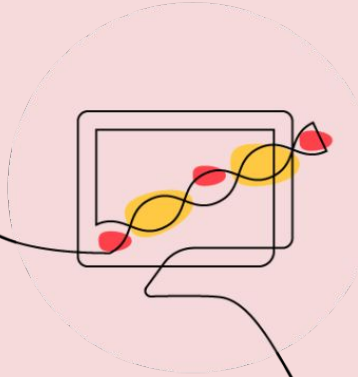
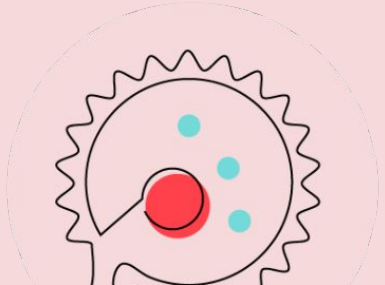
Plenary

# Five years of supporting open source software: What have we learned?

Carly Strasser & Kate Hertweck

CZI Open Science 2024 Meeting





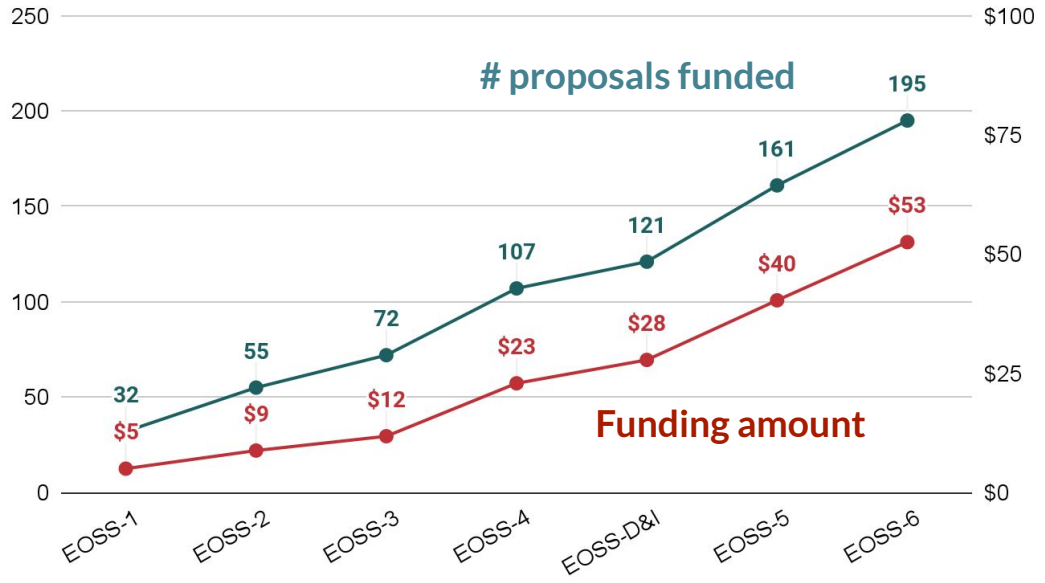
# Essential Open Source Software for Science

The majority of open source software for science  is undervalued and lacks funding for maintenance, growth, development, and community engagement—especially after the initial phase when it's linked to original research.

The Essential Open Source Software for Science program (EOSS) supports maintenance, growth, development, and community engagement for critical open source tools.

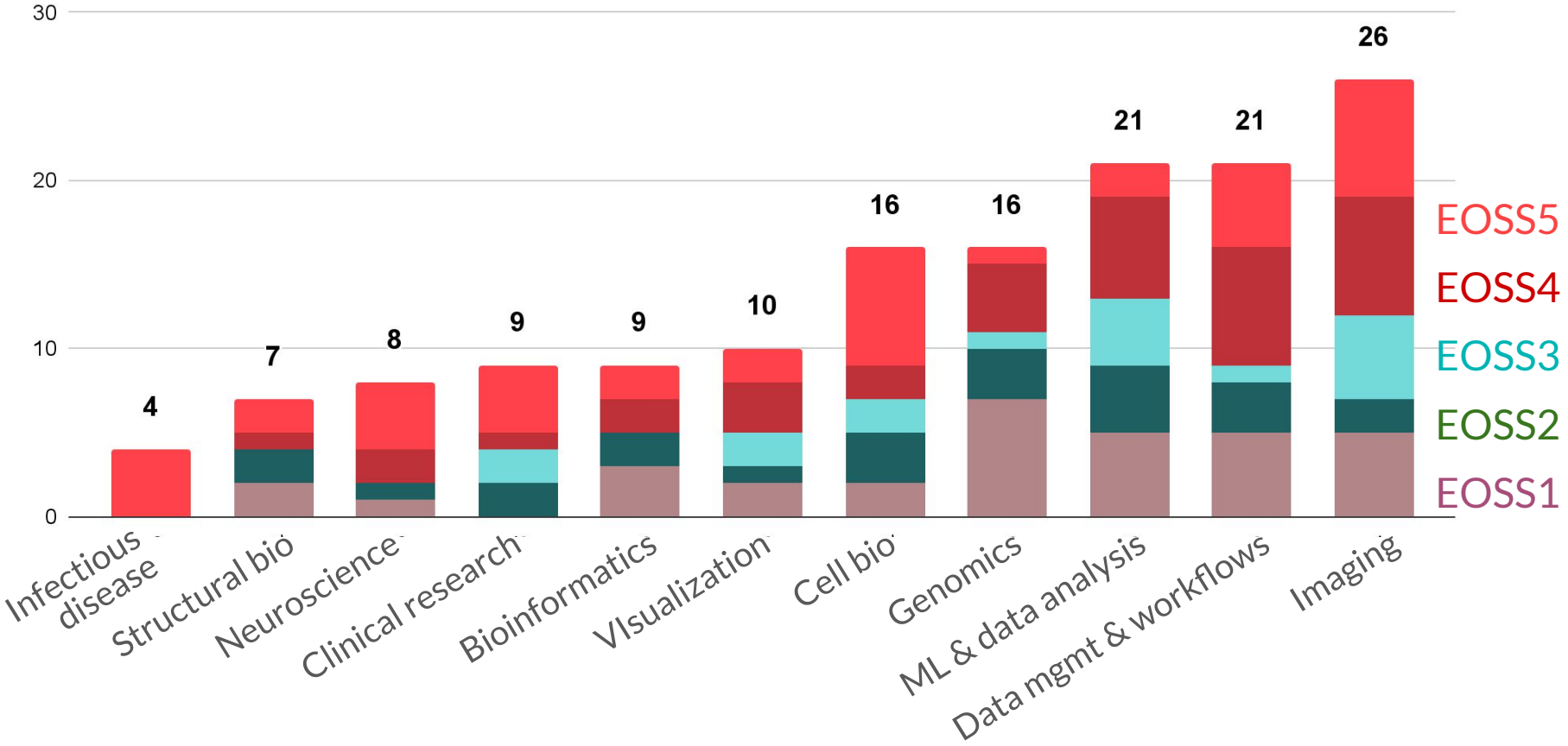
# EOSS program details

- Started in 2019; 6 cycles of funding
- Project budgets: $100K - $400K total for 2 years
- 195 proposals funded
- 233 projects supported
- $53M invested




czi.co/EOSS

# Number of proposals funded by category

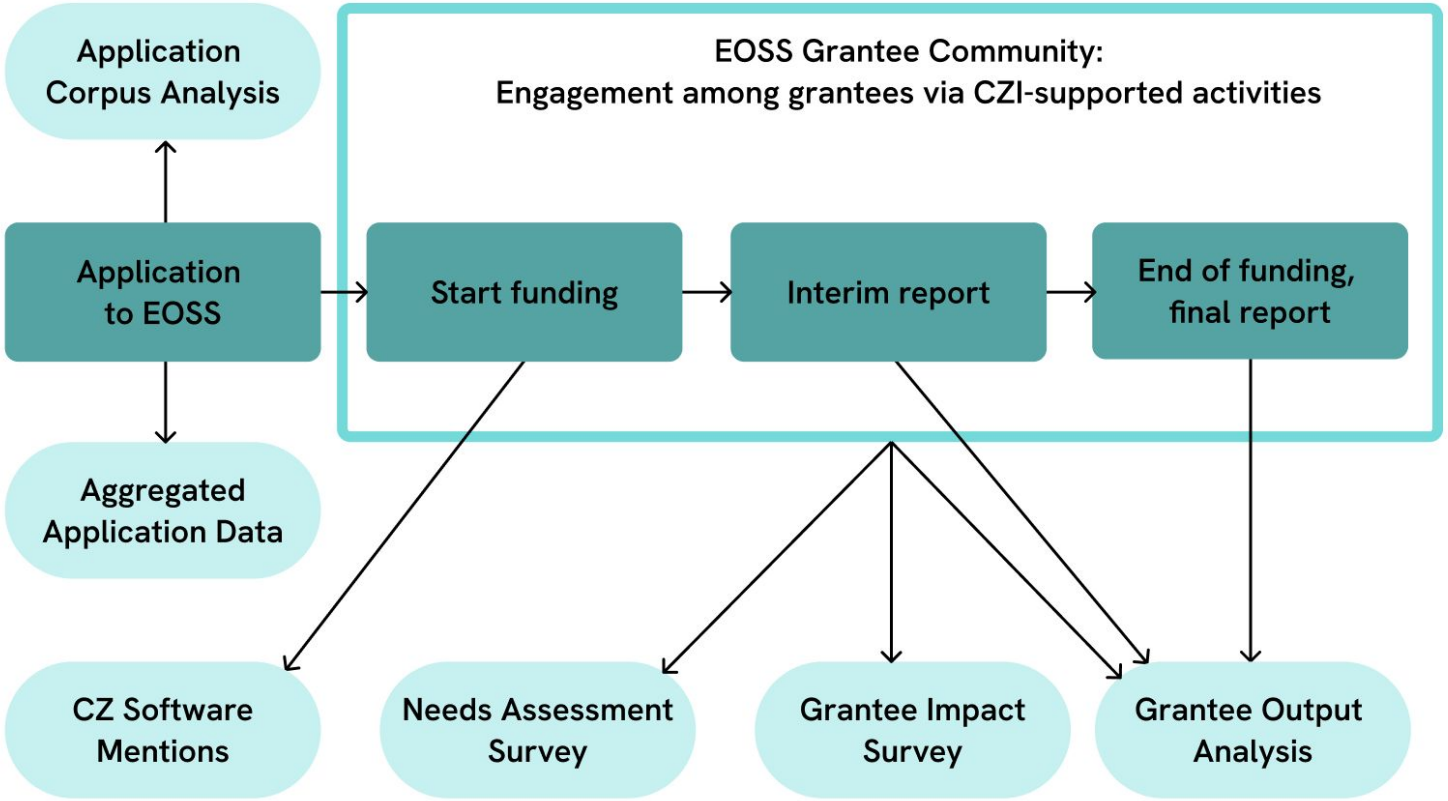


3D Slicer · alevin-fry · AmadeusGPT · AnnData · AnnDataR · Apache Arrow · Apollo · Array API Standard · ArviZ · Bactopia · bcbio-nextgen · BEDTOOLS · BigDataViewer · Binder · Bio-Formats · Bioconda · Bioconductor · biocViews · BioCypher · BioImage Model Zoo · Bokeh · BrainGlobe Atlas · brainreg · brainrender · BWA · Cardinal · Catalyst.jl · CCTBX · cellfinder · cgap-pipeline · cgap-portal · ChimeraX · CINeMA · CLIJ · cogent3 · conda-forge · conda-smithy · Cornerstone · CuPy · cuttlefish · CWL · Cytoscape · D4TOOLS · Dash · Dask · Datoviz · DeepBacs · DeepLabCut · demuxlet · DESeq2 · Deviser · DifferentialEquations.jl · DIPY · DL4MicEverywhere · dMRIPrep · doFuture · dynamo · dynast · dynverse · EDAM · eddymotion · edgeR · elastix · ETE Toolkit · FastSurfer · FEniCS · Fiji · fMRIPrep · freemuxlet · FreeSurfer · Funsor · future · Galata · Galaxy · GATK · genome-sampler · GGD · Giotto · Glimma · granite · GraphBin · GSVA · Hifiasm · HiGlass · Holoviz · HTSJDK · ICY · igraph · ilastik · ImageJ · ImgLib2 · Insight Toolkit · Integrative Modeling Platform · IPython · IQ-TREE · ISMAGS-in-Python · ITK-SNAP · ITK-WASM · ITKElastix · JBrowse · JupyterHub · JupyterLab · kallisto · KNIME · LIANA · libSBML · limma · LinkML · MACS · MACS3 · mamba · Matplotlib · MDAnalysis · memento · MetaCoAG · MetaInsight · MicrobiomeDB · MicroManager · Minimap2 · mixtools · MNE-BIDS · MNE-Python · MoBIE · ModelingToolkit.jl · MONAILabel · Monocle · moose · MRIQC · MSstats · MultiNicheNet · N5 · NanoJ · NanoPyx · napari · NetworkX · Neurodesk · neuroml · Nextflow · nf-core · NGFF · Nibabel · NicheNet · Nilearn · NiReports · nsdf · NumPy · NumPyro · ODK · OHDSI · OHIF Viewer · OMOP Common Data Model · Open DBDP · OHIF Viewer · OmniPath · Open Microscopy Environment · OpenBLAS · OpenCRAVAT · OpenFE · OpenFF · OpenFold · OpenHealthStack · OpenMM · OpenMRS · OpenMS · OpenPolscope · OpenPose · OpenRefine · OpenSim · OpenSPIM · Orange Data Mining · Pandas · Parsl · Percolator · PhasorPy · phy · pip · plyranges · progressr · Protégé · PsychoJS · PsychoPy · pufferfish · PyData Sphinx · Pygfx · PyMC · Pyro · qFit · QIIME2 · QUAST · quetz · QuPath · Read the Docs · ROBOT · Rocker · salmon · Scanpy · scikit-image · scikit-learn · SciML · SciPy · scvi-tools · seqproc · seqr · Seurat · Simbody · Slicer · SlicerAIGT · SlicerIGT · SMPLify-X · snakemake · Snippy · SPAdes · Spateo · SpatialData · Spyder · Stan · STAR · SymPy · Tibanna · tidybulk · tidyseurat · tidySingleCellExperiment · tolerance · TotalSegmentator · UCSC Xena · UpSet · VisPy · Vitessce · VTK · WESTPA · Wildmeshing · Xarray · Zarr · zellkonverter · ZeroCostDL4Mic

# Motivation

- Biggest investment in scientific OSS
- How do we measure the impact of the investment?
- What should we be asking grantees to understand the impact of their work?
- What are the patterns?

# Data sources



# Questions we addressed

What does the scientific open source community need?

What activities did the program fund?

How did the program impact funded software projects?

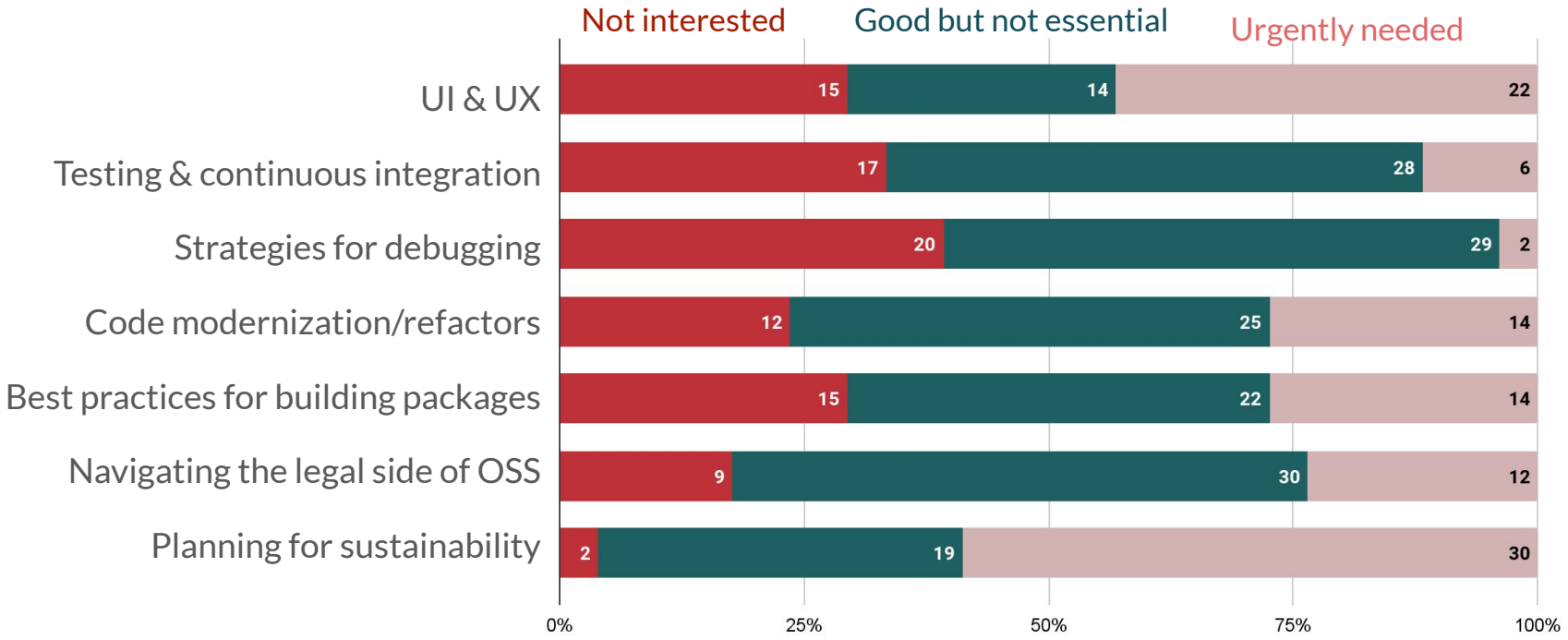
How did the program impact the open source community?

How did the program impact diverse participation in scientific open source?

How did EOSS-funded projects impact biomedical research?

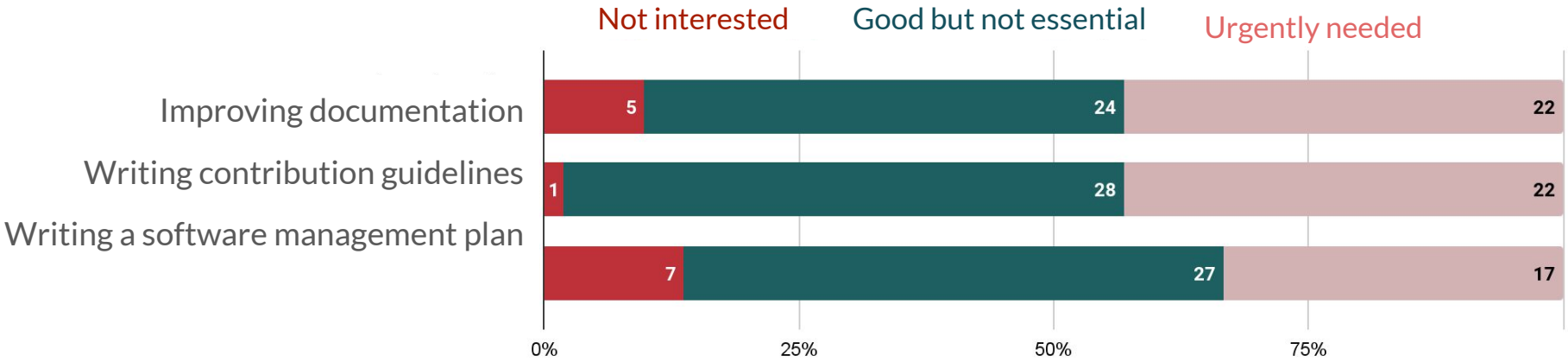
# What do scientific open source projects need?

## Gaps in **technical** capacity



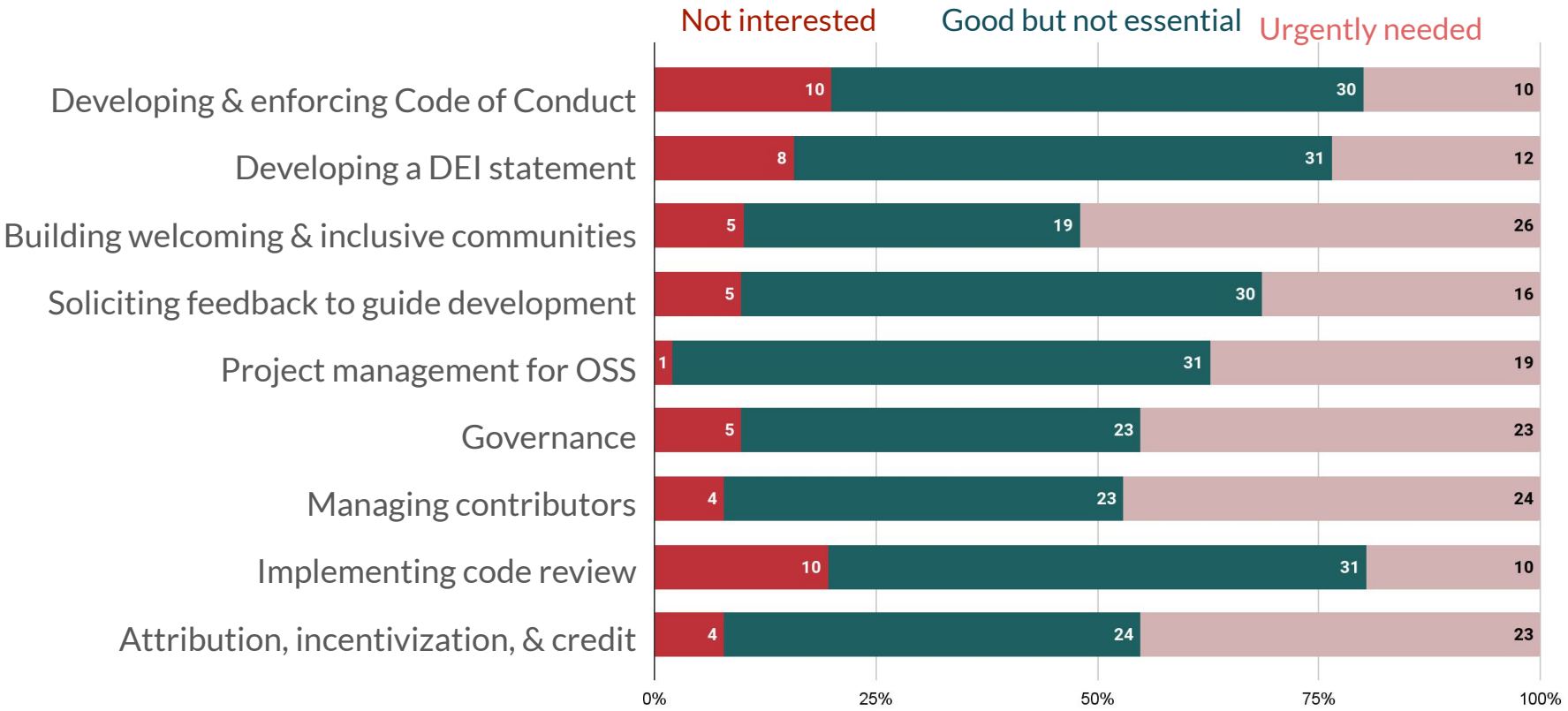
# What do scientific open source projects need?

## Gaps in **documentation** capacity



# What do scientific open source projects need?

## Gaps in **community-related** capacity

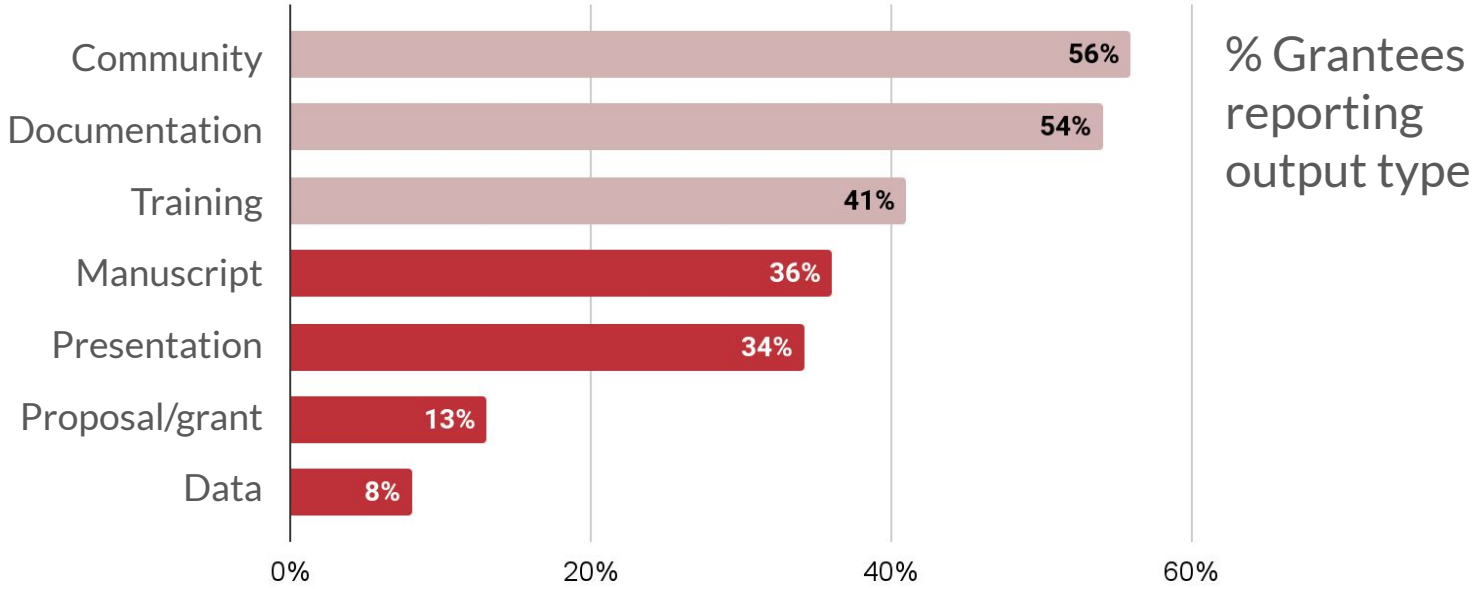


# What activities did EOSS grantees undertake? …or what outputs were produced?

| Output category | Activity |
|---|---|
| Technical | Updates to code/software |
| Community | Materials or events used to engage users or developers |
| Documentation | Any written material used as reference for users and/or developers |
| Training | Online or virtual course/workshop, mentorship |
| Manuscript | Preprint, submitted, or published manuscript |
| Presentation | Seminar or conference presentation (talk or poster) |
| Proposal/grant | Submitted proposal and/or awarded grant |
| Data | Curated dataset (e.g., for benchmark or testing) |

# What activities did EOSS grantees undertake?

## Non-technical activities & outputs reported by grantees



**But what does this mean?**

**Why does this work matter?**

# Impact: EOSS-funded projects

95% of survey respondents – EOSS extremely or very important to the success of their open source project.

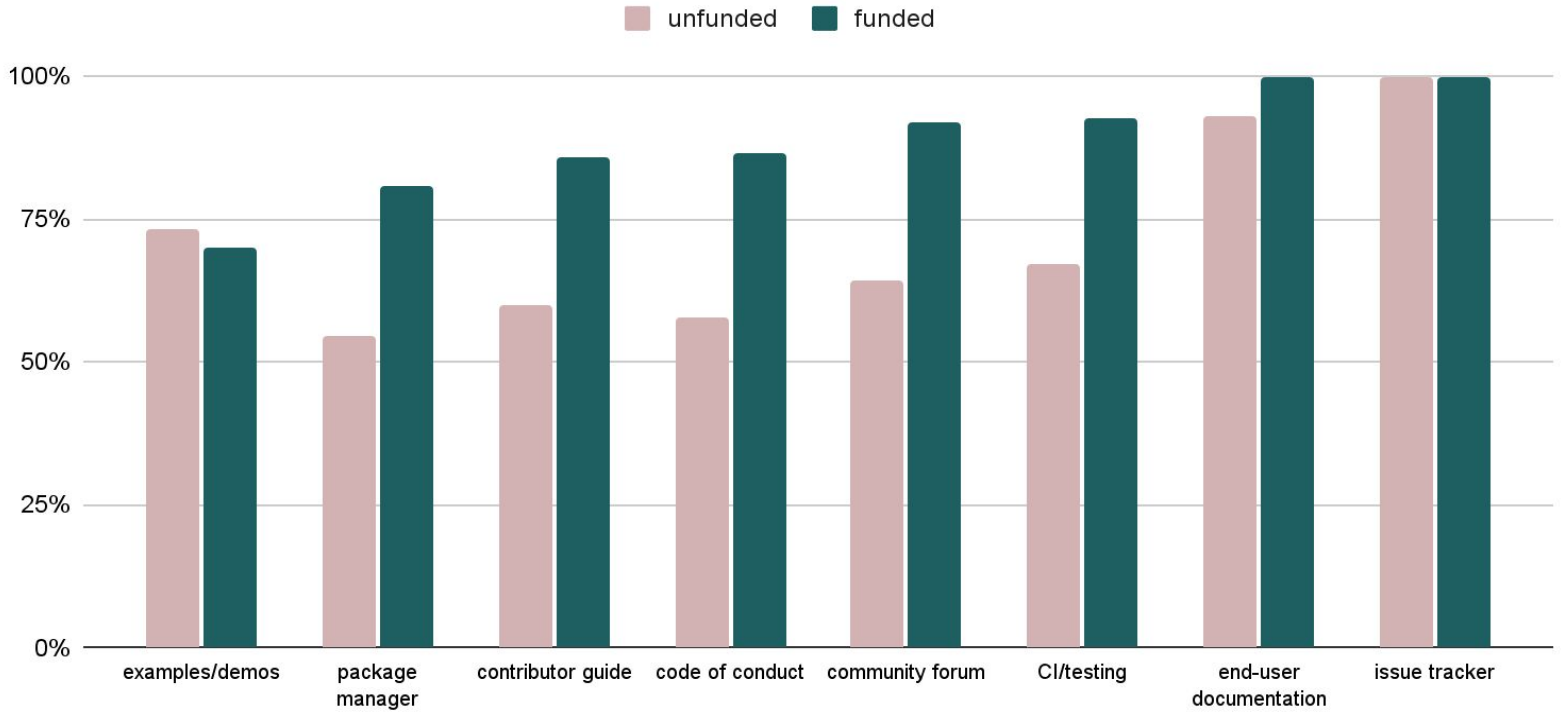
PsychoPy was able to professionalize tools and bridge a gap until their revenue model was fully developed and sustainable.

85% of survey respondents – EOSS extremely or very important for supporting the primary maintainers of the project.

NetworkX hired a community manager and focused on creating a developer community, the difference for maintainers is “really striking”.

# Impact: Open source community

EOSS projects demonstrate robust software development practices



# Impact: Diversity in OSS

59% of survey respondents – EOSS funding very or extremely important for supporting efforts to improve diversity and inclusion in their funded software projects.

56% of projects – activities to enhance diversity, equity and inclusion (DEI) in their projects.

# Impact: Diversity in OSS participation

## Activities to increase diverse participation

37% of projects – contributors from underrepresented groups

The Galaxy Project paid maintainers in areas for which other funding sources can't be used (e.g., Sri Lanka, Australia)

7% of projects – translation to additional human languages

nf-core recorded videos for foundational training in 5 languages.

25% of projects – outreach to new demographics and groups

Bokeh collaborated with Data Umbrella and engaged in virtual sprints at PyData Global and SciPy

# Impact: Biomedical research

Supporting the science and technology that will make it possible to cure, prevent, or manage all disease by the end of the century.

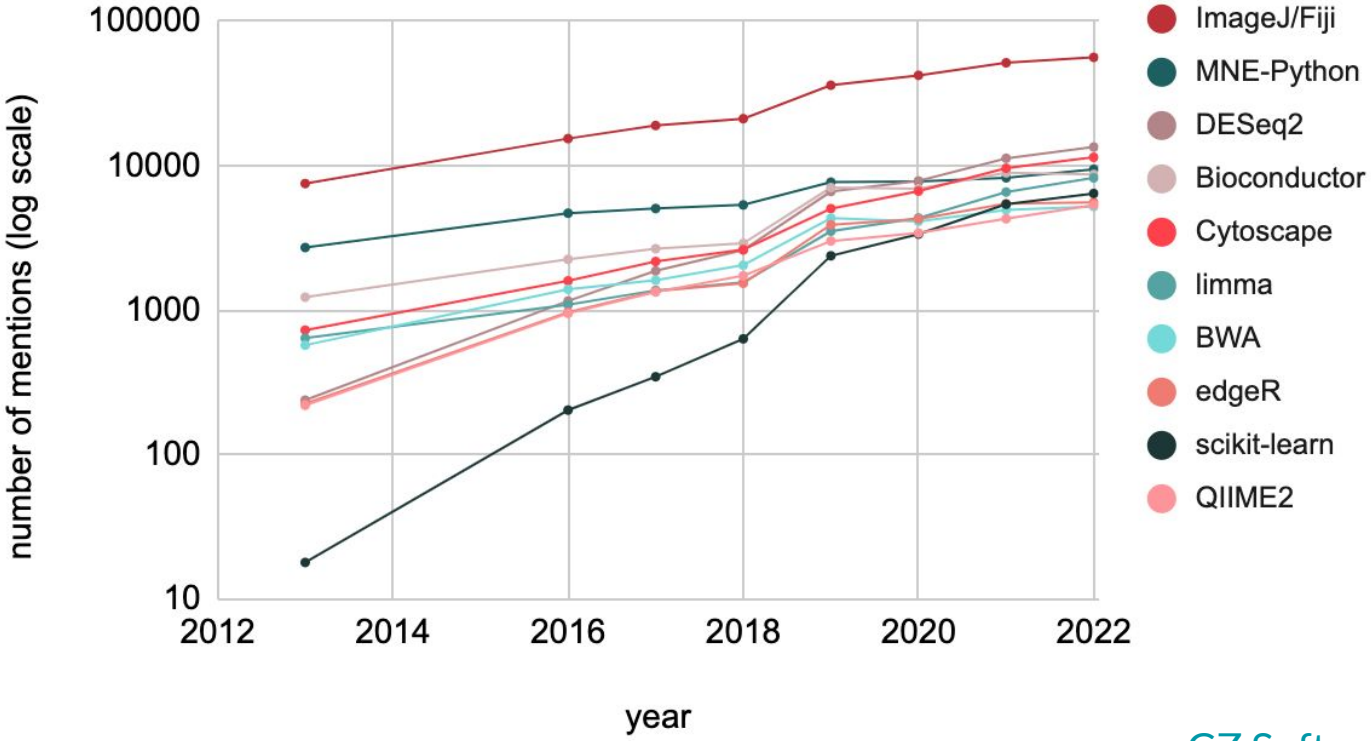
74% of survey respondents – EOSS funding very or extremely important to the uptake and use of their project in the biomedical research community.

Increase in general usability and use of software with wide user base

Specific use or expansion into biomedical research

# Impact: Biomedical research

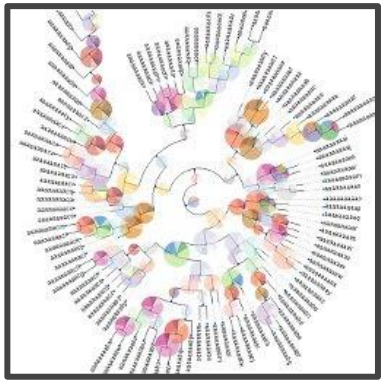
Top ten EOSS software projects by total mentions




CZ Software Mentions

# Impact: Biomedical research

**Expansion of features to support biomedical use cases**

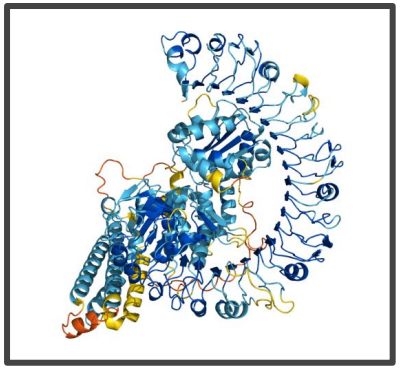


ETE Toolkit - features to support human microbiome studies

**Inclusion in platforms for biomedical researchers**

seqr - inclusion in NIH-NHGRI AnVIL cloud platform

**High-impact biomedical research outcomes**



SciPy, NumPy, scikit-learn, Matplotlib, pandas cited by AlphaFold (2021)

# What to expect: full report

**PDF + data available on Zenodo in July 2024**

## Activities

What support does the scientific open source community need?

What activities did the program fund?

## Impact

Funded software projects

Scientific open source community

Diverse participation in scientific open source

Biomedical research

# Parting thoughts

**Our ability to fund you is only as strong as the stories you share – tell your story and share it with us!**

There is no single data source for understanding the importance of scientific OSS and investments to support it–what data are relevant for your work?

The impact of scientific open source software occurs at many levels and can be viewed through multiple lenses.

**THANK YOU to those who answered our surveys and shared your stories.**

Have comments or concerns about the EOSS impact assessment? Use our form to share anonymous feedback: https://forms.gle/g8hjDeDh7J7dADZa9