

# Getty Vocabularies Reconciliation Service Using OpenRefine

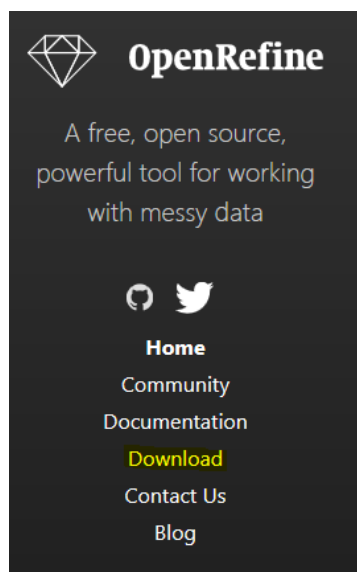
Step-by-step process from program installation to implementation

May 2019

Gregg Garcia, Getty Digital  
Documentation by Lindsey Gant

## Step 1. Download OpenRefine

OpenRefine is an open source tool that can be downloaded at [openrefine.org](https://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:

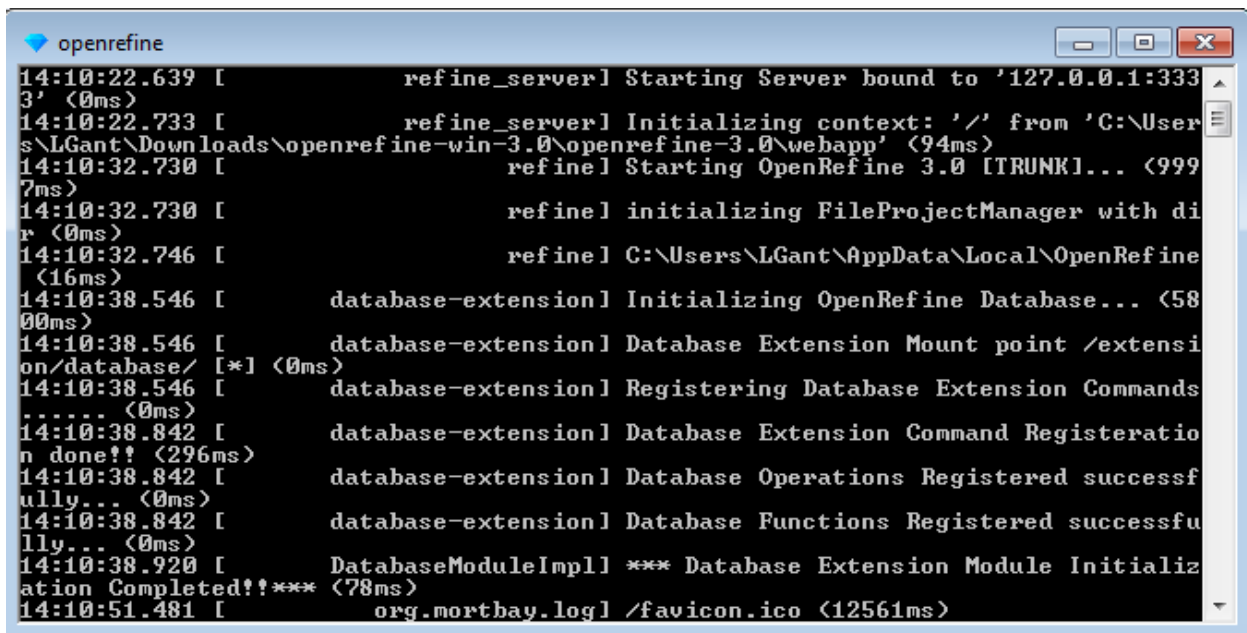


Select the Download option from the left column, as highlighted above. Choose the latest version for either the Windows, Mac, or Linux kit. For this demonstration, I am using Windows.

After installation is complete, unzip the file and save to your desktop. To open the program, select the `openrefine.exe` file (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
<b>openrefine.exe</b>	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

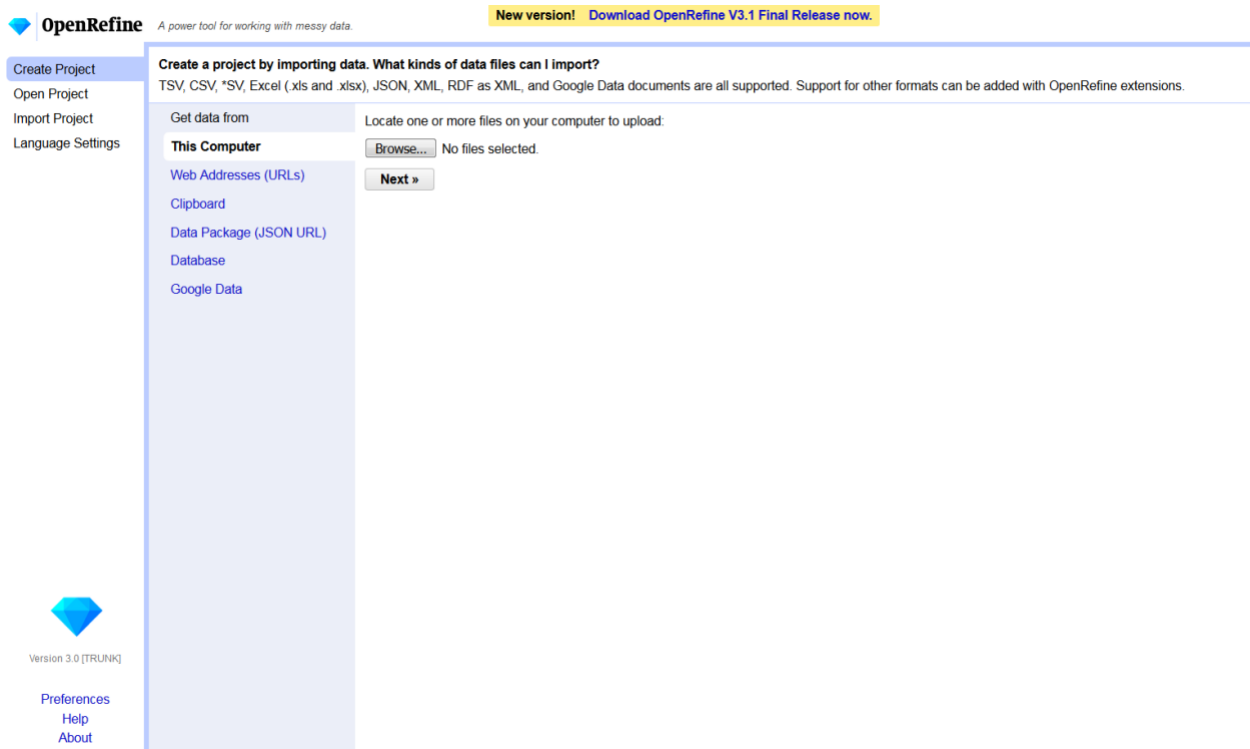
Once the application is selected, a command-line interface will open, initiating the web application. Keep this window open in the background; OpenRefine is a web-based application that works in your browser.



```
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.mortbay.log] /favicon.ico <12561ms>
```

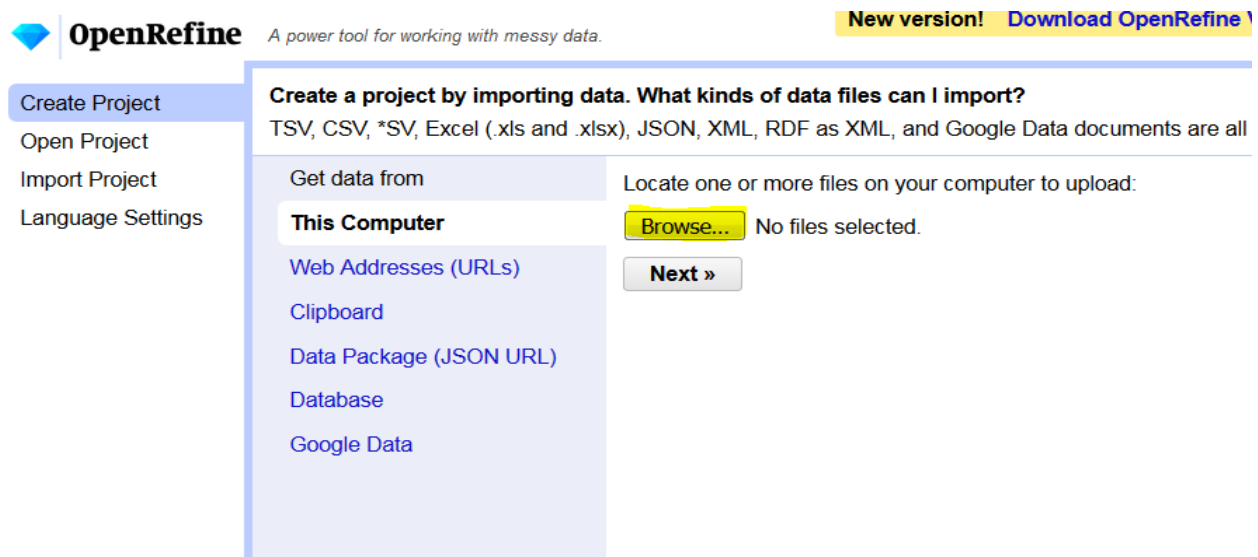
Wait a few moments for the application to open.

Once it opens, it should look like this:



## Step 2. Import spreadsheet

Click the Browse button and upload the existing data (may include Excel, CSV, XML, JSON, etc.).



Once uploaded in the Preview section, click Create Project on the top right side.

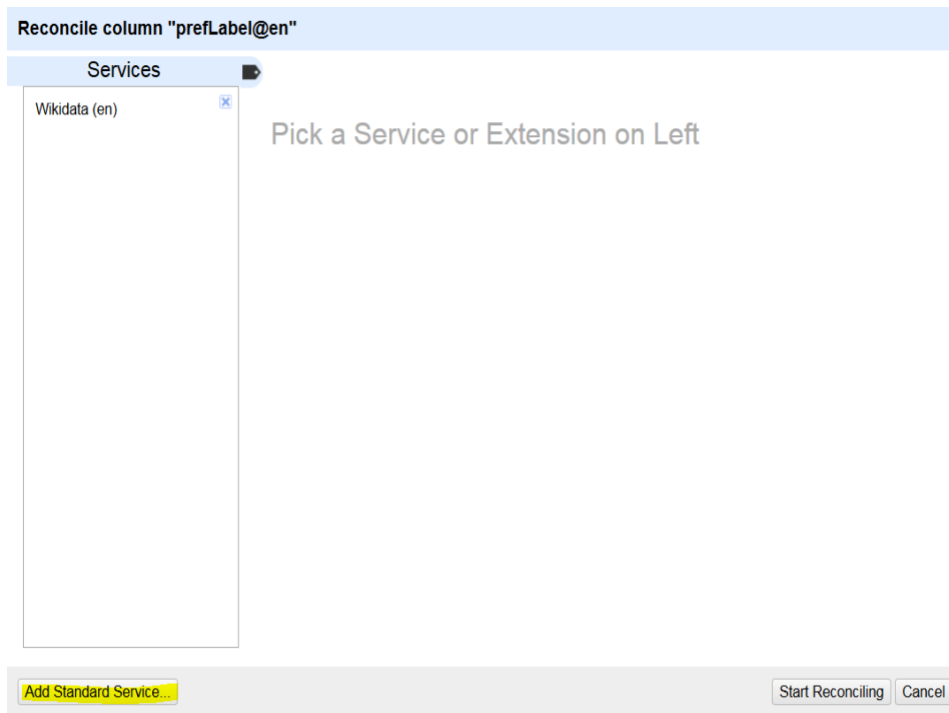
« Start Over		Configure Parsing Options		Project name IMJ.xls		Tags		Create Project »			
artistName	artistNonPrefName	artistBirth	artistDeath	artistNationality	artistBirthPlace 1	artistBirthPlace 2	artistBirthPlace 3	artistBirthPlace 4	artistBirthPlaceNation	artistRole	artist
1. Aaronson, Yakov	אהרונסון, יעקב	1924	2100	Israeli					Israel	31407/photographer	
2. Abadi, Abed	עבדי, עבד	1942	2100	Israeli					Israel	31261/painter	
3. Abadi, Shay	עבאדי, שי	1965	2100	Israeli	Jerusalem	Israel				31100/artist	
4. Abakanowicz, Magdalena	אבאקנוביץ' 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	אבט, 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	מאיי, אבלי	1904	1946	American					USA	31261/painter	3143

## Part 3. Connect to Getty Vocab reconciliation service

Choose the column for reconciliation (e.g., artistName). From the drop-down menu, hover over Reconcile and select Start reconciling.

19 rows											
Show as: rows records		Show: 5 10 25 50 rows									
All	artistName	artistNonPrefName	artistBirth	artistDeath	artistNationality	artistBirthPlace					
1.	Facet	אהרונסון, יעקב	1924	2100	Israeli						
2.	Text filter	עבדי, עבד	1942	2100	Israeli						
3.	Edit cells	עבאדי, שי	1965	2100	Israeli	Jerusalem					
4.	Edit column	אבאקנוביץ' מגדאלנה	1930	2017	Polish						
5.	Transpose	אבא, עירית	1953	2100	Israeli						
6.	Sort...	אבאסי, ריזה, (מיוחס ל)	1200	2100	Iranian						
7.	View	אבאסי, ריזה (אסכולה)	1200	2100	Iranian						
8.	Reconcile	אבאסי, ריזה	1565	1635	Iranian						
9.	Start reconciling...	אבט, ברניס	1898	1991	American						
10.	Facets	אמיל עבוד, ג'ומאנה		2100	Israeli	Shefaram					
11.	Actions	האחים עבדוללה	1858	1899	Turkish						
12.	Copy reconciliation data...	אבקסיס, רפאל		2100	Moroccan	Marrakesh					
13.	Use values as identifiers	אבלי, מאיי	1904	1946	American						
14.				2100	Cuban						
15.		אבלס, גבי	1200	2100							

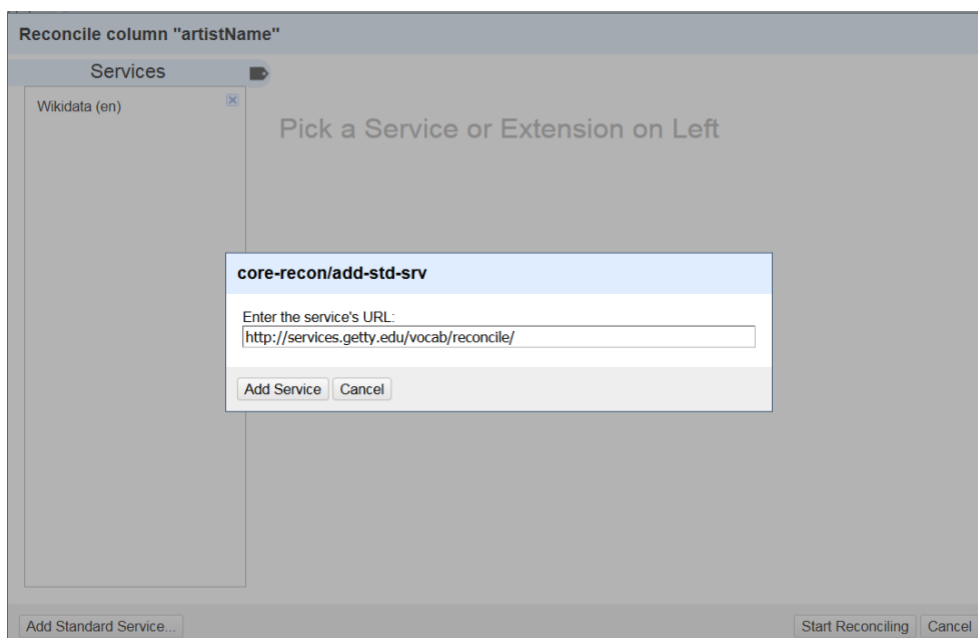
Click the Add Standard Service button at the bottom left of the page.



Input the following URL to connect to the Vocabularies service:

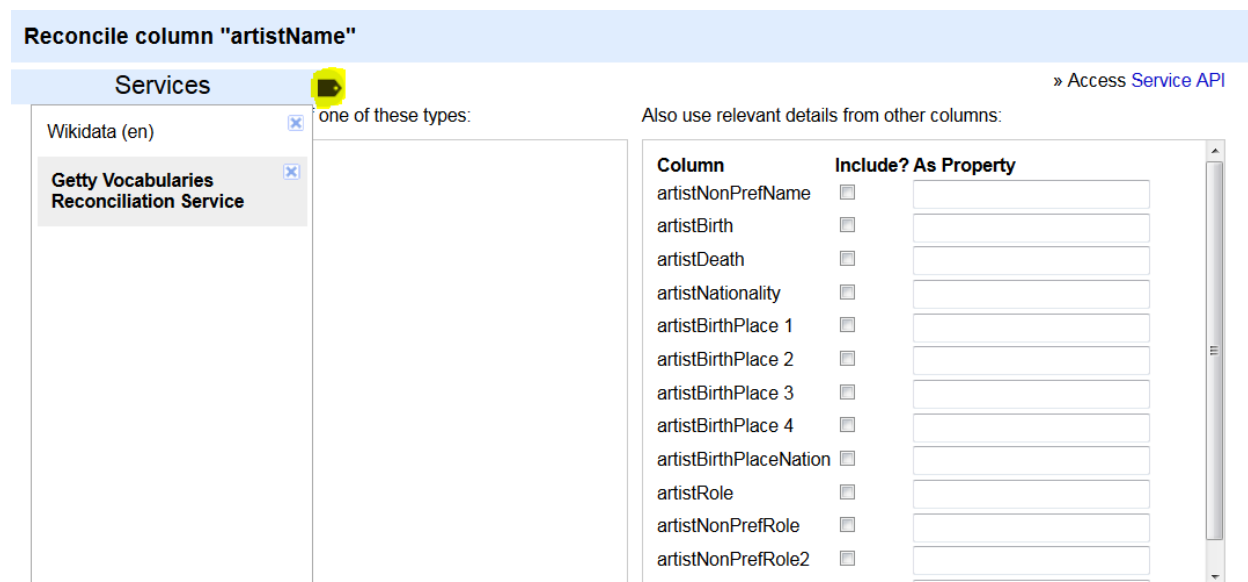
<http://services.getty.edu/vocab/reconcile/>

Click Add Service.



## Part 4. Select corresponding Vocabs properties

To view the Vocabularies options, click the tab symbol (highlighted below) to close the Services tab.



Reconcile column "artistName"

Services ▶ » Access [Service API](#)

Wikidata (en) ✕ one of these types:

Getty Vocabularies Reconciliation Service ✕

Also use relevant details from other columns:

Column	Include? As Property
artistNonPrefName	<input type="checkbox"/>
artistBirth	<input type="checkbox"/>
artistDeath	<input type="checkbox"/>
artistNationality	<input type="checkbox"/>
artistBirthPlace 1	<input type="checkbox"/>
artistBirthPlace 2	<input type="checkbox"/>
artistBirthPlace 3	<input type="checkbox"/>
artistBirthPlace 4	<input type="checkbox"/>
artistBirthPlaceNation	<input type="checkbox"/>
artistRole	<input type="checkbox"/>
artistNonPrefRole	<input type="checkbox"/>
artistNonPrefRole2	<input type="checkbox"/>

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

In the As Property column, add the corresponding Vocabularies property fields you wish to reconcile with existing columns in your spreadsheet, as demonstrated below. Please note the list of properties available for each Vocabulary type:

### For AAT

- note: the concept scope note
- broaderExt: any broader concept
- label: any variant term

### For ULAN

- agentType: the role of a person or organization
- nationality: the nationality of a person or organization
- gender: the gender of a person
- description: descriptive text about a person or organization
- birthPlace: a person's birthplace
- deathPlace: the place a person died
- birthDate: a person's date of birth

- deathDate: a person's date of death
- startDate: the date a person or organization began being active
- endDate: the date a person or organization stopped being active
- location: the location where a person or organization was active
- type: can be either schema:Person or schema:Organization
- parent: the parent or facet name

**For TGN**

- broaderExt: any broader place associated with the place being reconciled
- placeType: type of place being reconciled (nation, inhabited place, city, etc.)

**Reconcile column "artistName"** » Access [Service API](#)

Reconcile each cell to an entity of one of these types: Also use relevant details from other columns:

- ULAN search  
/ulan
- TGN search  
/tgn
- AAT search  
/aat

Column	Include?	As Property
artistNonPrefName	<input type="checkbox"/>	
artistBirth	<input type="checkbox"/>	birthDate
artistDeath	<input type="checkbox"/>	deathDate
artistNationality	<input type="checkbox"/>	nationality
artistBirthPlace 1	<input type="checkbox"/>	
artistBirthPlace 2	<input type="checkbox"/>	
artistBirthPlace 3	<input type="checkbox"/>	
artistBirthPlace 4	<input type="checkbox"/>	
artistBirthPlaceNation	<input type="checkbox"/>	
artistRole	<input type="checkbox"/>	agentType
artistNonPrefRole	<input type="checkbox"/>	
artistNonPrefRole2	<input type="checkbox"/>	

Reconcile against type:   
 Reconcile against no particular type  
 Auto-match candidates with high confidence

Maximum number of candidates to return

Note: Deselect “Auto-match candidates with high confidence” to prevent the reconciliation of terms to incorrect Getty Vocabularies records.

Click Start Reconciling at the bottom of the screen.

Resulting matches:

11 matching rows (19 total)						
Show as: <b>rows</b> records    Show: 5 10 25 <b>50</b> rows						
All	artistName	artistNonPrefName	artistBirth	artistDeath	artistNationality	
☆	3. Abadi, Shay <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Abadi, Shay (46) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Alkalay, Shay (22) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Abadi, Abed (20) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Peri, Shay Frisch (20) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Cleary, Shay (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Docking, Shay (17) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Shay, Art (16) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Abadi, Fritzie (15) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Shay, James (15) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Shay, Patricia (15) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	עבאדי, שי	1965	2100	Israeli	
☆	4. Abakanowicz, Magdalena <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Abakanowicz, Magdalena (43) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Hambrouch, Magdalena (20) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Wiecek, Magdalena (18) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Jetelová, Magdalena (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Pažourková, Magdalena (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Franciskovic, Magdalena (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Correa, Magdalena (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Pozas, Magdalena (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Rădulescu, Magdalena (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Jitrik, Magdalena (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	אבאקנוביץ' מגדלנה	1930	2017	Polish	
☆	6. Abbasi, Riza (attributed to) <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 Museum (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aqa Riza (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Shaykh 'Abbasi (22) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> To, Hiram (13) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Chiem, To (13) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> To, Johnnie (13) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Sze To (12) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> TO Design (12) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	אבאסי, ריזה, (מיחס ל')	1200	2100	Iranian	

## Part 5. Begin OpenRefine reconciliation

After the reconciliation process ends, follow standard OpenRefine guidelines for reconciliation.

Note: To view the full item record for each returned element, right-click on the highlighted name and select Open Link in New Tab.



For AAT reconciliation:

Click the highlighted link for a returned element to view a pop-up with a brief excerpt from the item scope note and its placement in the AAT hierarchy, as demonstrated below:

The screenshot shows the OpenRefine reconciliation interface. On the left, a list of terms is displayed with checkboxes. The term "string quilts" is highlighted in yellow. A pop-up window is overlaid on the right, containing the following information:

- Buttons: Match this Cell, Match All Identical Cells, Cancel
- Term: **string quilts**
- Description: Quilts similar to log cabin quilts in design and technique, except for the use of irregular shapes and a more random design.
- Location in Hierarchy: Objects Facet
  - \_\_\_\_\_ ...
  - \_\_\_\_\_ quilts
  - \_\_\_\_\_ <quilts bv pattern or motif>

For all Vocabularies:

To get Vocabularies IDs for the best match from the results list, select “Add column based on this column.”

The screenshot shows the OpenRefine interface with a table of 11 matching rows (19 total). The table has columns for 'All', 'artistName', 'artistNonPrefNa', 'artistBirth', and 'artistDeath'. A context menu is open over the 'artistName' column, with the option 'Add column based on this column...' highlighted. The table data is as follows:

All	artistName	artistNonPrefNa	artistBirth	artistDeath
3.	עבאד, שי		1965	2100
4.	Abakanowicz, Magdalena	אבא	1930	2017

The context menu for 'artistName' includes the following options:

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

The 'Edit column' sub-menu is open, showing the following options:

- Split into several 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

Create a new column name (i.e., ULAN ID). Enter cell values notation for ID:  
**cell.recon.best.id**

### Add column based on column artistName

New column name

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

Expression  Language  No syntax error.

**Preview** History Starred Help

row	value	cell.recon.best.id
3.	Abadi, Shay	ulan/500486311
4.	Abakanowicz, Magdalena	ulan/500084577
6.	Abbasi, Riza (attributed to)	ulan/500309512
7.	Abbasi, Riza (school of)	ulan/500309512
8.	Abbasi, Riza	ulan/500116631
9.	Abbott, Berenice	ulan/500020631
11	Abdullah Ersoy	ulan/500022285

The new column will contain the Vocabulary ID of any matched values.

Resulting spreadsheet:

11 matching rows (19 total)							
Show as: rows records		Show: 5 10 25 50 rows					
All	artistName	ULAN ID	artistNonPrefName	artistBirth	artistDeath	artistNationality	
<input type="checkbox"/>	3. Abadi, Shay <small>Choose new match</small>	ulan/500486311	עבאדי, שי	1965	2100	Israeli	Jerusalem
<input type="checkbox"/>	4. Abakanowicz, Magdalena <small>Choose new match</small>	ulan/500084577	אבאקנוביץ' מגדאלנה	1930	2017	Polish	
<input type="checkbox"/>	6. Riza 'Abbasi Museum <small>Choose new match</small>	ulan/500309512	אבאסי, ריזה, (מיזחס ל)	1200	2100	Iranian	
<input type="checkbox"/>	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 Museum (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aqa Riza (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Shaykh 'Abbasi (22) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> University of Wisconsin, School of Nursing (15) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Museum of Contemporary Art (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Tahmāsp I, Shah of Iran (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Dublin Institute of Technology. School of Architecture (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> University College, Dublin. School of Architecture (14) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	ulan/500309512	אבאסי, ריזה (אסכולה)	1200	2100	Iranian	
<input type="checkbox"/>	8. Abbasi, Riza <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Riza (44) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Riza 'Abbasi Museum (37) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Aqa Riza (24) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Imam Riza Shrine Museum (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Shaykh 'Abbasi (22) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	ulan/500116631	אבאסי, ריזה	1565	1635	Iranian	

Proceed with OpenRefine reconciliation for the remaining terms in the spreadsheet.