Please complete the following information for the Applicant (required):

The information entered should be for the individual submitting the application who will act as the main person responsible for the application and as its point of contact. Information about other Key Personnel on the proposal should be entered where requested in the Software Project Details part of the application. **Complete all fields in this task; all fields are required.** Many fields are auto-filled based on your answers in the LOI; please edit as needed.

To edit your name or email, please do so in your account information by clicking your name in the upper right corner and clicking My Account in the dropdown menu.

1. Name: Jaime Rodríguez-Guerra

2. Email: jrodriguez@quansight.com

---

3. Degrees (check all that apply)

Responses Selected:

Bachelor's degree or equivalent
Master's degree or equivalent
PhD or equivalent

---

4. Institution/Affiliation	Quansight Labs
5. Title/Position	Software Engineer
6. Country of Residence	Germany
7. Organization (Organization that would receive and distribute funding if a grant is successfully awarded. Note that this may be different from the Applicant's main affiliation/current employer.)	NumFOCUS

2. Organization Details

Completed - Jun 1 2022

Organization Details

Please complete all requested information. The information entered should be for the organization that would be directly receiving and distributing funding if a grant is successfully awarded (e.g. academic institution, fiscal sponsor). Note that this may be different from the Applicant's main affiliation.

1. Type of Organization (required):

Fiscal sponsor

2. Organization (required):

2a. Organization Name	NumFOCUS, Inc.
2b. Address	P.O. Box 90596
2c. City	Austin
2d. State/Province	Texas
2e. Country	United States
2f. Tax ID (9-digits; format: XX-XXXXXXX; enter 44-4444444 if not applicable)	45-4547709
2g. Organization website	https://numfocus.org

3. Organizational/Administrative Contact (required):

List the name and contact information for the administrative contact to discuss additional information needed, if selected for award.

3a. First Name	Nicole
3b. Last Name	Foster
3c. Title/Position	Operations Manager
3d. Email	nicole@numfocus.org

4. Signing Official (required):

List the name and contact information for the person authorized to sign on behalf of your organization.

3a. First Name	Leah
3b. Last Name	Silen
3c. Title/Position	Executive Director
3d. Email	leah@numfocus.org

5. Press Contact / Public Relations Official (required):

List the name and contact information for the person to discuss press releases and media.

3a. First Name	Nicole
3b. Last Name	Foster
3c. Title/Position	Operations Manager
3d. Email	nicole@numfocus.org

6. Institutional Approval Form (required):

Upload as a single PDF. This [form](#) should be reviewed and signed by a person authorized to sign on behalf of your organization agreeing to the stated institutional and investigator requirements and commitments on data, resource sharing, and publication policies, as well as endorsing/verifying your application materials and confirming their ability to receive funding for the proposal. In the event of an award, all funds will be awarded to the applicant institution as the prime institution, and the applicant institution will be responsible for ensuring compliance of all of the terms, including compliance of all partners/subcontract institutions. **These policies are non-negotiable so this form should only be signed if the organization is able to comply with the terms as stated.** While CZI does not require sign-off by all of your partner institutions, please refer to what your institution requires. **Note: digital signatures are permitted as long as the document is not encrypted or password-protected.**

[Cycle 5 Institutional Approval Form - signed.pdf](#)

Filename: Cycle 5 Institutional Approval Form - signed.pdf **Size:** 2.7 MB

3. Proposal Details

Completed - Jun 2 2022

Proposal Details

Please complete the following proposal information. ***All sections are required.***

1. Proposal Title: Transparent, open & sustainable infrastructure for conda-forge and bioconda

2. Previous Funding

Did you previously apply for funding for this or a related proposal under the CZI EOSS program?

Yes

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

No

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.

To improve conda-forge and bioconda's sustainability and transparency by adopting vendor-agnostic and secure infrastructure practices and developing comprehensive maintenance metrics and dashboards.

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)

Year One	199557
Year Two	198407
Total All Years	397964

5. Proposal Summary:

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

The core scientific principle of reproducibility is, in many ways, parallel to the core open-source tenets. In an open-source context, the scientific community can analyze every step of the process, building trust in its effectiveness and contributing to its robustness by identifying bugs when they arise. However, aiming for reproducibility is a complex task involving challenges regarding data provenance and deterministic development environments.

The Conda-forge and bioconda projects were founded in 2015 in response to frustrations scientific software users consistently faced when attempting to install system package dependencies. Installing open-source software packages with binary dependencies is frequently a multi-step process involving an intricate sequence of software compilation. The emergence of conda-forge massively reduced the scientific packaging toil by building on transparency, automation, compatibility and open-source principles. As a result, its community has grown exponentially, as has the number of artifacts hosted and downloaded (~18 thousand packages hosted, 927 thousand tarballs, and ~300 million package downloads/month). Such growth has significantly increased the pressure on its underlying infrastructure, tooling, and maintainers' workflow. Besides, this entire infrastructure and community rely on the anaconda.org service, which is not open-source.

To ensure the long-term sustainability of these projects, we propose migrating to fully open-source tooling as follows:

1. Reducing infrastructure technical debt

Conda-forge infrastructure and tooling are distributed across many GitHub repositories, external CI services (Azure DevOps, GitHub Actions, TravisCI, Drone.io, CircleCI), Heroku "dynos" and AWS instances. Many were built as ad-hoc fixes and currently lack documentation or risk mitigation plans. We plan to migrate the configuration and infrastructure provisioning to reproducible, vendor-agnostic tools such as Terraform, complemented with rigorous testing, vulnerability detection, and documentation strategies to enable better security, reliability, and recovery from adverse events.

2. Adopting an OCI-based mirroring strategy

Anaconda.org is the default and sole host for all published and installable scientific packages. Adopting vendor-neutral tooling and standards (such as OCI) will ensure we uphold the core principles of open source and aid the project's long-term sustainability. We also believe that using and building an infrastructure that follows these open principles are the right foundation for more productive and

impactful research and education.

3. Development of a maintenance dashboard on Quetz

There is no straightforward way to monitor the operational status of conda-forge's infrastructure. The existing conda-forge.org/status panel is far from giving a comprehensive view of ongoing maintenance tasks, bottlenecks or the overall health of the many bots and infrastructure pieces. Having a detailed picture of the infrastructure and automation tools will significantly improve the maintainers' workflow and aid with identifying critical risks— which is essential to keeping up with the increasing growth and demand from the community. Quetz is chosen as an open-source server for hosting conda packages, thus allowing for increased transparency and extensibility. This would result in the added benefit of centralizing the currently scattered-across-repositories packaging metadata in a canonical, API-first, performant-at-scale database, laying the foundation for further infrastructure automation and improvements to the building processes.

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)

This proposal aims to improve the long-term sustainability of the conda-forge project and its ecosystem through three main work items.

Reducing technical infrastructure debt

Since its emergence in 2015, the conda-forge project has seen explosive growth in contributors, maintainers, repositories, artefacts, and packages served. To serve such a vast ecosystem (and around 300 Million downloads per month), the core team has heavily relied on automation, Continuous Integration and Delivery platforms and in-kind donations from multiple infrastructure providers.

However, the underlying infrastructure and tooling development and implementation have been part of

an ad-hoc and responsive process, which has led to sub-optimal usage of specific platforms and tools. This lack of planning and the current lack of documentation and security best practices also introduce significant reliability, scalability, robustness, and security challenges.

To address these issues and ensure the long-term viability of the project and its ever-growing ecosystem, we propose the following actions:

1. Performing in-depth analyses and audits of the existing tools, bots, access tokens, and other critical infrastructure. These audits will enable the team to identify critical vulnerabilities and develop a comprehensive roadmap and architectural redesign of the conda-forge underlying infrastructure and tooling. Therefore, improving the reliability and performance of the packaging lifecycle.

2. Adopting a reproducible and vendor-agnostic infrastructure approach through Terraform. Thus allowing the team to recover from catastrophic incidents by using infrastructure as code tools. These changes will also support the open-source community's right to replicate and reuse.

3. Standardise automation workflows and reduce vulnerabilities at the conda-forge and user-level. This work proposes introducing DevSecOps practices such as policy as code, better access controls, reducing or eliminating the use of long-living credentials and leveraging tools such as GitBom for software identification and vulnerability management.

4. Develop comprehensive documentation on infrastructure, bots, automation processes, and related components. There is no documentation for this vital area of the project, which directly threatens the project's sustainability. In addition to this, we will standardise the onboarding and offboarding processes for new maintainers (including credential management, 2FA etc.).

5. Create a canonical database for metadata from conda-forge repositories (feedstocks), generated artefacts, workflows and published packages, and corresponding data schemas.

OCI based mirroring

OCI (Open Container Initiative) registries are a well-defined standard for versioned blob storage already implemented by many public cloud providers (i.e. GitHub). We plan to leverage the OCI registry on GitHub to develop a community maintained mirror for bioconda and conda-forge. This mirror will support the conda-forge and bioconda's sustainability by reducing the single dependency on the current implementation (Anaconda's commercial entity).

Besides, adopting an OCI-based strategy will provide additional flexibility for metadata expansion. With this in mind, we aim to produce and capture more comprehensive package metadata, which we will later use for: maintenance dashboards, package search functions, and security checks.

Maintenance dashboard on Quetz

The Quetz server has been built upon a plugin architecture (both on the front and backend components). This approach enables developers to quickly write front-end plugins to display aggregated data on the conda-forge repository. We aim to leverage this as well as the proposed work on improving conda-forge packages data collection by:

1. Deploying a Quetz instance on a public cloud server using Terraform and connecting this to the multiple conda-forge data sources.

2. Redesigning and modernising the existing conda-forge website. We will completely redesign the current conda-forge status and migration pages to provide more valuable and meaningful metrics to both the maintainers and end-users. We will work with the Quansight in-house UX and accessibility designers to develop the mockups and ensure that the site and dashboards adhere to accessibility standards.

3. Extending Quetz and its front-end components to display existing packages and dependencies and provide an interface to compare package variants and conda-forge pinnings.

4. Supplementing the package database with data from conda-forge repositories in GitHub. This will allow us to provide valuable information, such as the number of open Pull Requests and relevant issues. These data will be complemented with build logs and a quantitative health metric for the many conda-forge project repositories and their artefacts and workflows.

5. Finally, we have planned incremental improvements on both the data representation and integration and the UI components of the dashboard. Such an approach will facilitate an iterative design and development process centred around accessibility, end-user experience and user feedback.

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).

First 6 months:

The focus is on laying the foundation for the bulk of the infrastructure work, data collection and representation. The expected deliverables are:

1. An in-depth audit of the current infrastructure, tooling, and credentials in the conda-forge project. The audit outcomes will be made public (ensuring no critical security information is exposed) through the conda-forge organisation in GitHub. The team will also develop supporting documentation where relevant.

2. Generate a public roadmap for infrastructure improvements and database migrations with clearly outlined risks and mitigation plans.

3. Develop initial mockups of the conda-forge and Quetz maintainers dashboard.

4. Deploy a minimal public Quetz instance for the conda-forge community and maintainers. This instance will also enable the team to gather user feedback and use this to drive further improvements or redesigns throughout the grant.

5. Work with the broader packaging community to complete the implementation of oras-py (ORAS - OCI Registry As Storage, and oras-py is a Python library to interact with an OCI registry).

6. Run a scalable mirror of conda-forge on GitHub packages.

At 12 months:

The focus is to migrate the existing data sources and work on low-hanging-fruit infrastructure improvements identified through the audits:

1. Migration of existing data sources into a purpose-built database, including a suitable data schema and

schema validation.

2. Refactor existing Quetz components or add new ones to account for the new data sources and schemas.

3. Extend the deployed Quetz dashboard with newly designed interfaces for maintainers and integrate new data sources.

4. Integrate the notion of an OCI registry as a storage system into Quetz.

5. Improvements to conda and mamba to allow installation from OCI registries (notably the community mirror on GitHub)

6. Enhancements to feedstocks' verification and validation workflows primarily focused on performance and reliability.

7. Improvements to the conda-forge infrastructure as detailed in the public roadmap focusing on critical performance and security risks.

At 18 months:

The team will focus on improving the workflow and security of the multiple bots maintained and used within conda-forge to maximise its reusability in other organisations such as bioconda and expand Quetz data representation capabilities. Identifies deliverables are:

1. Rewrite bots and eliminate long-lived credentials, improve performance and reliability, and develop end-user and maintainer's documentation.

2. Initial migration to Terraform for critical infrastructure parts.

3. Integrate feedstock data (GitHub repositories) in the Quetz maintainer's dashboard. The team will focus on the integration of repository health metrics and comprehensive information on dependencies and artefacts.

At 24 months:

The last phase of this work will focus on security, reliability and documentation efforts.

1. Further improvement and adoption of DevSecOps practices across the conda-forge workflows.

2. Complete Terraform adoption and infrastructure testing and validation.

3. Develop end-user and maintainer documentation as needed, including security best practices for supply chain security and Quetz dashboard use.

4. Integrate a timeline view for events in conda-forge (such as uploads) to the Quetz maintainers' dashboard.

5. Integrate sigstore (open-source standard for signing packages and containers) in the mamba package manager.

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 outside of the EOSS program (maximum of 250 words).

The conda-forge project has received a grant as part of EOSS 4 under the title "Fast Software Package Management for Bio and Data Science" to improve the mamba package manager. (CZI - Grant Number: 2021-237432)

Voltron, Quansight-Labs, NVIDIA, Quantstack, Anaconda, Integrated Ocean Observing System: in-kind contribution collaboration with employees

Microsoft Azure, GitHub actions, Open Source Labs, Mac Stadium, OVH Cloud, Oracle Cloud, Travis CI: in-kind infrastructure contribution

9. Landscape Analysis:

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 LOI; update if needed)

While there are many more package managers available, the alternatives are either platform-specific (e.g. apt, brew, choco), language-specific (e.g. pip for Python, gem for Ruby), or workflow-specific (e.g. Spack for HPC).

Most of the widely available package managers provide little flexibility for users to specify and install specific (and often multiple) versions of a given package. Instead, they adopt patterns such as:

1. The distribution manager defines the versions of the packages on each release cycle (i.e. Linux package managers like apt). This prevents the user from choosing specific package versions and often

reduces access to newer releases and features.

2. The package manager constantly applies rolling releases or updates to the latest versions, making it extremely challenging for the user to roll back to previous versions (i.e. Arch Linux, Brew).

These approaches result in countless hours worth of human effort lost by researchers and developers who require a particular software configuration to conduct their investigations or verify other research work's reproducibility.

The conda ecosystem is possibly the only one to offer the following guarantees, essential for scientific reproducibility:

1. Choosing which version of a package to install, potentially with different supporting libraries (e.g. Numpy with openblas or mkl)

2. Built-in virtual environments (coexistence of multiple installations)

3. Guaranteed access to previous installations

4. A vast catalogue of available packages across operating systems, architectures and languages

10. Value to Biomedical Users:

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

Conda-based packaging has empowered researchers in many fields, allowing them to reduce the time needed to set up and share their working environment. In most cases, conda packages avoid the need to compile from source or ask the IT department to provide a specific library version.

Conda-forge and bioconda are two primary examples of community-driven responses to satisfy the needs of domain-specific packages in the conda ecosystem.
Bioconda alone provides ~9 thousand packages for the life sciences, including bioinformatics, genomics, medical imaging and molecular simulation. It relies on conda-forge to provide its supporting dependencies and provides packages for 9 thousand projects ready to install across various operating systems and architectures.

Ensuring that both conda-forge and bioconda are sustainable in the long-term is of paramount interest to the whole biomedical community. The proposed work aims to help with conda-forge's most pressing issues threatening its sustainability and ability to meet its vast and diverse community requirements. This work will also allow conda-forge and bioconda to support the principle of a user's right to replicate, i.e. using open source software for infrastructure, adopting vendor-agnostic tooling and APIs and having in-depth technical documentation. Consequently, the projects (and the broader community) will be able to port the infrastructure to any cloud provider (thus avoiding vendor lock-in), replicate the infrastructure and tooling, and adopt a more decentralized approach to scientific packaging and distribution. Therefore, benefiting the open-source, open education and open research ecosystems.

11. Category:

Choose the two categories that best describe the software project(s) audience

Category 1	Data management and workflows
Category 2	Bioinformatics

12. Diversity, Equity, and Inclusion Statement:

Advancing DEI is a core value for CZI, 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)

The conda-forge project is committed to creating an inclusive, safe, and equitable environment for community members. At a structural level, the project has created an internal Diversity and inclusion team focused on improving the representation of historically marginalised groups across all levels within the project, from contributors to leadership positions. Additionally, the project has adopted the NUMFOCUS Code of Conduct to establish a baseline for acceptable and desired behaviour and has a public governance process.

Also, the project has introductory contribution and development guidelines to improve the project's newcomers' experience. The team highly values bringing in new community members and tries to minimize any overburdensome gatekeeping that discourages new contributors.

Although the conda-forge team excels at geographical diversity, it struggles with gender and other dimensions of diversity, as does much of the open-source software world. Additional efforts to introduce and support marginalized and historically underrepresented groups have included participation in programmes such as the Outreachy internship (2021 and 2022) and organizing regular community meetings. Additionally, there have been considerable efforts to onboard new contributors through direct mentorship and support from partner organizations such as Quansight Labs.

4. Optional Attachments

Completed - Jun 1 2022

Attachments should be uploaded in a combined single PDF. This section can include figures, charts and tables, references for the proposal, or any additional material in support of the proposal (maximum of three pages). Uploading any additional information is optional.

[conda-forge-infra](#)

Filename: conda-forge-infra.pdf **Size:** 434.5 kB

5. CV of Applicant

Completed - Jun 1 2022

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)

[CV WOLF JAIME](#)

Filename: CV_WOLF_JAIME.pdf **Size:** 5.1 MB

6. Budget Description

Completed - Jun 2 2022

Upload in PDF format; budgets can be uploaded in a combined single PDF or one PDF for each software project (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 Chan Zuckerberg Initiative will work closely with successful applicants to arrive at a mutually acceptable budget after review.

[conda-forge final budget (1)](#)

Filename: conda-forge_final_budget_1.pdf **Size:** 31.7 kB

7. Open Source Software Project(s) Details

Completed - Jun 2 2022

Open Source Software Project Details

Provide details and metrics for each open source software project that will be supported by the grant to help us assess its impact and quality.

How many software projects involved in your proposal (up to five)? If multiple software projects are involved, details must be entered for all of them where indicated below.

4

SOFTWARE PROJECT #1

SOFTWARE PROJECT #1: Details:

Complete the following table for Open Source Software Project #1 of your proposal. All URLs should be in the format https://example.com and only one primary link should be provided.

Software Project name (required)	conda-forge
Main code repository (e.g. GitHub URL) (required)	https://github.com/conda-forge
Homepage URL (required)	https://conda-forge.org
Social media handles (if applicable)	https://twitter.com/condaforge
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)	conda-forge is a community-led collection of recipes, build infrastructure and distributions for the conda package manager. It relies on several key components: - Github hosts the organization and its 16k repositories where conda recipes are created and maintained - A set of build tools orchestrate and updates the configurations needed to instruct CI providers like Azure Pipelines or Github Actions to turn those recipes into 18k conda packages and growing - Anaconda.org hosts the conda-forge channel, where the freshly built conda packages are uploaded, amounting to a total of 892k artefacts at the time of writing - A community of 3.8k maintainers devoting their free time as volunteers to keep everything running

SOFTWARE PROJECT #1: List of Key Personnel:

Key personnel are people involved in the software project who will be supported by the grant if the application is successful.

Complete the following for the key personnel on the open source software project #1 (up to 5) (required); **enter n/a if any field is not applicable**. Personnel to be hired that have not been identified at this time can be listed in the budget section. You may need to use the scroll bar at the bottom of the table to scroll right to view and to complete all fields. Alternatively, you can tab to move through and complete the fields. **To add another person/row (up to five), click the box at the end of the row.**

	First name	Last name	Email address	Current employer /Affiliation	Job title	Developer username if applicable (e.g., GitHub handle)	Country of Residence	Add another person/row
1	Wolf	Vollprecht	wolf.vollprecht@quantstack.net	QuantStack	CTO	wolfv	Germany	✔
2	Jaime	Rodríguez-Guerra	jrodriguez@quansight.com	Quansight Labs	Software Engineer	jaimergp	Germany	✖

SOFTWARE PROJECT #1: Metrics- Quality (required):

Complete for the open source software project #1.

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?

Other (please specify): YAML, Bash, CMD, Python

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

https://numfocus.org/code-of-conduct

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

https://conda-forge.org/docs/user/00_intro.html

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

https://github.com/conda-forge/conda-forge.github.io/issues

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

https://gitter.im/conda-forge/conda-forge.github.io

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

https://conda-forge.org/index.html#contribute

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

No

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

A mixture of GitHub actions, Azure Pipelines and Circle CI (maybe others) See dashboard at
https://conda-forge.org/status/#azure for the usage of all services

SOFTWARE PROJECT #1: Metrics- Impact (optional):

Complete the following for the open source software project #1. **Providing metrics is optional and metrics can be approximate.** *For each metric, please provide a source, clarify how the metric was computed, and/or provide any other comments. For monthly metrics, please provide data from the most recent month for which the corresponding metric is available.*

1. SOFTWARE PROJECT #1: Complete the following table. List the number and explanation for each, if needed:

	Number	Comment
Scholarly paper(s) (including preprints) citing or mentioning the software project	350	Conda-forge has a Zenodo DOI: 10.5281/zenodo.4774216, but it has very few indexed citations. If you search Google Scholar for "conda-forge", you get 509 results. Dimensions.ai gives 359 publications. This is a lower bound.
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.)	336000000	The conda-forge channel serves ~336M downloads each month
Software projects that depend on the project	18000	18,000 projects publish their software on conda-forge See https://conda-forge.org/#about
Monthly visitors to project's website, discussion forum (e.g. Stack Overflow)	2000	conda-forge/conda-forge.github.io has around 2000 unique visitors a month (based on biweekly stats from Github repo insights). The Gitter room has received ~1,066 messages a month during 2022 (taken from the Gitter archives)

2. SOFTWARE PROJECT #1: List of software projects to which key personnel on this proposal are contributing.

Conda-forge - Jaime and Wolf

Conda-smithy - Jaime and Wolf

Quetz - Wolf

Bioconda -Wolf

SOFTWARE PROJECT #2

Complete the following table for Open Source Software Project #2 of your proposal. All URLs should be in the format *https://example.com* and only one primary link should be provided.

Software Project name (required)	bioconda
Main code repository (e.g. GitHub URL) (required)	https://github.com/bioconda
Homepage URL (required)	https://bioconda.github.io/
Social media handles (if applicable)	(No response)
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)	Bioconda is a channel for the conda package manager specializing in bioinformatics software. Bioconda consists of: - a repository of recipes hosted on GitHub - a build system turning these recipes into conda packages - a repository of packages containing over 7000 bioinformatics packages ready to use with conda install over 850 contributors and 570 members who add, modify, update and maintain the recipes The conda package manager makes installing software a vastly more streamlined process. Conda is a combination of other package managers you may have encountered, such as pip, CPAN, CRAN, Bioconductor, apt-get, and homebrew. Conda is both language- and OS-agnostic, and can be used to install C/C++, Fortran, Go, R, Python, Java etc programs on Linux, Mac OSX, and Windows.

SOFTWARE PROJECT #2: List of Key Personnel:

Key personnel are people involved in the software project who will be supported by the grant if the application is successful.

Complete the following for the key personnel on the open source software project #2 (up to 5) (required); **enter n/a if any field is not applicable***. Personnel to be hired that have not been identified at this time can be listed in the budget section. You may need to use the scroll bar at the bottom of the table to scroll right to view and to complete all fields. Alternatively, you can tab to move through and complete the fields.* **To add another person/row (up to five), click the box at the end of the row.**

	First name	Last name	Email address	Current employer /Affiliation	Job title	Developer username if applicable (e.g., GitHub handle)	Country of Residence	Add another person/row
1	Jaime	Rodríguez-Guerra	jrodriguez@quansight.com	Quansight Labs	Software Engineer	jaimergp	Germany	✔
2	Wolf	Vollprecht	wolf.vollprecht@quantstack.net	QuantStack	CTO	wolfv	Germany	✖

SOFTWARE PROJECT #2: Metrics- Quality (required):

Complete for the open source software project #2.

1. SOFTWARE PROJECT #2 : What is the software project license?

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

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

Other (please specify): YAML, Bash, CMD, Python

3. SOFTWARE PROJECT #2: Does the software project have a code of conduct?

https://github.com/bioconda/bioconda-recipes/blob/master/.github/CODE_OF_CONDUCT.md

4. SOFTWARE PROJECT #2: Does the software project have end-user documentation?

https://bioconda.github.io/user/index.html

5. SOFTWARE PROJECT #2 : Does the software project have an issue tracker?

https://github.com/bioconda/bioconda-utils/issues?q=is%3Aissue+is%3Aopen+sort%3Aupdated-desc

6. SOFTWARE PROJECT #2: Does the software project have a community engagement / Q&A forum (self-hosted, on Stack Exchange etc.)?

https://gitter.im/bioconda/Lobby

7. SOFTWARE PROJECT #2: Does the software project have contribution / coding guidelines?

https://bioconda.github.io/contributor/index.html

8. SOFTWARE PROJECT #2: Is there a corresponding package available in a package manager (PyPi, CRAN, etc.)?

No

9. SOFTWARE PROJECT #2: Does the software project support continuous integration for testing?

https://github.com/bioconda/bioconda-actions

SOFTWARE PROJECT #2: Metrics- Impact (optional):

Complete the following for the open source software project #2. **Providing metrics is optional and metrics can be approximate.** *For each metric, please provide a source, clarify how the metric was computed, and/or provide any other comments. For monthly metrics, please provide data from the most recent month for which the corresponding metric is available.*

1. SOFTWARE PROJECT #2 : Complete the following table. List the number and explanation for each, if needed:

	Number	Comment
Scholarly paper(s) (including preprints) citing or mentioning the software project	513	Nature Methods report 330 citations, while Google Scholar claims 513.
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.)	4300000	4.3 M downloads monthly. Calculated from https://github.com/bioconda/bioconda-stats/blob/main/package-downloads/anaconda.org/bioconda.json
Software projects that depend on the project	9254	9254 projects publish their software through bioconda Taken from https://anaconda.org/bioconda
Monthly visitors to project's website, discussion forum (e.g. Stack Overflow)	268	286 messages monthly According to their Gitter archives for 2022

2. SOFTWARE PROJECT #2: List of software projects to which key personnel on this proposal are contributing.

(No response)

SOFTWARE PROJECT #3

SOFTWARE PROJECT #3: Details:

Complete the following table for Open Source Software Project #3 of your proposal. All URLs should be in the format https://example.com and only one primary link should be provided.

Software Project name (required)	conda-smithy
Main code repository (e.g. GitHub URL) (required)	https://github.com/conda-forge/conda-smithy
Homepage URL (required)	https://github.com/conda-forge/conda-smithy
Social media handles (if applicable)	(No response)
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)	conda-smithy is a tool for combining a conda recipe with configurations to build using freely hosted CI services into a single repository, also known as a feedstock. conda-smithy is still a work-in-progress, but when complete, conda-smithy will: Create a git repo with a conda recipe and the files to run conda builds via CI services. Register the repo on GitHub and push it. Connect the repo to the CI services travis-ci.com, appveyor.com, circleci.com, dev.azure.com

SOFTWARE PROJECT #3: List of Key Personnel:

Key personnel are people involved in the software project who will be supported by the grant if the application is successful.

Complete the following for the key personnel on the open source software project #3 (up to 5) (required); **enter n/a if any field is not applicable**. *Personnel to be hired that have not been identified at this time can be listed in the budget section. You may need to use the scroll bar at the bottom of the table to scroll right to view and to complete all fields. Alternatively, you can tab to move through and complete the fields.* **To add another person/row (up to five), click the box at the end of the row.**

	First name	Last name	Email address	Current employer /Affiliation	Job title	Developer username if applicable (e.g., GitHub handle)	Country of Residence	Add another person/row
1	Jaime	Rodríguez-Guerra	jrodriguez@quansight.com	Quansight Labs	Software Engineer	jaimergp	Germany	✔
2	Wolf	Vollprecht	wolf.vollprecht@quantstack.net	Quantstack	CTO	wolfv	Germany	✖

SOFTWARE PROJECT #3: Metrics- Quality (required):

Complete for the open source software project #3.

1. SOFTWARE PROJECT #3 : What is the software project license?

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

2. SOFTWARE PROJECT #3 : What is the main programming language?

Python

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

https://conda-forge.org/docs/orga/governance.html#code-of-conduct

4. SOFTWARE PROJECT #3: Does the software project have end-user documentation?

https://conda-forge.org/docs/user/00_intro.html

5. SOFTWARE PROJECT #3 : Does the software project have an issue tracker?

https://github.com/conda-forge/conda-smithy/issues

6. SOFTWARE PROJECT #3: Does the software project have a community engagement / Q&A forum (self-hosted, on Stack Exchange etc.)?

No

7. SOFTWARE PROJECT #3: Does the software project have contribution / coding guidelines?

https://conda-forge.org/index.html#contribute

8. SOFTWARE PROJECT #3: Is there a corresponding package available in a package manager (PyPi, CRAN, etc.)?

https://anaconda.org/conda-forge/conda-smithy

9. SOFTWARE PROJECT #3: Does the software project support continuous integration for testing?

https://github.com/conda-forge/conda-smithy/actions

SOFTWARE PROJECT #3: Metrics- Impact (optional):

Complete the following for the open source software project #3. **Providing metrics is optional and metrics can be approximate.** *For each metric, please provide a source, clarify how the metric was computed, and/or provide any other comments. For monthly metrics, please provide data from the most recent month for which the corresponding metric is available.*

1. SOFTWARE PROJECT #3 : Complete the following table. List the number and explanation for each, if needed:

	Number	Comment
Scholarly paper(s) (including preprints) citing or mentioning the software project	0	Conda-smithy is an internal tool key to keep conda-forge operational. End users not involved in maintaining the infrastructure will not need to use it, hence why we expect few to no mentions in research articles.
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.)	25000	20-25K monthly downloads Obtained with `condastats overall --monthly --data_source conda-forge conda-smithy`
Software projects that depend on the project		The entire conda-forge ecosystem depends on conda-smithy
Monthly visitors to project's website, discussion forum (e.g. Stack Overflow)		~300 unique visitors a month in the repository

2. SOFTWARE PROJECT #3: List of software projects to which key personnel on this proposal are contributing.

(No response)

SOFTWARE PROJECT #4

SOFTWARE PROJECT #4: Details:

Complete the following table for Open Source Software Project #4 of your proposal. All URLs should be in the format https://example.com and only one primary link should be provided.

Software Project name (required)	quetz
Main code repository (e.g. GitHub URL) (required)	https://github.com/mamba-org/quetz
Homepage URL (required)	https://github.com/mamba-org/quetz
Social media handles (if applicable)	(No response)
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)	The quetz project is an open source server for conda packages. It is built upon FastAPI with an API-first approach. A quetz server can have many users, channels and packages. With quetz, fine-grained permissions on channel and package-name level will be possible. Quetz has an optional client (quetz-client) that can be used to upload packages to a quetz server instance.

SOFTWARE PROJECT #4: List of Key Personnel:

Key personnel are people involved in the software project who will be supported by the grant if the application is successful.

Complete the following for the key personnel on the open source software project #4 (up to 5) (required); **enter n/a if any field is not applicable**. Personnel to be hired that have not been identified at this time can be listed in the budget section. You may need to use the scroll bar at the bottom of the table to scroll right to view and to complete all fields. Alternatively, you can tab to move through and complete the fields. **To add another person/row (up to five), click the box at the end of the row.**

	First name	Last name	Email address	Current employer /Affiliation	Job title	Developer username if applicable (e.g., GitHub handle)	Country of Residence	Add another person/row
1	Wolf	Vollprecht	wolf.vollprecht@quantstack.net	QuantStack	CTO	wolfv	Germany	✖

SOFTWARE PROJECT #4: Metrics- Quality (required):

Complete for the open source software project #4.

1. SOFTWARE PROJECT #4 : What is the software project license?

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

2. SOFTWARE PROJECT #4 : What is the main programming language?

Other (please specify): Python, JavaScript, HTML

3. SOFTWARE PROJECT #4: Does the software project have a code of conduct?

https://numfocus.org/code-of-conduct

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

https://quetz.readthedocs.org/

5. SOFTWARE PROJECT #4 : Does the software project have an issue tracker?

https://github.com/mamba-org/quetz/issues?q=is%3Aissue+is%3Aopen+sort%3Aupdated-desc

6. SOFTWARE PROJECT #4: Does the software project have a community engagement / Q&A forum (self-hosted, on Stack Exchange etc.)?

https://gitter.im/mamba-org/Lobby?source=orgpage

7. SOFTWARE PROJECT #4: Does the software project have contribution / coding guidelines?

https://github.com/mamba-org/quetz/blob/main/CONTRIBUTING.md

8. SOFTWARE PROJECT #4: Is there a corresponding package available in a package manager (PyPi, CRAN, etc.)?

https://anaconda.org/conda-forge/quetz

9. SOFTWARE PROJECT #4: Does the software project support continuous integration for testing?

https://github.com/mamba-org/quetz/actions

SOFTWARE PROJECT #4: Metrics- Impact (optional):

Complete the following for the open source software project #4. **Providing metrics is optional and metrics can be approximate.** *For each metric, please provide a source, clarify how the metric was computed, and/or provide any other comments. For monthly metrics, please provide data from the most recent month for which the corresponding metric is available.*

1. SOFTWARE PROJECT #4 : Complete the following table. List the number and explanation for each, if needed:

	Number	Comment
Scholarly paper(s) (including preprints) citing or mentioning the software project		
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.)	300	Obtained with `condastats overall --monthly --data_source conda-forge quetz `
Software projects that depend on the project		
Monthly visitors to project's website, discussion forum (e.g. Stack Overflow)		

2. SOFTWARE PROJECT #4: List of software projects to which key personnel on this proposal are contributing.

(No response)

8. Equal Opportunity & Diversity

Completed - Jun 2 2022

Equal Opportunity & Diversity

CZI Science supports the science and technology that will make it possible to cure, prevent, or manage all diseases by the end of this century. Everyone is affected by disease, yet different communities are affected by or experience disease in different ways. Moreover, due to systemic barriers, the scientific enterprise itself is not a place where all voices and talents thrive. We believe the strongest scientific teams — encompassing ourselves, our grantees, and our partners — incorporate a wide range of backgrounds, lived experiences, and perspectives that guide them to the most important unsolved problems. To enable our