1898	1991	American					USA	31407/photographer	
10. Emil Aboud, Jumana	אמיל עבוד, ג'ומאנה	1971	2100	Israeli	Shefaram	Israel				31261/painter	
11. Abdullah Freres	האחים עבדוללה	1858	1899	Turkish					Turkey	31407/photographer	3126
12. Abecassis, Raphael	אבקסיס, רפאל	1953	2100	Moroccan	Marrakesh	Morocco				31261/painter	
13. Abel, Myer	מאיייר, אבל	1904	1946	American					USA	31261/painter	3143

2.3. Step 3: Connect to Getty Vocabularies Reconciliation Service

2.3.1. Choose the column for reconciliation

At the top of the column that you wish to reconcile, click the drop-down list; hover over 'Reconcile,' and select 'Start Reconciling.'

The screenshot shows the OpenRefine interface with a table of 2252 rows. The table has columns for 'pref_name', 'name direct orc', 'birth_year', and 'death_year'. A context menu is open over the 'birth_year' column header, with the 'Reconcile' option selected and 'Start reconciling...' highlighted. A mouse cursor is pointing at 'Start reconciling...'.

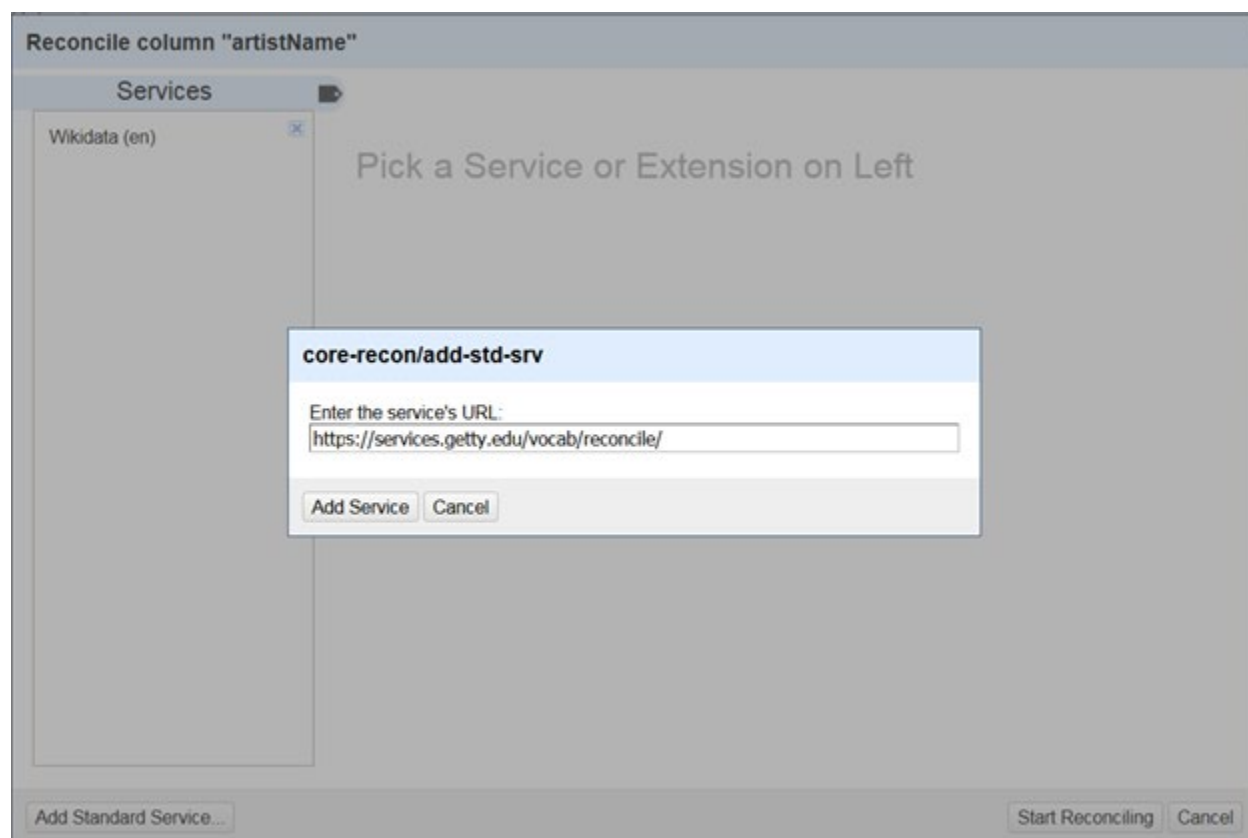
	All	pref_name	name direct orc	birth_year	death_year
9.			n McLaughlin	1919	
10.			Abbéma	1858	1927
11.			ce Abbott	1898	1991
12.			Phmps	Abbott	
13.		Abbott, Susan	Susan		

2.3.2. Choose the reconciliation service

From the page listing services, select 'Add Standard Service' at the bottom left of the page.

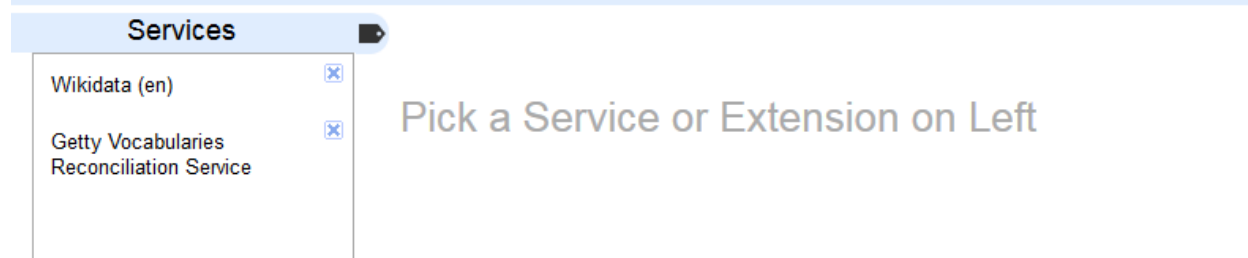
The screenshot shows a dialog box titled "Reconcile column 'prefLabel@en'". On the left, there is a "Services" panel with a list containing "Wikidata (en)". To the right of this panel, the text "Pick a Service or Extension on Left" is displayed. At the bottom of the dialog, there are three buttons: "Add Standard Service..." (highlighted in yellow), "Start Reconciling", and "Cancel".

In order to connect to the Getty Vocabulary Reconciliation Service, input this URL:
<https://services.getty.edu/vocab/reconcile/>
Then select 'Add Service.'



The Getty Vocabularies service will now appear in your Services list, as illustrated below. Every time hereafter that you use this same version of OpenRefine on this computer, the Getty Vocabularies service will be in your list.

Reconcile column "name direct order"



2.4. Step 4: Select the Vocabulary and Properties

2.4.1. Choose the Vocabulary

In the Services list, click on 'Getty Vocabularies Reconciliation Service.'

Select one of the Vocabularies: ULAN, TGN, or AAT.

Auto-matching: If there is a good correspondence between your source data set and the target Getty Vocabulary, you may wish to use 'Auto-match candidates with high confidence,' lower left. Auto-match tends to work better for ULAN than AAT or TGN, due to the typical variables (properties) in source data. See further discussion under [Tips](#) below.

Reconcile column "pref_name"

Reconcile each cell to an entity of one of these types:

- ULAN search
/ulan
- TGN search
/tgn
- AAT search
/aat
- Search all Vocab
/all

Also use relevant details from other columns:

NMWA_birth_ID	<input type="checkbox"/>	
death_place	<input type="checkbox"/>	
NMWA_death_place 1	<input type="checkbox"/>	
NMWA_death_reconcile_id	<input type="checkbox"/>	
NMWA_death_place 2	<input type="checkbox"/>	
NMWA_death_place 3	<input type="checkbox"/>	
NMWA_death_id	<input type="checkbox"/>	
nationality	<input checked="" type="checkbox"/>	nat

Select an item from the list:

- nationality
- nationalityPref
- nationalityNonPref

Reconcile against type:

Reconcile against no particular type

Auto-match candidates with high confidence **caution**

Maximum number of candidates to return

Add Standard Service... **Start Reconciling** Cancel

2.4.2. Optional: Choose properties

In most cases, you will have better results in matching if – in addition to matching on the name or term -- you can compare *properties* in the source data to *properties* in the Getty Vocabulary. If your dataset contains columns that can help identify or qualify the entities during reconciliation, OpenRefine reconciliation

results will take these columns into account in the matching scores. For example, for ULAN, perhaps you have a column for the nationality the artist.

List of properties available for each Getty Vocabulary:

For AAT

- note – the scope note or definition for the concept
- broaderExt – any broader concept
- label – any additional term for the concept (e.g., Used For Term, Alternate Descriptors)

For ULAN

- agentType – the role of a person or corporate body
- nationality – the nationality of a person or corporate body
- gender – the sex of a person
- description – display biography for the person or corporate body
- birthPlace – the place where a person was born
- deathPlace – the place where the person died
- birthDate – the date of birth for a person; in ULAN, this is estimated when necessary
- deathDate – the date of death for a person, in ULAN this is estimated when necessary
- location – the location of some activity performed by a person or the location of a corporate body
- startDate – for an activity, a date associated with the beginning of a span for the activity of a person or corporate body
- endDate – for an activity, a date associated with the end of a span for the activity of a person or corporate body
- type – can be either schema:Person or schema:Organization (corporate body)
- parent – the broader context for the record, a hierarchical ‘ancestor’ or a facet name

For TGN

- broaderExt – any broader context associated with the place being reconciled
- placeType – place type term describing the place being reconciled (*nation, inhabited place, city, etc.*)

Input the property selections under the column 'As Property.'

In the example below, the available columns from the source spreadsheet are displayed on the left under the heading 'Column.' Beside each column name is a box into which you can put the name of the property in the target Vocabulary against which you wish to reconcile. If you start typing a value, you can auto-fill by choosing suggestions provided for ULAN.

Reconcile column "pref_name" » Access Service API

Reconcile each cell to an entity of one of these types:

- ULAN search /ulan
- TGN search /tgn
- AAT search /aat
- Search all Vocabs /all

Also use relevant details from other columns:

Column	Include?	As Property
name direct order	<input type="checkbox"/>	
birth_year	<input checked="" type="checkbox"/>	birthDate
death_year	<input checked="" type="checkbox"/>	d

Select an item from the list:

- description
- deathPlace
- deathDate

xyz_new_nationality

bio_description

scope_note

pref_role

Reconcile against type:

Reconcile against no particular type

Auto-match candidates with high confidence

Maximum number of candidates to return

When you finish adding properties, click 'Start Reconciling.' You will be taken back to the spreadsheet view, where progress is tracked.

file xlsx [Permalink](#)

Reconcile cells in column pref_name to type /ulan
61% complete Cancel

2251 rows

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

All	pref_name	name direct orc	birth_year	death_year	birth_place	death_plac
☆	1. (Kolbrun Björgolfsdottir), Kogga	Kogga (Kolbrun Björgolfsdottir)	1952			

2.5. Step 5: Begin OpenRefine Reconciliation

2.5.1. Verify matches

After the reconciliation process ends, verify the matches. You may wish to check the matches by hand, if possible, particularly if the ranking is 40 or lower.

The possible matches (in this example to ULAN) appear in the column in blue under the source value being matched (*Aalto, Aino*), sorted by ranking (in parentheses). You can see information from the ULAN match by hovering over the blue name (hover in versions 3.2 and higher; in 3.1, you must click to see the ULAN record), when a window will display the ULAN data. In the example below, the name and birth and death dates match.

Click on 'match this cell.'

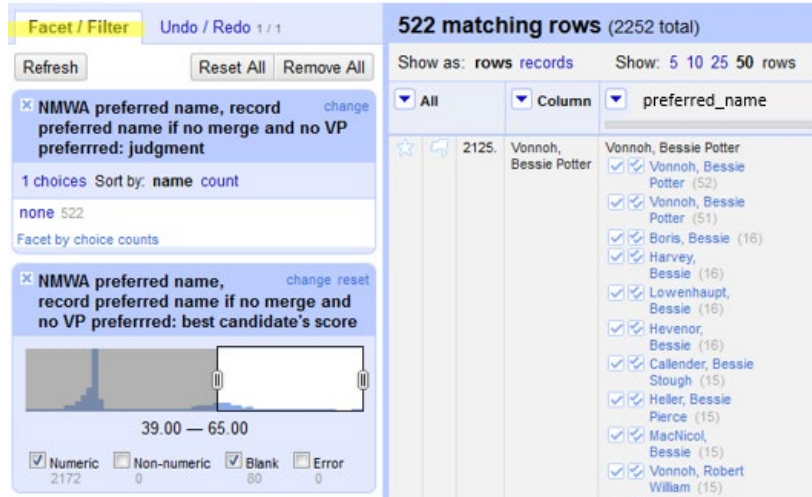
If you could have multiple values in this column that should be reconciled to the same Vocabulary record, you could click 'match all identical cells.' We advise that you use this option with caution, given the high possibility for homographs in the Vocabulary data.



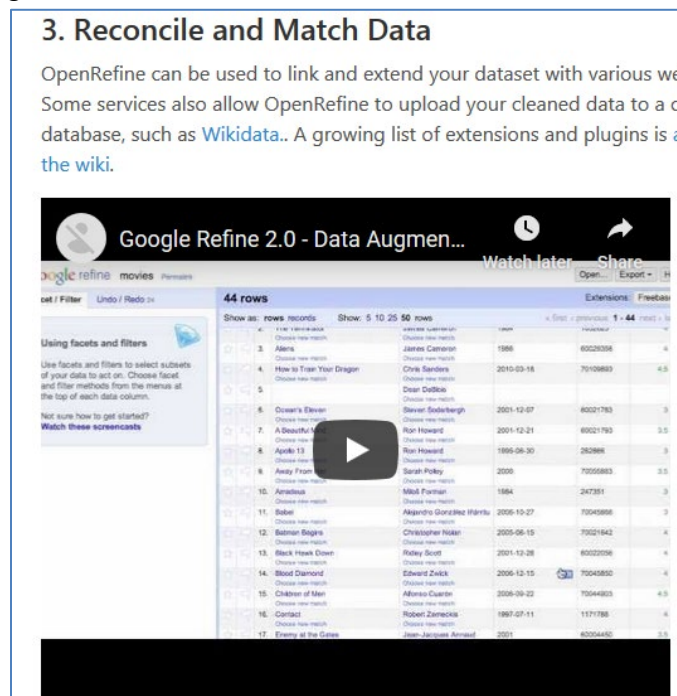
Do you need more information before deciding? Left click on the blue name to go to the full Web record for the Vocabulary entity.



You can use ranking, filtering, faceting, and other tools to review your matches.



Go to openrefine.org for general information about reconciling, filtering, and faceting.



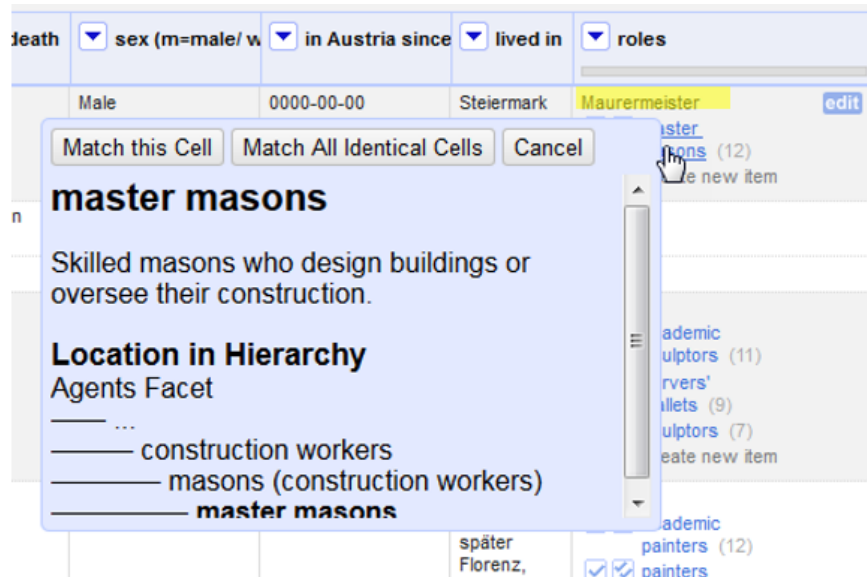
- **Note on scoring:** The Vocabularies Reconciliation Service for OpenRefine does not use normalized scaling for matching results. The search gives relative results based on how many data points match Vocabularies data (text matching, broader context, birth/death dates, etc.). In general, the better matches have higher scores and it is up to the user to determine what the acceptable threshold for an actual match should be. In this same manner, the service intentionally does not do auto-matching

because we did not want to assume any expertise about the source data and create bad matches.

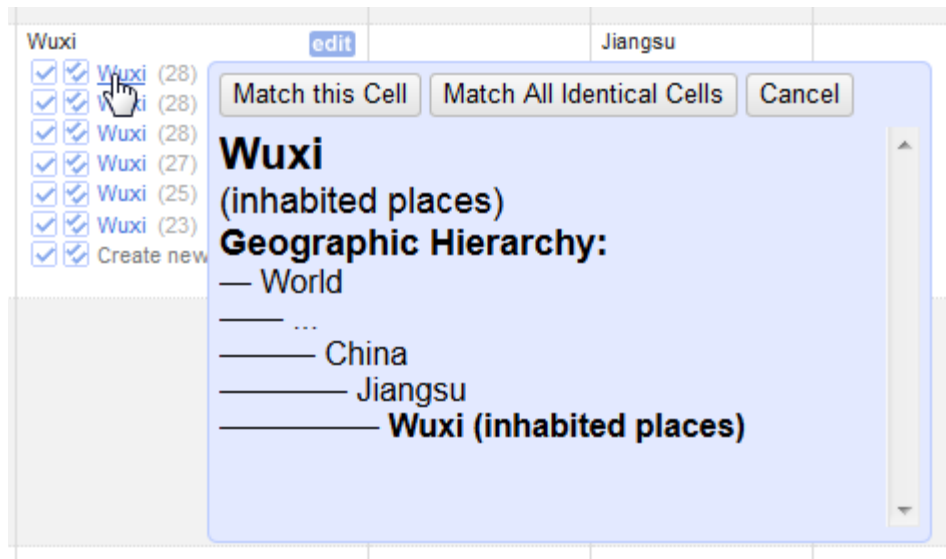
2.5.2. Examples in AAT and TGN

Most examples in this document have so far featured ULAN. However, the reconciliation service also works in the same way for AAT and TGN matching.

In the AAT example below, you can see in the AAT pop-up display how the artist role 'Maurermeister' in German has matched to the AAT record for 'master masons.' Left click the blue match to see the full AAT record online.



In the example for reconciling using TGN below, several homographs have nearly equal ranking. You can left click to go to the full records online in order to see the geographic coordinates and other information that could assist you in making a decision.

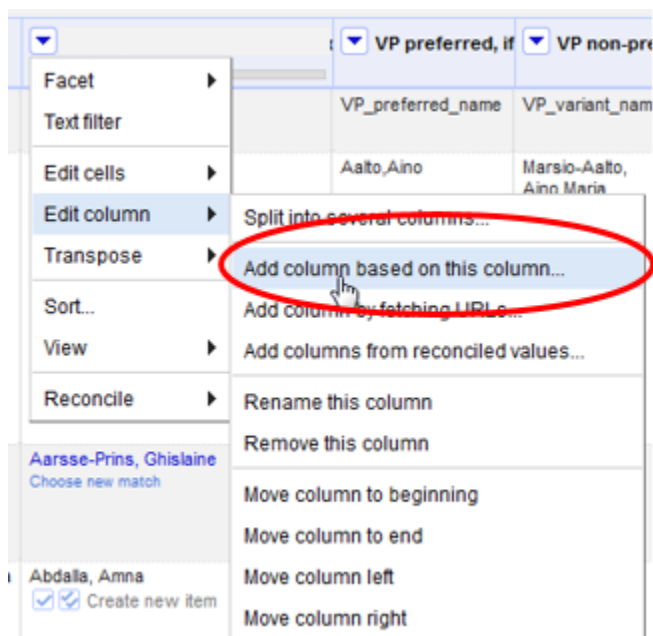


2.6. Step 6: Get Vocabulary identifiers

2.6.1. Get Vocabulary IDs for matches

After accepting the matches where appropriate, you will probably wish to get the unique identifiers for the matched Vocabulary records.

First, add a new column based on the column that you have reconciled. Give it an appropriate name, such as 'ULAN_ID.'

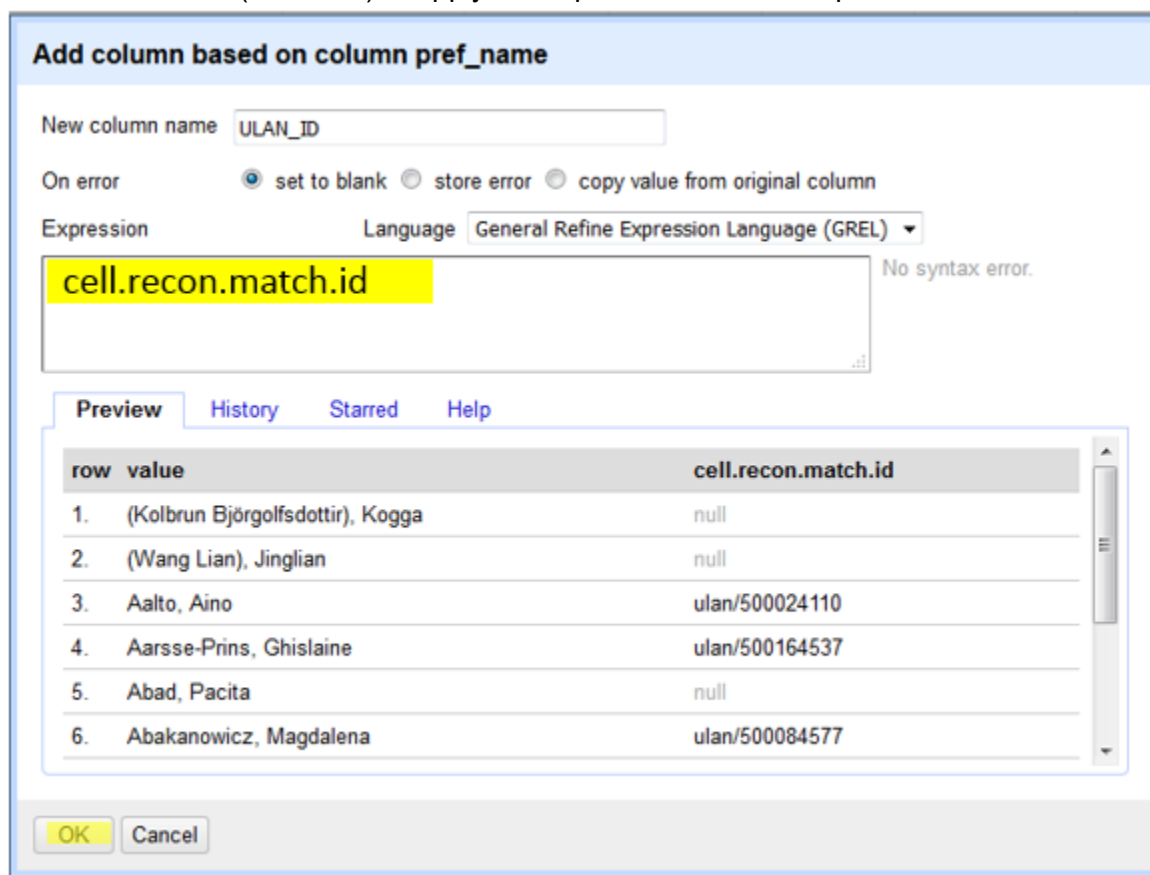


A window will appear, illustrated below, allowing you to define what will be included in the new column.

In the Expression area (illustrated below), input the expression 'cell.recon.match.id'

A sample set illustrating the results will appear. If there was no match, 'null' is inserted rather than the Vocabulary ID.

Click 'OK' (lower left) to apply the expression to the full spreadsheet.



The dialog box is titled "Add column based on column pref_name". It contains the following fields and options:

- New column name:
- On error: set to blank store error copy value from original column
- Expression: Language: No syntax error.

Below the expression field is a preview table with the following data:

row	value	cell.recon.match.id
1.	(Kolbrun Björgolfsdottir), Kogga	null
2.	(Wang Lian), Jinglian	null
3.	Aalto, Aino	ulan/500024110
4.	Aarsse-Prins, Ghislaine	ulan/500164537
5.	Abad, Pacita	null
6.	Abakanowicz, Magdalena	ulan/500084577

At the bottom of the dialog are "OK" and "Cancel" buttons.

2.6.2. Get IDs for Best Match?

Formerly this guide suggested using the Expression for 'best match' rather than 'match.' The expression for best match is 'cell.recon.best.id' With this expression, the ID is automatically inserted into the column before the match has been accepted by the editor.

After extensive testing, we advise against 'best match' because it often results in bad matches. However, it may be useful for big data sets where a certain

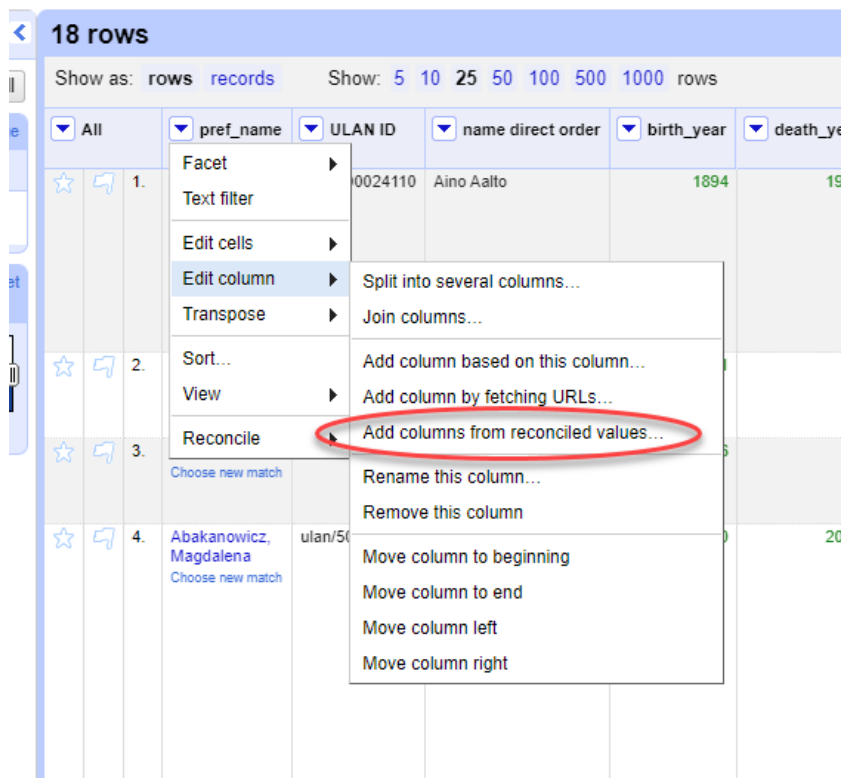
percentage of error is acceptable, or if you apply it to only those matches with exceptionally high ranking. For discussion, see [Tips](#).

2.7. Step 7: Add Additional Fields from Reconciled Values

Users can now add chosen data from reconciled Getty Vocabulary terms in OpenRefine by selecting the “Add Columns from Reconciled Values” function. This will add data from those Getty Vocabulary records in additional columns. The fields chosen for this function are the most commonly requested for each of the Getty Vocabularies.

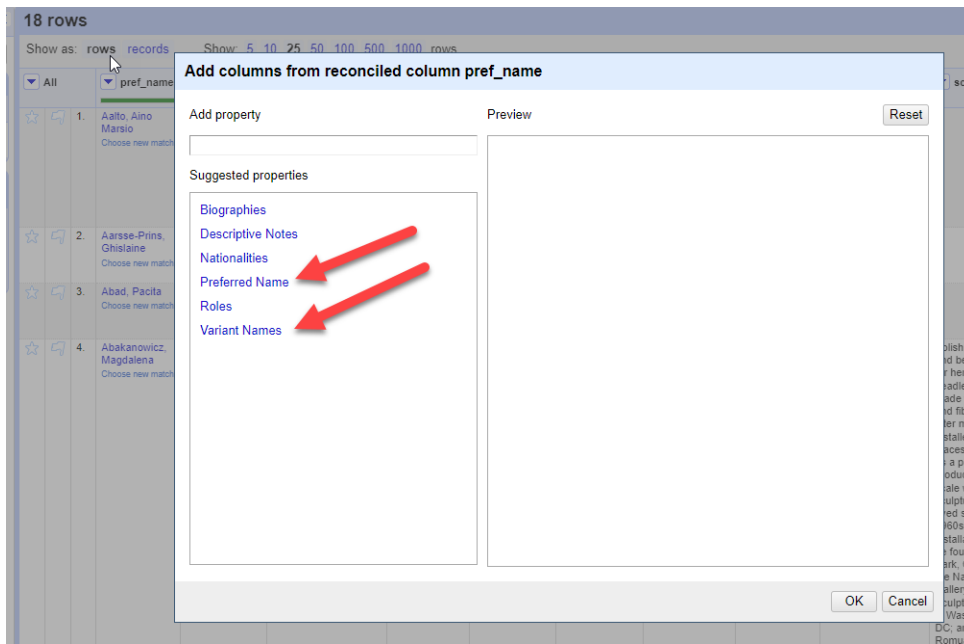
2.7.1 Adding additional fields from reconciled Vocabulary concepts

On the column with reconciled terms, go to Edit Column > Add columns from reconciled values.

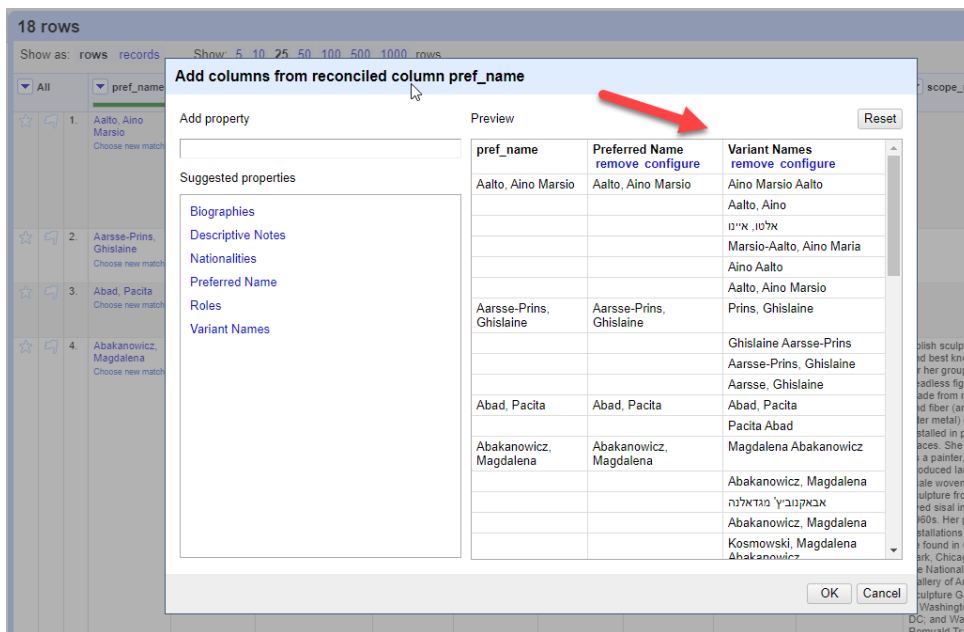


A box will appear with a select group of six ULAN fields (Biography, Descriptive Note, Nationality, Preferred Name, Role, Variant Name), the contents of which can be placed

in additional columns in your OpenRefine project. For example, let's say we would want to add columns that contain the Preferred Terms and Variant Terms.



Once selected, the contents of those fields corresponding to your reconciled terms appear to the right (you may remove or configure columns in this step).



Click "OK" and they will be added to your spreadsheet next to the column with reconciled terms.

59 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

All	pref_name	Preferred Name	Variant Names	ULAN ID	name direct order
☆ ↻ 1.	Aalto, Aino Marsio Choose new match	Aalto, Aino Marsio	Aino Marsio Aalto	ulan/500024110	Aino Aalto
☆ ↻ 2.			Aalto, Aino		
☆ ↻ 3.			אלטו, אינו		
☆ ↻ 4.			Marsio-Aalto, Aino Maria		
☆ ↻ 5.			Aino Aalto		
☆ ↻ 6.			Aalto, Aino Marsio		
☆ ↻ 7.	Aarsse-Prins, Ghislaine Choose new match	Aarsse-Prins, Ghislaine	Prins, Ghislaine	ulan/500164537	Ghislaine Aarsse-Prins
☆ ↻ 8.			Ghislaine Aarsse-Prins		
☆ ↻ 9.			Aarsse-Prins, Ghislaine		
☆ ↻ 10.			Aarsse, Ghislaine		
☆ ↻ 11.	Abad, Pacita Choose new match	Abad, Pacita	Abad, Pacita	ulan/500487777	Pacita Abad
☆ ↻ 12.			Pacita Abad		
☆ ↻ 13.	Abakanowicz, Magdalena Choose new match	Abakanowicz, Magdalena	Magdalena Abakanowicz	ulan/500084577	Magdalena Abakanowicz

The additional fields that can be added from TGN data are shown here (Coordinates, Descriptive Note, Parent Hierarchy, Place Type, Preferred Term, Variant Terms):

Add columns from reconciled column birth_place

Add property

Preview Reset

Suggested properties

- [Coordinates](#)
- [Descriptive Notes](#)
- [Parent Hierarchy](#)
- [Place Types](#)
- [Preferred Term](#)
- [Variant Terms](#)

Basco	[122.20.4167]	Batanes	inhabited places
		Philippines	provincial capitals
		Asia	
		World	
<not reconciled>			
Falenty	[20.924.52.137]	Mazowieckie	inhabited places
		Poland	villages
		Europe	
		World	
<not reconciled>			
<not reconciled>			
<not reconciled>			
<not reconciled>			
<not reconciled>			
<not reconciled>			
<not reconciled>			
<not reconciled>			

OK Cancel

Similarly, select AAT fields that can be added are shown here (Descriptive Note, Parent Hierarchy, Preferred Term, Variant Terms):

Add columns from reconciled column nationality

Add property

Preview Reset

Suggested properties

- [Descriptive Notes](#)
- [Parent Hierarchy](#)
- [Preferred Term](#)
- [Variant Terms](#)

nationality	Parent Hierarchy remove configure	Preferred Term remove configure
Finnish (culture or style)	Scandinavian	Finnish (culture or style)
	European regions	
	European	
	<styles, periods, and cultures by region>	
	Styles and Periods (hierarchy name)	
	Styles and Periods Facet	
<not reconciled>		
<not reconciled>		
<not reconciled>		
<not reconciled>		
<not reconciled>		
French (culture or style)	European regions	French (culture or style)
	European	
	<styles, periods, and cultures by region>	
	Styles and Periods (hierarchy name)	

OK Cancel

For those needing additional fielded data not shown here, please see the guidelines for parsing our full-record XML data in Section 4.

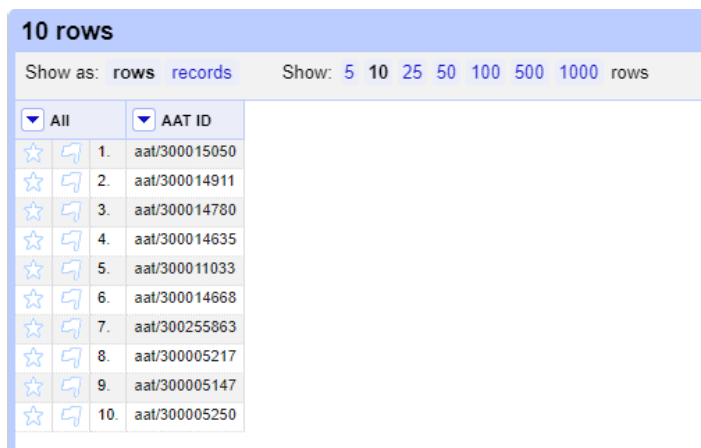
2.8. Step 8: Reconciling by Using Values (IDs) as Identifiers

Users can now reconcile (and expand) their Vocabulary data with our service simply by starting with a list of IDs.

2.8.1 Using Values as Identifiers

Rather than reconcile a list of terms, if a user solely has a list of Getty Vocabulary IDs, reconciling and grabbing additional information from the service is just as easy.

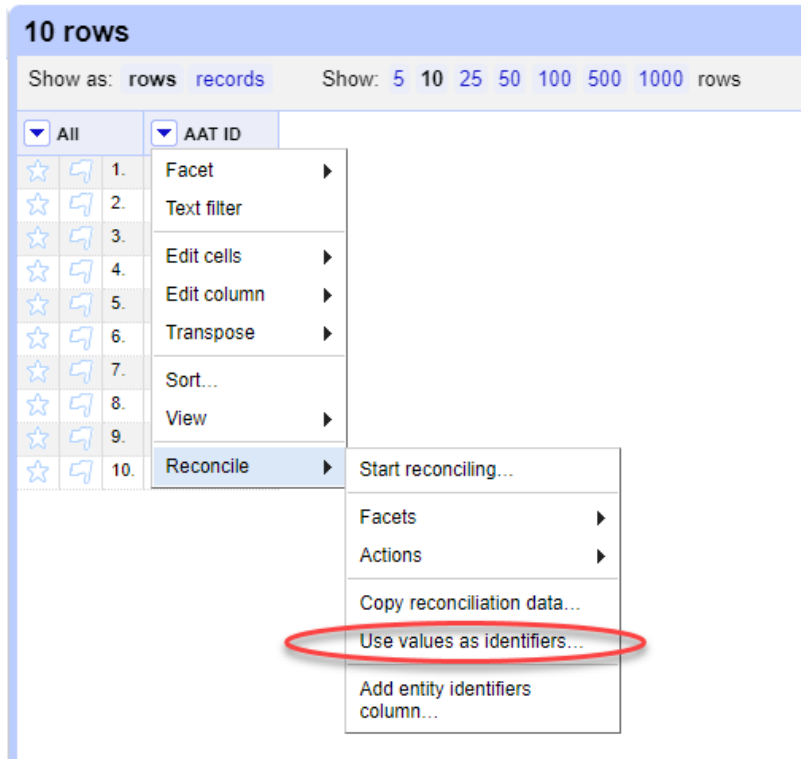
First, make sure your IDs have the proper prefix attached, based on the Vocabulary they are from (aat/, tgn/, ulan/):



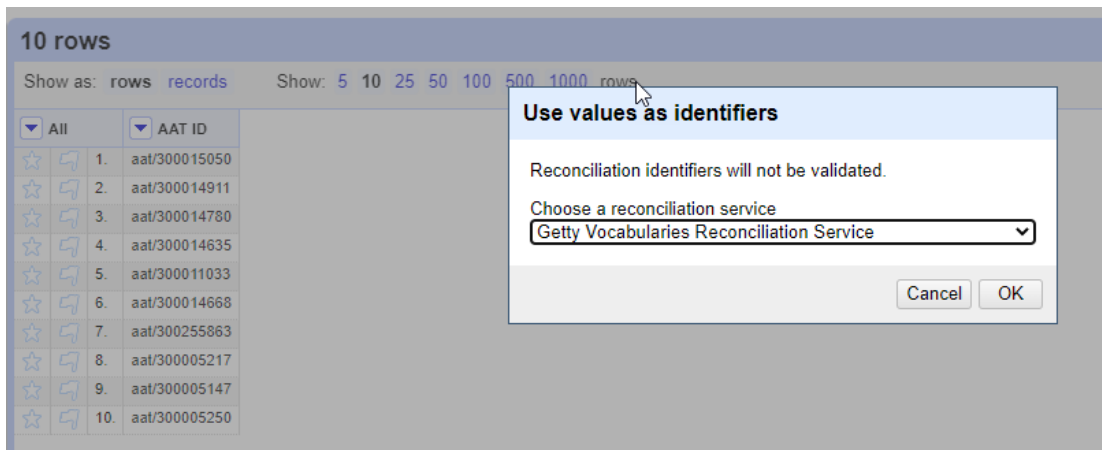
The screenshot shows a table interface with a header bar indicating '10 rows'. Below the header, there are options to 'Show as: rows records' and a 'Show:' dropdown menu with values 5, 10, 25, 50, 100, 500, 1000. The table has two columns: 'All' and 'AAT ID'. The 'AAT ID' column contains 10 rows of IDs, each preceded by a number from 1 to 10. Each row also has a star icon and a refresh icon.

All	AAT ID
★ ↻	1. aat/300015050
★ ↻	2. aat/300014911
★ ↻	3. aat/300014780
★ ↻	4. aat/300014635
★ ↻	5. aat/300011033
★ ↻	6. aat/300014668
★ ↻	7. aat/300255863
★ ↻	8. aat/300005217
★ ↻	9. aat/300005147
★ ↻	10. aat/300005250

Go to the column dropdown list and to Reconcile > Use values as identifiers:



Choose the Getty Vocabularies reconciliation service:



Click “OK” and it will change the IDs into reconciled data:

10 rows

Show as: **rows** records Show: 5 10 25 50 100 500 1000 rows

All		AAT ID	
☆	↶	1.	aat/300015050 <small>Choose new match</small>
☆	↶	2.	aat/300014911 <small>Choose new match</small>
☆	↶	3.	aat/300014780 <small>Choose new match</small>
☆	↶	4.	aat/300014635 <small>Choose new match</small>
☆	↶	5.	aat/300011033 <small>Choose new match</small>
☆	↶	6.	aat/300014668 <small>Choose new match</small>
☆	↶	7.	aat/300255863 <small>Choose new match</small>
☆	↶	8.	aat/300005217 <small>Choose new match</small>
☆	↶	9.	aat/300005147 <small>Choose new match</small>
☆	↶	10.	aat/300005250 <small>Choose new match</small>

To add additional Vocabulary data from these reconciled IDs, go to Edit Column > Add columns from reconciled values:

10 rows

Show as: **rows** records Show: 5 10 25 50 100 500 1000 rows

All		AAT ID	
☆	↶	1.	Facet
☆	↶	2.	Text filter
☆	↶	3.	Edit cells
☆	↶	4.	Edit column
☆	↶	5.	Transpose
☆	↶	6.	Sort...
☆	↶	7.	View
☆	↶	8.	Reconcile
☆	↶	7.	aat/300255863 <small>Choose new match</small>
☆	↶	8.	aat/300005217 <small>Choose new match</small>
☆	↶	9.	aat/300005147 <small>Choose new match</small>
☆	↶	10.	aat/300005250 <small>Choose new match</small>

- Facet
- Text filter
- Edit cells
- Edit column**
 - Split into several columns...
 - Join columns...
 - Add column based on this column...
 - Add column by fetching URLs...
 - Add columns from reconciled values...**
 - Rename this column...
 - Remove this column
 - Move column to beginning
 - Move column to end
 - Move column left
 - Move column right
- Transpose
- Sort...
- View
- Reconcile

This will bring up a box with numerous fields to choose from (see 2.7), the data for which you may bring up in separated columns. In this example of AAT IDs, let's choose Preferred Term and Parent Hierarchy:

The dialog box 'Add columns from reconciled column AAT ID' displays a list of 10 rows on the left. The 'Suggested properties' list includes:

- Biographies (ulan)
- Coordinates (tgn)
- Descriptive Notes (aat)
- Descriptive Notes (tgn)
- Descriptive Notes (ulan)
- Nationalities (ulan)
- Parent Hierarchy (aat)
- Parent Hierarchy (tgn)
- Place Types (tgn)
- Preferred Name (ulan)
- Preferred Term (aat)
- Preferred Term (tgn)
- Roles (ulan)
- Variant Names (ulan)
- Variant Terms (aat)

The 'Preview' table shows the following data:

AAT ID	Preferred Term	Parent Hierarchy
aat/300015050	oil paint (paint)	<paint by composition or origin>
		paint (coating)
		<coating by form>
		coating (material)
		<materials by function>
		materials (substances)
		Materials (hierarchy name)
		Materials Facet
aat/300014911	japan (enamel)	enamel (fused coating)
		<coating by composition or origin>
		coating (material)
		<materials by function>
		materials (substances)
		Materials (hierarchy name)
		Materials Facet
aat/300014780	pozzolana (clay)	<clay by composition or origin>
		clay
		inorganic material
		<materials by composition>

Click "OK" and it will place them into your OpenRefine project:

66 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

All	AAT ID	Preferred Term	Parent Hierarchy
☆ ↻ 1.	aat/300015050 Choose new match	oil paint (paint)	<paint by composition or origin>
☆ ↻ 2.			paint (coating)
☆ ↻ 3.			<coating by form>
☆ ↻ 4.			coating (material)
☆ ↻ 5.			<materials by function>
☆ ↻ 6.			materials (substances)
☆ ↻ 7.			Materials (hierarchy name)
☆ ↻ 8.			Materials Facet
☆ ↻ 9.	aat/300014911 Choose new match	japan (enamel)	enamel (fused coating)
☆ ↻ 10.			<coating by composition or origin>
☆ ↻ 11.			coating (material)
☆ ↻ 12.			<materials by function>
☆ ↻ 13.			materials (substances)
☆ ↻ 14.			Materials (hierarchy name)
☆ ↻ 15.			Materials Facet
☆ ↻ 16.	aat/300014780 Choose new match	pozzolana (clay)	<clay by composition or origin>
☆ ↻ 17.			clay
☆ ↻ 18.			inorganic material
☆ ↻ 19.			<materials by composition>
☆ ↻ 20.			materials (substances)
☆ ↻ 21.			Materials (hierarchy name)
☆ ↻ 22.			Materials Facet
☆ ↻ 23.	aat/300014635 Choose new match	filament (material)	<materials by form>
☆ ↻ 24.			materials (substances)
☆ ↻ 25.			Materials (hierarchy name)

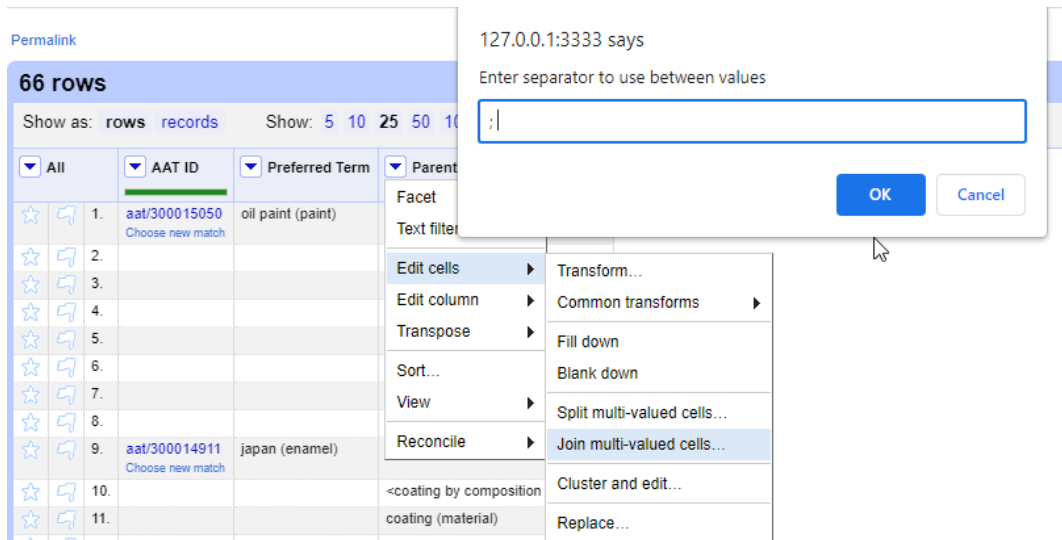
You'll notice the parent hierarchy is separated into cells. Very likely, you would want to create a single string to display a Parent Hierarchy. This can be done simply, by going to the dropdown, Edit cells > Join multi-valued cells:

66 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

▼ All	▼ AAT ID	▼ Preferred Term	▼ Parent Hierarchy
☆ ↻ 1.	aat/300015050 Choose new match	oil paint (paint)	Facet Text filter origin>
☆ ↻ 2.			Edit cells ▶ Transform...
☆ ↻ 3.			Edit column ▶ Common transforms ▶
☆ ↻ 4.			Transpose ▶ Fill down
☆ ↻ 5.			Sort... Blank down
☆ ↻ 6.			View ▶ Split multi-valued cells...
☆ ↻ 7.			Reconcile ▶ Join multi-valued cells...
☆ ↻ 8.			Cluster and edit...
☆ ↻ 9.	aat/300014911 Choose new match	japan (enamel)	Replace...
☆ ↻ 10.		<coating by composition	
☆ ↻ 11.		coating (material)	
☆ ↻ 12.		<materials by function>	
☆ ↻ 13.		materials (substances)	
☆ ↻ 14.		Materials (hierarchy name)	
☆ ↻ 15.		Materials Facet	
☆ ↻ 16.	aat/300014780 Choose new match	pozzolana (clay)	<clay by composition or origin>
☆ ↻ 17.		clay	
☆ ↻ 18.		inorganic material	
☆ ↻ 19.		<materials by composition>	
☆ ↻ 20.		materials (substances)	
☆ ↻ 21.		Materials (hierarchy name)	
☆ ↻ 22.		Materials Facet	
☆ ↻ 23.	aat/300014635 Choose new match	filament (material)	<materials by form>
☆ ↻ 24.		materials (substances)	
☆ ↻ 25.		Materials (hierarchy name)	

Add your preferred separator between cells (we'll choose a semi-colon):



Click “OK” and your OpenRefine project should look like this:

10 rows			
Show as: rows records		Show: 5 10 25 50 100 500 1000 rows	
All	AAT ID	Preferred Term	Parent Hierarchy
1.	aat/300015050 <small>Choose new match</small>	oil paint (paint)	<paint by composition or origin>; paint (coating); <coating by form>; coating (material); <materials by function>; materials (substances); Materials (hierarchy name); Materials Facet
2.	aat/300014911 <small>Choose new match</small>	japan (enamel)	enamel (fused coating); <coating by composition or origin>; coating (material); <materials by function>; materials (substances); Materials (hierarchy name); Materials Facet
3.	aat/300014780 <small>Choose new match</small>	pozzolana (clay)	<clay by composition or origin>; clay; inorganic material; <materials by composition>; materials (substances); Materials (hierarchy name); Materials Facet
4.	aat/300014635 <small>Choose new match</small>	filament (material)	<materials by form>; materials (substances); Materials (hierarchy name); Materials Facet
5.	aat/300011033 <small>Choose new match</small>	titanium	nonferrous metal; <metal by composition or origin>; metal; inorganic material; <materials by composition>; materials (substances); Materials (hierarchy name); Materials Facet
6.	aat/300014668 <small>Choose new match</small>	ribbon (material)	<materials by form>; materials (substances); Materials (hierarchy name); Materials Facet
7.	aat/300255863 <small>Choose new match</small>	acetate film	film (material by form); <materials by form>; materials (substances); Materials (hierarchy name); Materials Facet
8.	aat/300005217 <small>Choose new match</small>	branch banks	banks (buildings); financial institutions (buildings); commercial buildings; <single built works by function>; <single built works by specific type>; single built works (built environment); S Environment (hierarchy name); Objects Facet
9.	aat/300005147 <small>Choose new match</small>	commercial buildings	<single built works by function>; <single built works by specific type>; single built works (built environment); Single Built Works (hierarchy name); Built Environment (hierarchy name);
10.	aat/300005250 <small>Choose new match</small>	barbershops	service industry buildings; commercial buildings; <single built works by function>; <single built works by specific type>; single built works (built environment); Single Built Works (hierar Objects Facet

3. Frequently Asked Questions and Tips

This section will grow over time. Answers here may use OpenRefine or other tools available for the Getty Vocabularies, such as SPARQL and the APIs.

To submit a question for FAQ, please write to vocab@gettyl.edu. For questions about OpenRefine per se (not the Reconciliation), please see OpenRefine documentation: openrefine.org

3.1. List of FAQs

Q1: Basics of Open Refine

Q1a: What are the basic functions in Open Refine?

A: Basic OpenRefine instruction is available at other sites online. We will not reproduce this copious documentation here. We suggest these sites:

- General introduction
openrefine.org
- General OpenRefine Functions
<https://github.com/OpenRefine/OpenRefine/wiki/GREL-Functions>
- OpenRefine Reconciliation Service Documentation
<https://reconciliation-api.github.io/specs/0.1/>
- LibraryCarpentry Open Refine training
<https://librarycarpentry.org/lc-open-refine/>
- How do I access the Getty Vocabularies for reconciliation?
<http://services.getty.edu/vocab/reconcile/>

That said, here are a few tips for the beginner, to get you started.

Perform actions on a column or on the cells of a column.

At the top of the column that you wish to edit, click the blue dropdown arrow to see your options.



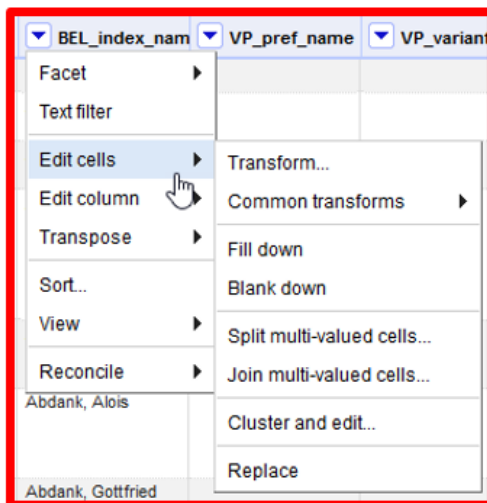
5000 rows

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

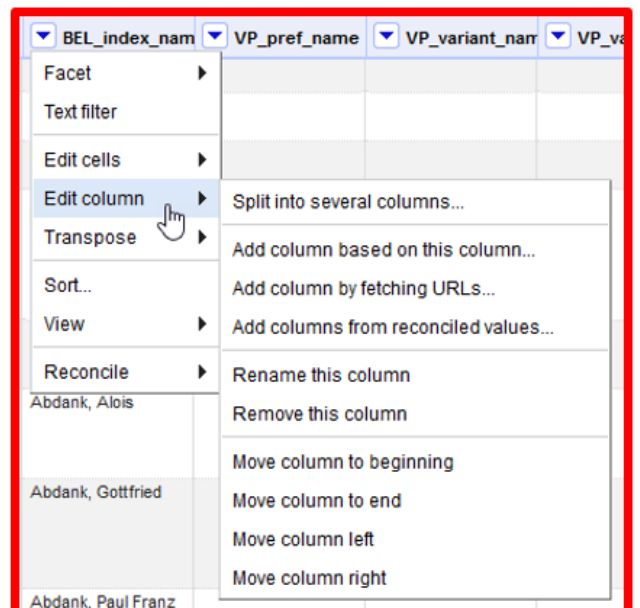
All	ID	BEL_legacy_ID	BEL_person_co	family name	given name	BEL_index_name
☆	1.	4965	4965	Person	Abandio	
☆	2.	4964	4964	Person	Abart	Franz
☆	3.	4963	4963	Person	Abb	Johann
☆	4.	4966	4966	Person	Abbati	Vincenz
☆	5.	4967	4967	Person	Abbiati	Julius

- Facet
- Text filter
- Edit cells
- Edit column
- Transpose
- Sort...
- View
- Reconcile

Edit cells, Edit column: You will likely often click “Edit cells” or “Edit column” as a first step in editing your data. Here are the options under each:



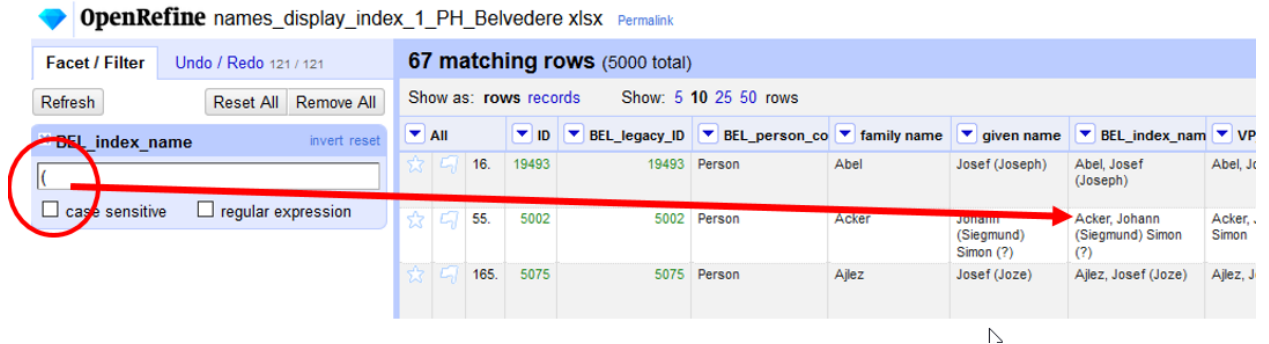
- Facet
- Text filter
- Edit cells
 - Transform...
- Edit column
 - Common transforms
- Transpose
 - Fill down
 - Blank down
- Sort...
- View
 - Split multi-valued cells...
- Reconcile
 - Join multi-valued cells...
 - Cluster and edit...
 - Replace



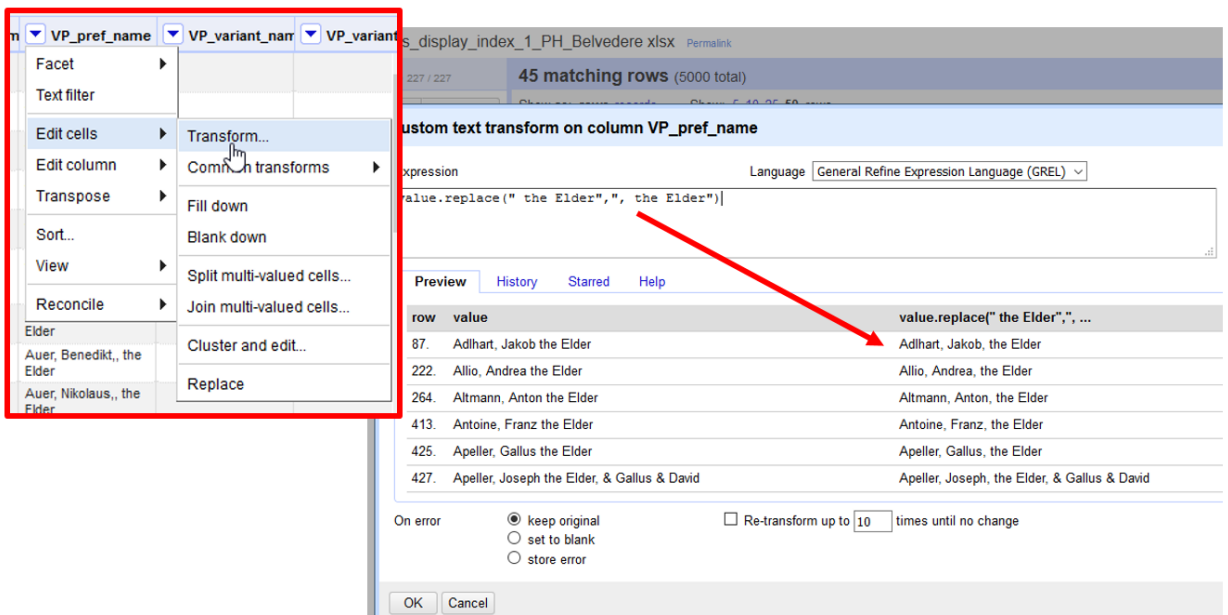
- Facet
- Text filter
- Edit cells
- Edit column
 - Split into several columns...
 - Add column based on this column...
 - Add column by fetching URLs...
 - Add columns from reconciled values...
- Transpose
- Sort...
- View
 - Rename this column
 - Remove this column
 - Move column to beginning
 - Move column to end
 - Move column left
 - Move column right

Facets and Filters: Facets and Filters are useful in isolating certain data that you wish to target for editing. In the example below, we want to find all values

that contain parentheses in a given column. Click the column that you wish to target, then in the left panel, in “Facet/Filter” tab, type the left parenthesis character. OpenRefine displays the 67 rows (out of 5,000) that have parentheses, making it easy for you to review or edit as you wish.



GREL expressions: You can use expressions to make changes to values in your spreadsheet. Using GREL is one way to do this. In this example, we want to add a comma to all values containing the phrase “the Elder” in our column. Click on the column, choose **Edit cells>Transform**. A screen will appear into which you can type your expression: `value.replace(" the Elder",", the Elder")` This means I want to replace the value “[space] the Elder” with “[comma] space the Elder”. You are provided with a column of sample values as they would appear using this expression; if this is what you need, click “OK” in order to apply to the entire spreadsheet.



Find sample GREL expressions in the FAQ of this tutorial, and learn how to construct them on the OpenRefine sites named above.

Q1b: What does the scoring mean in Reconciliation results?

The Vocabularies Reconciliation Service for OpenRefine does not use normalized scaling for matching results. The search gives relative results based on how many data points match Vocabularies data (text matching, broader context, birth/death dates, etc.).

In general, the better matches have higher scores and it is up to the user to determine what the acceptable threshold for an actual match should be. In this same manner, the service intentionally does not do auto-matching because we did not want to assume any expertise about the source data and create bad matches. (GG)

Q2: Creating concatenated name in ULAN.

In ULAN, how do I concatenate first and last names in one field?

A: For this question, let's start from the beginning:

1. Add a column based on Given Name, click on "edit column"; name the new column "BEL Index Name"
2. Go to the new column, which you would now like to alter, and click the arrow button on the column header; select the "Edit cells" option and then choose "Transform"
3. A pop-up window will appear with several sections, including a GREL expression input box and a preview panel. For more information on this window, refer to the Google Refine Expression Language (GREL) section.
4. The expression for combining column information is:
value + cells['Column'].value
"Value" indicates the values in the current column.
"Column" should match the name of the column whose values you would like to combine with the current column.
To combine column information AND additional strings, use a "+." For example:
value + "-" + cells['Column'].value
5. Adds a space (or punctuation) between the values of the current column and the values of the column to be combined with the current column.
6. This is the GREL statement: cells['family name'].value + ", " + cells['given name'].value

Note: GREL dislikes many special characters. E.g., the quotation marks cannot be “smart quotes” pasted from Word => change to “straight quotes” – below is example of an error message

Add column based on column given name

New column name:

core-views/addasdasd set to blank store error copy value from original column

Expression: Language:

Parsing error at offset 6:
Missing number, string,
identifier, regex, or
parenthesized expression

After fixing smart quotes and running the edit, here is the result:

	ID	family name	given name	BEL/VP index na	variant names
1.	4965	Abandio			
2.	4964	Abart	Franz	Abart, Franz	
3.	4963	Abb	Johann	Abb, Johann	
4.	4966	Abbati	Vincenz	Abbati, Vincenz	
5.	4967	Abbiati	Julius	Abbiati, Julius	

This method can be used to concatenate any fields in one’s spreadsheet. For example, creating a “biography” from existing data, one could concatenate fields with Nationality, Role, and birth and death dates to create a string (e.g. “Mexican painter, 1893-1955”).

Q2a: How to add a comma if name is already concatenated

For this one, it is easy to use Excel, then load your edited spreadsheet back into OpenRefine.

Make a column beside the column containing the names without commas. Place your cursor in the first empty cell of that column. Type in this formula, where the alpha-number is the cell that you wish to change, in this case C2 of our Excel spreadsheet. Drag the autofill handle down the column to apply this formula to cells.

=replace(C2,find(" ",C2),0,"")

Original AlphaSortName	Comma added column	Original AlphaSortName	Comma added column
Ashbee Charles Robert	=replace(C2,find(" ",C2),0,"")	Ashbee Charles Robert	Ashbee, Charles Robert
Abbott Berenice		Abbott Berenice	Abbott, Berenice
Amos Emma		Amos Emma	Amos, Emma
Andrews Benny		Andrews Benny	Andrews, Benny
Avedisian Edward		Avedisian Edward	Avedisian, Edward
Avedon Richard		Avedon Richard	Avedon, Richard

Q3: How to separate broader contexts to reconcile TGN.

To reconcile hierarchical data, when my source data includes broader contexts in one field, what do I do?

A: Separate the broader contexts in separate fields before reconciliation.

B: Go to Edit Column, choose “split into several columns” and then choose your delimiter.

place of birth	BEL birth place	BEL birth place
Schlinig (Tirol)	Schlinig	Tirol

C: When reconciling, you can use those additional fields as “broader context” (type ‘broaderExt’ into the field in question after choosing our service and the Vocabulary to reconcile) so the reconciliation service can retrieve more accurate results. This is particularly helpful with TGN data as there are many homographs.

Q4: Normalizing all caps in inconsistent ULAN data.

In the target data, the use of capitalization is inconsistent. How can I target only certain records for updates?

A: You could use Transform to convert to title case.

given name	variant names	artist group	date of birth	date of death	place
Achleitner	Simon				
Achternbusch	HERBERT				
Aachen	Hans von				

given name	variant names	artist group	date of birth	date of death	place
Facet	rt Schild		Wed Nov 23 00:00:00 PST 1938	0000-00-00	München
Text filter					
Edit cells	Transform...		0000-00-00	0000-00-00	
Edit column	Common transforms				
Transpose	Fill down				
Sort...	Blank down				
View	Split multi-valued cells...				
Reconcile	Join multi-valued cells...				
	Cluster and edit..				
	Replace				

However, note that doing this will also change all the correctly lowercase to title case. (*von*, *de*, etc.).

Instead, better to use GREL to target only the all uppercase.

Solution:

```
import re
val = value
m = re.search(r'[A-Z]{2,}', value)
if m is not None:
    val=val.replace(m.group(0),m.group(0).title())
return val
```

Q5: Getting adjectival form of Geographic place from noun form in ULAN.

I need to use the adjectival form of names for Nationality, but I have only the noun form of the nation. What do I do?

A: In TGN, the adjectival form is included for nations. Reconcile, then fetch it from there.

Solution:

1. Choose “Add column by fetching URLs ...”

2. Use this embedded SPARQL query in the GREL text box:

```
'http://vocab.getty.edu/sparql.json?query=select+(str(?aflabel)+as+?x)+{?s+skos:inScheme+tnn;+luc:term+""+escape(value,'url')+"";xl:prefLabel|xl:altLabel+?afterm.+?afterm+gvp:termPOS+%3Chttp://vocab.getty.edu/term/POS/Adjectival%3E;+xl:literalForm+?aflabel}'
```

3. Parse returned JSON using ‘Transform’ and the formula:
value.parseJson().results.bindings[0].x.value

Q6: Getting ULAN Nationalities

If I want to get records for artists having particular Nationalities from ULAN, should I use the Vocabulary Reconciliation Service?

A: Yes, see section 2.7 above to add Nationality and other select ULAN fields from reconciled values.

Q7: Getting full records XML.

How could I use OpenRefine to retrieve the full record XML for each ULAN ID that I need (or a batch)?

A: GREL for getting the XML where ‘value’ is the subject ID in the referenced column for ‘Add column by fetching URLs ...’ (NOTE: Be careful copying and pasting the single quote marks, a.k.a. neutral apostrophes! Should all be neutral apostrophes):

```
'http://vocabsservices.getty.edu/ULANService.asmx/ULANGetSubject?subjectId='+value
```

Parse nationality from XML using ‘Add column based on this column’ (again, make sure any copy/paste operations do not convert neutral apostrophes to single-quotes):

```
value.parseXml().select('Vocabulary')[0].select('Subject')[0].select('Nationalities')[0].select('Preferred_Nationality')[0].select('Nationality_Code')[0].ownText().replace(Λd{6}V/,")
```

Parsing of XML follows a logical flow from element to sub-element. Can be used to access any field in the full-XML.

Q8: Fetching Birth Year based on ULAN reconciliation.

If I have reconciled a person using ULAN, how can I fetch the birth year from ULAN?

A: First get the full XML using ‘Add column by fetching URLs ..’ and then use GREL to fetch the birth dates.

```
'http://vocabsservices.getty.edu/ULANService.asmx/ULANGetSubject?subjectId='+value
```

Parse birth date from XML in new column using “Add column based on column...”:

```
value.parseXml().select('Vocabulary')[0].select('Subject')[0].select('Biographies')[0].select('Preferred_Biography')[0].select('Birth_Date')[0].ownText()
```

See also below, **3.3.2. OpenRefine Data Enrichment.**

Q9: Getting Facet Code for AAT.

I want to use the Facet Code in AAT to sort results by type. This is a legacy field that formerly was used to create hierarchies, but now could have other functions.

A. Here’s the GREL for getting all facet codes for an ID value comma separated using “Add column by fetching URLs...”:

```
'http://vocab.getty.edu/sparql.json?query=select+(group_concat(?fcode;+separator=";"++)+as+?fcodes)+{aat:'+value+'+gvp:broaderPreferredExtended+?broader.+?broader+skos:notation+?fcode}'
```

Parse the JSON results using “Transform ...”:

value.parseJson().results.bindings[0].fcodes.value

Results:

- desc	AAT ID	AAT number	URL	Facet code	Google term
earthenware Choose new match	aat/300140803	300140803	<pre>{ "head": { "vars": ["fcodes"] }, "results": { "bindings": [{ "fcodes": { "type": "literal", "value": "V; V.V; V.VC" } }] } }</pre>	V; V.V; V.VC	earthenware

Term	AAT ID	Facet Codes	Broader
aba <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> abas (26) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	300209813	V; V.T; V.TE	Objects
abaco dinghy <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> abaco dinghy (33) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Bahama dinghies (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> bermuda dinghy (16) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> dinghies (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> sailing dinghy (13) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	300421656	V; V.T; V.TX	Objects
abacus <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> abaci (capital components) (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> abaci (sideboards) (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> abaci (calculators) (20) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> impost blocks (6) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	300001699	V; V.PJ	Objects
abdominal binder <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> abdominal binder (34) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> pamphlet binders (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> binders (costume) (13) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> binding warp (13) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> binder clip (12) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	300421658	V; V.T; V.TH	Objects

Q10: Getting Coordinates for TGN.

How can I get a list of TGN authority records with geographic coordinates?

A: You can use our API and SPARQL query page (SPARQL link at the top of the page next to "Queries," our page of sample queries): <http://vocab.getty.edu>

We don't have a pre-written TGN SPARQL query that brings precisely what you're looking for, but this will work. It may time out after 2 minutes, so you may need to add other variables.

```
select ?s ?lat ?long {?s skos:inScheme tgn.; foaf:focus ?place. ?place wgs:lat ?lat. ?place wgs:long ?long}
```

B: To retrieve coordinates from reconciled TGN terms, see section 2.7.

Q11: Getting Arabic terms in TGN.

How can I get the Arabic terms currently in TGN?

A: Using SPARQL, try this:

```
select ?s ?term ?text ?engTerm {?term dcterms:language gvp_lang:ar.
?s xl:prefLabel|xl:altLabel ?term.
?term xl:literalForm ?text.
optional {?s gvp:prefLabelGVP [ xl:literalForm ?engTerm ]}}
```

Q12: Qualifying or constraining reconciliation criteria.

How do I constrain or qualify the Reconciliation using other than name?

A. Use the "properties." See [section 2.4](#) above.

Reconcile column "pref_name"

» Access Service API

Reconcile each cell to an entity of one of these types:

- ULAN search
/ulan
- TGN search
/tgn
- AAT search
/aat
- Search all Vocab
/all

Also use relevant details from other columns:

NMWA_birth_ID	<input type="checkbox"/>	
death_place	<input type="checkbox"/>	
NMWA_death_place 1	<input type="checkbox"/>	
NMWA_death_reconcile_id	<input type="checkbox"/>	
NMWA_death_place 2	<input type="checkbox"/>	
NMWA_death_place 3	<input type="checkbox"/>	
NMWA_death_id	<input type="checkbox"/>	
nationality	<input checked="" type="checkbox"/>	nat

Select an item from the list:

- nationality**
- nationalityPref
- nationalityNonPref

Reconcile against type:
 Reconcile against no particular type
 Auto-match candidates with high confidence ← *caution*
 Maximum number of candidates to return

3.2 Practical tips for using OR and the Reconciliation Service

While we recommend that you consult the voluminous online OpenRefine training and help resources for basic OpenRefine skills, we include a few tips here which we found helpful in our clean-up work.

3.2.1. Reconcile data sets of manageable size and content

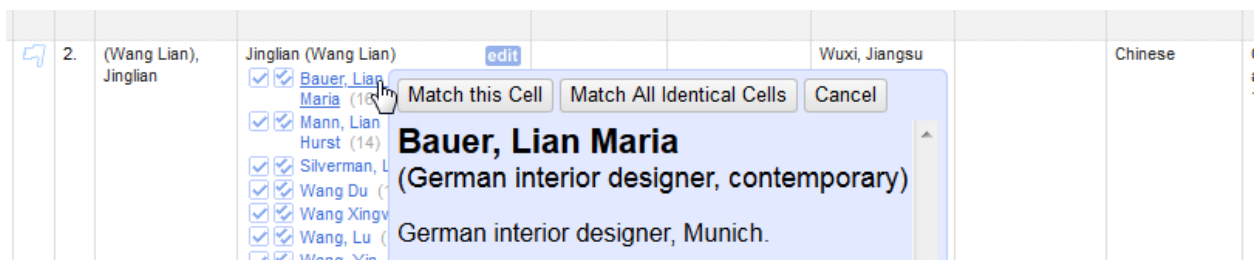
We recommend that you break up big data sets into smaller files of 5,000 or fewer rows, if you are doing detailed reconciliation. Reason is, it is faster and more efficient to curate by hand and to undo errors if the data set is small enough for a human to edit.

If you are using reconciliation on a large data set and your project can tolerate a certain percentage of errors, you may decide it is more efficient to reconcile the big data set and absorb the errors.

3.2.2. Use auto-match with care

Given the vast possibilities of variation in expressing information surrounding art vocabularies, and the frequency with which such data is unknown or estimated, it is unlikely that the source data set (your data) and the target data set (the Getty Vocabulary) will match well enough to form secure auto-matches 100% of the time. Even with good correspondence between the two data sets, it is likely that an error rate of 20% will occur with auto-match. Perhaps that is good enough, if your project can tolerate a percentage of error or if the auto-match is a first step prior to human curation.

In the example below, the parentheses in the source data name and the lack of birth or death dates for the source artist data have caused the highest ranked ULAN match to be obviously false. However, the ranking of this match was only '17,' so would likely have been eliminated from consideration even with auto-match.



If the source data set and target Getty Vocabulary share similar rules for data values, the percentage of reliable auto-matches is increased. To compare the Getty Vocabulary editorial rules with the practice used for the source data, please see

<http://www.getty.edu/research/tools/vocabularies/guidelines/index.html> For example, for the source artist data set, if the death date is unknown or the artist is still living, what value is used for death date? The rule for estimating death dates in ULAN is to allow a life-span of 120 (add 120 years to the birth date). In ULAN, these estimated dates are not visible to end users, but are utilized only for matching and retrieval. In the example above, for the source data, the practice is to leave birth and death date blank when it is unknown, rather than estimating a value; thus dates are unhelpful or harmful in the matching algorithms. This rule for unknown dates is only one of dozens of such rules for estimating unknown or ambiguous data in ULAN, TGN, and AAT.

3.2.3. Normalize your data prior to reconciliation

It is optional, but recommended, to normalize and clean your data prior to reconciliation. You are likely to get better results if your data is consistent and in a form more readily matched to the Getty Vocabulary data.

In the example below, you can see that some values in the artistBirth column are numbers (green) and some values are characters (black). Maybe some dates in this column include alphabetic phrases, such as 'ca.' or 'after.' If you wish to reconcile or otherwise work with this data, or load it into a system, it would be best to normalize and edit the data for consistency throughout the column. In this example, we would recommend that you follow instructions for general OpenRefine (at openrefine.org) to change all the values to characters (because ULAN stores values representing years as characters), and to delete the alphabetical strings such as 'ca.'

« Start Over		Configure Parsing Options		Project name: IMJ.xls		Tags		Create Project »			
artistName	artistNonPrefName	artistBirth	artistDeath	artistNationality	artistBirthPlace 1	artistBirthPlace 2	artistBirthPlace 3	artistBirthPlace 4	artistBirthPlaceNation	artistRole	artist
1. Aaronson, Yakov	אהרונסון, יעקב	1924	2100	Israeli					Israel	31407/photographer	
2. Abadi, Abed	עבד, עבד	1942	2100	Israeli					Israel	31261/painter	
3. Abadi, Shay	עבאדי, שי	1965	2100	Israeli	Jerusalem	Israel				31100/artist	
4. Abakanowicz, Magdalena	אבאקאנוביץ' מגדאלנה	1930	2017	Polish					Poland	34222/weaver	3126
5. Abba, Irit	אבא, עירית	1953	2100	Israeli					Israel	31557/designer	3322
6. Abbasi, Riza (attributed to)	אבאסי, ריזה (מיחס ל)	1200	2100	Iranian					Iran	31261/painter	
7. Abbasi, Riza (school of)	אבאסי, ריזה (אסכולה)	1200	2100	Iranian					Iran	Miniatures painter	
8. Abbasi, Riza	אבאסי, ריזה	1565	1635	Iranian					Iran	31261/painter	
9. Abbott, Berenice	אבוט, ברניס	1898	1991	American					USA	31407/photographer	
10. Emil Aboud, Jumana	אמיל עבד, ג'ומאנה	1971	2100	Israeli	Shefaram	Israel				31261/painter	
11. Abdullah Freres	האחים עבדוללה	1858	1899	Turkish					Turkey	31407/photographer	3126
12. Abecassis, Raphael	אבקסיס, רפאל	1953	2100	Moroccan	Marrakesh	Morocco				31261/painter	
13. Abel, Myer	מאיייר, אבן	1904	1946	American					USA	31261/painter	3143

In the example below, the pref_name column includes multiple artist names in the same column, separated with parentheses. This may impair the reconciliation when seeking appropriate matches in ULAN. Using general OpenRefine editing procedures, you can parse out the different names into separate columns in order to reconcile or import data (ULAN typically does not include parenthetical names in the

same name field, e.g., ‘Kogga’ would be in one field and ‘Kolbrun Bjorgolfsdottir’ would be in a separate field).

	▼ pref_name	▼ name direct or c	▼ birth_year	▼ death_year	▼ birth_place	▼ death_place	▼ nationality
1.	(Kolbrun Björgolfsdottir), Kogga	Kogga (Kolbrun Björgolfsdottir)	1952				Icelandic
2.	(Wang Lian), Jinglian	Jinglian (Wang Lian)			Wuxi, Jiangsu province		Chinese

You may edit the column as-is. However, our practice is to always leave the original data intact (here, *pref_name*); we would add new columns based on the original column, with new column headings, where we would edit the data (e.g., *VP_preferred_name*). This allows us to see at a glance what has been edited and compare it to the original data, thus noticing editing errors or miscalculations immediately. OpenRefine allows you to easily step back (undo an edit) if you find it has yielded a bad result, but it is important to find errors immediately or you may lose good edits when you Undo an older error.

Below is another example where parsing source data into separate columns is helpful. Here the source data contains multiple place names in the column *birth_place*, where some names represent broader contexts for the focal place. In order to better find matches in TGN, the values have been parsed out in separate columns *birth_place_1*, *birth_place_2*, *birth_place_3*; with this method, you can reconcile against TGN using the broader contexts as Properties for more accurate matches (in TGN, the broader contexts are separate records, linked hierarchically, not combined in one name).

▼ birth_place	▼ birth_place_1	▼ TGN_ID	▼ birth_place_2	▼ birth_place_3
New York, New York	New York	7007567	New York	
Etampes, Seine-et-Oise, France	Etampes		Seine-et-Oise	France
Springfield, Ohio	Springfield	7014533	Ohio	
Germany	Germany	7000084		

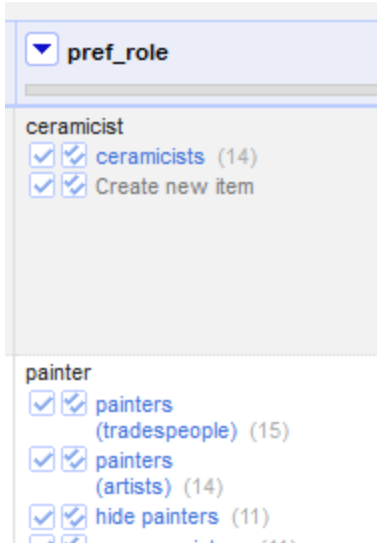
3.2.4. Inserting IDs: Use Best Match with caution

We recommend caution in using the expression 'cell.recon.best.id' to automatically insert the Vocabulary ID in a column based on the auto-matched best ranked result, but before the match has been vetted by the editor.

After extensive testing, we advise against 'best match' because it often results in bad matches. However, it may be useful for big data sets where a certain percentage of error is acceptable, or if you apply it to only those matches with exceptionally high ranking.

Here is an example of a bad match being made automatically, where an incorrect ID has been inserted.

In the illustration below, a human editor who knows the data understands that the match for source data term 'painter' should be made to the second on the list, 'painters (artists)' for this data set. However, since 'painters (tradespeople)' has a higher ranking, this is the match that would be made automatically.



Even if, after having gathered IDs in the AAT_ID column using 'best match,' the editor then manually chooses 'painters (artists)' aat/ 300025136, the ID in the AAT_ID column will remain the same, that of the incorrect 'painters (tradespeople)' aat/300235851.

It is more reliable to accept the correct matches first, then go back to get the AAT_ID through the expression for 'match' rather than 'best match.'



Using 'best match' with ULAN reconciliation will be more successful than with AAT or TGN, due to the greater number of properties that can aid in matching; however there will still likely be errors. In the example below, the match has been made to the ULAN record for the museum, rather than to the artist. The source value '*Abbasi, Riza (school of)*' is an attribution statement containing a named artist and the attribution 'school of' to an unknown person or persons in his school. This is not a museum; but the match to the

museum has the highest ranking and thus was linked incorrectly as ulan/500309512 with 'best match.'

	artistName	U
3.	Abadi, Shay Choose new match	ulan/500486311 שי, י
4.	Abakanowicz, Magdalena Choose new match	ulan/500084577 אלנה
6.	Riza 'Abbasi Museum Choose new match	ulan/500309512 (ר ל
7.	Abbasi, Riza (school of) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Riza 'Abbasi Museum (37) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Riza (31) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Imam Riza Shrine	ulan/500309512 (נולה

3.2.5. When to format cells in Excel prior to loading in OpenRefine

In a few cases, we find that it is easier to format some data in the spreadsheet using Excel or Notepad prior to opening it in OpenRefine.

Example: How to normalize when dates and text are in same column

In OpenRefine

0000-00-00	vor 1770-00-00	
1765-02-00	0000-00-00	E
1677-00-00	1743-08-07	
1873-10-29	Sun Feb 13 00:00:00 PST 1916	V
0000-00-00	1564-02-14	

In this example, the spreadsheet contains birth and death dates in different formats, date and text. Our goal is to create a new column containing only 4-digit years, in text format. When imported in OpenRefine on our system, the date format that appeared "2/13/1916" in the original data set is translated by default into an unwieldy form.

In OpenRefine, you could construct multiple Grel statements to deal with this and all the other oddities in this column. or you could normalize a bit prior to import in OpenRefine, which we find easier.

3.2.6. Workflow for Adding and Reconciling New Terms

In our reconciliation pre-preprocessing of Vocabulary contributions, if the term is not already in the Vocabulary, we need to add it to the target Vocabulary; then reconcile again. Thanks to graduate intern, Devon Murphy, for drafting a workflow for this process.

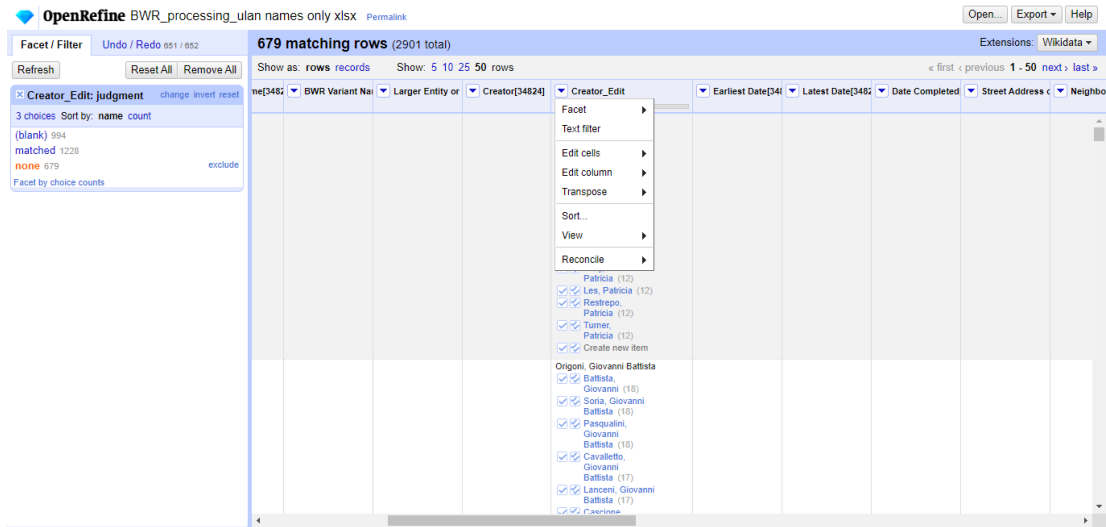
Workflow for Adding and Reconciling New Terms

Before Entering in the Data:

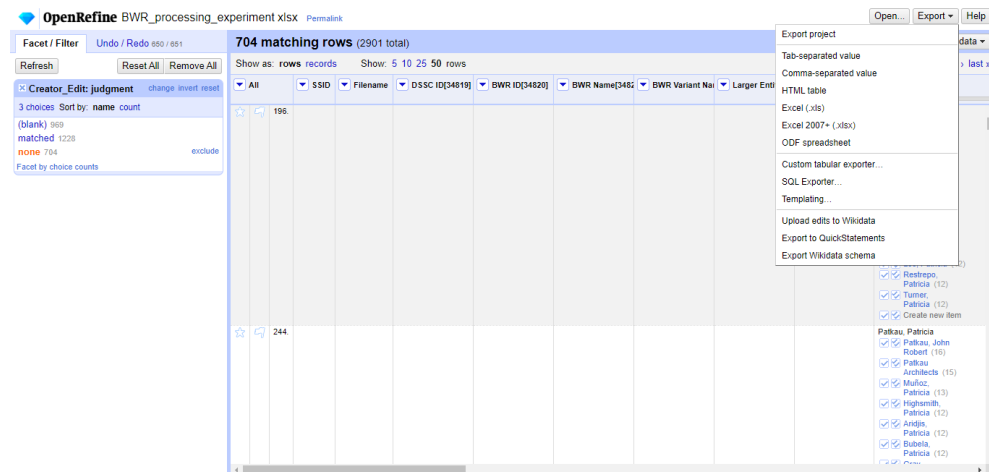
- **Clean up the data to ensure best matches for reconciling names:**
This can include faceting, clustering, slicing, etc. For further information on data cleaning functions, please see Open Refine's Github wiki: <https://github.com/OpenRefine/OpenRefine/wiki>
- **Reconcile the names to ULAN:**
Best results are achieved if this process is finished first. For further information on how to reconcile to the Getty Vocabularies, please see this guide: [Getty Vocabularies OpenRefine Tutorial](#)
- Keep your facet of reconciled results open:
this is how you will determine what rows to export (Reconcile -> Facet -> By judgement; include "none" results, exclude "matched" or "blank").
- Make sure you are viewing your results as rows and not records:
this filters out records that have multiple names attached where some are reconciled and some are not.

The screenshot displays the OpenRefine interface for a file named 'BWR_processing_experiment.xlsx'. The main table shows 704 matching rows out of a total of 2901. The table columns include 'All', 'SSID', 'Filename', 'DSSC ID[34819]', 'BWR ID[34820]', 'BWR Name[34821]', 'BWR Variant Name', 'Larger Entity or', and 'Creator[34824]'. The 'All' column shows two rows: one with ID 196 and another with ID 244. On the left, a facet titled 'Creator_Edit: judgment' is active, showing counts for 'blank' (969), 'matched' (1228), and 'none' (704). On the right, a facet titled 'Creator_Edit' is open, displaying a list of names with checkboxes, such as 'Patkau, Patricia', 'Patkau, John', 'Robert (16)', 'Patkau Architects (15)', 'Muñoz, Patricia (13)', 'Highsmith, Patricia (12)', 'Arida, Patricia (12)', 'Bubela, Patricia (12)', 'Gray, Patricia (12)', 'Les, Patricia (12)', 'Residpo, Patricia (12)', 'Turner, Patricia (12)', and 'Create new item'.

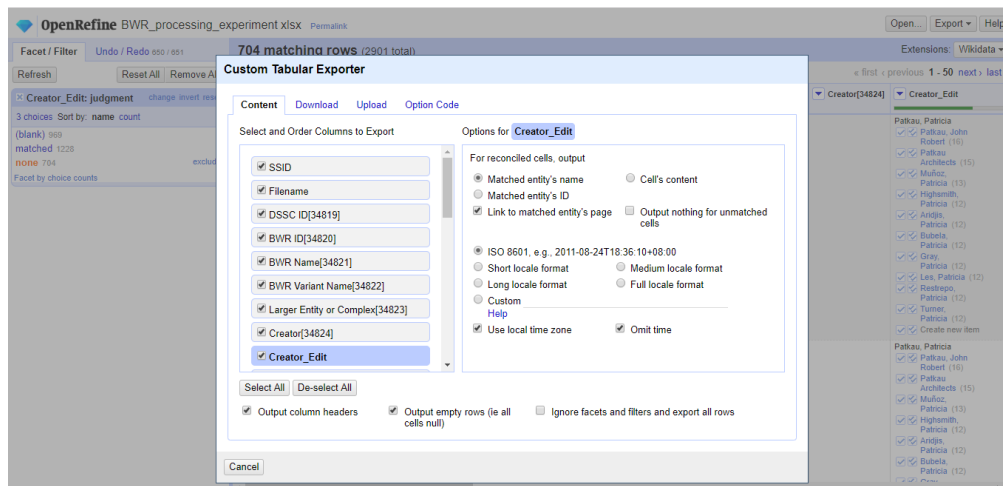
- Remove duplicate names:
(Edit Cells -> Blank down).



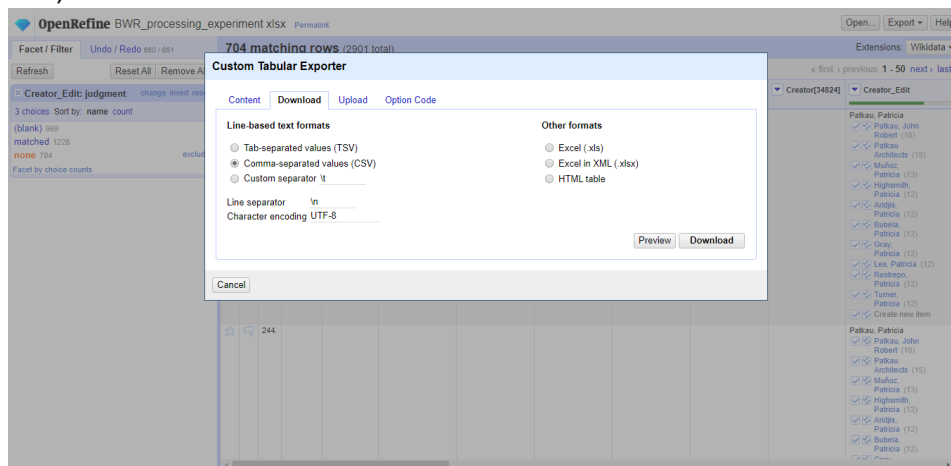
- Export a file that contains all unreconciled names:
 - Go to Export at the top right of the OpenRefine menu.
 - Choose “Custom Tabular Explorer.”



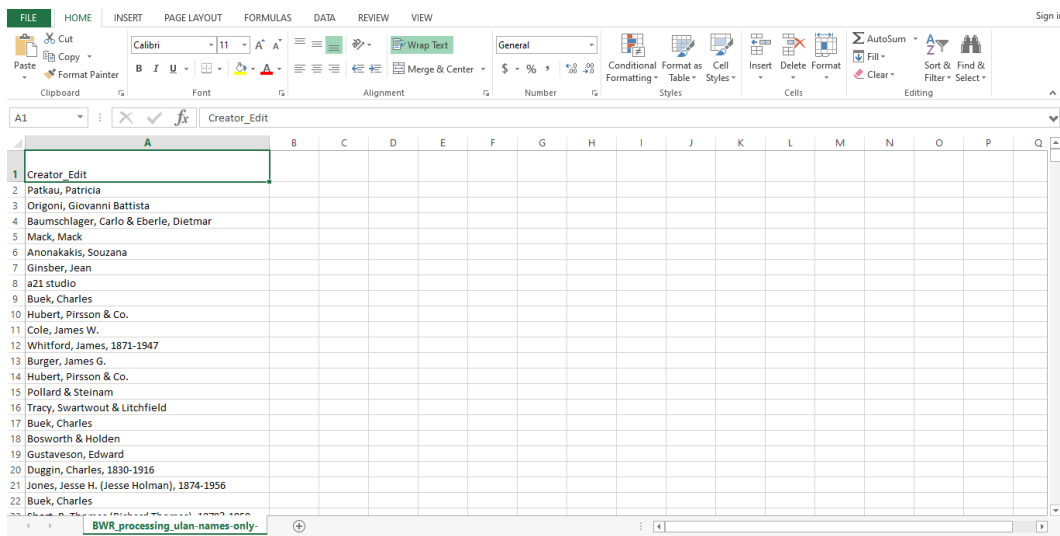
- Click on the left to select which column you wish to filter by (in this case, your reconciled column). You can also remove columns that are not germane to your needs.



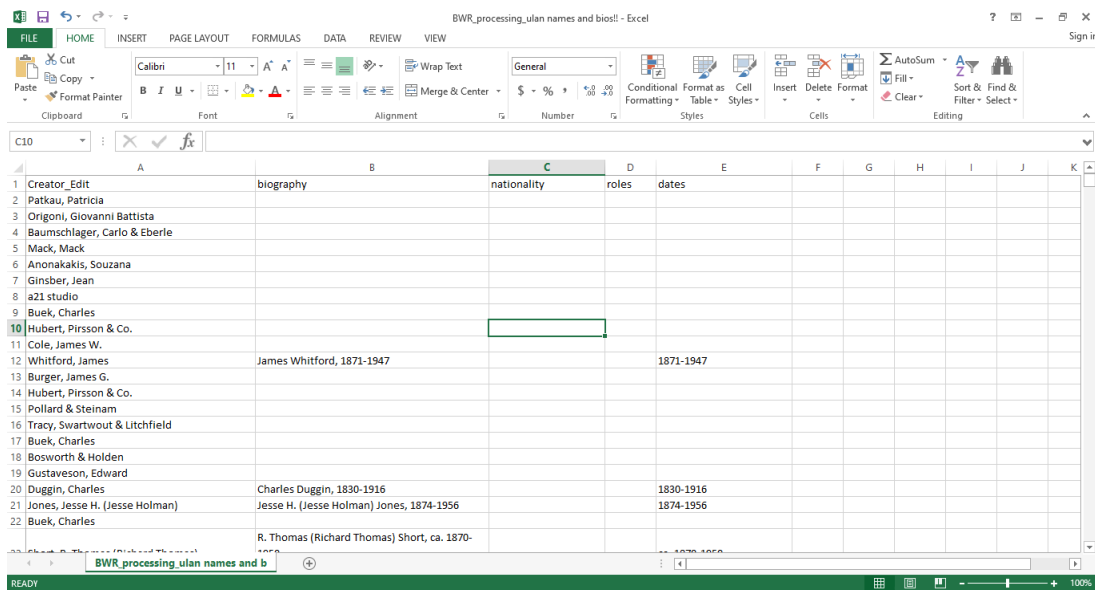
- On the right menu, uncheck “output nothing for unmatched cells.”
- Click on the Download tab to set your download format preference (XML, CSV etc.) and click “Download.”



- Undo the blank down command to retain your full records in your original spreadsheet: Your results will be a record of all names that have not been reconciled. For this particular project, only the name fields are needed, as the master spreadsheet of other included fields persists as an Open Refine project.



- Here is a view that includes the unreconciled names in .csv format.



- Here is a view that includes names, biographical information, and other related fields of data that were contained in the original name field entered in by the contributor in .csv format. You can export various columns based on your needs.

After Entering in the Data:

- Reconcile the name column a second time to match all names.
- Use the reconciliation service to also retrieve the ULAN IDs for each name.

In Excel

0000-00-00	0000-00-00
0000-00-00	vor 1770-00-00
1765-02-00	0000-00-00
1677-00-00	1743-08-07
1873-10-29	2/13/1916
0000-00-00	1564-02-14

Using Excel and Notepad, we were able to more easily format the string as we needed it and transform it to text format rather than data type = date.

0000-00-00	0000-00-00	vor 1770-00-00	vor 1770-00-00
1765-02-00	1765-02-00	0000-00-00	0000-00-00
1677-00-00	1677-00-00	1743-08-07	1743-08-07
1873-10-29	1873-10-29	1916-02-13	1916-02-13
0000-00-00	0000-00-00	1564-02-14	1564-02-14

In OpenRefine

vor 1770-00-00	vor 1770-00-00	1770
0000-00-00	0000-00-00	0000
1743-08-07	1743-08-07	1743
Sun Feb 13 00:00:00 PST 1916	1916-02-13	1916
1564-02-14	1564-02-14	1564

Then in OpenRefine, we opened the spreadsheet and edited the data using faceting to eliminate words, such as “vor” and using Grel to trim all values to the right of the 4-digit year.

Faceting in OpenRefine

Grel expression to truncate string

Preview	History	Starred	Help
15.	1963-06-02		1963
16.	1818-10-07		1818
17.	0000-00-00		0002

4. Advanced OpenRefine Techniques Using the Getty Vocabularies

by Gregg Garcia

41. OpenRefine ID Sync Example Using XML Web Services

1. Create a new column using “Add column by fetching URLs...”

The screenshot shows the OpenRefine interface with a table of records. The columns are: DISPLAYNAME, ULAN ID, Pref Term, FIRSTNAME, LASTNAME, BEGINDATE, and EN. A context menu is open over the ULAN ID column, showing options like Facet, Text filter, Edit cells, Edit column, Transpose, Sort..., View, and Reconcile. The 'Add column by fetching URLs...' option is highlighted.

DISPLAYNAME	ULAN ID	Pref Term	FIRSTNAME	LASTNAME	BEGINDATE	EN
Maxwell Hamilton Maxwell, Hamilton (24) Hamilton, Maxwell, Inc. (23) Fry, Edwin Maxwell (14) Bates, Maxwell (14) Maxwell, John (14) Create new item		ll, Hamilton	Maxwell	Hamilton		0
Charles Aubry Aubry, Charles (34) Aubry, Charles (33) Aubry, Emile (18) Aubry, Louis (18) Aubry, René Marcel (17) Create new item						1811
Fédèle Azari Azari, Fedele (19) Azari Isfarayini, Hamzah ibn 'Ali Malik (5) Create new item	500103519	Azari,				1895
František Vobecký Vobecký, František (43) Schmoranz, František (17) Fišer, František (17) Kadlec, František (17) Kysela, František (17) Create new item	500028103	Vobec				1902

2. Define column name and input GREL expression:
['http://vocabsservices.getty.edu/ULANService.aspx/ULANGetSyncSubjectID?subjectID='+value](http://vocabsservices.getty.edu/ULANService.aspx/ULANGetSyncSubjectID?subjectID=)

Add column by fetching URLs based on column ULAN ID

New column name Throttle delay milliseconds

On error set to blank store error Cache responses

HTTP headers to be used when fetching URLs: [Show](#)

Formulate the URLs to fetch:

Expression Language

Preview History Starred Help

row	value	'http://vocabsservices.getty.e ...
1.	500299270	http://vocabsservices.getty.edu/ULANService.aspx/ULANGetSyncSubjectID?subjectID=500299270
2.	500029715	http://vocabsservices.getty.edu/ULANService.aspx/ULANGetSyncSubjectID?subjectID=500029715
3.	500103519	http://vocabsservices.getty.edu/ULANService.aspx/ULANGetSyncSubjectID?subjectID=500103519
4.	500028103	http://vocabsservices.getty.edu/ULANService.aspx/ULANGetSyncSubjectID?subjectID=500028103

Resulting XML



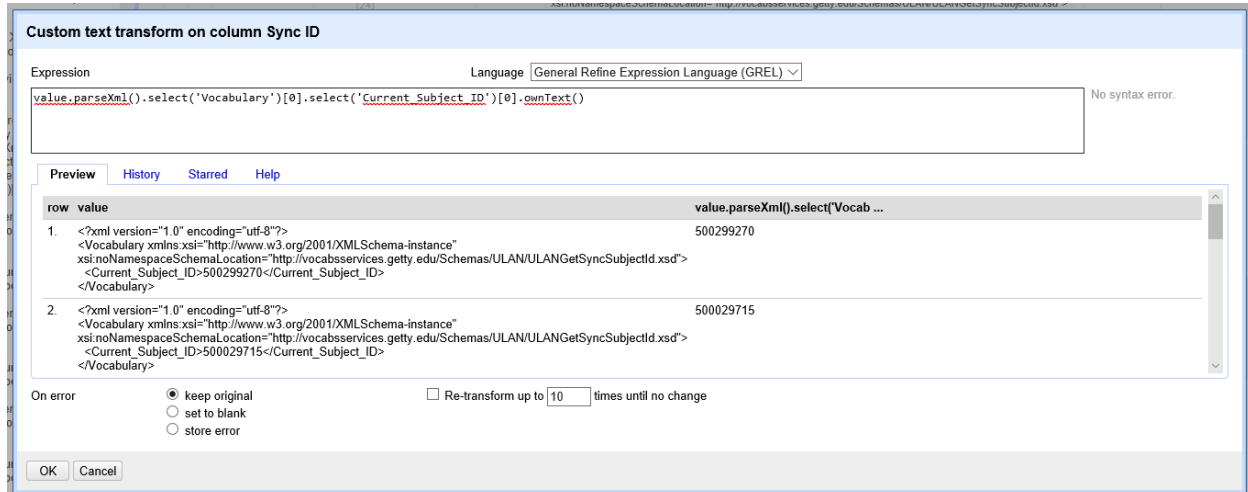
DISPLAYNAME	ULAN ID	Sync ID
Maxwell Hamilton <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Maxwell, Hamilton (24) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Hamilton, Maxwell, Inc. (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Fry, Edwin Maxwell (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Bates, Maxwell (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Maxwell, John (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500299270	<?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://vocabsservices.getty.edu/Schemas/ULAN/ULANGetSyncSubjectId.xsd"> <Current_Subject_ID>500299270</Current_Subject_ID> </Vocabulary>
Charles Aubry <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Charles (34) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Charles (33) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Emile (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Louis (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, René Marcel (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500029715	<?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://vocabsservices.getty.edu/Schemas/ULAN/ULANGetSyncSubjectId.xsd"> <Current_Subject_ID>500029715</Current_Subject_ID> </Vocabulary>
Fédèle Azari <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Azari, Fedele (19) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Āzarī Isfarāyīnī, Hamzah ibn 'Alī Malik (5) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500103519	<?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://vocabsservices.getty.edu/Schemas/ULAN/ULANGetSyncSubjectId.xsd"> <Current_Subject_ID>500103519</Current_Subject_ID> </Vocabulary>
František Vobecký	500028103	<?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

3. Parse XML using "Transform..."

DISPLAYNAME	ULAN ID	Sync ID
Maxwell Hamilton <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Maxwell, Hamilton (24) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Hamilton, Maxwell, Inc. (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Fry, Edwin Maxwell (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Bates, Maxwell (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Maxwell, John (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500299270	<?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://vocabsservices.getty.edu/Schemas/ULAN/ULANGetSyncSubjectId.xsd"> <Current_Subject_ID>500299270</Current_Subject_ID> </Vocabulary>
Charles Aubry <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Charles (34) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Charles (33) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Emile (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Louis (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, René Marcel (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500029715	<?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://vocabsservices.getty.edu/Schemas/ULAN/ULANGetSyncSubjectId.xsd"> <Current_Subject_ID>500029715</Current_Subject_ID> </Vocabulary>
Fédèle Azari <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Azari, Fedele (19) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Āzarī Isfarāyīnī, Hamzah ibn 'Alī Malik (5) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500103519	<?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://vocabsservices.getty.edu/Schemas/ULAN/ULANGetSyncSubjectId.xsd"> <Current_Subject_ID>500103519</Current_Subject_ID> </Vocabulary>
František Vobecký	500028103	<?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

4. Transform GREL expression:

`value.parseXml().select('Vocabulary')[0].select('Current_Subject_ID')[0].ownText()`



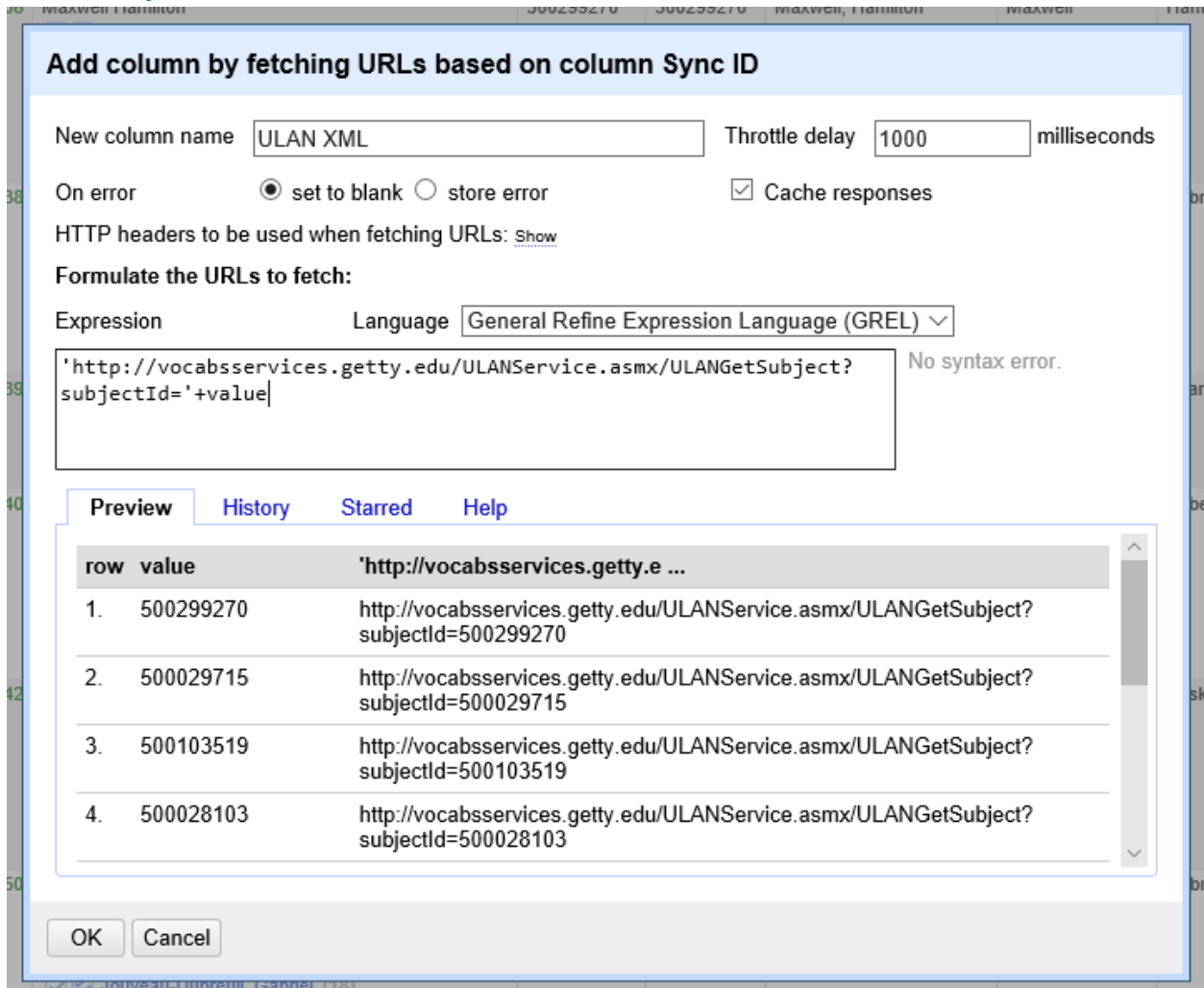
Sync Column becomes current ULAN ID



▼ DISPLAYNAME	▼ ULAN ID	▼ Sync ID
Maxwell Hamilton <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Maxwell, Hamilton (24) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Hamilton, Maxwell, Inc. (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Fry, Edwin Maxwell (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Bates, Maxwell (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Maxwell, John (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500299270	500299270
Charles Aubry <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Charles (34) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Charles (33) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Emile (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Louis (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, René Marcel (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500029715	500029715
Fédèle Azari <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Azari, Fedele (19) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Āzarī Isfarāyīnī, Ĥamzah ibn 'Alī Malīk (5) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500103519	500103519
František Vobecký <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Vobecký, František (43) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Schmoranz, František (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Fišer, František (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Kadlec, František (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Kysela, František (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500028103	500028103

4.2. OpenRefine Data Enrichment Example Using Full XML from Web Service

1. Get full XML by adding column based on synced ULAN ID, and input formula: ['http://vocabsservices.getty.edu/ULANService.asmx/ULANGetSubject?subjectId='+value](http://vocabsservices.getty.edu/ULANService.asmx/ULANGetSubject?subjectId=)



Add column by fetching URLs based on column Sync ID

New column name: Throttle delay: milliseconds

On error: set to blank store error Cache responses

HTTP headers to be used when fetching URLs: [Show](#)

Formulate the URLs to fetch:

Expression: Language: No syntax error.

Preview History Starred Help

row	value	'http://vocabsservices.getty.e ...
1.	500299270	http://vocabsservices.getty.edu/ULANService.asmx/ULANGetSubject?subjectId=500299270
2.	500029715	http://vocabsservices.getty.edu/ULANService.asmx/ULANGetSubject?subjectId=500029715
3.	500103519	http://vocabsservices.getty.edu/ULANService.asmx/ULANGetSubject?subjectId=500103519
4.	500028103	http://vocabsservices.getty.edu/ULANService.asmx/ULANGetSubject?subjectId=500028103

OK Cancel

Full XML in OpenRefine Cell

DISPLAYNAME	ULAN ID	Sync ID	ULAN XML
Maxwell Hamilton <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Maxwell, Hamilton (24) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Hamilton, Maxwell, Inc. (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Fry, Edwin Maxwell (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Bates, Maxwell (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Maxwell, John (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500299270	500299270	<pre> <?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://vocabsservices.getty.edu/Schemas/ULAN/ULANGetSubject.xsd"> <Subject Subject_ID="500299270"> <Parent_Relationships> <Preferred_Parent> <Parent_Subject_ID>500000002</Parent_Subject_ID> <Relationship_Type>Parent/Child</Relationship_Type> <Historic_Flag>Undetermined</Historic_Flag> <Parent_String>Persons, Artists [500000002]</Parent_String> <Hier_Rel_Type>Instance-BTI</Hier_Rel_Type> </Preferred_Parent> </Parent_Relationships> <Descriptive_Notes> <Descriptive_Note> <Note_Text>Comment on works: landscape</Note_Text> <Note_Language>English</Note_Language> <Note_Contributors> <Note_Contributor> <Contributor_id>2500000011/PROV</Contributor_id> </Note_Contributor> </Note_Contributors> <Note_Sources> <Note_Source> <Source> <Source_ID>2100043316/Provenance Index Databases, Authority file (1985-)</Source_ID> </Source> </Note_Source> </Note_Sources> </Descriptive_Notes> </Descriptive_Notes> <Record_Type>Person</Record_Type> <Merged_Status>Merged</Merged_Status> <Terms> <Preferred_Term> <Term_Text>Maxwell, Hamilton</Term_Text> <Display_Name>Index</Display_Name> <Historic_Flag>N/A</Historic_Flag> <Vernacular>Undetermined</Vernacular> <Term_ID>1500668205</Term_ID> <Term_Languages> <Term_Language> <Language>70001/undetermined</Language> </pre>

2. Parse preferred biography from full XML:

```
value.parseXml().select('Vocabulary')[0].select('Subject')[0].select('Biographies')[0].select('Preferred_Biography')[0].select('Biography_Text')[0].ownText()
```

Add column based on column ULAN XML

New column name:

On error: set to blank store error copy value from original column

Expression Language: No syntax error.

```
value.parseXml().select('Vocabulary')[0].select('Subject')[0].select('Biographies')[0].select('Preferred_Biography')[0].select('Biography_Text')[0].ownText()
```

Preview History Starred Help

row	value	value.parseXml().select('Vocab ...
1.	<?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://vocabs: <Subject Subject_ID="500299270"> <Parent_Relationships> <Preferred_Parent> <Parent_Subject_ID>500000002</Parent_Subject	Scottish painter and architect, 1830-1923

OK Cancel

Preferred Biography in OpenRefine cells

DISPLAYNAME	ULAN ID	Sync ID	Pref Bio
Maxwell Hamilton <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Maxwell, Hamilton (24) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Hamilton, Maxwell, Inc. (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Fry, Edwin Maxwell (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Bates, Maxwell (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Maxwell, John (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500299270	500299270	Scottish painter and architect, 1830-1923
Charles Aubry <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Charles (34) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Charles (33) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Emile (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Louis (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, René Marcel (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500029715	500029715	French photographer and designer, 1811-1877
Fédèle Azari <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Azari, Fedele (19) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Āzarī Isfarāyīnī, Ĥamzah ibn 'Alī Malik (5) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500103519	500103519	Italian photographer, 1895-1930
František Vobecký <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Vobecký, František (43) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Schmoranz, František (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Fišer, František (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Kadlec, František (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Kysela, František (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500028103	500028103	Czech photographer, 1902-1990
Eugen Wiskovsky <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Wiškovský, Eugen (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Kahler, Eugen von (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Ržibek, Eugen (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Nevan, Eugen (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Osswald, Eugen (13) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500031206	500031206	Czech photographer, 1888-1964

4.3. OpenRefine Example: Parse All Variant Terms from Full XML

1. Add column and using the full XML, parse all variant names:

```
join(forEach(value.parseXml().select('Vocabulary')[0].select('Subject')[0].select('Terms')[0].select('Non-Preferred_Term'),term,term.select('Term_Text')[0].ownText()),", ")
```

Add column based on column ULAN XML

New column name

On error set to blank store error copy value from original column

Expression Language No syntax error.

```
join(forEach(value.parseXml().select('Vocabulary')[0].select('Subject')[0].select('Terms')[0].select('Non-Preferred_Term'),term,term.select('Term_Text')[0].ownText()),", ")
```

Preview History Starred Help

row	value	join(forEach(value.parseXml(). ...
1.	<?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://vocabs:<Subject Subject_ID="500299270"> <Parent_Relationships> <Preferred_Parent> <Parent_Subject_ID>500000002</Parent_Subject	Hamilton Maxwell, H. Maxwell, Maxwell

OK Cancel

Variant Names, comma-separated, in OpenRefine cells

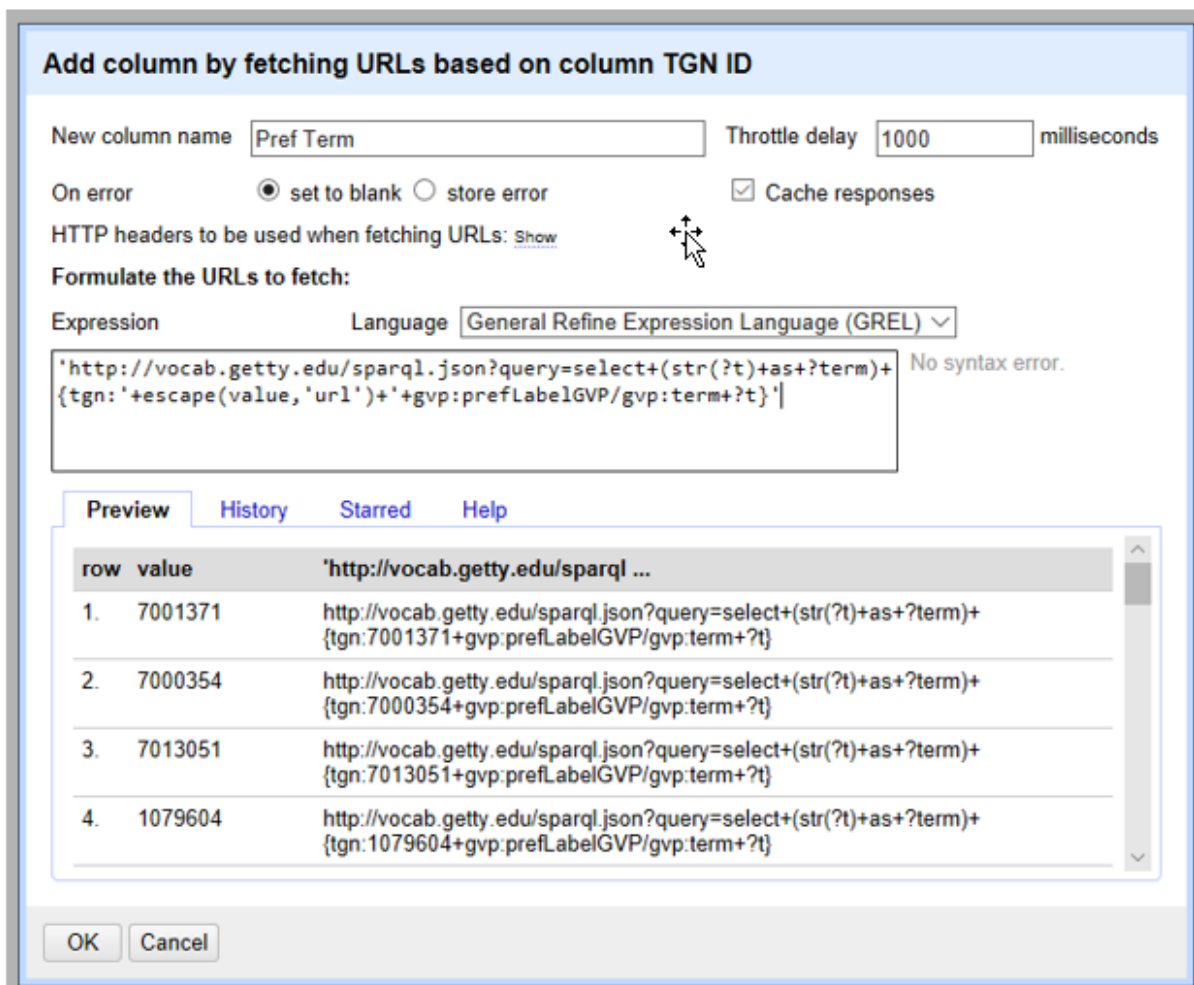
DISPLAYNAME	ULAN ID	Sync ID	Variants	Pref Bio	Pref Term
Maxwell Hamilton <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Maxwell, Hamilton (24) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Hamilton, Maxwell, Inc. (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Fry, Edwin Maxwell (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Bates, Maxwell (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Maxwell, John (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500299270	500299270	Hamilton Maxwell, H. Maxwell, Maxwell	Scottish painter and architect, 1830-1923	Maxwell, Hamilton
Charles Aubry <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Charles (34) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Charles (33) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Emile (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, Louis (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aubry, René Marcel (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500029715	500029715	Charles Aubry, Aubry, Charles-Hippolyte	French photographer and designer, 1811-1877	Aubry, Charles
Fédèle Azari <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Azari, Fedele (19) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Āzarī Isfarāyīnī, Hamzah ibn 'Alī Malik (5) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	500103519	500103519	Fedele Azari	Italian photographer, 1895-1930	Azari, Fedele
František Vobecký <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Vobecký, František (43) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Schmoranz, František (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Fišer, František (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Kadlec, František (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Kysela, František	500028103	500028103	Vobecky, Frantisek, František Vobecký, Vobecký, Frantisek	Czech photographer, 1902-1990	Vobecký, František



4.4. OpenRefine Example: Get TGN Place Coordinates from SPARQL Endpoint

1. Get preferred term by adding column by fetching URLs based on TGN ID, and input formula:

```
'http://vocab.getty.edu/sparql.json?query=select+(str(?t)+as+?term)+  
{tgn:'+escape(value,'url')+'+gvp:prefLabelGVP/gvp:term+?t}'
```



Add column by fetching URLs based on column TGN ID

New column name Throttle delay milliseconds

On error set to blank store error Cache responses

HTTP headers to be used when fetching URLs: [Show](#)

Formulate the URLs to fetch:

Expression Language

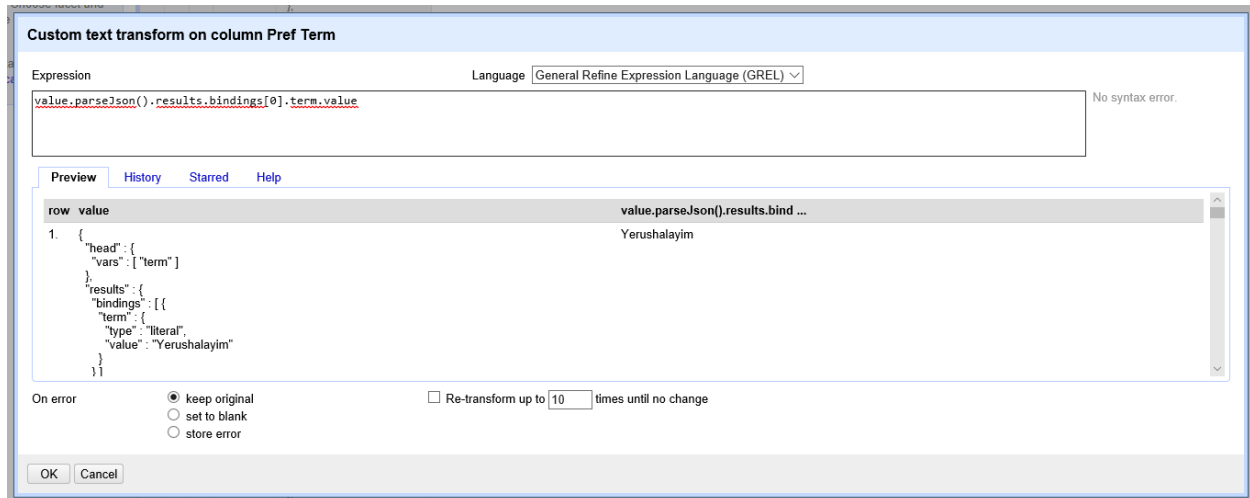
```
'http://vocab.getty.edu/sparql.json?query=select+(str(?t)+as+?term)+  
{tgn:'+escape(value,'url')+'+gvp:prefLabelGVP/gvp:term+?t}'
```

No syntax error.

Preview History Starred Help

row	value	'http://vocab.getty.edu/sparql ...
1.	7001371	http://vocab.getty.edu/sparql.json?query=select+(str(?t)+as+?term)+ {tgn:7001371+gvp:prefLabelGVP/gvp:term+?t}
2.	7000354	http://vocab.getty.edu/sparql.json?query=select+(str(?t)+as+?term)+ {tgn:7000354+gvp:prefLabelGVP/gvp:term+?t}
3.	7013051	http://vocab.getty.edu/sparql.json?query=select+(str(?t)+as+?term)+ {tgn:7013051+gvp:prefLabelGVP/gvp:term+?t}
4.	1079604	http://vocab.getty.edu/sparql.json?query=select+(str(?t)+as+?term)+ {tgn:1079604+gvp:prefLabelGVP/gvp:term+?t}

2. Transform resulting JSON: `value.parseJson().results.bindings[0].term.value`



3. Get coordinates by adding column by fetching URLs based on TGN ID, and input formula:

```
'http://vocab.getty.edu/sparql.json?query=select+?lat+?long+{tgn:'+'+escape(value,'url')+'+'+foaf:focus+?place.+?place+wgs:lat+?lat.+?place+wgs:long+?long}'
```

Add column by fetching URLs based on column TGN ID

New column name Throttle delay milliseconds

On error set to blank store error Cache responses

HTTP headers to be used when fetching URLs: [Show](#)

Formulate the URLs to fetch:

Expression Language

```
'http://vocab.getty.edu/sparql.json?query=select+?lat+?long+{tgn:'+'+escape(value,'url')+'+'+foaf:focus+?place.+?place+wgs:lat+?lat.+?place+wgs:long+?long}'
```

No syntax error.

[Preview](#) [History](#) [Starred](#) [Help](#)

row	value	'http://vocab.getty.edu/sparql ...'
1.	7001371	http://vocab.getty.edu/sparql.json?query=select+?lat+?long+{tgn:7001371+foaf:focus+?place.+?place+wgs:lat+?lat.+?place+wgs:long+?long}
2.	7000354	http://vocab.getty.edu/sparql.json?query=select+?lat+?long+{tgn:7000354+foaf:focus+?place.+?place+wgs:lat+?lat.+?place+wgs:long+?long}
3.	7013051	http://vocab.getty.edu/sparql.json?query=select+?lat+?long+{tgn:7013051+foaf:focus+?place.+?place+wgs:lat+?lat.+?place+wgs:long+?long}
4.	1079604	http://vocab.getty.edu/sparql.json?query=select+?lat+?long+{tgn:1079604+foaf:focus+?place.+?place+wgs:lat+?lat.+?place+wgs:long+?long}

4. Parse resulting JSON:

`value.parseJson().results.bindings[0].lat.value+',`
`'+value.parseJson().results.bindings[0].long.value`

Custom text transform on column LatLong

Expression Language General Refine Expression Language (GREL)

`value.parseJson().results.bindings[0].lat.value+', '+value.parseJson().results.bindings[0].long.value` No syntax error.

Preview History Starred Help

row	value
1.	{ "head": { "vars": ["lat", "long"] }, "results": { "bindings": [{ "lat": { "datatype": "http://www.w3.org/2001/XMLSchema#decimal", "type": "literal", "value": "31.7667" } } } }

On error keep original Re-transform up to times until no change set to blank store error

OK Cancel

Resulting data enrichment:

All	TGN ID	Pref Term	LatLong	
☆	1.	7001371	Yerushalayim	31.7667, 35.2333
☆	2.	7000354	Marrakech	31.633, -8.009
☆	3.	7013051	Montréal	45.5, -73.6
☆	4.	1079604	Qiryat Tiv'on	32.7167, 35.1333
☆	5.	7010273	Sankt-Peterburg	59.8833, 30.25
☆	6.	7001369	Tel Aviv-Yafo	32.0833, 34.7667
☆	7.	1079614	Rishon LeZiyyon	31.95, 34.8
☆	8.	7007567	New York	40.7, -74
☆	9.	7023856	Shanghai	31.222222, 121.458056
☆	10.	7001370	Ramat Gan	32.0833, 34.8167
☆	11.	1079547	Kinneret	32.7167, 35.5667
☆	12.	1079641	Umm el Fahm	32.5167, 35.15
☆	13.	7001371	Yerushalayim	31.7667, 35.2333
☆	14.	7001362	Hefa	32.8167, 34.9833
☆	15.	7012178	Dushanbe	38.6333, 68.85
☆	16.	7004443	Düsseldorf	51.216667, 6.766667
☆	17.	7012333	Mulhouse	47.738082, 7.347932
☆	18.	7001371	Yerushalayim	31.7667, 35.2333
☆	19.	7001369	Tel Aviv-Yafo	32.0833, 34.7667
☆	20.	7001369	Tel Aviv-Yafo	32.0833, 34.7667
☆	21.	7001369	Tel Aviv-Yafo	32.0833, 34.7667
☆	22.	1079466	Holon	32.0167, 34.7667

4.5. OpenRefine Example: Query Reconciliation Service Using HTTP

Send JSON via HTTP Request

JSON Request Example 1:

```
queries=  
{  
  "q0": {  
    "query": "rhyta",  
    "type": "/aat"  
  }  
}
```

Curl request:

```
curl -X POST -d 'queries={"q0":{"query":"rhyta","type":"/aat"}}'  
"http://services.getty.edu/vocab/reconcile/"
```

Parse JSON Response

```
{
  "q0": {
    "result": [
      {
        "id": "aat/300198841",
        "match": false,
        "name": "rhyta",
        "score": 19.114761,
        "type": .... ,
      }
      {
        "id": "aat/300265000",
        "match": false,
        "name": "dimidiating rhyta",
        "score": 4.626431,
        "type": ...
      }
    ]
  }
}
```

Send JSON via HTTP Request

JSON Request Example 2:

```
queries=  
"q2": {  
    "query": "berlin",  
    "type": "/tgn",  
    "limit": 10,  
    "properties": [  
        {  
            "p": "place type",  
            "pid": "placeType",  
            "v": "inhabited places"  
        },  
        {  
            "p": "broader place",  
            "pid": "broaderExt",  
            "v": "Germany"  
        }  
    ]  
}
```

Curl request:

```
curl -X POST -d  
'queries={"q2":{"query":"berlin","type":"/tgn","limit":10,"properties":[{"p":"place  
type","pid":"placeType","v":"inhabited places"},{"p":"broader  
place","pid":"broaderExt","v":"Germany"}]}'  
"http://services.getty.edu/vocab/reconcile/"
```


Parse JSON Response

```
{
  "q2": {
    "result": [
      {
        "id": "tgn/7209905",
        "match": false,
        "name": "Berlin",
        "score": 18.897251,
        "type": ...
      },
      {
        "id": "tgn/7029437",
        "match": false,
        "name": "East Berlin",
        "score": 18.348005,
        "type": ...
      },
      {
        "id": "tgn/7003712",
        "match": false,
        "name": "Berlin",
        "score": 17.77391,
        "type": ...
      }
    ]
  }
}
```

Correct match: <http://vocab.getty.edu/page/tgn/7003712>

Online Resources

- Vocab Reconciliation Tutorial
http://www.getty.edu/research/tools/vocabularies/obtain/getty_vocabularies_open_refine_tutorial.pdf
- General OpenRefine Functions
<https://github.com/OpenRefine/OpenRefine/wiki/GREL-Functions>
- OpenRefine Reconciliation Service Documentation
<https://reconciliation-api.github.io/specs/0.1/>
- LibraryCarpentry Open Refine training
<https://librarycarpentry.org/lc-open-refine/>
- Vocabularies XML Web Services
http://www.getty.edu/research/tools/vocabularies/vocab_web_services.pdf
- Vocabularies SPARQL Queries Example Page
<http://vocab.getty.edu/queries>

5. Library of GREL, SPARQL, SQL, and other expressions

(under construction)

5.1 GREL

- To fill in blank/no artist columns to fit standard of “unknown ‘culture’ artist”
`"unknown " + value.replace(cells["ULAN_Artist"].value, (cells["culture"].value)) + " artist"`
- To change contributed gender info “W” to standard of F for female
`value.replace("W", "F")`
- To fill in death dates fitting ULAN standard of 120 + birth date
`if(isNonBlank(value), value, cells["birth_year"].value+120)`
- to fill in birth dates if not known (put in year based on ULAN guidelines)
`if(isNonBlank(value), value, "1900")`
- To change date from year-month-day to 4-digits representing year
`value.toDate('yyyy-MM-DD').toString('yyyy')`
- to get rid of unnecessary parentheses, semicolons, etc.
`value.replace("(", "")`
- To change certain abbreviations to ULAN editorial terms
`value.replace("b.", "born")`
- To fill in “artist” when artist role field is blank
`if(isNonBlank(value), value, "Artist")`
- To concatenate two columns in a third column
`cells["family name"].value + ", " + cells["given name"].value`
- To add column values together; this can be lengthened or iterative; good for cases where there is no biographical information or if citation information is spread across columns
`cells["col1"].value + ", " + cells["col2"].value`
- To put artist roles in blank biographical descriptions, iterative. Notice patterns in the data, e.g., “American”
`value.replace(", ", " artist,")`
- To create a column with stars or flags. Can be useful if need to deal with a certain set of parameters
`if(row.starred, "yes", "no")`
- To add a column by fetching URLs based on reconciled ULAN column
`'http://vocabsservices.getty.edu/ULANService.asmx/ULANGetSyncSubjectID?subjectID='+value`
- To fetch the ID for the reconciled names
`cell.recon.match.id`
- To retrieve values in a separate OpenRefine project/spreadsheet based on a shared attribute (what columnToMatch refers to). Projects do not have to be running in OpenRefine for this to work. E.g., retrieving artist roles from a separate OpenRefine project based on the same subject ID shared in both projects
`cell.cross(otherProject, columnToMatch)[0].cells[columnToAdd]`

- To retrieve *yes* or *no* values based on information located in OpenRefine projects (can be iterative). E.g.,. Are there death dates greater than 2000? Or checking for the presence of particular string values. Can iterate to check across spreadsheets
`if(cell.cross(table, column)[0].value > 2000, "yes", "no")`
`if(cell.cross(table, column)[0].value = "stringvalue", "yes", "no")`
- Replacing characters; can be useful to make changes to separators or terms across columns. Also can be useful to create standard separators before using the slice() function
`replaceChars(value, ";;() ", "|")`
- To find particular string values, can also be used to find particular numerical values, ex. `value.find("number")`. This is case sensitive.
`value.find("string")`
- To split data based on a common separator, string, or number, useful if a field has information not germane to the field (but you want to capture it elsewhere). The numbers refer to which item in the series you are editing.
`slice(split(value, ","), 0, 2).join(",")`
 E.g., Original data = "United States"
`slice(split(value, ""), 0, 1).join("") = "United"`
- To reverse order if names are first name, last name
`value.match(/(.*) (.*)/).reverse().join(" ")`
- Similar to adding data in columns together as above, but checks first if the column value is blank or not. Useful if you do not want to change existing values, but need to edit some results in a particular column.
`if(isNonBlank(value), value, cells["col1"].value + ", " + cells["col2"].value + ", " + cells["col3"].value)`

5.2. SPARQL

- Query using SPARQL and see the 90 sample queries here: <http://vocab.getty.edu/doc/queries/> . Additional queries are below.
- To find a value, such as a source's ID or URL, in the "page" field. In this example, the Nomenclature ID can be found by querying on the source page field, or document part in the LOD profile.
`select * {?vp agvp:Subject; dcterms:source/bibo:locator ?nom
 FILTER(contains(str(?nom), 'nomenclature.info'))}`

5.3. SQL

- To find Scope Notes of a certain language (Dutch here) that are missing, but include the English Scope Note for reference (AAT)
`select subject_id, note_text from scope_notes where
subject_id not in (select subject_id from scope_notes where language_code=70261)`
- To find names from a certain contributor where the names have the flag “local use”, employing a *join* (ULAN)
`select distinct contrib_rels_term.subject_id from term, contrib_rels_term where
term.historic_flag = 'LU'
and contrib_rels_term.contrib_id = 2500000298`
- To find a value in the Display Dimensions field (CONA)
`select distinct subject_id from other_displays where
display_dimensions like '%braccia%'`
- To find IDs in the Iconography Authority (which is in the CONA data model)
`select distinct iconography_id from iconography_authority`
- To find records that are new from a particular time range (any vocab)
`select distinct rh.subject_id from revision_history rh, subject sub where
rh.date_time between '1 Aug 2019' and '1 Feb 2020'
and (rh.action like 'moved' or (rh.action like 'created' and rh.record_type like 'S'))
and sub.subject_id=rh.subject_id and sub.candidate_stat = 'N'`
- To find terms with a specific source (any vocab)
`Select subject_id from source_rels_term where source_id = 2100039329`
- To find records where a specific birth or death place is used in a biography (ULAN)
`select distinct.subject_id from biography where biography.birth_place like '4540060004'
or biography.death_place like '4540060004'`
- To find people that have a birth or death place of London, England, that are NOT British (ULAN)
`select subject_id from biography
where (birth_place = 4601330462 or death_place = 4601330462)
and biography not like 'British%'`
- To find people erroneously flagged as Male, based on words in name, e.g., “Mrs” (ULAN)
`select distinct.subject_id from term t, biography b
where t.subject_id = b.subject_id
and b.sex like 'M' and t.normalized_term like '%MRS%'`
- To find records which have a specific start date in the events field (ULAN, CONA)
`select subject_id from event_rels where start_date = '1840'`
- To search for records with no value in a certain field (any Vocab)
`select subject_id from subject where scope is null or scope = '' [two single quotes]`
- To find records only of Czech painters (ULAN)
`select subject_id from nationality N, ptype_role_rels R
where N.nationality_code = '905800'
and R.ptype_role_id = 31261
and N.subject_id = R.subject_id`
- To create an alpha-sorted list with preferred biographies that have “British artist” in them (ULAN)

```

select bio.subject_id from biography bio, term tm
where bio.preferred = 'P'
AND bio.biography like '%British artist%'
AND tm.subject_id = bio.subject_id
AND tm.preferred = 'P'
order by tm.normalized_term

```

- To find Corporate Bodies erroneously in Person facet (ULAN)
[to find Persons erroneously placed in Corp Body facet, NB, Corporate Body ID is 500000003]

```

Select distinct ptype_role_rels.subject_id from ptype_role_rels, subject
where subject.candidate_stat='N'
and subject.parent_key=500000002
and ptype_role_rels.subject_id=subject.subject_id
and ptype_role_rels.ptype_role_id=30003

```
- To find Non-Candidates with a specific birth or death date (-2000, 9999, etc) (ULAN)

```

select distinct biography.subject_id from biography, subject
where subject.candidate_stat = 'N'
and biography.subject_id = subject.subject_id
and biography.birth_date = -2000
and biography.preferred = 'P'

```
- To find Non-Candidates with Nationality of “undetermined,” (ULAN)

```

Select distinct nationality.subject_id from nationality, subject
where subject.candidate_stat = 'N'
and nationality.subject_id = subject.subject_id
and nationality.national_id = 900010

```
- To find Non-Candidates that are not monogrammists or masters and that have gender/sex = undetermined

```

select distinct sub.subject_id from subject sub, nationality nat, ptype_role_rels prr, term tm
where sub.candidate_stat = 'N' and nat.subject_id=sub.subject_id
and nat.nationality_code = '900010'
and prr.subject_id=sub.subject_id and prr.ptype_role_id in (10010, 10000)
and tm.subject_id=sub.subject_id and tm.term not like 'Monogram%'
and tm.term not like '%Master%'

```
- To find birth and death dates that are not 9999, and greater than 2090, in any Preferred display biography, in any record except those for Corporate Bodies (ULAN)

```

Select distinct biography.subject_id from biography, subject
where subject.candidate_stat = 'N'
and subject.record_type = 'P'
and biography.subject_id = subject.subject_id
and biography.preferred = 'P'
and (biography.birth_date > 2090
or biography.death_date > 2090)
and (biography.birth_date not like '9999'
or biography.death_date not like '9999')

```
- To find problem punctuation in submitted or merged terms (brackets, parens, symbols) (all Vocab)

```

Select distinct subject_id from term

```

where subject_id in (select subject_id from vnoncandidate_subjects)
and term like '%!%'

- To find scope notes without period at end of sentence (all Vocabs)
Select subject_id, scope from subject where rtrim(scope) not like '%.' and rtrim(scope) not like '%.'
- To find all records with certain place types where record type is erroneously “administrative” or erroneously “physical” (TGN)
Select ptr.subject_id from ptype_role_rels ptr, subject sub
where ptr.ptype_role_id in (21486, 21105, 21430)
and sub.subject_id=ptr.subject_id
and sub.record_type='A'
- To find Non-Candidates where Preferred term does not have English or American English as a language (AAT)
Select subject_id from term where preferred='P'
and term_id not in (select term_id from language_rels
where language_code in ('70051', '70052'))
and subject_id in (select subject_id from vnoncandidate_subjects)
- AAT: Report To find records with English or American English as language, but neither flagged as preferred (AAT)
Select subject_id from term where preferred = 'P'
and term_id in (select term_id from language_rels
where language_code in ('70051', '70052') and preferred='N')
and subject_id in (select subject_id from vnoncandidate_subjects)
- To find terms which have adjectival form as preferred (AAT)
Select distinct term.subject_id from term, language_rels
where term.term_type='A'
and term.term_id=language_rels.term_id
and language_rels.preferred='P'
- To find null scopes in Non-Candidates (AAT)
Select distinct subject_id from subject
where candidate_stat='N' and record_type='C'
and subject_id not in (select subject_id from scope_notes)
- To find one-word names under a given parent where display bio has a certain word in it (ULAN)
select * from term where term not like '% %' and preferred='P' and subject_id in
(select subject_id from contrib_rels_subject where contrib_id=2500000011)
and subject_id in
(select subject_id from biography where preferred='P' and biography like '%century%')
and subject_id in
(select subject_id from subject where parent_key=500355043)
order by subject_id
- To find counts of all nationalities linked to records (ULAN)
select lv.code, lv.description, count(*) as nat_count from nationality nat, lookup_values lv
where lv.list_name='nat_nationality_code' and lv.code=nat.nationality_code and
nat.subject_id in (select subject_id from vnoncandidate_subjects)
group by lv.code,lv.description order by nat_count desc,lv.description
- Find terms in specific languages (all Vocabs)
select distinct.language_rels.subject_id from language_rels, term
where language_rels.language_code = '70049'
and language_rels.subject_id = term.subject_id

- To find a role that is Preferred in records in the Non-Artists facet (ULAN)
`select ptr.subject_id from ptype_role_rels ptr, subject sub
where ptr.ptype_role_id in (38131)
and ptr.preferred = 'P'
and sub.subject_id=ptr.subject_id
and sub.parent_key='500299802'`