Documentation GLAMhack 2024

Antoin Sevruguin

Antoin Sevruguin was an iranian-armenian photographer mainly active in the late 19. c. in Qajar Iran. The Museum Rietberg (MRZ) as well as the Ethnographic Museum at the UZH (VMZ) both have a number of albums with photos mainly ascribed to Sevruguin in their collections. In total there are around 500 photos in these albums.

Since most of the original negatives of Sevruguins burned down in the beginning of the 20. c. it is difficult to reconstruct which photos are actually by Sevruguin and which by other Iranian photographers (for example Abdullah Qajar (1849–1908/1912) or laymen photographers (for example Ernst Hoelzler (1835–1911).

Adding those photos to wikimedia commons could greatly increase the availability for researchers and would expand the already existing images by Sevruguin on wikimedia commons – uploaded by other institutions from around the world.

This challenge for the GLAMhack 24 was about learning about the best practice for GLAMs to upload photos from their collection to Wikimedia Commons, learn about Wikidata, learn about the use of OpenRefine, do a first batch upload, and get in a discussion about the best ways of describing the photos, taking into account the historic descriptions in the albums.

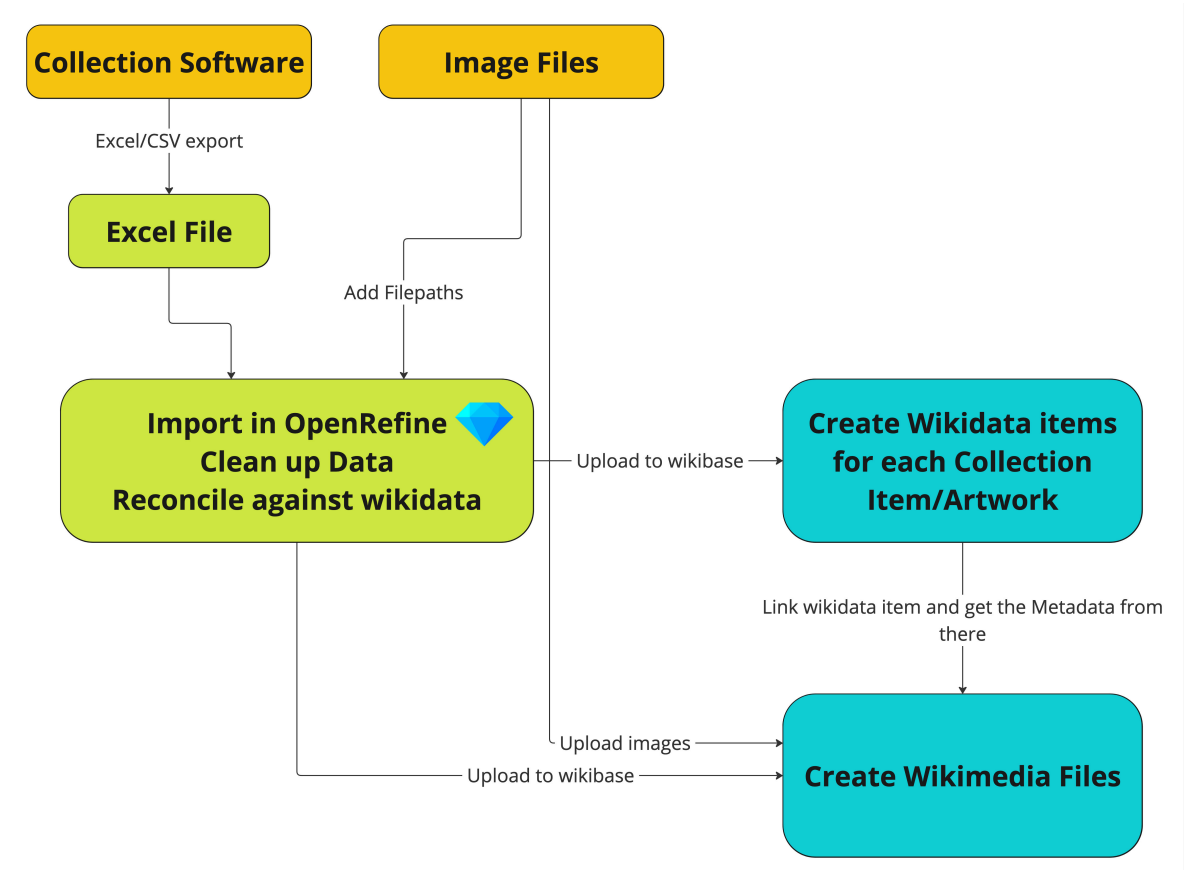
The members of our group had various backgrounds: Being active in the wikivers for decades, being curators for westasia and photos, coming from documentation or digital humanities.

Learning about Wikimedia Commons and the Wikiverse

A first step was learn about Wikimedia Commons and the Wikiverse in general. Thanks to our longtime wikimedians Diego and Martin, everyone started to understand the distinction between the different sister projects Wikidata, Wikimedia Commons, and Wikipedia and how they are connected and enriching each other.

Of course Wikimedia Commons was the focus: We learned how to upload a single (or up to 25) images directly with the Upload Wizard. This way of uploading is what is mainly used wikimedians like Diego and Martin, which also makes sense: Their usecase is to upload only a few images at a time, adding the metadata about the image and the depicted topics manually at the same time as uploading it. This way one has control over the single image, and can really enrich the wikitext and structured data of the few images being uploaded. It is also possible to copy the information of the first image uploaded and use it for the next batch of photos that have the same metadata.

From the view of GLAMs this way is maybe to slow and might double the work: usually an institution has a big batch of images (more than 25 at a time) and has already at least some metadata documented in their documentation software which in turn can be exported as an excel. It is here where the process with OpenRefine has the big impact – for only uploading a few images without already existing metadata it might be too much work.



Content consideration

Before we actually uploaded images to wikimedia commons, there was also a lively discussion about the content, content warnings as well as titles. These thoughts were specific about the images form Sevruguin, but can be expanded to other images of course as well.

Titles

Since the images of by Sevruguin don't have original titles given by Sevruguin, it was very difficult to come up with a good solutions for the titles. The photographs did have some titles given by the collectors from around 1900, and while these serve as a historical source, the museums don't want to reproduce racial stereotypes and wrong attributions, or at least mark it as a historical naming. This of course is a bigger discussion that is also discussed in different working groups from different swiss institutions. Here the ETH published guidelines on how they work on problematic titles or descriptions of objects in their collections.

In the case of the photographs in the collection of the VMZ, we decided to use the title from the album, but also clearly mark it as being a historical attribution:

> Example: The legend in the album for VMZ.523.01.037 was "Armen. [Armenische] Nonnen in Djoulfa". For the wikidata item we only used the inventory number as the label, and added a description with the given legend: "Fotografie von Antoin Sevruguin, Bildunterschrift auf Albumseite: "Armen. [Armenische] Nonnen in Djoulfa", aus der Sammlung des VMZ". In this way, it is clearly marked as being historical. The downside is that the description of the wikidat item is not shown on wikimedia commons, and only the inventory number is given as a title. To counter that we decided to write the same line as a description on wikimedia commons.

The filename on wikimedia commons was another point of discussion. Since it is very difficult to change afterwards, it needs to be as correct as possible from the start. While in the guidelines on uploading files with OpenRefine state that it might be helpful to name the files with a more descriptive name, we concluded that it could actually create more problems down the road (as well having the same issues as above). We decided to name it with the inventory number and the name of the creator. But also this turned out to be not a good decision: While working on it, we got the information that one of the photos already uploaded beforehand by the MRZ was wrongly attributed to Sevruguin and is actually shot by another photographer. While the metadata is easy to change, the title will still say "Sevruguin".

Sensitive Content

Images can also contain sensitive content, i.e. showing violence or sick and dead people. While there are guidelines on nudity on wikimedia commons, there are no guidelines on images depicting violence. After some discussion we decided to not upload one image of hanged people, even though it can be found on various websites. We don't know if this is the way to go forward, since it hinders a complete overview. Putting an editorial blurr the image could be an option and has been done on various museum collection websites, but we don't know if this could also be solution for wikimedia commons.

OpenRefine Process

Daniela from the VMZ brought with her as a testing file an excel export from the museums database. This excel was a simple export of certain fields without any processing of the data at the time of the export. This is probably the most common case. If it is possible to already prepare the export better, the whole process can get even smoother. The following description is not a full tutorial on OpenRefine, but focusses on the process for GLAMs. There are already some good tutorials online, either on youtube to follow along, or on the wikiversity.

Here is the header of the file, with a sample line:

	Inv.nr.	Beschreibung	Anmerkung zur Beschreibung	ist Teil von	Datierung	Zusatz Datierung	Fotograf:in	Material/ Technik	Masse Foto in cm_HxB	Land	Ort	Bemerkungen geografische Zuordnung	Bemerkungen zum Objekt	Beschriftungen im/auf Bild
1	VMZ.523.01.001	"Persischer Dorfbewohner"	handschriftliche Notiz auf Albumseite	Album Iran 1 (VMZ 523.01)	1880-1900	ca.	Sevruguin, Antoin	Albuminpapierabzug	20 x 13	Iran		früher: Persien		97

+ Neu

This already gives almost all of the necessary information for an upload on wikidata, but a few columns needed to be changed or added. For the purpose of learning to use OpenRefine it was perfect!

1. Spliting columns: The column "Datierung" as well as "Masse Foto in cm_HxB" needed to be split into two separate columns. This can be done by going *to Edit column -> Split into several columns...* and then specify the seperator. In this case it was "-" for the dating and " x " for the dimensions (take note of the extra spaces being part of the seperator!)
2. We could also edit the column which had the content "früher: Persien": We learned, that it was actually called "Qajar Iran" at the time of

Sevruguin. By editing one cell and then clicking on "Apply to all identical cells", this was easily done for the whole file.

3. We also changed the column "Beschreibung" to the above mentioned longer description. We used *Edit cells -> Transform* and then added the extra content before and after the existing values: "Fotografie von Antoin Sevruguin, Bildunterschrift auf Albumseite: "+value+", aus der Sammlung des VMZ".
4. Since the image files were named with the inventory number, but had underscores instead of dots, we already needed to peak a little into GREL, but just a little. We added a new column on the base of the inventory number: *Edit Column -> Add Column based on this column*. We then replaced the dots by underscores: value.replace('.','_')

These are the most basic editing tools, and a combination of them will get you very far already! GREL can get you very far, but it also gets very complicated quickly.

Reconciling

Reconciling is a way turning your text into links to wikidata. It is also needed to create new entries in wikidata as well as wikimedia commons. Most of the statements of a wikidata item are linking to an other wikidata item, while measurements or dates are not necessary to link, of course.

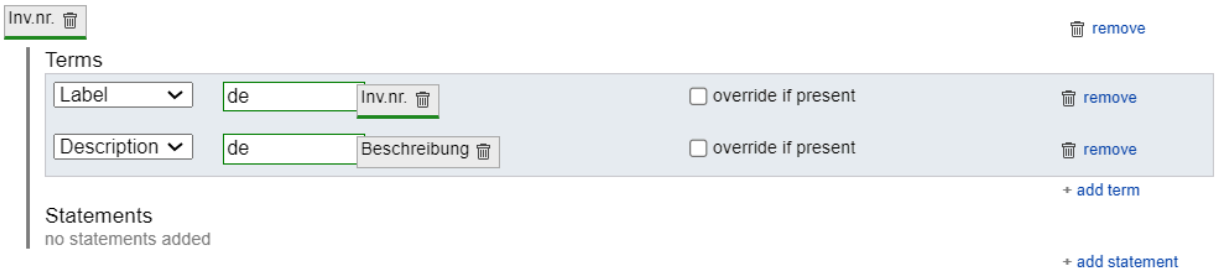
1. Reconcile all the columns you want to use for the wikidata item, like "Material/Technik", "Fotograf:in" or "Land", by going to *Reconcile -> Start reconciling*. It will then ask you to chose which wikibase you want to use, we chose wikidata to create the links to the wikidata items. Sometimes it already recognises a type and suggest to only search for items with that type (for example "human" for the column "Fotograf:in"), otherwise it just searches all of wikidata. Start reconciling!
2. Matching the links: sometimes (for example for Sevruguin) it found directly the correct wikidata item - the entry in the cell turns blue and by hovering over it you can see a preview of the wikidata item linked. Othertimes, for example with "Iran", it finds multiple items which are a close match: you have to choose the correct one. By clicking on the two check marks, you choose it for all the identical cells. There will be a dark green line indicating how many percent of that colunn has been reconciled. If none of the given option is the correct one, you can also search for the correct one.
3. To create new items for each artwork or collection item, you need to have one columns marked for that. The easiest way is to use the column for the label, in this case the "Inventarnummer". If you reconcile this column, it wont find any wikidata item to match (hopefully!). it will give you the option to create a new item based on that cell. If you go to *Reconcile -> Actions -> Create a new item for each cell* you can do that for the whole column. There will be a bright green line indicating that this is the one to create new items.

Be sure to reconcile all of the data that you need, before you go on to the next stage. If you already building you schema and then change the strucure of the data, OpenRefine has a hard time to update the schema to the new structure. Usually, you have to delete the schema and start anew.

Building the schema

By going to *Extension: Wikibase -> Edit Wikibase Schema* the view changes, now it is time to "build" the wikidata item. These schemas can be saved – which is speeding up the process tremendously for any further uploads. if you always have the same export with the same labels, you only need to build the schema once.

By clicking on "Add item" you tell OpenRefine to create a new item. You can then add the column headings to the correct spot: to actually add a new item, the reconciled column where we chose "create new item" needs to be added in the top spot. After that you can add terms and add statements. Every wikidata item needs to have a label and a description. You will need to reuse the label-column here:



By adding statements you can build your wikidata item. There a only a few rules or guidelines on what statements are need. Here is a list of statements that make sense for photos in a collection of a GLAM institute:

- instance of: should be "photograph" (Q125191), carefull to not choose "photography" (Q11633)!
- inception: the dating model on wikidata is a bit confusing. Use the overarching century/millenium and add the qualifiers "earliest date" and "latest date"
- creator
- made from material
- height
- width
- collection
- inventory number: since the inventory number isn't a general fact about the object, it needs the qualifier "collection"
- location of creation:

If you are not sure if you do it correctly, you can switch to the preview, where it shows you how it will look like.

By the way: Don't get too confused by many issue that OpenRefine throws at you! There are three kind of issues, and they vary greatly in the severity:

1. Information: This usually just informs you about something, for example, that you indeed will create items!
2. Warnings: Wikidata loves sources for each statement. In many case, this is not neccessary, or can be added at a later date. You can generally ignore these warnings.
3. Errors: Sometimes there is indeed an error – you might have forgotten to add a label to a wikidata object. These errors need to be corrected before uploading.

Upload to Wikidata

Now comes to fun and magic part, uploading the data to wikidata. Go to *Extensions: Wikibase ->Upload edits to Wikibase*. You will need to login again with your credentials. It again shows you all of the issues. Unless there is an Error showing, you can proceed. Give a short summary of you edit, a few words are enough. By clicking on Upload edits OpenRefine goes on to create wikidata items!

Wikimedia Commons

Now we need to switch our minds to wikimedia commons. Many of the processes are the same though!

Filepath

OpenRefine needs to know where your images are, to be able to upload them. If you use OpenRefine on PAWS, you have to upload them to the PAWS Server first, follow these steps: Running OpenRefine online. If you are running OpenRefine locally, you will need absolute file paths. In both cases, you want to create a column with the filepath.

We tried both versions, and both worked fine, althoug we didn't do a stresstest with thousands of hi-res files. On PAWS, you only have 3 GB of diskspace, so that might limit you eventually.

Filename

On wikimedia, the filename has to be unique for every file! For GLAMs this can easily be achieved if you put in your inventory number and you institutions name or acronym: for example "VMZ.523.01.003.jpg" in our case. We decided to also add the creator name, although as described earlier, this can lead to problems later on and we wouldn't recomend it. In cases of photoalbums, one could use the descripitive title of the album, as the Golestan Palace Library did: "Golestan Palace Album No. 208-5.jpg" We haven't yet found any good solution.

The filename is also the column that we need to reconcile, but now we need to choose wikimedia commons as the wikibase. You need to get **no** match – otherwise there is already a file on wikimedia commons with the same file name. As before with the label, we need to choose "Add new item for each cell".

Wikimedia Commons is not available out of the box as a wikibase, you need to manually add it the first time you use OpenRefine. At *Extensions: Wikibase -> Manage Wikibase instances* you can choose wikimedia commons and add it easily.

Wikitext

Usually all the information about an image on Commons is stored in wikitext. This has its own syntax and can get quite complex. Every file on wikimedia commons needs to have some wikitext, but if you do the upload this way with wikidata and OpenRefine, the wikitext can be minimalized. There are basically three things that need to be in there:

1. == {{int:filedesc}} ==

Here you need to add the template name, in the case of historical photographs, it is {{Art photo}}.

2. == {{int:license-header}} ==

Here you need to state the licence, in our case with Sevruguin, who died in 1933, we can state this like this: {{PD-Art|PD-old-auto-expired|deathyear=1933}}

3. Adding categories: all images on wikimedia commons need to be in at least one category. You need to state them in wikitext, or then afterwards with different tools. We used for example the following category: [[Category:Photographs by Antoin Sevruguin]]

We found it helpfull to look at uploaded files to see what kind of categories are already in use around that topic.

You should also add a category called [[Category:Uploaded by {your institution}]], so that you are able to easily track your uploads.

In the end it might look like this:

== {{int:filedesc}} == {{Art photo}}

== {{int:license-header}} == {{PD-Art|PD-old-auto-expired|deathyear=1933}}[[Category:Antoin Sevruguin]]

[[Category:Qajar Iran]]

[[Category:Photographs by Antoin Sevruguin]]

[[Category:Black and white photographs by Antoin Sevruguin]]

[[Category:Photographs by Antoin Sevruguin in the collection of Völkerkundemuseum der Universität Zürich]]

[[Category:Landscape photographs by Antoin Sevruguin]]

[[Category:Architecture photographs by Antoin Sevruguin]]

Building the Schema

Now it is already time to build your schema, this time you need to choose "Wikimedia Commons" as your Target Wikibase instance. The other steps are similiar to those above. This time around though, you will add a new item with column for the filename (which includes the file extension!).

As described above, we added the description as a caption.

There are some statements required as well. These statements will show up in the structured data on wikimedia commons:

1. digital representation of: Here you need to add the link to your wikidata item. Since you still working in the same sheet, and already have created the item, you can drag that column here (it was "Inventarnummer" in our case). This connects the wikidata item to the image on commons. In theory, an item can have mulitple representations. Think about a 3D-Object, but maybe even a photonegative, which you upload once as thruthful negative, and once as positive.
2. Urheberrechtsstatus
3. Lizenz

Uploading

Now it's time to upload everthing to wikimedia commons. It is the same process as above, but this time it probably will take considerably longer, since it is actually uploading images.

Connecting the File to the wikidata item

As a last step it is helpful to connect the wikidata item also with the image. This is easy, since you already have everything you need: a column with your wikidata item, and a column with your wikimedia commons file.

Just switch back to wikidata as you Target wikibase instance and add a new item based on the wikidata column. Add the statement "image" (P18) and drag the filename column there. Then upload the edits to wikidata!

And now?

Congratulations, you successfully uploaded images on wikimedia commons! These are now available for everyone, and are a valuable resource for researchers around the world.

Categorizing: Cat-a-lot

Now you can start categorize the images even more. Jonas was doing a lot of categorizing during those two days and sorted the already uploaded images by institution, as well as started to sort them by topic: landscape or portrait photography. He used a very handy tool called Cat-a-lot

Embed on Wikipedia Pages

You can now start add the uploaded photos on Wikipedia pages! Enrich Wikipedia by adding those photos any article where it might be relevant. You can do that in any language. But be aware to not add too many "unneccessary" images. If an article is very short, adding 10 images makes it unbalanced.

Benefits and Statistics

While we were busy uploading, Dominic and Francesca did a lot of Spraql queries around the topic and did some statistics. We tried to figure out how many GLAMs in Switzerland actually worked like this, when it comes to historic photographs:

```
SELECT ?collection ?collectionLabel ?item ?itemLabel ?value ?valueLabel

WHERE

{

  ?collection wdt:P31 wd:Q1497649.

  ?collection wdt:P17 wd:Q39.

  ?item wdt:P195 ?collection.

  ?item wdt:P31 ?value.

  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],mul,en". } # Helps get the label in your language, if not, then default for all languages, then en language

}
```

This gives a long list of objects in swiss collections with a wikidata item and looks at the type. As it turns out, for historical photographs only the MRZ and the VMZ have been doing it this way. This can have various reasons which would need further research: Is it even neccessary to create a wikidata item for every photograph? But oftentimes, as Diego pointed out, did the institution upload their photographs a longer time ago, when this method was not yet available at all.

But it does show the benefit of doing it that way: with sparql as a powerfull query language, it makes the images on Wikimedia Commons more easily searchable, beyond the use of categories. By simply adding "?item wdt:P18 ?pic." to the query above, it shows you the images of those items (do add "LIMIT 100" add the end of the query!).

Geographic References of swiss collection items

Francesca was also interested in geographic references for the objects in swiss collections and created from queried data following maps:

https://public.tableau.com/app/profile/francesca.altorfer/viz/DEPICTSWikidataObjectsofEuropeanGLAMInsti...

https://public.tableau.com/app/profile/francesca.altorfer/viz/DEPICTSWikidataObjectsofEuropeanGLAMInstitutionsbydepics/Sheet1

https://public.tableau.com/views/WikidataobjectsofEuropeanGLAMInstitutionsbycountryoforigin/Sheet1?:la...

https://public.tableau.com/views/WikidataobjectsofEuropeanGLAMInstitutionsbycountryoforigin/Sheet1?:language=de-DE&:sid=&:redirect=auth&:display_count=n...

They show an interesting discrepancy between where objects are coming from originally and what places they are depicting. It is clear that there are many european paintings which depict european place uploaded to wikidata. Further research and analysis could be done on this - especially considering that many historical photographs dont have an wikidata item to beginn with. How would this distribution change?

Standards for describing collections artworks

It would be further interesting to compare how different institutions describe their objects on wikidata. Dominik did start on a complex Excel analysis, which did not end fruitful so far. This remains to be done.