[CZI Diversity grant budget](#) - $50,000 budgeted for technical contracting
- To spend before end of grant period, i.e. August 31, 2023 (unless we ask for an extension)
- Sandra: Let's aim to spend around 45K of this budget; we will likely spend a bit of this budget on additional travel costs and various other expenses too

📄 20210330 Full application - EOSS Diversity and Inclusion Grants
📄 Interim report CZI EOSS Diversity grant - OpenRefine - deadline August 1

Excerpts from work plan:

**(Work plan includes:) Tackling cultural biases in the tool.** With fewer cultural biases, we expect to see more users and contributors with diverse backgrounds coming to the project. Our Project Director and Technical Lead (hired thanks to the EOSS-4 grants) will ensure a smooth onboarding and develop relationship with those new users. This work will be proposed as internship subjects, as this front consists in many relatively small and independent tasks, such as:

- Improvements to date parsing with better support for non-Western date formats
- Improvements to number parsing with locale support
- Generalization of the reconciliation API to expose service-defined features
- Localization of parts of the UI which cannot be translated yet for architectural reasons
- Implementation of clustering heuristics suitable for non-latin alphabets
- Better support for legacy encodings

**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.

- By December 2022: Assessment of cultural biases in OpenRefine with the community
- By September 2023 correction of at least three biases identified previously.

**(Outcomes include:)** Initiating exploration into the tool's cultural biases by:

- Educate our community on bias in open source processes, software, and user experience
- Engage our user and contributor community to identify and develop strategies to address bias in OpenRefine

# Table of Content

# Vote Here !!!!

Give points to your top 5 favorite ideas, giving a 5 to your most favorite and 1 to the one you'd prioritize least (and other numbers in between)

| Idea | Antonin | Jan | Martin | Sandra | Totals |
|---|---|---|---|---|---|
| [Debian packaging](#) | 1 | | Maintenance | 1 | 2 |
| [Windows packaging](#) | 3 | | Maintenance | 2 | **5** |
| [Accessibility review](#) | 4 | 4 | 4 Diversity | y | **12** |
| [Onboarding wizard](#) | 2 | | Scale community | | 2 |
| [Research cultural/linguistic diversity priorities](#) | 5 | 5 | 5 Diversity | 5 | **20** |

| | | | | | |
|---|---|---|---|---|---|
| [Improve translation infrastructure](#) | | 3 | 3 Diversity Scale community | 4 | **10** |
| [Consultants find more diverse committee members](#) | | | Meh | | |
| [Training on biases in data, open source, software, UX](#) | | | 1 Scale community Diversity | | 1 |
| [Set up volunteering framework](#) | | 1 | 2 Scale community | 3 | **6** |
| [OR website renewal](#) | | | Maintenance | | |
| [New explainer videos (internationalized)](#) | | 2 | Scale community | y | 2 |
| | | | | | |
| | | | | | |

# Project proposals - add yours too!

- Feel free to comment upon and endorse (or criticize) any proposal below
- And/or add your ideas (see [template at the bottom which you can copy and fill](#))
- **Deadline: Sunday, Aug 14, 2022 .** After this date, OpenRefine's advisory committee will regroup and take first decisions. We may get back to you with questions or follow-ups!
- Some proposals are general 'to dos' that would be valuable for OpenRefine in general, but perhaps not very diversity-related. They are listed as alternative suggestions to spend part of the budget, in case we don't collect enough strictly diversity/development related proposals.

## Debian packaging, longer term support over various years

- Software development
- Suggested by: Antonin
- Cost: 20K?
- Who will work on this? Freexian (existing contractor)
- What? Continue Debian (includes also Ubuntu and other Linux) packaging work and pay for it in advance for several of the upcoming years
- Arguments pro:

- - ○ Very little overhead and management effort, as we continue working with existing contractors who are quite autonomous (Antonin)
    - ○ Unearths underlying structural problems (Antonin)
  - ● Arguments con:
    - ○ Looking at end users: serves only a small part of our user base (Sandra)
    - ○ Packaging improvements to other OSes will equally unearth underlying structural problems? (Sandra)

## Windows packaging, eliminating the existing Java problems

- ● Software development
- ● Suggested by: Antonin, endorsed by Sandra
- ● Cost: 10-15K?
- ● Who will work on this? New contractor (to be recruited)
- ● What: Improve our Windows packaging so that the ongoing issues with Java inclusion are eliminated
- ● Arguments pro:
  - ○ Serves a large part of our user base, including (probably) many users in regions we currently don't serve very well yet (Africa, Asia, Latin America) (Sandra)
  - ○ Resolves a problem that currently causes a lot of frustration from end users and consumes a lot of time for trainers and anyone who generally supports/helps OpenRefine users (Sandra)
- ● Arguments con:
  - ○ It'll take us (mainly Antonin) quite a bit of time and effort to find contractors to do this, to onboard and manage them (Antonin)
    - ■ (Sandra) I suggest to counter this by also budgeting management time for whoever manages (i.e. probably Antonin in this case)

## Accessibility review of OpenRefine

- ● Research
- ● Suggested by: Antonin, endorsed by Jan
- ● Cost: ?
- ● Who will work on this? External consultant
- ● What? Do WCAG2 review of OR
- ● Arguments pro:
  - ○ Will give us insight in what our biggest gaps are
  - ○ Fairly easy to hire for
- ● Arguments con:
  - ○ Not core to the grant, but bad accessibility also hinders diversity in who are able to participate

# Design and develop an onboarding wizard for OpenRefine

- UI/UX design and software development project
- Suggested by: Lozana, endorsed by Sandra
- Cost: 5-10K?
- Who will work on this? UX designer (Lozana or new hire) + developer (Antonin or someone else / new hire)
- What? A step by step wizard inside OpenRefine that introduces new users to OpenRefine's features (similar to what you sometimes see in computer games)
  - Must be designed and built with internationalization/translatability as core characteristic
- Arguments pro:
  - I think this is an excellent idea in general, and if translatability is central in its design, then it will help international and diverse audiences use our tool (Sandra)
- Arguments con:
  - Not core to the grant / not extremely diversity focused (Sandra)
  - We need to find/recruit suitable people, which may be time consuming (Sandra)
  - Management efforts needed, probably mostly from Antonin (Sandra) - management and mentoring time to be budgeted too

# Research project, including user survey, to clarify and outline cultural and linguistic diversity priorities in OpenRefine

- Research project (no software development)
- Suggested by: Sandra (to fulfill deliverable in D&I project)
- Cost: 5-10K or more?
- Who will work on this? (To be recruited) academic(s) in the combined field of linguistics / data science
- What? Produce a written report (can potentially become as big as a set of academic articles / even evolve into a thesis over the longer term…) that outlines and prioritizes areas in which tools like OpenRefine need to support language and cultural variety, and tackle cultural biases in datasets. Based on:
  - desk research (through knowledge of characteristics of the world's languages, datasets in diverse languages, encoding formats, scripts… and general areas of typical cultural bias)
  - 1 or more surveys
    - user survey among OpenRefine users, prioritizing those with a non-Western background
    - potentially also among data scientists with non-Western backgrounds who use datasets in non-Western languages, scripts and formats and who are not acquainted with OpenRefine yet
- Arguments pro:
  - We promised in the D&I application to do a user survey focused on diversity and cultural biases (Sandra)

- ○ I have the impression we need this and it's actually very foundational for us; I feel that our existing insight into what's actually urgent and needed is rather anecdotal, but please contradict me if I'm wrong! (Sandra)
  - ○ I think this is generically extremely interesting research, which could be interesting for academics in linguistics and data science in the broad sense, so we'd contribute to broader goals than just ours (Sandra)
- Arguments con:
  - ○ Not a software development project, although it should be written so that it can clearly inform future software development (Sandra)
  - ○ Probably quite labor-intensive to formulate this sharply and to find good people to do this (Sandra)
  - ○ Needs management time on our side too - 2-4 hrs/week of Sandra's time - to be budgeted as such

## Improving OpenRefine's interface translation infrastructure and growing its translators' community

- Community and editorial project (no software development)
- Suggested by: Sandra
- Cost: 5-10K
- Who will work on this? Language / UI translation specialist / community manager (can be mostly Sandra, as addition to her current contract, and/or a new person)
- What? See work plan draft:
  - 📄 Improving OpenRefine's UI translation infrastructure and growing its translator co…
- Arguments pro:
  - ○ In our 2022 survey, actually many people expressed interest to translate OpenRefine's interface, but our community of translators is still small and can use a boost (Sandra)
  - ○ Weblate itself is not very intuitive for newcomers and I really think we can improve the user experience through some dedicated activity on our side (Sandra)
- Arguments con:
  - ○ Not a software development project (Sandra)
  - ○ Will need management - to be budgeted separately (Sandra)
  - ○ If not (fully) done by Sandra: finding/hiring and managing someone will be time consuming (Sandra)

## Contract consultants / agency that helps us recruit more diverse members for OpenRefine's governance bodies

- Governance/consultancy project (no software development)
- Suggested by: Sandra; CS&S

- Cost: 5-10K?
- Who will work on this? External consultants; we can tap into the CS&S network for recommendations
- What? Hire an agency which helps us find 3-4 international / diverse new members for OpenRefine's steering and advisory committee (aligned with diversity grant)
- Arguments pro:
  - Although we do our best to tap into our networks, finding good new people in areas/geographies that we don't serve well yet is not easy. A well-chosen agency can really help us here (Sandra)
  - Helps us achieve targets for the D&I grant which are, without help, difficult to achieve (Sandra)
- Arguments con:
  - Not a software development project at all, so perhaps a stretch to fit this into the development part of the D&I grant (Sandra)
  - Will need dedicated management time from Sandra; to be budgeted as well, e.g. 2-3 hrs/wk?
  - We already have tons of people to reach out to after the survey / Dataharvest contacts (Antonin)

# Guest lectures / webinar series on biases in open source processes, software, and user experience

- Community education project
- Suggested by: Sandra, based on a deliverable we formulated/promised in the D&I application ('Educate our community on bias in open source processes, software, and user experience')
- Cost: 10K
- Who will work on this? Sandra or (preferably) contractor, hopefully with help from CS&S and its network
- What? 3-5 guest lectures / webinars
  - Focus can also be narrowed per lecture e.g. focusing on diversity in date/time formats/notation, intricacies of data in non-alphabetic scripts,...
- Arguments pro:
  - If well-chosen / organized: can be very inspiring for OpenRefine's community (Sandra)
  - We promised to educate OpenRefine's community in these areas in the D&I grant application (Sandra)
  - Trainings that can become part of a mandatory onboarding process (Martin)
- Arguments con:
  - Not a software development process, although, if well designed, this should be very inspiring for our software development prioritization (Sandra)
  - Will be time-consuming to design and organize an interesting lecture series that is engaging/attractive for our community (especially to attract technical folks) + to find and engage the right lecturers / guest speakers (Sandra)

- ○ There's probably significant overlap with other programs that CS&S is already doing (this may also be a strength/opportunity if CS&S wants to work together on this) (Sandra)

# Set up a formal volunteering framework

- Community
- Suggested by: Sandra
- Cost: 5 K to set up, but needs longer term investment
- Who will work on this? Sandra (as addition to current contract / separate project) ? Other contractor is good too
- What? Develop a more formal framework / system to attract, mentor and retain non-technical volunteers. Modeled like the volunteering support of e.g. cultural organizations, events - i.e. more formal than usually in open source projects, with the goal to attract more diverse people. The support system would include a dedicated web page, ongoing explicit job openings, remuneration / perks / reward system, volunteer support/training/meetings… One potential activity can include the proactive recruitment of (paid) volunteers from underrepresented backgrounds.
- Arguments pro:
  - ○ Will make possible non-technical contributions to our project more visible and explicit for people who would otherwise not even consider contributing (Sandra)
  - ○ Similar to list of project for outreachy and GOS where the community can write a wishlist of what needs to be addressed
- Arguments con:
  - ○ Not a development project (Sandra)

# OpenRefine website renewal

- Communications / PR
- Suggested by: Sandra
- Cost: 5-10K
- Who will work on this? Some contractor
- What? Use some of the D&I money to actually fully redesign and rebuild our website
- Arguments pro:
  - ○ A renewed website can be created more with international audiences (and beginners) in mind (Sandra)
- Arguments con:
  - ○ Not a development project (Sandra)

# New OpenRefine explainer videos in 3-4 languages

- Communications / PR
- Suggested by: Sandra
- Cost: 5-10 K

- Who will work on this? Contractor / video producer; also pay for translators and multiple voice overs
- What? New official OR explainer videos, replacing the ones currently on OR's homepage, actually mentioning OpenRefine (not Google Refine) + produced with translatability and international audiences in mind
- Arguments pro:
  - Would be great and useful to have (Sandra)
- Arguments con:
  - Not a development project (Sandra)

## Template: My Amazing Idea

- Pick category/ies: Software development / research / community / other…
- Suggested by: Moi
- Cost: NN K
- Who will work on this? The right people
- What? My elevator pitch goes here
- Arguments pro:
  - My positive view here
- Arguments con:
  - Here I have my Negative Nancy hat on