

**I am applying for:** \* OpenRefine

**Amount Requested:** USD 50,000

**Project Title\*** Enhancing OpenRefine Developer Ecosystem by improving OpenRefine extension supports.

**Project Purpose:\* What issue of privacy, agency, bias, or transparency in the data life cycle does your project attempt to address?**

OpenRefine is a free, open source power tool for working with messy data and improving it. OpenRefine's point-and-click interface empowers domain experts with a deep understanding of data context, enabling them to manipulate large datasets without requiring extensive programming expertise.

OpenRefine democratizes data manipulation for librarians, data journalists, cultural sector professionals, semantic web enthusiasts, and researchers. [Our 2022 user survey](#) revealed that over 50% of OpenRefine users rely on the tool for tasks such as dataset comprehension, analysis, and enhancement. These tasks encompass data cleaning, normalization, transformation, reconciliation, and integration with third-party sources. Through that process, domain experts can detect potential bias in a dataset. OpenRefine runs directly on a user's computer, ensuring data privacy.

With the support of this grant, we want to improve the experience for OpenRefine extension developers by helping them integrate OpenRefine with their applications and in-house datasets. Such extensions offer a seamless experience to their non-technical users and avoid any data offloading between applications. Extensions encompass data reconciliation and round-trip integration involving data export and import with other applications—such as databases, cloud platforms, or knowledge graphs.

Our project aims to empower developers to create a more cohesive and interconnected data ecosystem, ensuring users can exercise greater control and agency over their data-related processes.

**Concept Summary:\* Describe your project and how it addresses this issue (200 words or less).**

After successfully releasing the Wikidata/Wikibase and Wikimedia Commons extensions, our project aims to respond to the growing demand within the community for expanded integrations with diverse datasets and applications (e.g., IIIF, Flickr). However, given the complex and varied nature of these workflows and use cases, it is not feasible for the core OpenRefine team to single-handedly develop and maintain such extensions.

OpenRefine's existing support for extensions and plugins provides a foundation for extending functionality. Our recent interviews with current and past extension developers underscored the importance of offering improved documentation, comprehensive tutorials, enhancements to OpenRefine's backend infrastructure to include more hooks and API endpoints, and possibly introducing a native extension manager to enhance discoverability and installation process.

In 2020, the transition from org.json to Jackson and Ant to Maven introduced backward incompatibilities. Our upcoming [4.0 version](#), which leverages Spark as a data engine, will necessitate further updates to existing extensions. A revamped extension development guide will also be crucial to accommodate these changes.

By addressing these challenges head-on, our project seeks to empower developers to extend OpenRefine's capabilities, fostering a more integrated and agile environment for their users ultimately contributing to enhancing agency and control for data professionals across various domains.

**Beneficiaries:\* List the main beneficiaries of your work (100 words or less).**

The primary beneficiaries of our work are professionals and individuals engaged in data-related tasks across a diverse spectrum of fields. Our 2022 user survey sheds light on the broad user base, with participants including librarians, cultural sector professionals, Linked Open Data enthusiasts, researchers, and data scientists.

According to our survey, OpenRefine serves both professional and personal contexts, with 15% of users leveraging it in their free time. The survey underscores the global nature of OpenRefine's user base, with 64% using the English version and 36% using one of the ten existing translations.

Close to 50% of OpenRefine users installed at least one extension. Popular [extensions](#) include the Wikidata extension, several extensions to produce RDF and integrate with SPARQL endpoints, an integration with FAIR data point, and GeoJSON export. Reconciliation services such as Wikidata, Getty, or VIAF are widely utilized, and close to 10% of the users rely on OpenRefine to reconcile against an in-house dataset.

**Do you have a public repository of code or a demo for this project? If yes, please provide the URL.**

**OPTIONAL: Please share any demographic information with us here (100 words).**

OpenRefine support a diverse contributor base, reflecting our commitment to inclusivity. Since 2020, we have partnered with Outreachy, welcoming at least two interns each summer. This collaboration fosters a welcoming environment and drives engagement from underrepresented groups, enriching our developer community.