

Wikimedia-OpenRefine training and sustainability project, 2023-24

Until end October 2022, with support from [a Wikimedia grant](#), the popular open source data wrangling tool [OpenRefine](#) (which is already [widely used in the Wikidata community](#)) has been extended with [functionalities for batch editing and uploading media files to Wikimedia Commons](#), with a focus on structured data. OpenRefine is now a usable Wikimedia Commons tool which serves more advanced users and large-scale content partners, and currently provides the Wikimedia movement with an alternative for the simpler Pattypan (which does not support structured data yet) and (the now discontinued) GLAMwiki Toolset.

However, OpenRefine is new to many Wikimedians and partner organizations. Since deployment of OpenRefine's Wikimedia Commons features, dozens of individual GLAMs and Wikimedians have requested some level of support and training in using the tool, and are providing feedback for improving both the Wikimedia Commons features and OpenRefine usability in general. With this grant, we invest in:

1. Training and empowering a group of **GLAM-Wikimedia-OpenRefine liaisons and trainers** (roughly 4 to 7 people) who will channel and organize such requests from now on.
2. This group of liaisons will produce and improve **documentation and training materials**, do **outreach** in their own communities, and will teach a **"train the trainer" program** for OpenRefine to GLAM-wiki practitioners.
3. The Wikimedia Foundation's Culture and Heritage team, and other Wikimedia organizations with an interest in OpenRefine, will engage in conversations around **maintenance and upkeep** of OpenRefine's Wikimedia features, so that they will remain available and up for Wikimedians and their partner organizations for the next five years, and will also engage in more strategic long-term partnership conversations with the OpenRefine team.
4. This grant also provides the OpenRefine team with the financial resources to follow up on first-level bug fixes after the software release in October 2022. Prioritization of these software improvements will be done in close collaboration with the Wikimedia community liaisons, and after consulting and observing end users. The OpenRefine team will continue to collect higher-level pain points, larger feature requests, and strategic priorities.

Budget: 50,000 USD, part of the Wikimedia Foundation's 2022-23 Annual Plan.

Stakeholders and partners (potentially to be extended):

- OpenRefine
- The [Wikimedia Foundation's Culture and Heritage team](#) (contact: Fiona Romeo)
- Wikimedia Sverige / Content Partnerships Hub (John Andersson / André Costa)

- And slightly more removed, but relevant:
 - Wikimedia Deutschland / Wikidata and Wikibase teams
 - Wikibase Stakeholder Group (which also has an interest in the media file upload and editing features)
 - TIB Hannover
 - Others tbd

To-dos and deliverables

1. Recruitment and establishment of a group of GLAM-Wiki-OpenRefine

liaisons/trainers. At the end of this grant period, the Wikimedia movement and its content partners will be helped and trained by a clearly defined, and publicly visible and approachable, group of 4 to 7 individuals.

- This group will be recruited through an open call and 1:1 outreach, aiming at diverse representations of various backgrounds and skill sets: a mix of Wikimedians and Wikimedia-savvy GLAM staff, from varying geographical / regional and linguistic backgrounds, ideally representing major subcommunities of the GLAM-Wiki space.
- Three liaisons will be funded for a flat fee of USD 5,000 via this grant. Up to four additional interesting candidates can participate with support from a GLAM, a Wikimedia affiliate or other related capacity-building organization.
- During this grant period, these liaisons will be onboarded and will work (with financial compensation) on the tasks mentioned below (points 2-3): improvements to documentation and training materials, translation of these materials, holding a train the trainer course.
- At the end of the training period, the liaisons will be formally and publicly listed as certified trainers and GLAM-Wikimedia-OpenRefine contacts, similar to https://meta.wikimedia.org/wiki/Education/Reading_Wikipedia_in_the_Classroom/Meet_our_Certified_Trainers and https://meta.wikimedia.org/wiki/Wikibase/Consultants_and_Support_Providers.
- For at least the upcoming two years after ending this grant, this group of individuals will:
 - i. be sufficiently confident in their usage of OpenRefine to be able to work with user requests autonomously;
 - ii. keep in touch with their peer group of fellow OpenRefine liaisons through an established set of communications channels, and train/onboard other peers;
 - iii. continue to channel and communicate user needs and bug reports to the OpenRefine forum and issue tracker

2. Training materials and documentation. Creation of generic and re-usable training materials + cleanup of legacy documentation and training materials

- Basic courses in English on the <https://learn.wiki> platform, which can be either used for scheduled courses (with trainer and assignments) or can be done

autonomously at the learner's own pace. Topics will include at least (highest priority first):

- i. Commons file uploading with SDC with OpenRefine
- ii. SDC editing with OpenRefine
- iii. Teaching OpenRefine to Wikimedians and GLAM staff (i.e. a train the trainer course/program)

If time allows and if the liaisons group collectively plans this, to be extended with:

- iv. Wikidata editing with OpenRefine
 - v. Basics of OpenRefine data wrangling and cleaning for Wikimedians; this can build upon, the (CC-BY licensed) [Library Carpentry course for OpenRefine](#)
 - Translation of these courses into at least 3 major languages that are not English. Preferably Spanish + two other languages TBD in consultation with the Wikimedia community, the liaisons' group (see above), and Wikimedia chapters/affiliates.
 - As time allows: cleanup of old, outdated documentation training materials on Wikimedia wikis and updating OpenRefine's current info pages on Commons and Wikidata; setting them up for translation.
3. **Scheduled OpenRefine courses for GLAM staff and Wikimedians who engage in larger-scale Wikimedia content partnerships.** At least 3 formal, scheduled, guided trainings by members of the liaison group established in point 1. (using the above online courses at <https://learn.wiki>, with graded homework and personal mentoring)
- The goal is to train at least 50 Wikimedians using at least one of the developed training modules. The 'train the trainer' course specifically should result in at least 10 certified Wikimedia-OpenRefine trainers of diverse international and linguistic backgrounds, at least representing Spanish, Arabic, and preferably several East Asian, Indic and African languages.
4. **Software development.** General maintenance, and bug fixes, to Wikimedia Commons functionalities in OpenRefine
5. **Strategic partnership discussions on sustainability of OpenRefine's Wikimedia features.** Continued conversations to address reliable support for, and longer-term sustainability of OpenRefine's Wikimedia / Wikibase features.
- Continuation of a monthly (or other frequency) meeting between OpenRefine leadership and Wikimedia stakeholders, potentially in a new format
 - This step will ideally result in the negotiation and signing of MoUs (Memoranda of Understanding) between OpenRefine leadership and various Wikimedia stakeholder organizations.