

1. Proposal Title

Supporting Biomedical communities in using OpenRefine

2. Previous funding

Have you previously received funding for this proposal under the CZI EOSS program?

EOSS5-0000000404, EOSS-Diversity: ORefine: 21 SVCF - 2021-237149 OR for Everyone

Progress Report: If the grant is still active, provide a short summary of progress towards the deliverables (maximum of 250 words).,

The OpenRefine project currently has two grants open with CZI: the EOSS-Diversity and EOSS-5 grant. OpenRefine project director's resignation in December 2022 delayed both grants. OpenRefine hired a new project manager in June 2023, to resume our progress.

With the EOSS-Diversity grant, our goal is to diversify representation in OpenRefine governance. After extensive consultation with the community and external partners, we decided to change our strategy. Moving to a bottom-up model, OpenRefine, a successful project with hundreds of contributors, needs tools and processes to identify key contributors and cultivate them as leaders. To support this, we identified two new deliverables. Firstly, we will clarify OpenRefine's mission, vision, and values to ensure the community works towards common goals. Then, we will develop a [Mountain of Engagement framework](#) or [Contributor Pathways](#), a strategy to grow leadership from within user communities.

Significant steps were taken by 2023 to address OpenRefine's design limitations identified in the previous interim report for the Diversity grant. We established a formal design practice by recruiting an Outreachy intern in June of that year. Subsequently, we hired a part-time designer funded through the EOSS-5 budget. The lead designer will continue to improve our design practice and enhance OpenRefine's reproducibility.

For the EOSS-5 grant, we made progress toward improving OpenRefine reproducibility, specifically regarding preventing data loss and preserving pagination settings. However, reviewing changes to the code base has been challenging for the volunteer-run developer community, and funding may need to be allocated to support this effort.

3. Proposal Purpose:

Describe the purpose of the proposal in one sentence (maximum of 200 characters including spaces). Example: To develop a comprehensive, validated atlas of the human kidney at single-cell resolution open to the entire scientific and clinical community.

Enhance OpenRefine through community engagement, developing trainer, advocate, documenter, and technical contributor pathways, with a focus on engaging biomedical communities.

4. Amount Requested

Enter the amount requested per year between \$50,000 USD and \$200,000 USD per year, including indirect costs), as well as the total budget requested for all years in USD (between \$100,000 USD and \$400,000 USD total, including indirect costs). These numbers should match those described in the Budget Description to follow. Enter whole numbers only (no dollar signs, commas, or cents)

We are requesting a total of USD 400,000, which will be split into USD 200,000 per year for a period of two years.

5. Proposal Summary - Scope of Work (500 words)

Provide a short summary of the application (maximum of 500 words) (auto-filled from LOI; update if needed)

OpenRefine is free, open-source software providing diverse data cleansing features. It plays a critical role in biomedicine by aiding researchers in data cleaning and normalization, ensuring data quality, reconciling datasets, and preparing taxonomies. Informed by the EOSS-Diversity grant, we identified Contributor Pathways as optimal frameworks for a diverse contributor base and community leadership. The EOSS-6 grant will support three contributor pathways: providing better support for advocates and trainers, translating OpenRefine documentation to bridge language barriers in the scientific community, and securing developers' time to maintain the project.

1. Improve Support for Advocates and Trainers

Recognizing the critical role of trainers in OpenRefine's ecosystem, we aim to formalize and enhance their role while ensuring better support for the biomedical community. Through our survey, community forum, and direct outreach, we've found that trainers act as ambassadors, advocates, and educators. Despite their significance, they often operate independently, without sufficient support.

By hiring a community manager, we will establish a formal communication channel between the trainers and the OpenRefine developers. This will enable the developers to keep trainers informed about new features and upcoming releases, while trainers can provide valuable feedback on software usage and feature requests and help guide the roadmap. This initiative will help strengthen their connection with the OpenRefine community.

We also want to offer direct financial support for advocates and trainers to present OpenRefine at conferences and organize webinars. In-person events provide valuable opportunities to observe how people use OpenRefine, collect feedback and feature requests.

2. Translate OpenRefine documentation and training curriculum

English is considered the lingua franca of science. However, many trainers have reported that although their audience is comfortable using software with an English interface, using English documentation or training materials creates more barriers. [The manifold costs of being a non-native English speaker in science](#) and [Scientific publishing has a language problem](#) publication demonstrates that using a second language can increase the cognitive load for both trainers and trainees. Given that more than one-third of our users use OpenRefine in languages other than English, we want to allocate funds to support the translation of our user documentation and existing training curriculum, promoting inclusivity and accessibility.

3. Technical Position for GitHub Repository Maintenance

The renewal of the lead developer role is crucial for maintaining the project. Over the past four years, OpenRefine has undergone substantial enhancements, propelled by the support from the EOSS-1 and EOSS-5 grants. These improvements encompass critical features tailored for the biomedical community, including robust support for larger datasets and streamlined functionality for reproducibility and workflow automation.

Focusing on 2025 and 2026, The lead developer will ensure that developed features seamlessly integrate into user workflows. The lead developer will also document pathways for core and extension developers, and implement new features based on trainer community feedback.

6. Work Plan

A description of the proposed work for which funding is being requested, including resources the applicants will provide that are not part of the requested funding. For software development-related work (e.g., engineering, product design, user research), specify how the work fits into the existing software project roadmap. For community outreach related activities (e.g., sprints, training), specify how these activities will be organized, the target audience, and expected outcomes (maximum of 750 words)

We propose a comprehensive plan organized around three axes to enhance OpenRefine's impact on the biomedical research community. Each axis represents a strategic focus area,

ensuring a holistic approach to community engagement, translation of efforts, and sustained developer support.

1. Improve Support for Advocates and Trainers

Our goal is to leverage the EOSS-Diversity grant to establish a robust foundation to support the OpenRefine community. During the first six months of 2024, we will document OpenRefine's mission, vision, and values. Preparing for the EOSS-6 grant, we plan to initiate the community manager hiring process in the second quarter of 2024.

Via the EOSS-Diversity grant, we aim to establish six different contributor pathways by the end of 2024. These pathways will include (1) developer, (2) designer, (3) documentation and training, (4) translation, (5) user support, and (6) advocacy and outreach tracks. Each pathway will consist of a series of activities that one can undertake to contribute to OpenRefine, with the level of commitment and impact on the community increasing with each activity. The idea is to encourage people to start with small steps and progress toward leadership positions. The EOSS-6 grant will be used to sustain the documentation and training pathways, as well as the advocacy and outreach track for the biomedical community.

The community manager plays a crucial role in ensuring the smooth administration of contributor pathways. They help new contributors get on board different tracks and facilitate their growth as leaders. Additionally, we identified the following tasks that will be funded by the EOSS-6 grant and led by the community manager.

1. Offer financial support covering travel expenses to present OpenRefine at biomedical conferences or to organize online events such as webinars. In return, we ask for formal feedback through an online form to capture what the person learned at the conference and how we can improve OpenRefine. We want to establish a process to formally capture this information and share it with the developer and designer community.
2. Organize monthly communication with trainers and advocates via calls, blogs, or forums to get their feedback on software or community improvements. This cultivates community belonging and ensures their voices are heard.
3. Create a trainer directory to recognize and promote knowledgeable instructors.
4. Curate tutorials and cheat sheets for the biomedical community, providing access to quality content to advocates and trainers.
5. Organize a yearly in-person contributor event to share ideas on improving the OpenRefine community and software, acknowledge contributions, and foster collaboration and engagement.

The community manager is also responsible for granting and revoking rights as defined in the pathways and contributor guide. To supplement this role, the part-time project manager position will be renewed. The project manager will handle administrative tasks such as budget management, and coordinating with external partners. The project manager will be responsible

for managing grants and administrating the conference support programs, as well as hiring and contract review.

2. Translate OpenRefine documentation and training curriculum

We have received feedback from our non-English speaking users and partners in Europe, Latin America, and Africa that learning and teaching in English as a second language can be challenging and increase cognitive load. Providing training materials and documentation in their native language can significantly improve their learning experience by reducing barriers. Therefore, we are committed to translating our user documentation and creating new training materials in non-English languages.

In 2024, we will conduct our bi-annual survey to identify priority languages. We also want to work with organizations already offering OpenRefine courses to determine how we can help maintain existing non-English courses or help translate new ones. Finally, we intend to dedicate a budget to translate OpenRefine user documentation.

3. Continue to sustain developer support for the project.

Our goal is to continue to fund the technical lead position to maintain the OpenRefine code base and support our community. The technical lead is responsible for maintaining the OpenRefine GitHub repository, triaging tickets, reviewing pull requests, publishing releases, responding to security advisories, and participating in internship programs.

The lead developer will be responsible for documenting the developer pathways and improving the documentation for extension development. They will focus on enhancing the development environment, creating documentation, and maintaining an example extension that follows recommended best practices. The example extension will include automatic updates, end-to-end testing, localization, and automated release processes.

Finally, the lead developer will prioritize the development of feature requests and issue tickets submitted by our trainer as part of the feedback process.

7. Milestones and Deliverables

List expected milestones and deliverables, and their expected timeline. Be specific and include where possible any goals for metrics the software project(s) are expected to reach upon completion of the grant. Please use a third-person voice (maximum of 500 words).

From January 2024, we will work on EOSS-Diversity grant deliverables as stepping stones for the EOSS-6 and document OpenRefine Mission, Vision, Values along with our six contributors' pathways.

Deliverables for Advocates and Trainers:

- Starting April 2024, with the grant announcement:
 - Identify target biomedical conferences.
 - Provide an online form for trainers to offer post-event feedback, aiming for at least 20 submissions per year.
 - Start the recruitment process for the community manager with a start date no later than September 2024.
- Starting October 2024 until June 2026, organize monthly events with the trainer community focusing on the biomedical community. We will confirm the format between a community call, webinar, newsletter, or forum posts.
- Starting October 2024, project and community managers will accept conference stipend applications continuously throughout the grant. We plan to fund at least five events per year.
- By end-2024, create the Trainer Directory with a goal of at least 20 trainers by the grant's end.
- Starting in 2024, organize a yearly in-person event for 20 to 30 contributors.

Deliverables to translate OpenRefine documentation and training curriculum

- Starting in June 2024 the project manager will start establishing formal partnerships with the maintainers of at least three courses that we want to help maintain or translate. Some of the courses we are interested in include The [Carpentries](#), the [Programming Historian](#), the [Global Biodiversity Information Facility \(GBIF\)](#), along with courses developed by the University of Toronto, UC San Francisco, the University of Texas, the University of Illinois, and the Griffith Library. We also plan to leverage the community of trainers we developed through the [Wikimedia Commons Train-the-trainer program](#). Furthermore, we are in early discussions with [MetaDocencia](#) to assist us with the contextualization of our documentation and training materials.
- By January 2025,
 - Set up translation workflows with Docusaurus for OpenRefine official documentation.
 - Create a call for proposals to translate the documentation or curriculum or maintain existing courses in partnership with the course maintainer.
- By March 2025, select individuals performing the documentation and course translation along with curriculum maintenance.
- By June 2026 we want to have at least two of the following
 - OpenRefine documentation translated into two languages;
 - At least one curriculum translated into two languages;
 - At least three non-English courses updated to the latest version of OpenRefine
- By June 2026, we want to document our learning and update the process for contributing to OpenRefine translation and documentation. This aligns with our efforts to develop contributor pathways.

Deliverable to continue to sustain developer support for the project.

- At least two minor releases per year.
- At least five bugs or feature requests reported by the trainer community addressed per year
- By June 2026
 - Create a well-maintained extension that follows recommended best practices including automatic updates, end-to-end testing, localization, and automated release processes.
 - Welcome at least ten new contributors.

8. Existing Support

List active and recently completed (previous two calendar years) financial or in-kind support for the software project(s), including duration, total costs in USD, and source of funding. Include any previous funding for these software projects received from CZI, Wellcome, and/or Kavli outside of the EOSS program (maximum of 250 words).

We publish a list of our funding sources on our website: <https://openrefine.org/funding>

2023 NFDI

- Status: Awarded - Current
- Amount: EUR 10,000
- Source: [NFDI4Culture](#)
- Start date: August 2023
- End date: December 2023

2023 Wikimedia Foundation

- Status: Awarded - Current
- Amount: USD 50,000
- Source: [Wikimedia Foundation Funds](#)
- Start date: June 2023
- End date: June 2024

2022 NFDI

- Status: Awarded - Completed
- Amount: EUR 10,000
- Source: [NFDI4Culture](#)
- Start date: July 2022
- End date: December 2022

2022 EOSS-5

- Status: Awarded - Current
- Amount: USD 310,100 (USD 155,050 per year for two years)
- Source: [Chan-Zuckerberg Initiative](#)
- Start date: January 2023
- End date: December 2024

2021 EOSS-Diversity

- Status: Awarded - Current
- Amount: USD 240,000 (USD 120,000 per year for two years)
- Source: [Chan-Zuckerberg Initiative](#)
- Start Date: September 2021
- End Date: December 2024 (extended from August 2023)

Unrestricted Individual Donations

- 2023: USD 830
- 2022: USD 570

9. Landscape Analysis (250 words)

Describe the other software tools (either proprietary or open source) that the audience for this proposal primarily uses. How do the software project(s) in this proposal compare to these other tools in terms of user base size, usage, and maturity? How do existing tools and the project(s) in this proposal interact? (maximum of 250 words). (auto-filled from LOL; update if needed)

Data cleansing tools can be categorized as follows:

1. Spreadsheet software provides an approachable interface to data manipulation, which now includes more data cleaning features, but lacks reproducible workflows and does not scale for the large datasets commonly used in science contexts.
2. Programming languages like Python and R offer flexibility and reproducibility but have a steep learning curve for people with no prior coding experience.
3. Software like KNIME or SAS provides a data workbench including data analytics, reporting, and integration using visual components.
4. Self Service Data preparation software fills the gap between these categories. With a powerful GUI, this category of software can be easily mastered by non-programmers.

Free and open-source data manipulation software makes this functionality available to communities and user groups with few resources as well. Solutions include:

- Orange <http://orange.biolab.si/> - focus on data visualization
- Workbench <https://github.com/CJWorkbench> - focus on data journalism; ceased operations in 2021

- QSV <https://qsv.dathere.com/> released in 2023 currently offers limited functionality.

OpenRefine stands out in the data preparation category, bridging the gap between spreadsheet software and programming languages. With its robust GUI, OpenRefine is designed for non-programmers while supporting the handling of large datasets. OpenRefine is the most advanced interface for data reconciliation and linkage within the biomedical research context. Overall, OpenRefine provides a low-code interface, making most data exploration, cleaning, and preprocessing tasks user-friendly.

10. Value for Biomedical user (250 words)

Describe the expected value of the proposed work to the biomedical research community (maximum of 250 words). (auto-filled from LOI; update if needed)

Our [2022 user survey](#), which included 50 participants with a research background (out of a total of 178 answers), along with relevant literature reviews focusing on publications from 2022 and 2023, highlights the importance of OpenRefine for biomedical researchers.

According to the survey results, 75% of researchers use OpenRefine monthly for at least a year.

Researchers employ OpenRefine for diverse tasks, with 94% utilizing it to clean and normalize data—a critical role in maintaining data quality. This is evidenced in publications such as

- [Patients' Severity States Classification based on Electronic Health Record \(EHR\) Data.](#)
- [Assessing spatial and temporal patterns and risk factors for tick acquisition.](#)
- [Analyzing Autism Prevalence Among Original Medicare Beneficiaries.](#)

60% of users rely on OpenRefine to prepare taxonomies and thesaurus before importing data into other software. The publication below provides an example of such workflow..

- [An Ontology-Based System for Cancer Registry Data.](#)
- [Semantic resource responding to Open Science principles: The meat thesaurus.](#)

Half of the respondents use OpenRefine for reconciliation against other datasets, a speeding up thesaurus and bibliometric analyses, evident in publications below:

- [Antifungals from plants: a bibliometric analysis](#)
- [Potential use of microalga *Dunaliella salina* for industrial bioproducts](#)
- [Modeling COVID-19 Transmission Dynamics: A Bibliometric Review.](#)
- [Coral restoration patents and academic research disconnection.](#)
- [Assessments of two-pore channel 2 in the human MDAMB-231 breast cancer cell line.](#)

11. Category

Data Management and workflow
Machine learning and data analysis.

12. Diversity, Equity, and Inclusion Statement:

Advancing DEI is a core value for this program, and we are requesting information on your efforts in this area. Describe any efforts the software project(s) named in this proposal have undertaken to increase diversity, equity, and inclusion with respect to their contributors and audience. Please see [examples](#) from applications funded in previous cycles (maximum of 250 words)

OpenRefine has a broad user community, remarkably diverse in fields of expertise, as well as geographical, gender, and racial backgrounds. We value the diverse backgrounds and opinions of our community, which helps us create better technology and support our users. We prioritize diverse voices in OpenRefine's governance and community. We aim to continuously evaluate our governance and community strategy (in collaboration with experts) to improve OpenRefine's accessibility and inclusive practices.

OpenRefine has participated in several Outreachy and Google Summer of Code internship programs, mentoring interns and new code contributors from China, India, Cameroon, and Kenya. When hiring new contractors and staff, we consciously reach out to communities that are traditionally underrepresented in open-source development; for the current project on the integration of Wikimedia Commons features, we work with developers from Cameroon and Costa Rica. To support diversity in OpenRefine, we formalized and published a Code of Conduct in 2020:

https://github.com/OpenRefine/OpenRefine/blob/master/CODE_OF_CONDUCT.md Early 2022, OpenRefine hired a project director (financed via CZI's EOSS Diversity and Inclusion grant) who is tasked to increase diversity in our community and governance bodies. Via the same EOSS DEI grant, we are improving our code base and clustering algorithms to increase OpenRefine's usefulness for datasets in non-Western languages and to generally eliminate cultural biases from the tool. OpenRefine is fiscally sponsored by Code for Science and Society (CS&S). CS&S is an equal-opportunity employer committed to hiring a diverse workforce at all levels of the organization

13. CV of Applicant

Upload in PDF format; include current and recent employment, education history, and references to any major publications, software contributions, or other relevant outputs (maximum of two pages)

14. Budget Description

Upload in PDF format; budgets can be uploaded in a combined single PDF or one PDF for each software project; font must be 11 point or larger and margins must be at least one-half inch (top, bottom, left and right) for all pages (letter size required, one page per software project maximum).

- Description of the costs to be funded by this grant at a high level and in narrative or tabular form, outlining costs for personnel (including names, if known), supplies, equipment, travel, meetings/hackathons/sprints, subcontracts, other costs, and up to 15% indirect costs (excluding equipment and subcontracts).
- Indirect costs are limited to up to 15% of direct costs and are included within the annual budget total. Indirect costs may not be assessed on capital equipment or subcontracts, but subcontractors may include up to 15% indirect costs of their direct costs. Non-charitable entities must include a clear allocation and explanation for any indirect costs included in a proposed budget.
- Budget should be requested in US dollars.
- International grantees must use all grant funds exclusively for activities conducted outside the United States of America. Travel expenses to the United States (including round-trip tickets) should not be covered from the requested grant funds.
- Application budgets must reflect the actual needs of the proposal. The Funders will work closely with successful applicants to arrive at a mutually acceptable budget after review.

Item	June 24 - May 25	June 25 - May 26	TOTAL	Notes
CS&S Fee (15% on cash received)	\$30,000	\$30,000	\$60,000	Note 1
Developer	\$45,000	\$90,000	\$135,000	Note 2
Project Manager		\$15,000	\$15,000	Note 3
Community Manager	\$30,000	\$30,000	\$60,000	Note 4
Contributor Conference		\$15,000	\$15,000	Note 5
Travel Stipend for advocate	\$10,000	\$10,000	\$20,000	Note 6
Documentation and Training Translation	\$47,500	\$47,500	\$95,000	Note 7
TOTAL	\$162,5000	\$237,5000	\$400,000	Note 8

Note 1: This budget line covers the contribution by Code for Science & Society, OpenRefine's fiscal sponsor, providing financial administration, accounting, administrative and strategic support.

Note 2: This should cover approximately 0.5 FTE, which is sufficient for the activities outlined in the proposal. Combined funding from the EOSS-5 grant and the Institut für Angewandte Informatik funds this position full-time until December 2024. Through 2024, we will look for funding opportunities for the remaining 18 months.

Note 3: Part-time position for approximately 10 hours per week. This position corresponds to Martin Magdinier's current contract as a Project Manager. The Project Manager position will be covered in 2024 by the EOSS-Diversity grant, and for 2025, by the EOSS-5 grant.

Note 4: Part-time position for between 15 and 20 hours per week.

Note 5: The 2024 event will be supported by EOSS-Diversity, while EOSS-5 and EOSS-6 grants will support the 2025 and 2026 editions respectively.

Note 6: Estimate for 5 events with a budget of USD 2,000 per event to cover travel expenses and conference registration.

Note 7: As a reference point, in 2021, we spent USD 33,000 to overhaul and restructure OpenRefine's documentation completely. A budget of USD 95,000 will cover the cost of developing courses and documentation in three to four languages.

Note 8: We are asking for the maximum funding of USD 200,000 per year. Fund not spent in year 1 will be allocated for year 2 activities as described in our budget.

15. Open Source Software Project(s) Details

- Software Project name (required): **OpenRefine**
- Main code repository (e.g. GitHub URL) (required)
<https://github.com/OpenRefine/OpenRefine>
- Homepage URL (required) **<https://openrefine.org/>**
- Social media handles (if applicable)
 - <https://twitter.com/openrefine>
 - <https://fosstodon.org/@OpenRefine>
 - <https://www.linkedin.com/company/openrefine/>
- Do you or software project key personnel have commit rights to the code repositories for this software project? (required) **YES**
- Short description of software project (200 words maximum) (required)

OpenRefine is a power tool to clean up messy data. Requiring no knowledge of a programming or query language, it lets users find and fix inconsistencies interactively, match their data to external databases, pull additional data from these, and many other useful operations. The resulting workflows can be extracted and applied to other datasets.

OpenRefine, originally known as "Freebase Gridworks", was created by Metaweb as an Export-Transform-Load tool to populate Freebase. Following Google's acquisition of Metaweb, it briefly became a Google product. Since 2013, a team of devoted users and developers have been taking care of the project under the name OpenRefine. Thanks to the support of various funders, a small core team has been able to maintain the project's sustainability since 2019. Many universities, newsrooms, and libraries have active user communities that run training workshops about the tool. Our committed team of trainers provides quick and high-quality support through our forum, StackExchange, and social media.

SOFTWARE PROJECT #1: List of Key Personnel:

- First Name: **Martin**
 - Last Name: **Magdinier**
 - Email Address: **martin@openrefine.org**
 - Current employer/Affiliation: Code for Science and Society, RefinePro Company
 - Role in project **Project Manager**
 - Developer username if applicable (e.g., GitHub handle):
<https://github.com/magdmartin>
 - Country of Residence: **Canada**
-
- First Name: **Antonin**
 - Last Name: **Delpeuch**
 - Email Address: **antonin@delpeuch.eu**
 - Current employer/Affiliation: **Institut für Angewandte Informatik**
 - Role in project **Technical lead**
 - Developer username if applicable (e.g., GitHub handle): **<https://github.com/wetneb>**
 - Country of Residence: **Germany**

1. SOFTWARE PROJECT #1 : What is the software project license?

Permissive license (e.g. **BSD 3-Clause**, MIT, Apache 2.0)

2. SOFTWARE PROJECT #1 : What is the main programming language?

Java

3. SOFTWARE PROJECT #1: Does the software project have a code of conduct?

Yes. Link (optional):

https://github.com/OpenRefine/OpenRefine/blob/master/CODE_OF_CONDUCT.md

4. SOFTWARE PROJECT #1: Does the software project have end-user documentation?

Yes. Link (optional): <https://openrefine.org/docs>

5. SOFTWARE PROJECT #1: Does the software project have an issue tracker?

Yes. Link (optional): <https://github.com/OpenRefine/OpenRefine/issues>

6. SOFTWARE PROJECT #1: Does the software project have a community engagement / Q&A forum (self-hosted, on Stack Exchange etc.)?

Yes. Link (optional): <https://forum.openrefine.org/>

7. SOFTWARE PROJECT #1: Does the software project have contribution / coding guidelines?

Yes. Link (optional): <https://openrefine.org/docs/technical-reference/contributing>
<https://github.com/OpenRefine/OpenRefine/blob/master/CONTRIBUTING.md>

8. SOFTWARE PROJECT #1: Is there a corresponding package available in a package manager (PyPi, CRAN, etc.)?

Yes. Link (optional): : <https://packages.debian.org/bookworm/openrefine>,
<https://formulae.brew.sh/cask/openrefine>,
<https://snapcraft.io/install/openrefine/ubuntu>

9. SOFTWARE PROJECT #1: Does the software project support continuous integration for testing?

Yes. Comment (optional): Java unit tests (TestNG) and Cypress UI tests

9. SOFTWARE PROJECT #1: Metrics

	Number	Comment
Scholarly paper(s) (including preprints) citing or mentioning the software project	6,630	According to Google Scholar as of 2023-12-04 searching for OpenRefine OR "Google Refine" OR "Freebase Gridworks"
Monthly users, if applicable (based on one or more of the following: monthly downloads from websites, monthly downloads from package managers, monthly unique requests for updates, etc.)	32,801	32,801 downloads OpenRefine 3.7.6 since its release on 2023-10-09
Software projects that depend on the project		As a web-based tool designed for end-users, OpenRefine is not meant to be used as a software dependency.
Monthly visitors to project's website, discussion forum (e.g. Stack Overflow)	10,000	Average of 10,000 monthly visitors on our forum between September 1st and November 30, 2023. Average of 130 post on our forum per month for the same period. 2,908 visits to OpenRefine's GitHub repository between November 21 and December 4, 2023 (statistics provided through GitHub Insights, which gives info about the last 14 days). We do not track visits to our website.