

Interim Report - 2019-207403, EOSS-0000000332 OpenRefine for Everyone

[Grant Overview](#)

[List of key personnel directly funded by this EOSS award \(including the lead\):](#)

[Financial Overview](#)

[Budget narrative](#)

[Progress Overview](#)

[Brief summary of outcomes and results for this reporting period:](#)

[Brief summary of progress towards learning goals for this reporting period:](#)

[Impact:](#)

[Challenges:](#)

[Key Outputs and Project Recognition](#)

[Recognition and milestones for work or personnel associated with this grant:](#)

Grant Overview

Grant title: OpenRefine for Everyone	
Grant number: 2019-207403, EOSS-0000000332	Grant amount: 240k total, split over 120k USD per year for two years
Project(s): OpenRefine	
Grant term covered by this report: September 1, 2021 through August 31, 2023	Report term frequency: Annual
Lead applicant: Sandra Fauconnier (replacing Antonin Delpeuch)	Grant recipient organization: Code for Science & Society

List of key personnel directly funded by this EOSS award (including the lead):

First name	Last name	Open source project	Organization	Email Address	Code contributor handle (e.g. @dartar)	ORCID
Sandra	Fauconnier	OpenRefine	Code for Science & Society	sandra@openrefine.org	@trnstlntk	0000-0002-3303-9644

Financial Overview

Provide original estimated and actual spending for each category for the reporting period. You can also submit this information as a separate spreadsheet.

Category	Original estimate (USD)	Actual (USD)	Notes
Personnel - salary	45,000	0	(Note 1)
Personnel - fringe	55,000	13,000	(Note 2)
Equipment (over \$5,000)	0	0	
Travel and convenings	1,300	392	(Note 3)
Other	18,700	18,945	(Note 4)
Total	120,000	32,337	

Budget narrative

Briefly explain any changes to the original budget, any challenges impacting spending; unspent budget; and plans for use of remaining funds.

Until now, we have underspent for this grant, mainly due to the longer time it has taken to hire and onboard a Project Director in this project. Now that the Project Director is hired, we expect to fully catch up on spending in the second year of this grant.

- Note 1: As the result of an [open application procedure](#), [Sandra Fauconnier has been hired as Project Director of OpenRefine](#). She started working for the project on February 1, 2022. Until the end of October 2022, she is available to dedicate 10 hours per week to this position; after that, her employment will increase to 20 hours per week. For administrative purposes, Sandra's first months of employment have been fully covered

by another grant (Structured Data on Wikimedia Commons support in OpenRefine, funded by a Project Grant from the Wikimedia Foundation), which explains why this budget line mentions zero spending at this moment. The 90K budgeted in the context of this (EOSS D&I) grant will be fully spent on Sandra's salary between November 2022 and August 31, 2023.

- Note 2: In May-August 2022, [two Outreachy interns work on diversity-related projects for OpenRefine](#); this budget line covers their internships.
- Note 3: The OpenRefine team held an in-person meeting in April 2022 with [Lozana Rossenova](#), UX designer and member of OpenRefine's steering committee. This meeting covered plans for general improvements of OpenRefine's user interface, which is a popular request by OpenRefine users and an improvement which will make the tool accessible to more diverse users.
- Note 4: This budget line covers:
 - Fiscal sponsorship fee by Code for Science & Society - administrative and strategic support (USD 18,000)
 - Code of Conduct training for three OpenRefine community members (USD 945)

Progress Overview

Description of original deliverables: (for your reference, based on the original proposal)

1. **Milestone: Hire and onboard a project director from the OpenRefine user community by October 2021.** Deliverables for this grant focus on diversifying the new contributor pipeline. This role would also be funded by the EOSS 4 (if successful) and deliverables for that portion of the project directors role focus on the roadmap, governance, and partnerships.
 - a. **Deliverable: Regular participation in programs that develop OpenRefine's contributor pipeline.** Project director and technical lead (hired thanks to the EOSS-4) will work together to manage OpenRefine's participation in at least three rounds of internship programs (GSOC, Outreachy, or others) with a goal of cultivating experienced interns who return as volunteer contributors or for contract work. This will support junior developers from diverse backgrounds by mentoring them through the process of making a significant contribution to an open source project. We want to encourage as many of these interns to keep contributing to OpenRefine afterwards but that is not in itself our primary goal, since our alumni can be active in other projects, so their training is a service to the community. **Timeline:** One to two rounds of internships per year for the duration of the grant, each lasting for a few months.
 - b. **Deliverable: Increased diversity along geographic, racial, and ethnic axes in governance.** The new project director will develop OpenRefine's existing

governing bodies over the next 24 months, yielding increased diversity in multiple axes on OpenRefine governing bodies. Part of this deliverable include updating and maintaining our governance, code of conduct and contributing documents to create a safe and welcoming space for contributors. This deliverable links to the EOSS 4 proposed deliverable of increasing stakeholder representation in governance. The project director will be supported by the existing committees to increase diversity in multiple axes on OpenRefine governing bodies.

2. **Milestone: Have tackled at least three cultural biases in the tool.** These will be tackled either during the internships or by independent efforts. Timeline: flexible, as we want to leave our interns a choice in the projects they tackle.
 - a. By December 2021: Assessment of cultural biases in OpenRefine with the community
 - b. By September 2023 correction of at least three biases identified previously.

Brief summary of outcomes and results for this reporting period:

Provide a short description, in a bulleted list, of milestones/deliverables achieved for this reporting period.

- **Milestone: Hire and onboard a project director from the OpenRefine user community by October 2021** - This milestone was achieved, but later than planned: Sandra Fauconnier started working as OpenRefine's project director on February 1, 2022. The hiring period took longer than expected due to a Code of Conduct incident (which was resolved).
 - **Deliverable: Regular participation in programs that develop OpenRefine's contributor pipeline.** This deliverable is on track. OpenRefine has participated in one round of Outreachy internships (May-August 2022 cohort), welcoming developers from Kenya and Cameroon to work on diversity-related projects.
 - **Deliverable: Increased diversity along geographic, racial, and ethnic axes in governance.** First groundwork is laid for this deliverable. Through [OpenRefine's two-yearly survey](#), a cohort of people has expressed interest in participating in OpenRefine's governance committees and working groups; some of these potential new members can bring expertise in right-to-left and Semitic languages (Hebrew and Arabic). Proactive outreach and recruitment is needed to bring in additional geographic, racial and ethnic diversity, which will be the focus of OpenRefine's project director in the next year. Due to the project director's later hiring and onboarding, and her limited employment (10 hours per week), this deliverable is progressing more slowly than planned but it is top priority for the upcoming time.
 - Three OpenRefine community members participated in a [Code of Conduct Enforcement training](#) in late 2021.

- **Milestone: Have tackled at least three cultural biases in the tool.**
 - A community assessment of existing cultural biases in OpenRefine has not fully taken place yet, partly due to the project director's later hiring and onboarding and her limited employment in this first period. For the first time, OpenRefine's user survey has included several questions in this direction, but this has not yet produced conclusive results. In the second year of this grant, this topic will be prioritized. We are starting this process by asking OpenRefine's community to propose and comment upon software development prioritization in July 2022.
 - One of OpenRefine's May-August 2022 Outreachy interns works on [back-end localization of OpenRefine's code](#), which will make it possible for end users / translators to translate all of OpenRefine's relevant interface text strings. This project addresses one out of the three cultural biases we'd like to tackle as part of this grant.
 - One of the hurdles to more diverse adoption of OpenRefine is its user interface, which - as we repeatedly learn from feedback in our user surveys and through teaching OpenRefine in workshop settings - is perceived as complex, and daunting to newcomers. While OpenRefine's UI and UX are quite consistent and unique, the UI is also becoming increasingly 'dated'; many of our users are asking for a more contemporary design. As part of another grant ([Wikimedia Commons support in OpenRefine](#), funded by the Wikimedia Foundation) we have initiated a [general UI/UX focused project](#), aiming to make OpenRefine's design-related issues more visible, and to help attract new contributors with a UX/UI/interaction design background (rather than a purely technical background). We dedicated one day of a team offsite to further discuss and brainstorm this approach, including the creation of general UI/UX design toolkits, and an interactive onboarding wizard integrated in the tool. We have also established contacts with the [Open Source Design](#) community, an active community of FLOSS-enthusiastic UI/UX designers, as a step to further grow our community of contributors in this area.

Description of original learning goals: (for your reference, based on the original proposal)

We want the project director to be genuinely able to emancipate from the current advisory committee which hires them in the first place. This is necessary for the project director to be in a real leadership position. This means proactively countering the initial subordination relation which might be strengthened by the fact that we want to hire someone from a less technical background than the current advisory committee. With the support of Code for Science and Society, we will review this situation 3, 6, 12 and 18 months after the hire, to assess

- which decision-making responsibilities they have been able to take over and what prevents them from becoming more independent.
- progress made against the Diversity and Inclusion assessment plan defined at the beginning of the contract.

Internships will be reviewed following the processes laid out by the internship programmes themselves, which provide sufficient oversight on the progress of the intern and the working relationship with their mentor. In addition to that, we want to review our own processes around the internship programmes, by having a conversation between interns, mentors and administrators at the end of each round. In 2020, such a conversation helped us identify that it is better to have a single mentor per intern, to avoid dilution of responsibilities.

For the cultural biases in the tool, we will conduct a survey advertised to our user community to let them surface the cultural barriers in the tool that we should tackle first. For instance, the dialogue with our Japanese community made it clear that biases in the application logic (clustering, number parsing) are more problematic than lack of translation support in some parts of the user interface. We believe an online survey could let us surface similar feedback from other backgrounds.

Brief summary of progress towards learning goals for this reporting period:

- Due to the late hire of OpenRefine's project director, an initial formal assessment of their decision-making responsibilities has not taken place yet, but it is on the agenda for the upcoming period, as the project director is now onboarded.
- The first round of Outreachy internships funded by this Diversity and Inclusion grant is currently in progress. We are applying lessons learned from previous editions and indeed plan to review and evaluate the current round after it has completed (end August 2022), so that we can again improve in next internship programs.
- We have not yet conducted a dedicated survey for our users focused on cultural barriers and biases. However, our [ongoing \(two-yearly\) general user survey](#) contained several new questions asking about language-specific usage of OpenRefine and this will serve as a starting point for further surveying and research.

Impact:

What was the expected impact of activities undertaken during this reporting period related to the participation, retention and leadership progression of contributors from underrepresented groups? What was the actual impact?

- We are roughly on track with our planned Outreachy internships, which introduce code contributors from traditionally underrepresented backgrounds to our project. We are currently working with interns from Cameroon and Kenya. Furthermore, we have hired a recent Outreachy alumnus as a contractor on our grant from the Wikimedia Foundation, and are making space in that contract for learning and training opportunities.
- We have not yet succeeded in renewing our governance bodies, mostly due to time constraints. In the upcoming year, we plan to renew OpenRefine's steering committee

and create various smaller working groups (focused on OpenRefine's roadmap or Code of Conduct). OpenRefine's two-yearly user survey has resulted in a first selection of candidates for these committees.

Challenges:

Describe any challenges (scientific and/or operational) that you have encountered during the reporting period, as well as any mitigation efforts. Have any of these challenges resulted in changes to the scope or project plan?

- As mentioned above, hiring OpenRefine's project director was delayed by four months, because of a Code of Conduct issue. While unpleasant, the issue activated OpenRefine's Code of Conduct committee, helped us gain experience in this area and address a long-standing problem in our community. This issue was ultimately addressed in a way that has a positive effect on the project at large.
- Because of other engagements, the project director currently works 10 hours/week (moving to 20 hours per week in November 2022). This lower number of hours causes this project to develop more slowly than originally planned.

Key Outputs and Project Recognition

Provide a list of outputs (preprints, journal or conference publications, methods/protocols, datasets, documentation, and/or code repositories) generated as a result of this grant. You can also include other notable outputs such as events organized, accepted/invited talks, community calls, and publications in progress.

Identifier (URL or DOI)	Output type Please use an option from the following list: Code Preprint Peer-reviewed publication Method/Protocol Dataset Documentation Event or talk Community call Other	Title	Notes (optional)
https://groups.google.com/g/openrefine-dev/c/pLG_OOxMpXQ/m/2bOgmknDAgAJ	Community call	Monthly OpenRefine development community call	This monthly call was initiated in January 2022.
https://github.com/OpenRefine/SparglExtension	Code	SPARQL Extension for OpenRefine	New repository created as part of Outreachy internship
https://sosdesign.sustainoss.org/quests/rossenova	Other (Podcast)	Sustainable Open Source Design podcast: Lozana Rossenova on the intersection of design	

		research, open source communities and GLAMs	
https://openrefine.org/blog/2022/06/28/2022-survey-results.html	Other (Blog post)	Summary of OpenRefine's 2022 user survey	
https://dataharvest22.sched.com/event/10QPw/clean-data-like-a-pro-with-openrefine	Talk (Workshop)	Clean data like a pro with OpenRefine	Participation in Dataharvest conference + OpenRefine workshop for data journalists