OpenRefine // Structured Data on Wikimedia Commons //
Design report by Lozana Rossenova

Summary of user testing session results
June 2022

From 21.06–01.07.2022, we carried out five user testing sessions with the aim of testing how users navigate the new reconciliation and file upload options for Wikimedia Commons introduced in the OpenRefine 3.7 snapshot. The participants in the sessions included four GLAM staff members, and one user with a Wikimedia community background. The participants also represented five different country/ language contexts (Belgium, Germany, India, Netherlands, and the US). Three of the participants were seasoned OpenRefine users. One had used OpenRefine before, but mostly for data cleaning, not data upload (even to Wikidata). And one user was fairly new to OpenRefine, but experienced with Wikimedia Commons. This range of skill sets helped us to identify a range of issues with the current workflows – ranging from errors and bugs, to feature improvement requests, to simply need for more specific documentation around parts of the workflows. Below is an outline of the core workflow issues and potential solutions discussed with participants during these sessions.

Table of contents:

What worked well

Installation

Most users were able to successfully install the latest 3.7 snapshot of OpenRefine.

Reconciliation

Users were able to successfully reconcile data against Wikidata and (with some guidance) to reconcile files against WM Commons.

Users appreciated being able to search and find new reconciliation services directly from the reconciliation dialog menu.

Upload

All users were able to perform media file uploads to Commons (with some guidance).
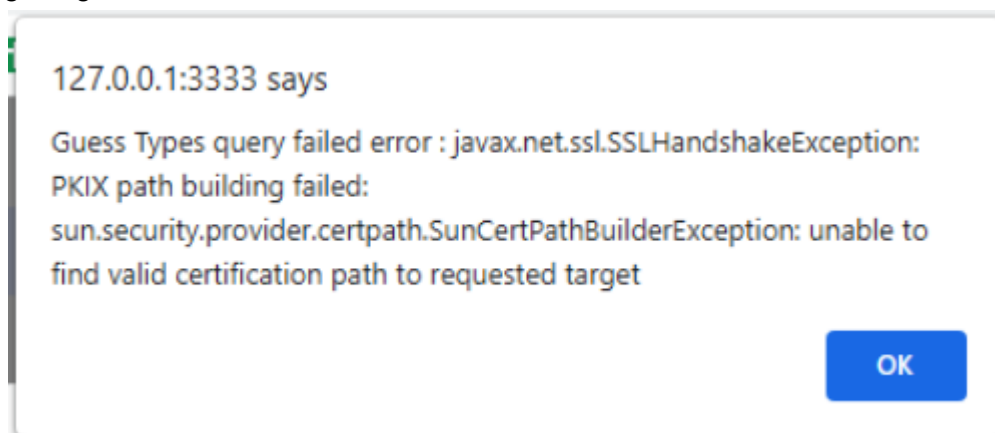

What didn't work

Installation

Issue: Admin access restrictions
One user was unable to install the 3.7 snapshot on their work computer and had to perform the tests from a personal laptop (likely reason: non-admin user account restrictions). Another user also couldn't install the snapshot on their work machine, but was able to run it from a USB drive.

Issue: Network access restrictions
One user was unable to get the reconciliation service to WM commons installed and was getting this error:

127.0.0.1:3333 says

Guess Types query failed error : javax.net.ssl.SSLHandshakeException: PKIX path building failed: sun.security.provider.certpath.SunCertPathBuilderException: unable to find valid certification path to requested target

OK

(likely reason: restrictions on the institutional network access)

This user said they usually prefer using the PAWS OpenRefine service, in order to avoid such network restrictions.

Project creation

Issue: Create project button is too small [feature]
One user reported continuously struggling with the small, off-centre position of the 'Create project' button and asked for a redesign.

Possible solution: Move the whole block of functions related to project creation to the left & increase button size (in line with existing pull request proposal #4957).

Issue: Getting file paths for files on local computer is cumbersome [feature] [documentation]
Users reported using tools like Pattypan to get a list of the file paths from their local machine folder. Ideally this functionality can be incorporated in OpenRefine directly.

Possible solution: Either provide a new workflow within OpenRefine that can generate a list of file paths based on a selection of a folder, OR include detailed documentation with instructions how to obtain such a list on Mac, Windows, etc. See #13.

Reconciliation

Issue: Records disappearing after reconciliation [bug]
On some occasions when users reconciled files against Wikimedia Commons, the records would all disappear from the main screen area, and users would have to delete several of the facets generated during the reconciliation process in order to bring back the row view of records.
See video snippet
Documented on Github: See #5157.

Issue: Differentiating columns reconciled against Wikidata vs Wikimedia Commons [feature]
Some users struggled with identifying which columns need to be reconciled to Wikidata and which to Wikimedia Commons.

After individual columns were reconciled, users struggled remembering which column was reconciled to what service and made mistakes when dragging and dropping columns during schema building.

Possible solution: Adding logo icons for the different services next to the green bar indicating reconciliation completion status (see #4824).

Issue: Reconciliation before new item creation [feature] [documentation]
Not obvious that it's necessary to first reconcile media files against Commons before uploading them.

Possible solution: Make the action to create new items from each cell (from the Reconcile menu) possible to apply directly before first running a reconciliation process (see [#4985](#)). Provide additional instructions on this in the documentation.

Data extension

Issue: Errors in preview during data extension [bug]
In some cases, during data extension users were getting error messages inside the preview of the dialog window. If then they hit Reset and switched to another property, and if that property worked, they could continue with the data extension process. However, if they clicked OK despite the error message, then no new columns were added, so the error wasn't only a preview issue, but a genuine issue with accessing and requesting that data.
See video snippet 1
See video snippet 2
Documentation on Phabricator: See [T296298](#).

Issue: Internal server error in hover state [bug]
On some occasions, when data was extended from Commons, the reconciled items in the newly created columns returned 'internal server error' message in the hover state preview. However, upon clicking on the item users could reach the correct location in Wikidata.
See video snippet 1
See video snippet 2

Possible solution: See [#75](#). See also [T314743](#).

Issue: Differentiating columns with data extended from Wikidata vs Wikimedia Commons [feature]
When users extended data from the different Wikibase instances, similar to reconciliation, there was no clarity which columns were sourced from which service, leading to errors during schema building.

Possible solution: Adding a prefix to the columns title, e.g. *wd:* for Wikidata and *wmc:* for Wikimedia Commons (see [#5130](#)).

Issue: Caption is not easy to find as an option in the properties list [feature] [documentation]
Extracting existing captions for media files can be useful when users want to add more structured data to existing media, or if they want to add captions in other languages, etc. At the moment, they have to enter the caption as a special property in the "Add property" field, but without instruction or documentation that's not obvious to new users.

Possible solution: Add 'Caption' as a default option in the 'Suggested properties' sidebar with option for users to configure the language settings. See [T314745](#).

Issue: Selecting more than one property for data extension – not obvious [feature]
Users indicated that it is not obvious that more than one properties can be selected at the same time.

Possible solution: Add tick boxes in the 'Suggested properties' sidebar. See #5159.

Issue: Search box to add more properties – not obvious [feature]
One user pointed out that the 'Add properties' field is not obviously a search box for discovering more properties that can be linked to the reconciled item.

Possible solution: UI improvements are needed in the 'Add column from reconciled values' dialog window to direct users' attention to the option to add more properties from the reconciliation service. See #5159.

Schema building

Issue: Confusion where to start from [feature]
Since the schema building screen is empty to begin with and the '+ add media' / '+ add statement' calls to action are rather small and off centre, it is easy to miss what to do next. When asking users only to add statements to existing media files, they were also unsure what CTA to click on, because '+ add media' suggests only uploading files, rather than adding data to existing files.

Possible solution: Add more central calls to action. Rethink how to indicate working with media files. Schema templates can also help with this issue. See #5050 and #5103.

Issue: Confusion what entity to add first [feature]
Since the '+ add media' button conforms to the behaviours typical when uploading items to Wikidata, it was not clear to users that they need to drop the column with the file name at the top of the schema, some thought that the Wikidata item needs to go there.

Possible solution: Change the UX pattern when starting from '+ add media' to be more clear users need to add file name as lead entity. Remove duplication of file name afterwards. See #5050 and #47.

Issue: Wikitext building can be improved [feature] [documentation]
Users liked the option to add categories separately from Wikitext, but that raised the issue as to whether we can provide further help in building Wikitext. Based on discussions over the wireframes, for the moment, it seems better to not provide separate elements that are normally part of Wikitext (e.g. categories) as a standalone field, while also providing a field for manually entered Wikitext.

Possible solution: One idea that came out of the sessions was to provide a complete set of "building blocks" for Wikitext as dropdown option menus - i.e. Wikitext template (such as Information, Artwork, etc), licence, categories. In addition, we can retain the option for users to add manually entered Wikitext. See #48.

Issue: Wikitext (and possibly Captions, too) should be possible to override [feature]
A lot of users have old Wikitext (pre-dating SDC) that can be improved with the new simpler templates and SDC.

Possible solution: Users need the option to override previous Wikitext via a simple tick box element on the interface (similar to how Labels and Descriptions can be overridden during Wikidata schema building). See #4989.

Issue: Not obvious what elements are mandatory to fill out when adding new media files [feature] [documentation]
We can improve communicating to users what elements are mandatory when adding files vs when adding SDC to existing files.

Possible solution: Consider using * elements & helptext; OR a new workflow that splits the UI into different views depending if the user chooses to "Add new items" or "Add data to existing items". Alternatively, emphasise the role of schema templates. See #5050 and #5103.

Issue: Users found it generally unclear how to do the data modelling for media files [feature] [documentation]
Questions included: what properties to use to add basic data about the media files, how much data from Wikidata is worth repeating, and where to find more guidance.

Possible solution: Use schema templates (all users welcomed this idea as per wireframes design). See #5103. Provide clear documentation, helptext and more onboarding guidance.

Upload

Issue: Missing file name extension [documentation]
Uploading files without the file name extension did not work, but there wasn't a meaningful error message to users that their file name column must retain the .jpg, .png, or another file name extension.

Possible solution: Explain this issue in the official documentation. Add an appropriate error message in the Issues tab.

Issue: Looking for the upload button [feature]
Users not too familiar with the Wikibase extension kept looking for the upload button elsewhere. Looking for the upload button under the extension menu was not obvious to them and also the extension itself was still saying 'Wikidata' (not ' WM Commons') - adding to the confusion.

Possible solution: Changing the extension name from Wikidata to Wikimedia Commons might help (see #4525), but we can also think about making that dropdown menu more prominent, or even adding a shortcut to the upload button directly in e.g. Preview tab.

Issue: Users are not aware how to undo edits after upload [feature] [documentation]
Users were not aware of the EditGroups tool. Additionally the editing can only undo SDC edits, not uploading new files, which should also work.

Possible solution: Making it possible to undo file uploads is important. We could also highlight the EditGroups tool better and make a link to it possible directly via the Extensions menu, or via a new tab, as per suggestion in #5032.

Conclusion

The main take-aways from this round of user testing sessions are:
- Given the challenges of running a local instance in an institutional setting, it is important to communicate with (and support) the maintainers of the **PAWS instance** so that they are able to **keep up with new releases** of OpenRefine.
- **Improving the schema building experience** and providing templates for data modelling guidance should be a priority for the remaining work on the SDC extension.
- Since the workflows introduced in the SDC extension are new to most users (even though the workflows are built upon previous UX patterns from the Wikidata extension), there is a strong need for **detailed documentation, easily accessible guidance and onboarding materials** for both existing and new users.