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

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 September 30, 2023	**Report term frequency:** Annual
Lead applicant: Jan Ainali (replacing Sandra Fauconnier)	**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
Martin	Magdinier	OpenRefine	Code for Science & Society	martin@openrefine.org	@magdmartin	0009-0005-1057-220X

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	$ 90,000	$ 41,958	(Note 1)
Personnel - fringe	$ 110,000	$ 30,000	(Note 2)
Equipment (over $5,000)	$ 0	0	
Travel and convenings	$ 2,600	$ 673	(Note 3)
Other	$ 37,400	$ 36,945	(Note 4)
Total	$ 240,000	$ 109,700	

Budget narrative

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

- Note 1: In February 2022, we hired Sandra as the project director to lead the work for the EOSS Diversity grant and other projects. Sandra's position was initially funded through another grant from February to July 2022. It was only in July 2022 that we began utilizing funds from the EOSS-Diversity grant, resulting in a delay of approximately one year from the original schedule. Unfortunately, Sandra resigned from her director position on

December 1st, 2022, leaving the project needing a new project manager. We hired Martin Magdinier as Project Manager in June 2023 (please refer to the attached [announcement](#)). Martin was formerly on our Advisory Committee board. His arrival has brought stability to the project, and we are confident we are on track to achieve our objectives.

- Note 2:
    - In May-August 2022, [two Outreachy interns work on diversity-related projects for OpenRefine](#); this budget line covers their internships.
    - In May-August 2023, [two Outreachy interns](#) worked on improving the designer onboarding experience and improving our reconciliation feature.

- 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.
    - Martin Magdinier met with Antonin Delpeuch to discuss OpenRefine in person in July 2023.

- Note 4: This budget line covers:
    - Fiscal sponsorship fee by Code for Science & Society - administrative and strategic support (USD 18,000 per year)
    - 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 more complicated to achieve than initially planned due to the fact that we had to hire a new project director in June 2023.
    - **Deliverable: Regular participation in programs that develop OpenRefine's contributor pipeline.** The project participated in two Outreachy programs in 2022 and 2023, welcoming a total of four students from Kenya, Cameroon, India, and Nigeria
    - **Deliverable: Increased diversity along geographic, racial, and ethnic axes in governance.** During the period from August 2022 to August 2023, the community engaged in discussions regarding different governance models aimed at bringing new contributors onboard and promoting their development as leaders. Forum discussions [1], [2] and [3] were held during this time, and Martin Magdinier had separate conversations with over 20 users and contributors. As a

result of these discussions, the decision was made to implement a bottom-up model.This model will allow the identification of key contributors within the community and enable them to grow as leaders. Two new deliverables were identified to achieve this goal: defining OpenRefine's Mission, Vision, and Values, and creating a Contributor's Handbook to support the [Mountain of Engagement framework as defined in Mozilla Open Leadership training](#) or [Contributor Pathways (the Turing Way).](#)

- **Milestone: Have tackled at least three cultural biases in the tool**.
  - As of 2023, we have taken significant steps to address the design limitations of OpenRefine that were identified in our previous interim report. We accomplished this by establishing a formal design practice. In June of that year, we began our efforts by recruiting an outreach intern, followed by hiring a part-time designer who is being funded through the EOSS-5 budget to further advance this work.
  - Based on the feedback received from our [2022 user survey](#), we made several improvements to our platform.
    - We moved from a mailing list to discourse to increase engagement from non-developers.
    - We implemented a drag and drop column feature to improve the user experience.
    - We had a conversation with Johan Huijkman, an accessibility engineer at Q42, who helped us identify key accessibility points.
    - A volunteer contributor worked on addressing accessibility by creating a keyboard-accessible GUI and enabling the option to use OpenRefine through keyboard only.

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 toward learning goals for this reporting period:

- We faced challenges in hiring and keeping a project director due to the difficulty of navigating OpenRefine, which lacks a clear direction for newcomers. Based on this, we have decided to prioritize the identification of OpenRefine's mission, vision, and values in 2024. In addition, creating a contributor handbook will help uncover the hidden practices and tribal knowledge within the community.
- After two successful outreachy rounds, we are enhancing our onboarding process for new contributors, extending it to cover design, documentation, training, and translation.
- The change in project manager caused a delay in the user survey about cultural bias. We will include questions about cultural bias in our 2024 user survey.

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 I mentioned earlier, the hiring and retention of our project director posed a challenge that resulted in a delay in our project. Additionally, we faced difficulties in identifying the appropriate model for developing an OpenRefine contributor pipeline. However, we are optimistic that our new approach, based on the success of other communities, will be effective in addressing these challenges.

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://www.youtube.com/watch?v=e9fB8NQV6Nc	Webinar	Data Umbrella - [90] Intro to OpenRefine for Data Cleaning and Reconciliation (Martin Magdinier)	
https://forum.openrefine.org/t/openrefine-contributors-meetup/836/17	Community Call	Monthly OpenRefine development community call	This monthly call was initiated in January 2022 and kept going since
https://openrefine.org/blog/2022/09/30/windows-macos-packaging	Code	Bounty Program to improve our packaging for MacOS and Windows #5205 #3224	
https://forum.openrefine.org/ and https://openrefine.org/blog/2022/09/20/discourse	Other	Migrated our mailing list to discourse	