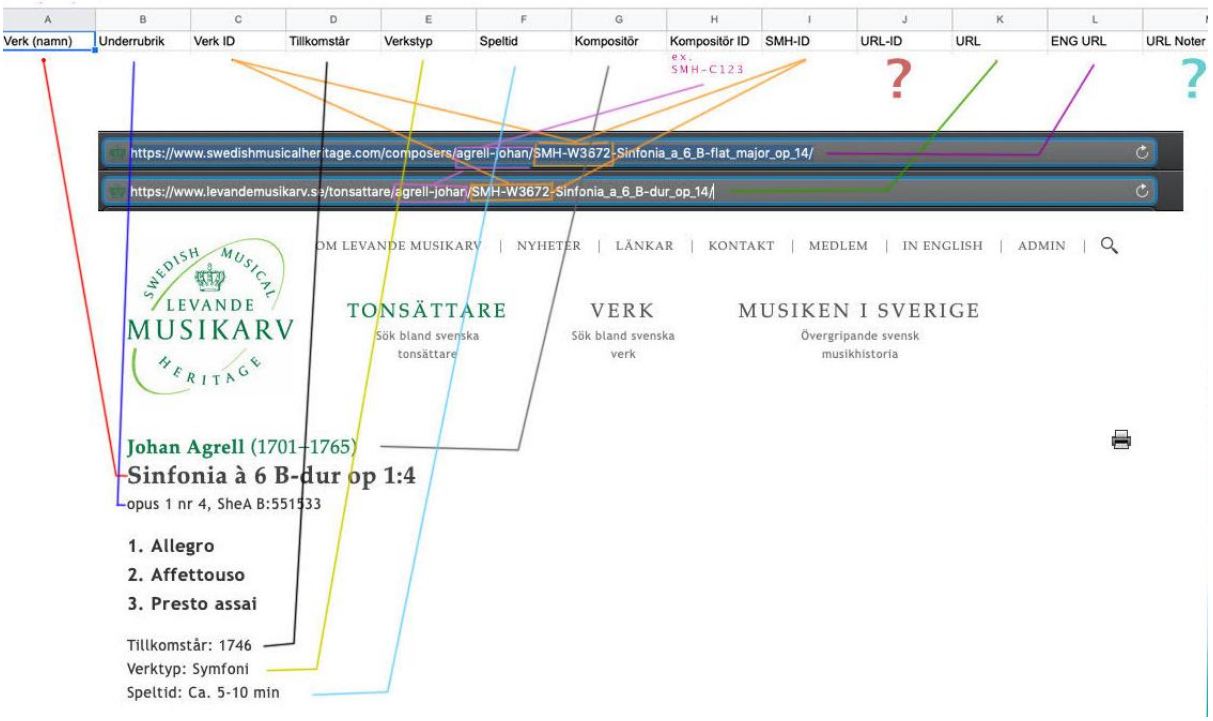
Testing OpenRefine 2021

This is some notes from testing OpenRefine in the project that we called Free music on Wikipedia, Fri musik på Wikipedia in Swedish. I had not used openrefine or similar tools before but I have some computer skills and have tested both Pattypan and Quickstatements.

The dataset I used came from Swedish Musical Heritage and contained over 6000 objects of musical works from Swedish composers. I first matched the columns to names that also could be properties in Wikidata. In Swedish I had following columns to work with:

Verk (namn)	The name of the musical work
Underrubrik	Subheading
Verk ID	Musical work ID (identifier)
Tillkomstår	Year of creation
Verkstyp	Type of work
Speltid	Playing time
Kompositör	Composer
Kompositör ID	Composer ID
URL	Swedish landing page – URL
Eng URL	English landing page – URL
URL Noter	URL to PDF for musical sheets

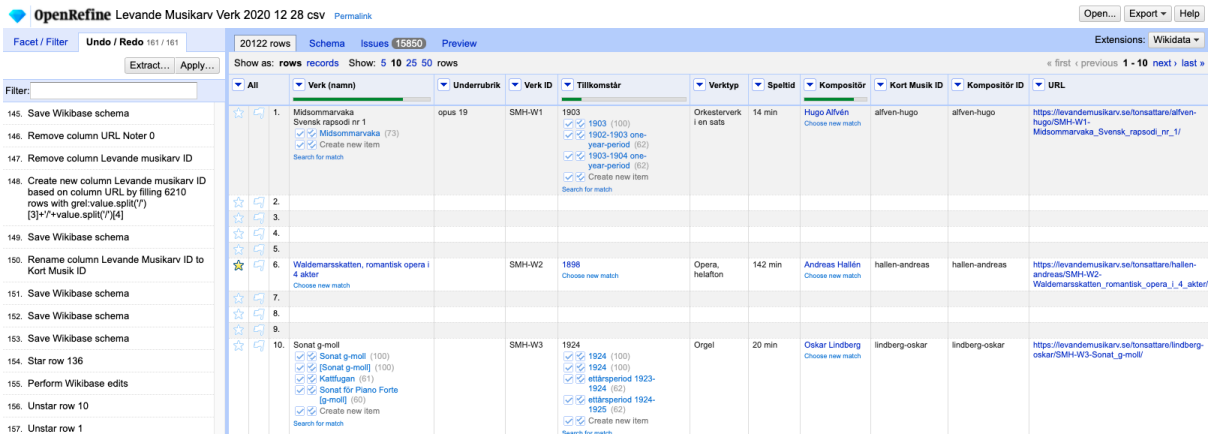
And this is how we could match fields from the web site to columns in a sheet:



I was downloading and running OpenRefine on a MacBook Pro and the program runs in the local web browser. I imported the dataset quite straight forward and started to test different possibilities in the tool.

It is possible to see all the changes you make and in this way also be able to roll back to a previous version of the process. That's a good feature to work with in the beginning. It is in the right place in this image.

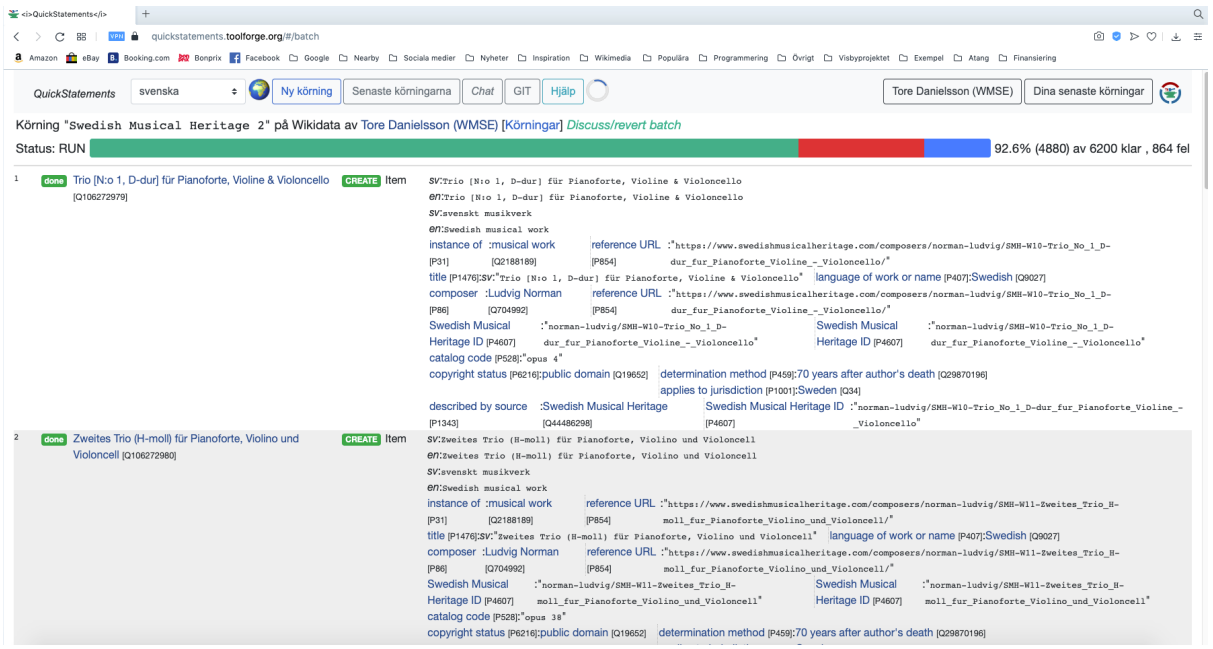
One of the tab views, rows, in OpenRefine:



The columns in OpenRefine are in the same order that it is from the uploading dataset. Now there is a possibility to start to clean and reconcile data. It can also be done against Wikidata and see if some cells in our dataset match with data in Wikidata.

There are also several ways to clean data both with built-in functions and with writing program code. In this way you can extract new data and create new columns from previously entered data.

The last step is to upload the new clean and structured data to Wikidata. Here I used the tool QuickStatements. There is also a possibility to export directly from OpenRefine but I was more familiar with using this extra step. It is an easy process to export a QuickStatement file from OpenRefine.



The results are now in Wikidata as you can see in the upload batch for musical works from Swedish Musical Heritage.

There is also a possibility to try OpenRefine with PAWS. This means that you can work with OpenRefine without downloading the program to your computer. I tested PAWS with a batch of farms at Gotland.

This is an image from preparing the tool Pattypan and the uploading of phonographs recording from the Swedish National Library. It would be nice if there was a way to reconcile the titles with OpenRefine and to try to match with objects when there is very little information. This is one problem with this dataset – there is more information in the link (example Ack, Värmland du sköna) that is not structured in the sheet.

Filename	Title	SMDBURL
ZS_V03-0012.mp3	Upplands regementes paradmarsch	http://smdb.kb.se/catalog/id/001560988
ZS_V03-0015.mp3	Romans ur Mignon	http://smdb.kb.se/catalog/id/001560991
ZS_V03-0016.mp3	Per Svinaherde	http://smdb.kb.se/catalog/id/001560992
ZS_V03-0017.mp3	Tiggarstudenten	http://smdb.kb.se/catalog/id/001560993
ZS_V03-0018.mp3	Kungl. Göta livgardes paradmarsch	http://smdb.kb.se/catalog/id/001560994
ZS_V78-0002.mp3	Capture of Santiago	http://smdb.kb.se/catalog/id/001451689
ZS_V78-0004.mp3	Rigoletto. Instrumental	http://smdb.kb.se/catalog/id/001451690
ZS_V78-0005.mp3	Hjärterosen	http://smdb.kb.se/catalog/id/001451691
ZS_V78-0008.mp3	Lektion	http://smdb.kb.se/catalog/id/001451694
ZS_V78-0011.mp3	Medley of popular airs	http://smdb.kb.se/catalog/id/001451697
ZS_V78-0012.mp3	Flora Waltz	http://smdb.kb.se/catalog/id/001451698
ZS_V78-0016.mp3	La donna e mobile	http://smdb.kb.se/catalog/id/001451702
ZS_V78-0017.mp3	Dancing in the dark	http://smdb.kb.se/catalog/id/001451703
ZS_V78-0018.mp3	Selections from "Fra Diavolo"	http://smdb.kb.se/catalog/id/001451704
ZS_V78-0019.mp3	René Waltz	http://smdb.kb.se/catalog/id/001451705
ZS_V78-0020.mp3	[At a Georgia Camp meeting]	http://smdb.kb.se/catalog/id/001451706
ZS_V78-0021.mp3	Kungssången	http://smdb.kb.se/catalog/id/001451707
ZS_V78-0022.mp3	Dagen därpå	http://smdb.kb.se/catalog/id/001451708
ZS_V78-0023.mp3	The Laughing Song	http://smdb.kb.se/catalog/id/001451709
ZS_V78-0026.mp3	Finska rytteriets marsch	http://smdb.kb.se/catalog/id/001451712
ZS_V78-0028.mp3	Hapsburg March	http://smdb.kb.se/catalog/id/001451714
ZS_V78-0029.mp3	Love's Dream After the Ball	http://smdb.kb.se/catalog/id/001451715
ZS_V78-0030.mp3	Du gamla, du friska	http://smdb.kb.se/catalog/id/001451716