The impact of Ceph on open source at UC Santa Cruz and beyond

Carlos Maltzahn, 7/9/24
HotStorage Keynote

Who am I?

1991 Diplom Informatik at Passau University
Advisor Matthias Jarke

1992 Joined Clarence “Skip” Ellis’ research group at CU Boulder
Workflow systems / groupware

1997 Switched to web proxy server design
Advisors Kathy Richardson, Dirk Grunwald
Funded by DEC

1999 PhD in web proxy performance management

1999 Performance engineer at Netapp
Switched to storage systems

2005 Joined UC Santa Cruz as Research Scientist

2005 Co-founded Ceph project

2008 Became adjunct faculty

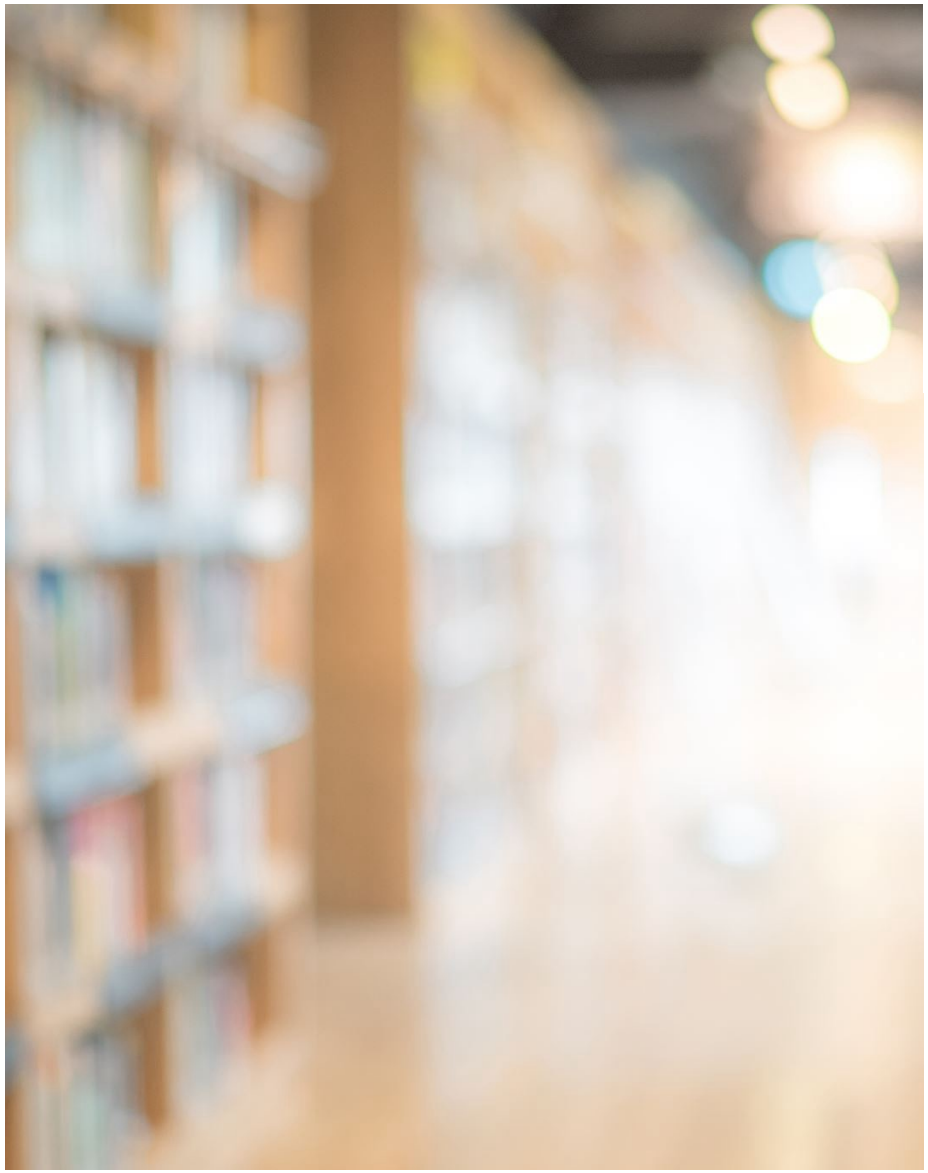
Graduated 10 PhD, 12 MS students,

140 publications

Founded & led CROSS and OSPO UC Santa Cruz

Raised millions from government, industry, and non-profits

Recently retired from UC Santa Cruz



Goal & Disclaimers

- Goal
  - Encourage exploration of how to amplify research impact
  - Potential of open source
  - Institutional support
- Disclaimers
  - My perspective
  - Anecdotal – this is not a research talk

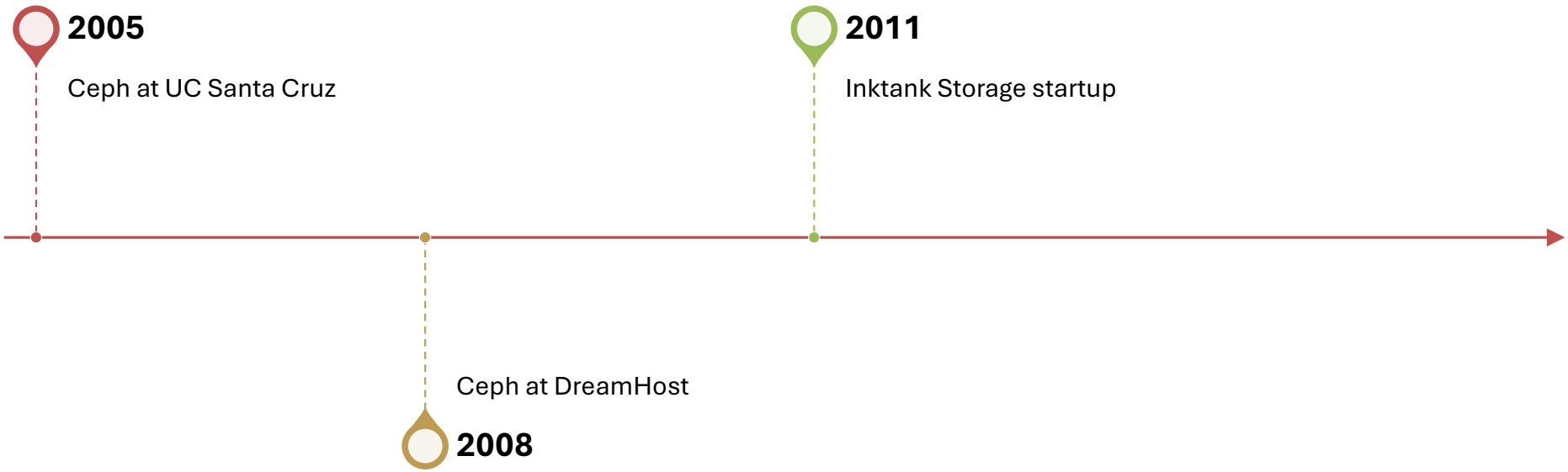
Outline

History of Ceph

Impact of Ceph on open source at UC Santa Cruz

... and beyond

History of Ceph



2005 Ceph at UC Santa Cruz: the plan

Summer project to build working prototype

- Integrate OBSD work at UC Santa Cruz
- Submit journal publication in fall
- Started out as team effort with PhD students working with Scott Brandt, Darrell Long, Ethan Miller
- One of the students was Sage Weil (with Scott)

Scott Brandt,
Associate VP for
Research & Innovation,
UCOP

What actually happened (1)

Sage Weil quickly built “Ceph”, working prototype file system

- First time running: 5/3/2005 12:39AM

Revealed a number of uninvestigated design gaps

Trilabs demo on 5/25/2005 resulted in funding

- Gary Grider (LANL), Bill Loewe (LLNL), and Lee Ward (Sandia)
- Funded by Advanced Simulation and Computing Program (ASC)

Sage chose Ceph as his PhD topic

Sage Weil,
CEO Civic Media

What actually happened (2)

2006: Published in OSDI ‘06 and SC ’06

- Unified redundancy, failure management, and cluster expansion/contraction
- Pushed into OSD layer
- CRUSH: Calculated placement across failure domains
- Ceph code available under LGPL

2007: Sage graduates with PhD

Sage Weil,
CEO Civic Media

What didn’t happen

Journal paper planned for 2005

2008 Ceph at DreamHost: the plan

Sage co-founded DreamHost in 1996

Storage was web hosting key cost center

Replace expensive, proprietary storage with Ceph

Turn Ceph into production-ready storage system!

- Small team at DreamHost dedicated to Ceph

Sage was able to continue his work

- Vast majority of research products are abandoned after graduation

What worked

2010 Ceph Linux kernel client included in 2.6.34 (blog post)

- Required advocacy from early fans well-known within kernel community
- Significant increase in visibility, credibility, and convenience
- Upstreamed client still maintained by DreamHost's Ceph team

2010 "Ceph RADOS block device" (rbd) (blog post)

- Allowed Ceph to become dominant storage system in OpenStack
- OpenStack significantly increased Ceph's community
- Initial implementation by Christian Brunner (outside DreamHost)

openstack

What didn’t work

DreamHost is web hosting, not storage

- Difficult for government and industry to justify R&D funding
- Difficult to attract storage talent

2011 Inktank Storage startup: the plan

DreamHost spin-off

- Founded by Sage Weil (CTO) and Bryan Bogensberger (CEO)
- Bryan was VP Business Strategy at DreamHost

Products:

- Proprietary enterprise grade Ceph storage management console
- Professional service and training

What worked

Attracted funding, people, and partnerships

- Mark Shuttleworth (Ubuntu) invests $1m

2014 Red Hat acquires Inktank for $175m

- Ceph has now strategic value for industry, especially component makers

2019 IBM acquires Red Hat for $34b

- Organizations who know how to leverage open source are valuable!

2021 Sage Weil retires from Ceph, creates Civic Media

2022 Ceph becomes part of IBM Storage

- Neha Ojha (M.S. 2017 from UC Santa Cruz) is leading Ceph at IBM and in the open source community

IBM Storage Ceph

What didn’t work (so far)

Ceph “OSDs” run on servers, not devices

- Computational storage: offloading to *individually acting* devices
- Still need: offloading to *collectively acting* devices
  - Examples: reliability, data management
  - Lots of good ideas out there

Outline

History of Ceph

Impact of Ceph on open source at UC Santa Cruz

... and beyond

Impact on open source at UC Santa Cruz

Ceph as research platform

Institutional support for open source

Center for Research in Open Source Software (CROSS)

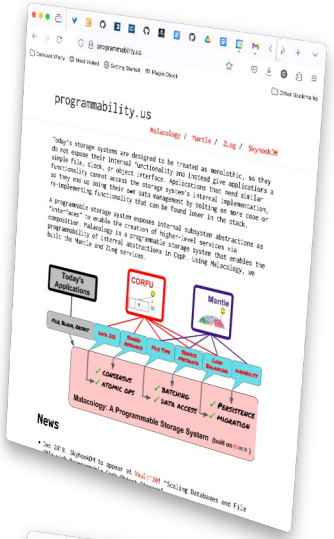
Open Source Program Office (OSPO) UC Santa Cruz

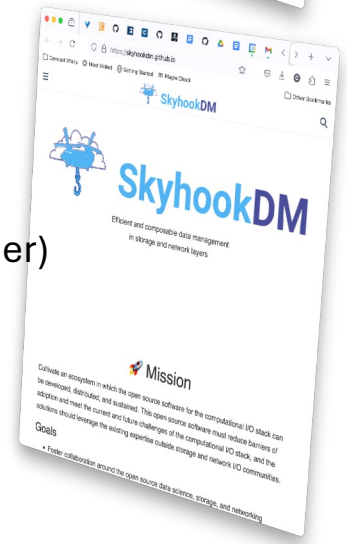
Ceph as a research platform

- Prototyping platform is production-quality software
- Students have access to large open source community
  - Help
  - Research collaboration
  - Summer jobs
  - Job opportunities
- Familiarity with Ceph is valued in job market
- Ceph has mature extensibility mechanism (CLS)

What worked

- Raised significant research funding
- Lots of publications at high-impact conferences
- Programmable Storage
  - Surfacing of internal abstractions to compose higher-level services
  - Programmability.us
  - Mantle at SC 2016, Malacology at EuroSys 2017, DeclStor at HotStorage 2017, Cudele at IPDPS 2018, Tintenfisch at HotStorage 2018
- SkyhookDM
  - Efficient and composable data management in storage & network layers
  - 2022 upstreamed Ceph extension into Apache Arrow (CCGrid'22)
  - Extending composable data services into SmartNICs (COMPSYS 2023, Best Paper)
- Graduated 6 Ph.D. students
  - Dimitris Skourtis, Latchesar Ionkov, Michael Sevilla, Noah Watkins, Ivo Jimenez, Jinshian Liu
- Upstreamed research products into Ceph releases





What didn’t work

- Most research products are abandoned
- Lack of institutional support for open source

Institutional support for open source

- 2014 Red Hat acquisition of Inktank
  - Scott Brandt and I asked Sage for a donation to UC Santa Cruz
  - Sage wants to support a “structure” to enable other students to have a similar career at UC Santa Cruz
  - Immediately resonated: undergrad course teaching how to submit a Linux kernel patch
    - Idea by Andrew Shewmaker, PhD 2016 from UC Santa Cruz
  - Result: $2m for “structure” + $500k for Chair for Open Source Software

- Don’t be shy asking alumni for donations – and listen to their ideas!

Amplify research impact via open source

Research results

Open Source Project

Institutional support of open source in universities: Connect Sponsors, Scientists, Students

Scientists

Bring in new, excited developers to your open source communities!

Funding

Open Source Projects

Sponsors

Engage with researchers and new student talent on exciting open source technologies!

Mentorship

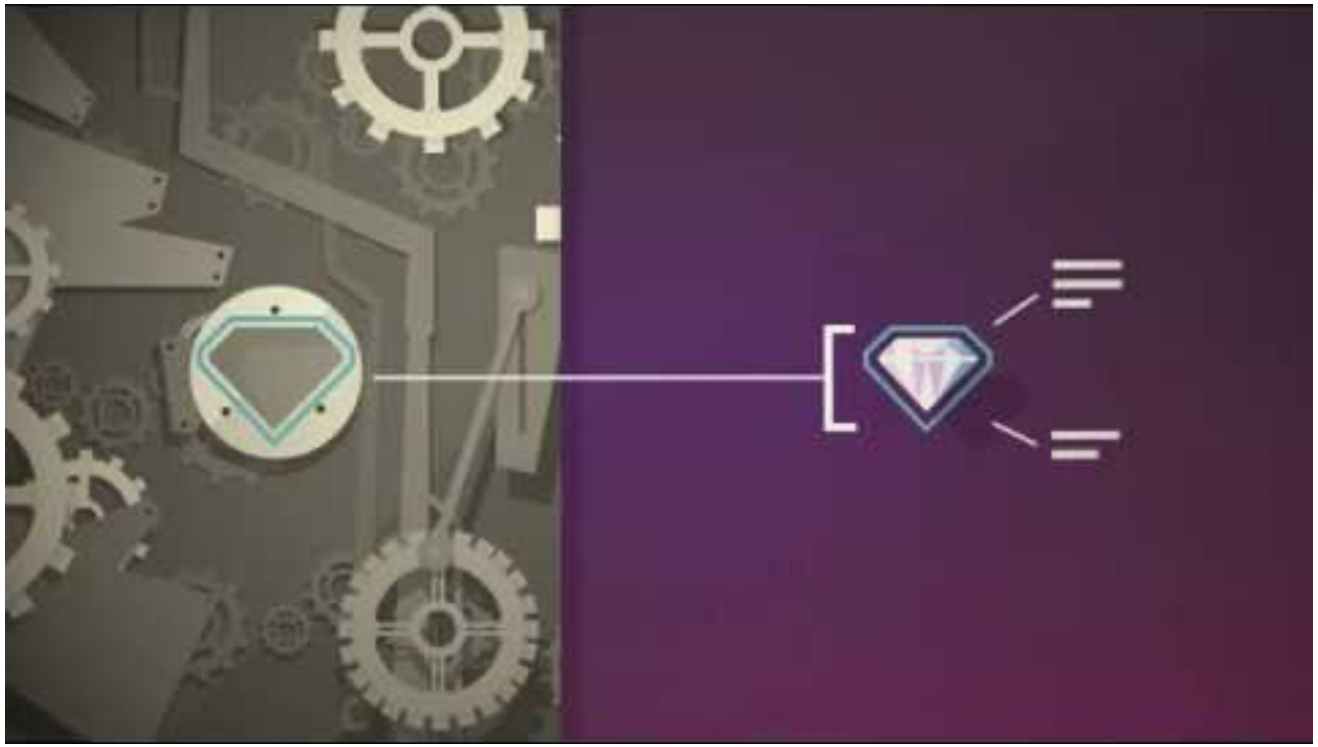
Talent

Students

Spend your summer writing code for exciting open source projects!

CROSS

CENTER FOR RESEARCH IN OPEN SOURCE SOFTWARE



cross.ucsc.edu est. 2015

UC SANTA CRUZ | Baskin Engineering

https://youtu.be/TceTFrDStVE

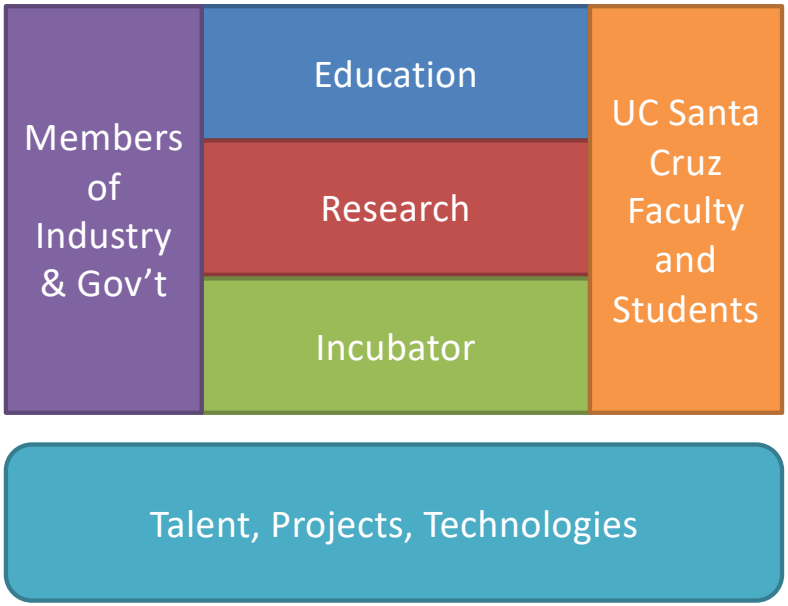
CROSS | CENTER FOR RESEARCH IN OPEN SOURCE SOFTWARE

- Bridges gap between student research & open source projects
- Structured as an Industry / University Collaborative Research Center (I/UCRC)
- Within Baskin Engineering

Teach students on how to productively engage in open source communities

Fund high-impact research with plausible path to successful open source projects

Incubate developer communities around research prototypes



cross.ucsc.edu est. 2015

CROSS | CENTER FOR RESEARCH IN OPEN SOURCE SOFTWARE

We thank Sage Weil and
our past and present sponsors!

KIOXIA SEAGATE

HUAWEI SAMSUNG FUJITSU

SK hynix Micron Western Digital

CROSS | CENTER FOR RESEARCH IN OPEN SOURCE SOFTWARE

Advisory Committee

James Davis (Professor of Computer Science), Founding Director of UCSC Center for Entrepreneurship, sits on advisory councils for startups and nonprofits.

Doug Cutting (Founder, Apache Lucene & Hadoop), Founder of numerous groundbreaking open-source projects, including Lucene, Nutch, Avro, Hadoop.

Nithya Ruff (Head of Open Source Program Office, Comcast & Chair, Linux Foundation Board of Directors)

Karen Sandler (Executive Director, SF Conservancy), Former executive director of GNOME Foundation, former general counsel of Software Freedom Law Center

Nissa Strottman (VP, Products and Operations, VISA), Years of experience in legal and M&A in the open-source software space.

Sage Weil (Chief Architect of Ceph, Red Hat), Creator of WebRing, founder of DreamHost, founder and CTO of Inktank, recipient of O'Reilly Open Source Award

CROSS | CENTER FOR RESEARCH IN OPEN SOURCE SOFTWARE

Measures of success

- Raised $2,600,000 in membership fees since Sept 2015
- Raised funding for 24 summer internships, incl. 10 GSoC projects
- Supported 14 research fellows (Ph.D. students)
- Supported 5 incubator fellows (Postdocs)
  - Oskar Elek (Polyphorm)
  - Jeff LeFevre (SkyhookDM)
  - Ivo Jimenez (Popper)
  - Kate Compton (Tracery)
  - Claudia Wehrhahn (Nimble)
- Created new undergrad course on Open Source Programming
- 46 research publications, 4 awards

CROSS | CENTER FOR RESEARCH IN OPEN SOURCE SOFTWARE

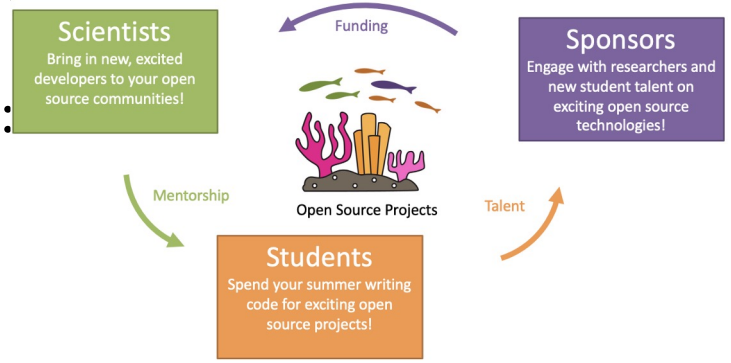
Lessons

- **Successful projects**: powerful catalysts for industry engagement
- **Open source programs**: great reachout opportunity for inclusiveness
- **Sponsors**:
  - Prioritize research over open source
  - For software: prefer open source licensing
- **CROSS fellows** (Ph.D. students and postdocs):
  - Have great project ideas
  - Are great student mentors
  - Are ambassadors for best open source practices
- **Students**:
  - Seed open source communities around research prototypes
  - Want to know learn how to engage in communities

CROSS | CENTER FOR RESEARCH IN OPEN SOURCE SOFTWARE

Lessons

- **Successful projects**: powerful catalysts for industry engagement
- **Open source programs**: great reachout opportunity for inclusiveness
- **Sponsors**:
  - Prioritize research over open source
  - For software: prefer open source licensing
- **CROSS fellows** (Ph.D. students and postdocs):
  - Have great project ideas
  - Are great student mentors
  - Are ambassadors for best open source practices
- **Students**:
  - Seed open source communities around research prototypes
  - Want to know learn how to engage in communities



CROSS | CENTER FOR RESEARCH IN OPEN SOURCE SOFTWARE

Gaps

STUDENT OPEN SOURCE TRAINING

VALUE OF OPEN SOURCE

BROAD INDUSTRY ENGAGEMENT

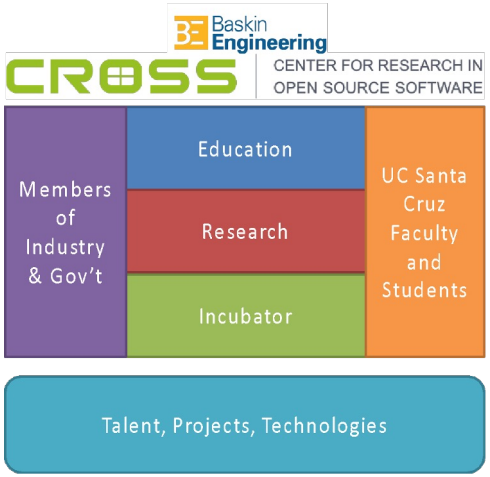
OSPO UC SANTA CRUZ

ospo.ucsc.edu est. 2022

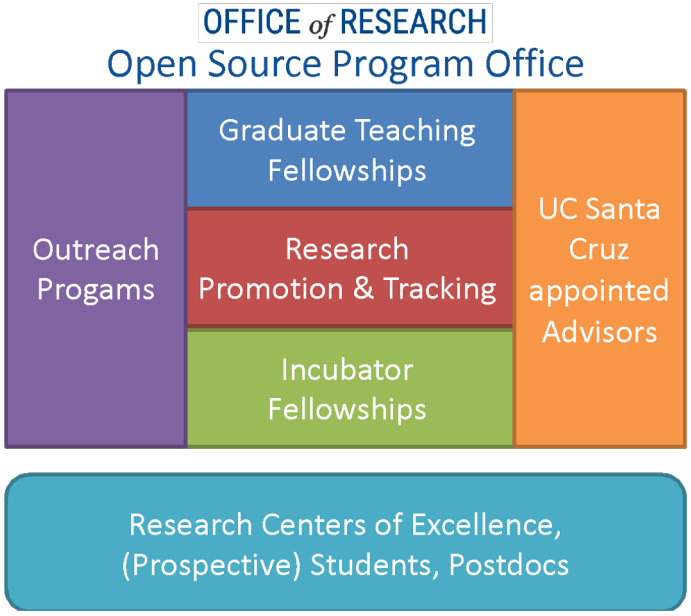
ALFRED P. SLOAN FOUNDATION

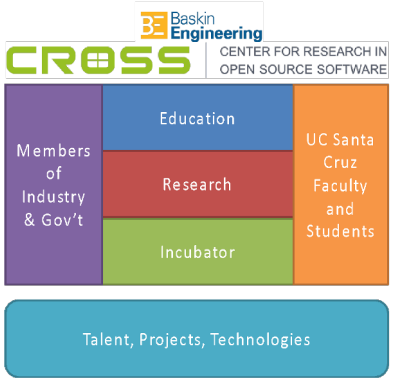
https://youtu.be/FTI245WyklgI

Expand CROSS programs to campus and beyond



Expand CROSS programs to campus and beyond





Expand CROSS programs to campus and beyond

OFFICE *of* RESEARCH

Open Source Program Office

Outreach Progams

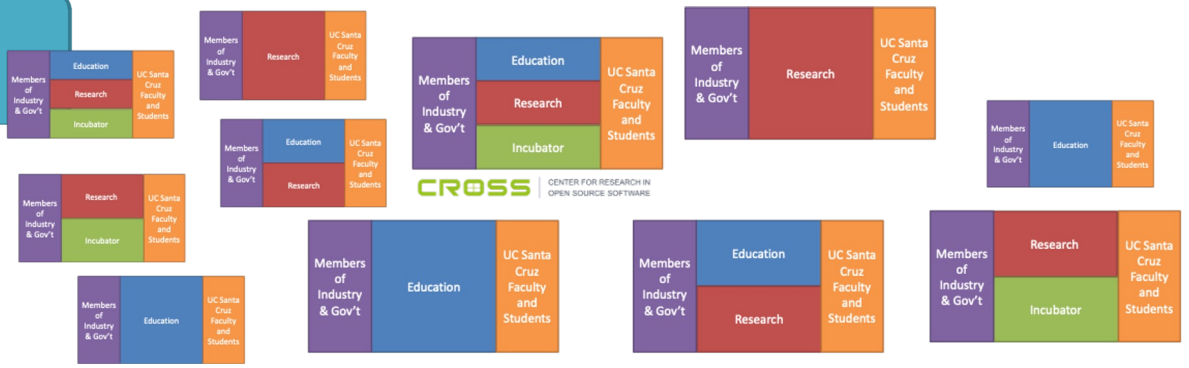
Graduate Teaching Fellowships

Research Promotion & Tracking

Incubator Fellowships

UC Santa Cruz appointed Advisors

Research Centers of Excellence, (Prospective) Students, Postdocs



OSPO UC SANTA CRUZ

An emerging playbook

WITHIN OFFICE OF RESEARCH

GRADUATE TEACHING FELLOWSHIP

MARKETPLACE OF PROJECT IDEAS

INCUBATOR FELLOWSHIPS

OUTREACH PROGRAMS

OSPO UC SANTA CRUZ

An emerging playbook

WITHIN OFFICE OF RESEARCH

GRADUATE TEACHING FELLOWSHIP

MARKETPLACE OF PROJECT IDEAS

INCUBATOR FELLOWSHIPS

OUTREACH PROGRAMS

OSPO Programs

filling

gaps

Graduate Teaching Fellowship

Marketplace of Project Ideas

Incubator Fellowship

Student open source training

Value of open source

Broad industry engagement

OSPO Programs

filling

gaps

Graduate Teaching Fellowship

Courses designed by grads & postdocs

Student open source training

Marketplace of Project Ideas

Value of open source

Incubator Fellowship

Broad industry engagement

OSPO Programs

filling

gaps

Graduate Teaching Fellowship

Marketplace of Project Ideas

Incubator Fellowship

Courses designed by grads & postdocs

Mentor/student matching

Student open source training

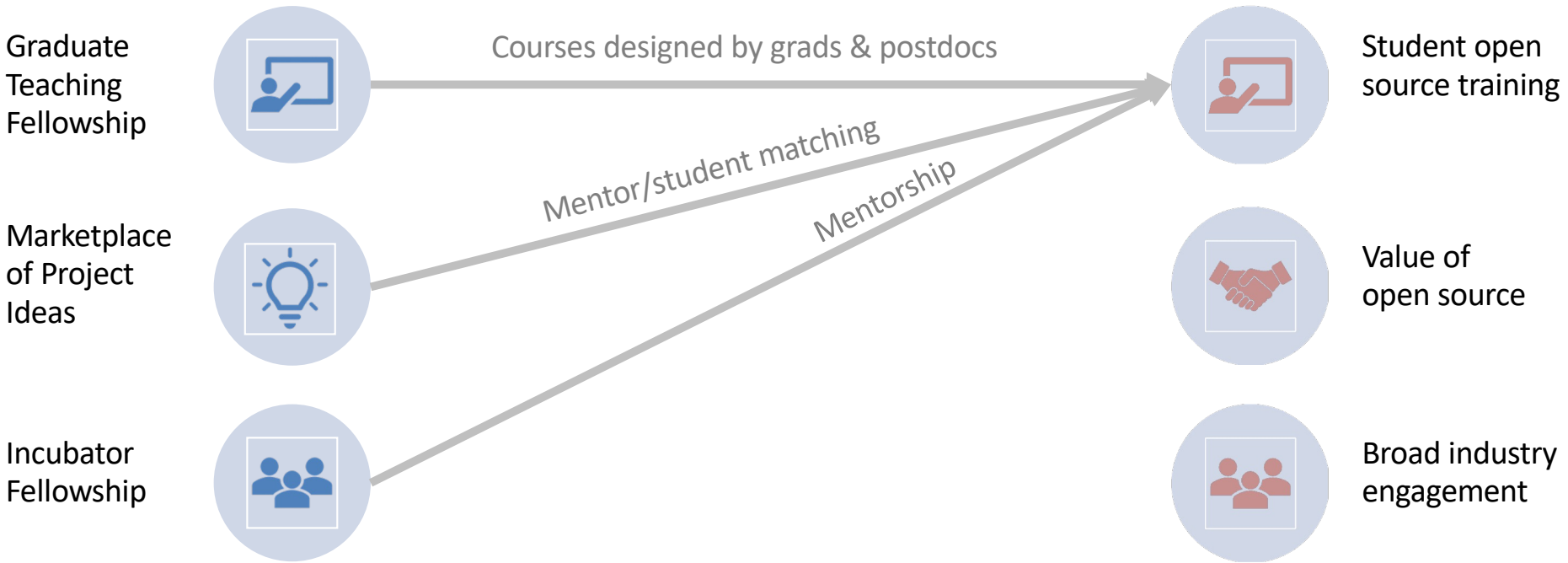
Value of open source

Broad industry engagement

OSPO Programs

filling

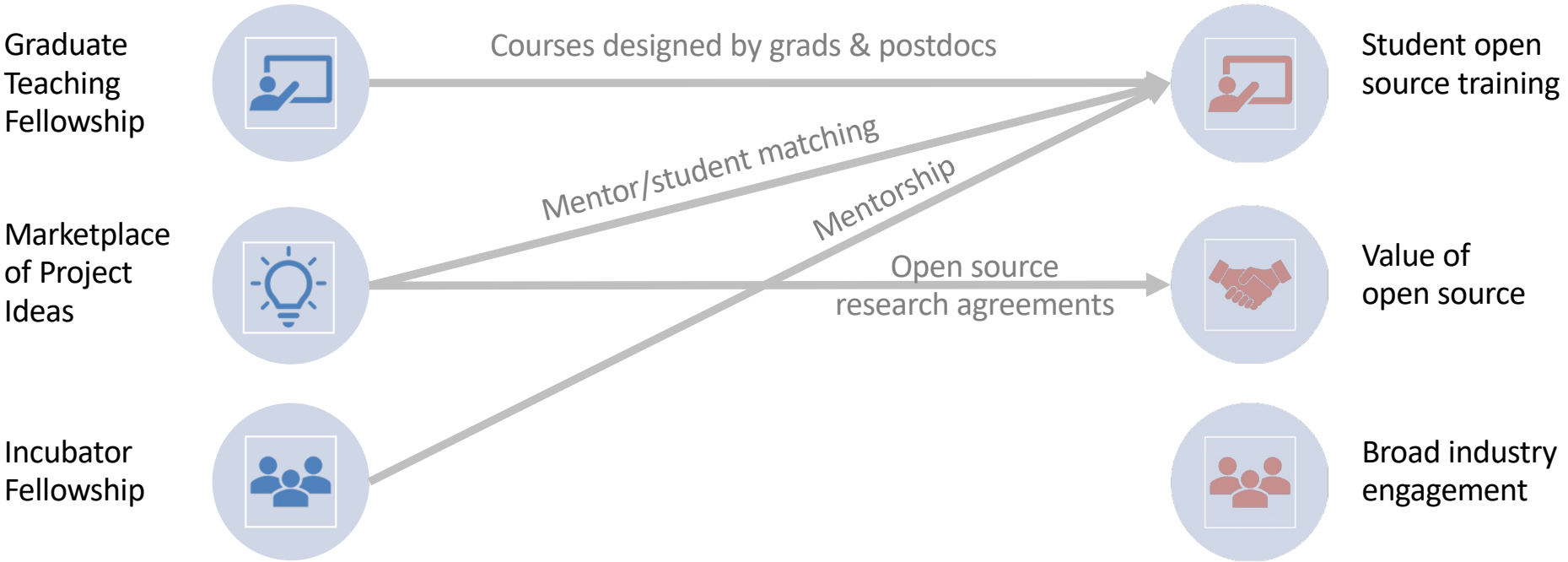
gaps



OSPO Programs

filling

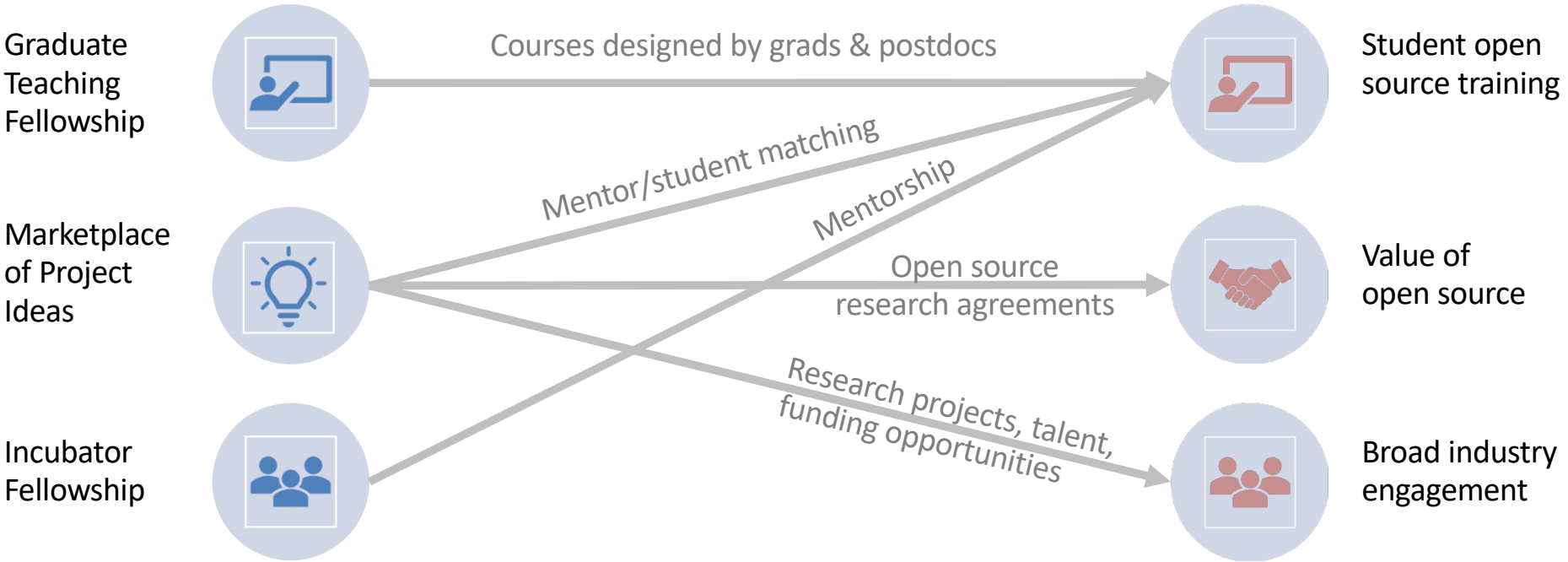
gaps



OSPO Programs

filling

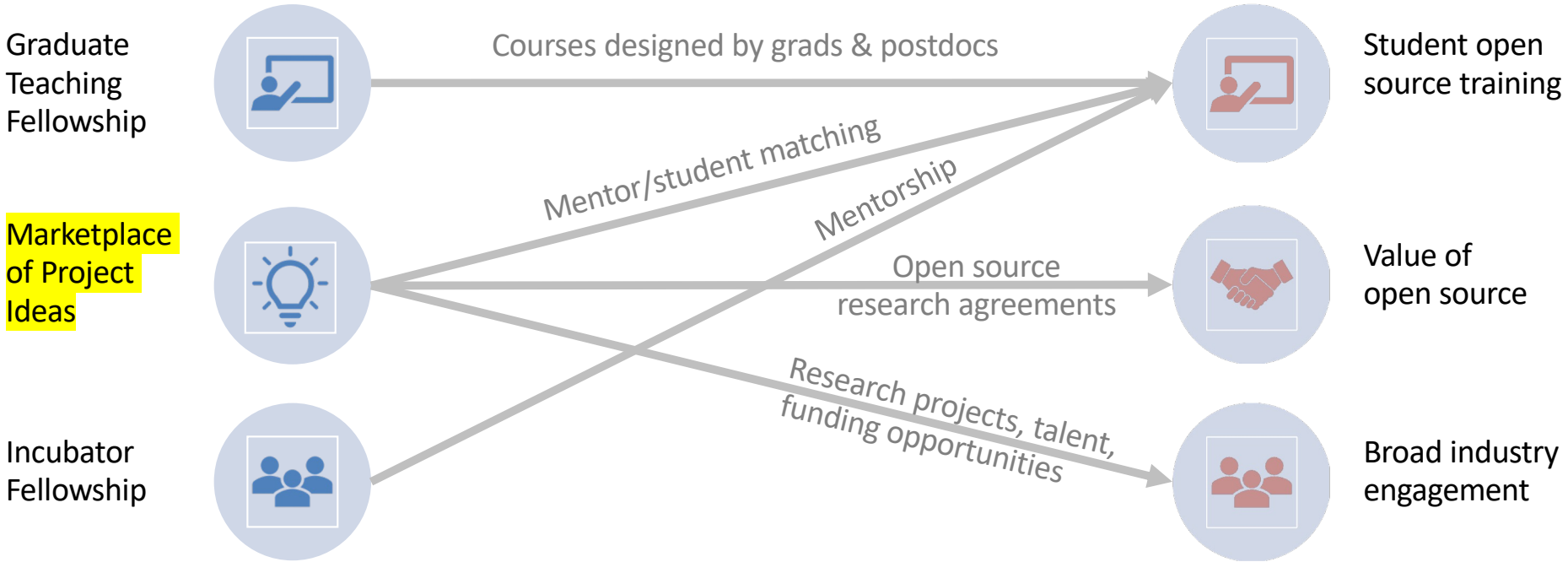
gaps



OSPO Programs

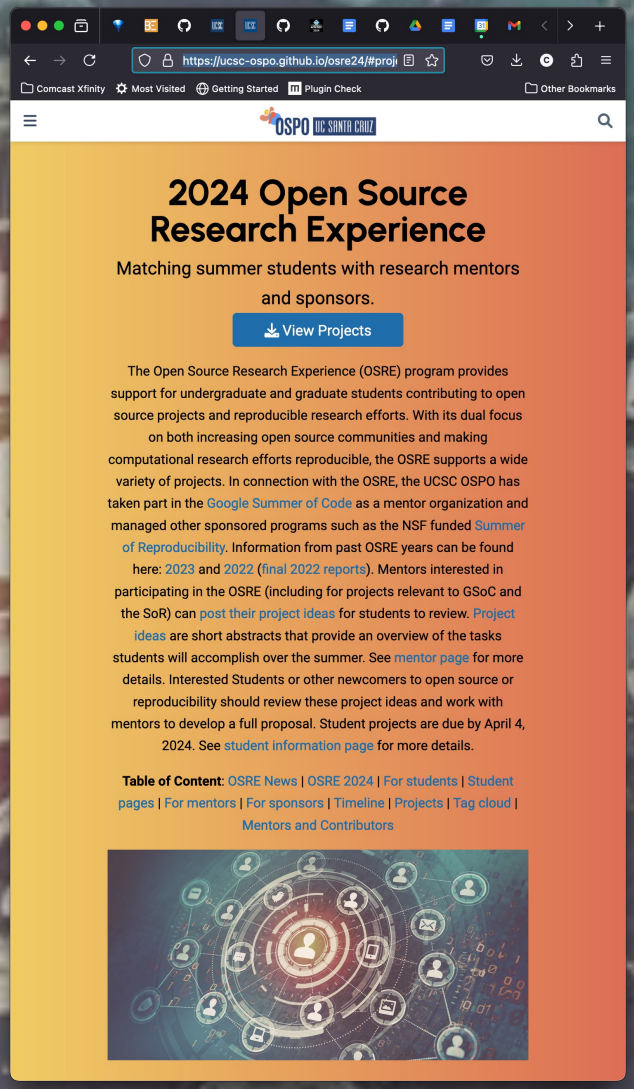
filling

gaps

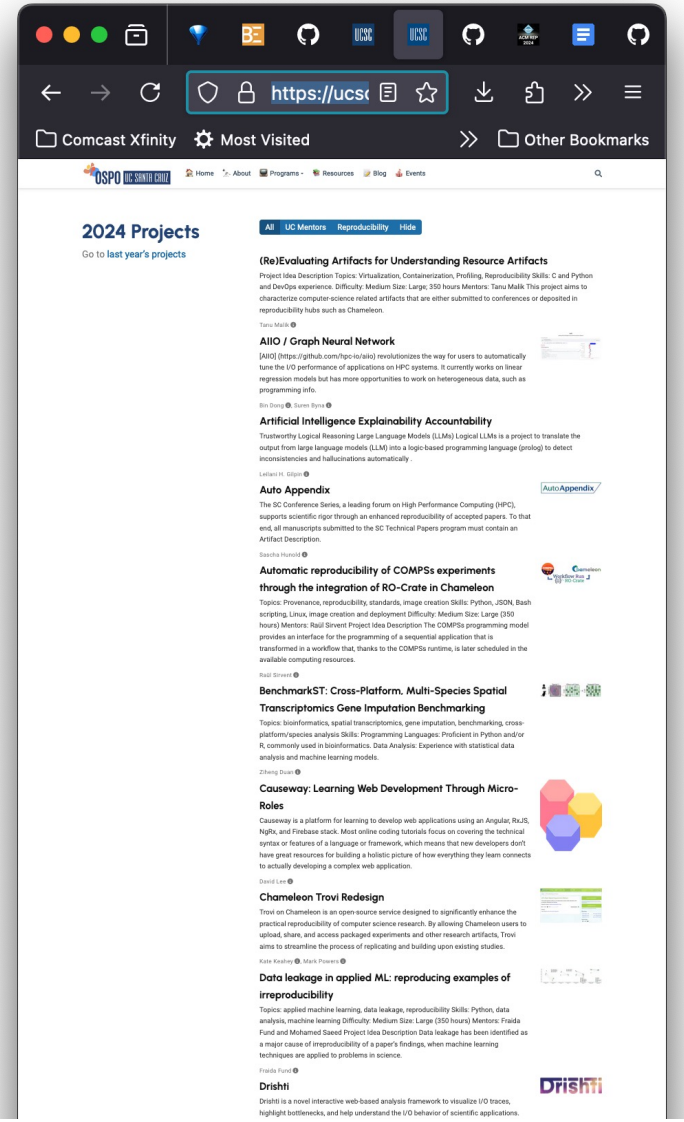


Open Source Research Experience

- A scalable marketplace for project ideas
- Jumpstarts contributor communities
- Students work with top scientists
- Leverages global outreach programs



ospo.ucsc.edu/osre24

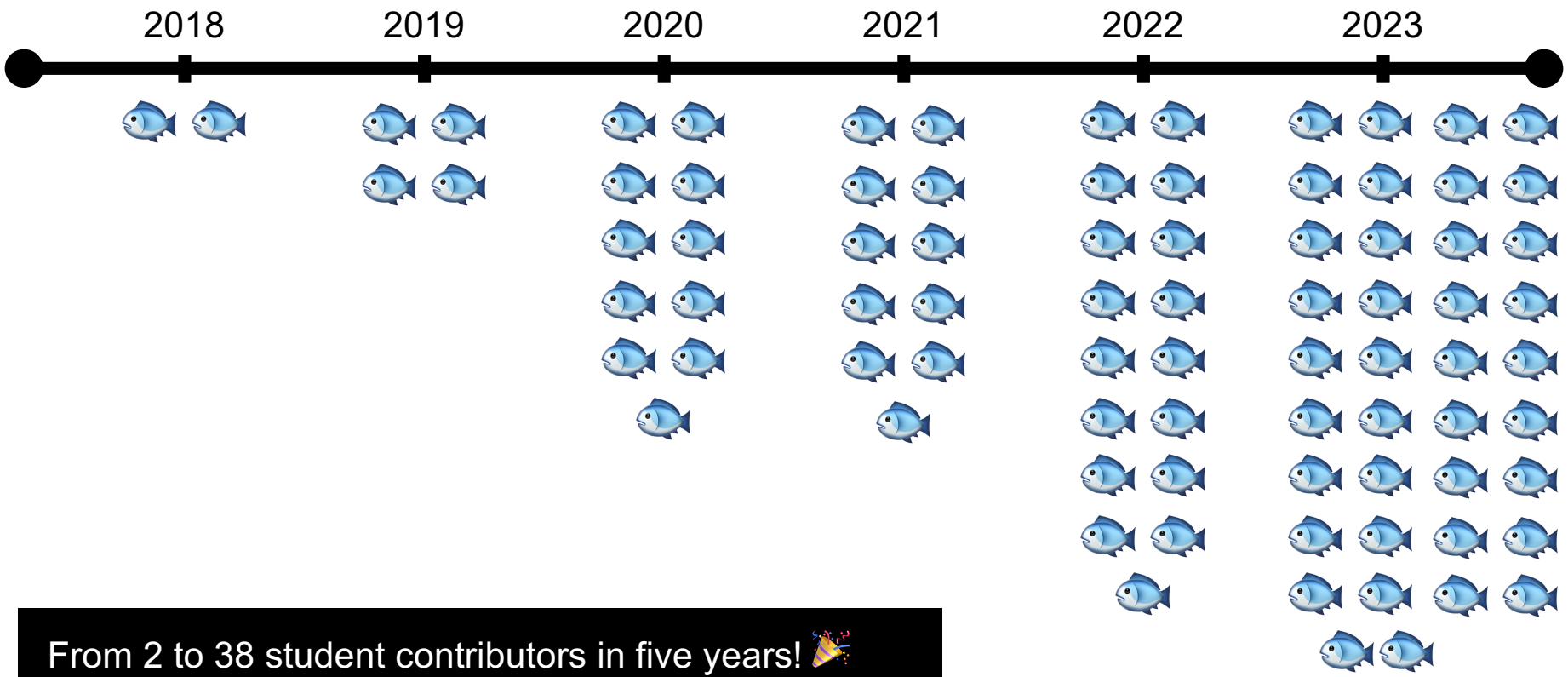


Project Ideas Catalog

1. Mentors post project ideas
2. Student and mentors work closely together to build proposals
3. Students send in proposals
4. Mentors pick favorites
5. Slots are assigned
6. Then … let the coding begin!

Check out the ideas page
https://ucsc-ospo.github.io/osre24/#projects

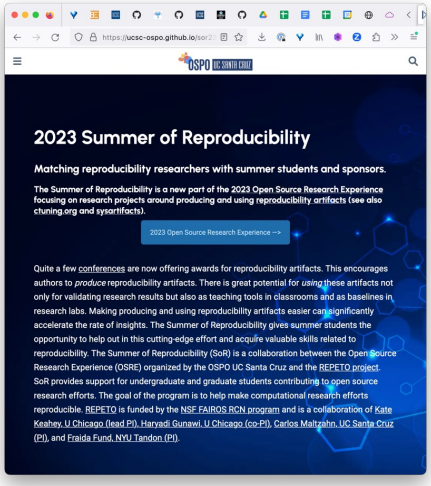
OSRE Program Evolution



From 2 to 38 student contributors in five years! 🎉

New OSRE Components in 2023

- **Catalyst**: Connecting UC Santa Cruz to HBCUs via open source
- **Summer of Reproducibility**: Creating an exchange for reproducibility artifacts



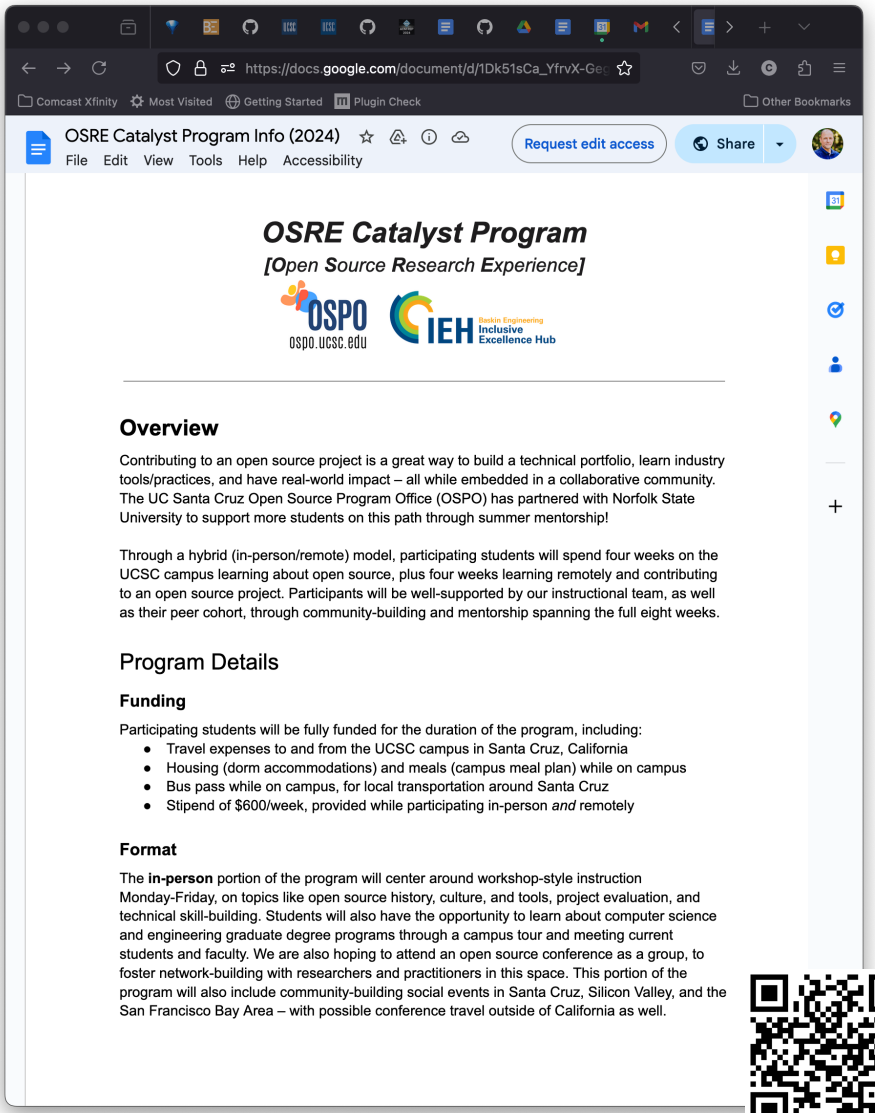
- **OSRE Catalyst Pilot Program - June –August 2023**
  - Aimed supporting more diverse contributor/student participation
  - Sponsored by the Baskin Engineering Inclusive Excellence Hub (IEH) as part of greater UC-HBCU partnership
  - In-person for community building, remote to simulate our typical OSRE or OS community interaction
  - Two layers of mentorship – one layer from lead instructor and other from OS community
  - Envisions expansion to more partner schools and tapping alumni as peer mentors

Dr. Emily Lovell

- **Incubator Fellow**
  OSPO UC Santa Cruz
- **Lead Mentor/Instructor**
  OSRE Catalyst Program
- **Previous Experience**
  Faculty at Berea College
  MIT Media Lab Researcher
- **Current Research**
  Strengthening Underserved Segments of the Open Source Pipeline

OSRE 2024 Contributor Catalyst

- Focus on cohort-based learning within larger communities of practice
- Collaboration with HBCUs
- Work-in-progress program repo: https://github.com/emmet0r/contributor-catalyst



OSRE Catalyst Program Info (2024)

File Edit View Tools Help Accessibility

Request edit access | Share

OSRE Catalyst Program

[Open Source Research Experience]

OSPO ospo.ucsc.edu | IEH Baskin Engineering Inclusive Excellence Hub

Overview

Contributing to an open source project is a great way to build a technical portfolio, learn industry tools/practices, and have real-world impact – all while embedded in a collaborative community. The UC Santa Cruz Open Source Program Office (OSPO) has partnered with Norfolk State University to support more students on this path through summer mentorship!

Through a hybrid (in-person/remote) model, participating students will spend four weeks on the UCSC campus learning about open source, plus four weeks learning remotely and contributing to an open source project. Participants will be well-supported by our instructional team, as well as their peer cohort, through community-building and mentorship spanning the full eight weeks.

Program Details

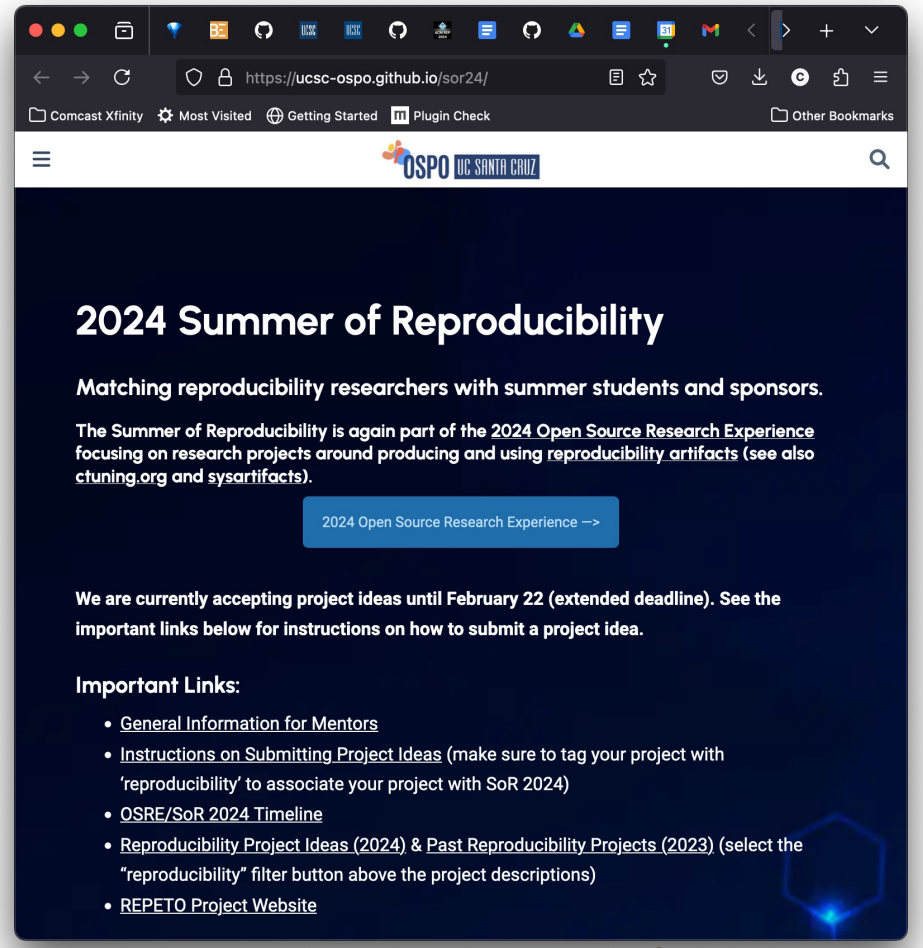
Funding

Participating students will be fully funded for the duration of the program, including:

- Travel expenses to and from the UCSC campus in Santa Cruz, California
- Housing (dorm accommodations) and meals (campus meal plan) while on campus
- Bus pass while on campus, for local transportation around Santa Cruz
- Stipend of $600/week, provided while participating in-person *and* remotely

Format

The **in-person** portion of the program will center around workshop-style instruction Monday-Friday, on topics like open source history, culture, and tools, project evaluation, and technical skill-building. Students will also have the opportunity to learn about computer science and engineering graduate degree programs through a campus tour and meeting current students and faculty. We are also hoping to attend an open source conference as a group, to foster network-building with researchers and practitioners in this space. This portion of the program will also include community-building social events in Santa Cruz, Silicon Valley, and the San Francisco Bay Area – with possible conference travel outside of California as well.



Summer of Reproducibility

- Part of REPETO project
  - https://repeto.cs.uchicago.edu/
  - Led by Kate Keahey
- Creating a marketplace for reproducibility artifacts
- Opportunity for contributors to acquire open source *and* reproducibility skills
- Allows OSRE to work with broader range of students and mentors

https://ospo.ucsc.edu/sor24

Outline

History of Ceph

Impact of Ceph on open source at UC Santa Cruz

... and beyond

Network of OSPOs

Connect innovative open source activities
Develop coordinated approach
Develop locally, act collectively; leveraging other campuses
Expand communities of practice, institutional support
Document in playbooks

ALFRED P. SLOAN FOUNDATION

OSPO NETWORK

Funded! Started in Jan 2024.

Main themes

University Open Source Landscapes & Metrics

Training, Development, Mentorship

Open Source Project Sustainability

ALFRED P. SLOAN FOUNDATION

OSPO NETWORK

Funded! Started in Jan 2024.

Institutional support for open source 2015

Institutional support for open source

2022

Institutional support for open source

2024

Institutional support for open source 2024

OSPO UC SANTA CRUZ

Baskin Engineering

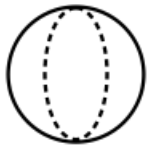
CROSS

CENTER FOR RESEARCH IN OPEN SOURCE SOFTWARE

UC Open Source Repository Browser

Institutional support for open source

Program



UC Open Source
Repository Browser

Discovery

Production

CROSS

CENTER FOR RESEARCH IN
OPEN SOURCE SOFTWARE

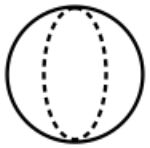
Deep integration with the mission: teach, research, public service

Programs

OSPO

Programs: OSRE, Incubator fellowship, …
Center of competency for open source
Defining & implementing strategies and policies

Discovery, many sources
Characterization of risk, value
FAIR OS artifacts
Curated guides
Research instrument



UC Open Source Repository Browser

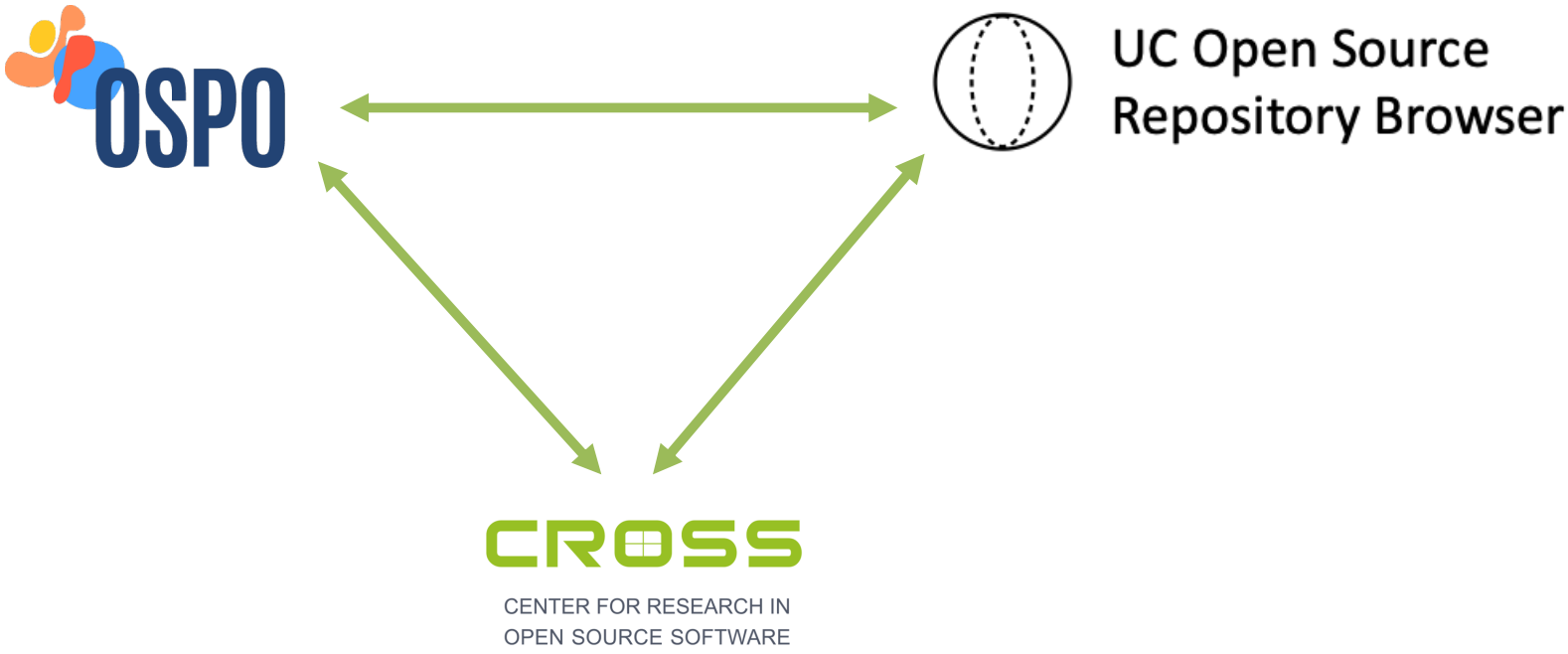
Discovery

Production

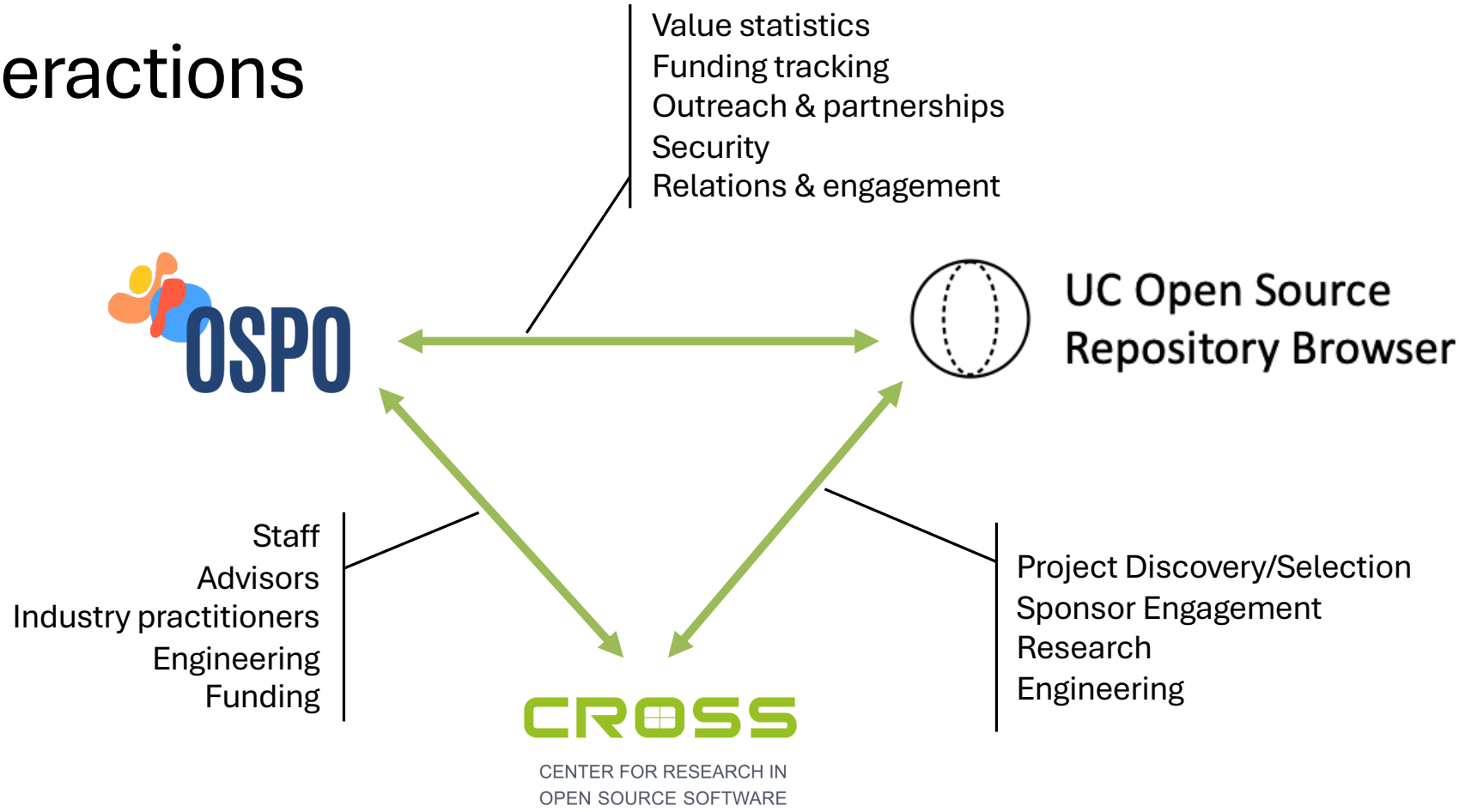
CROSS
CENTER FOR RESEARCH IN OPEN SOURCE SOFTWARE

Teaches open source literacy
Trains/**certifies** mentors (with CITL)
Provides **research** fellowships
Provides sponsored/embedded **engineering**
Engages with sponsors

Interactions



Interactions

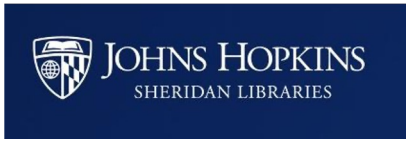


Outline

History of Ceph

Impact of Ceph on open source at UC Santa Cruz

... and beyond



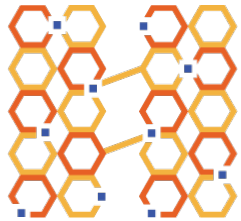
First 6 Sloan Supported University OSPOs

2022

Now 12 Sloan Supported University OSPOs

2023

Understanding Open Science / Open Source in University Setting

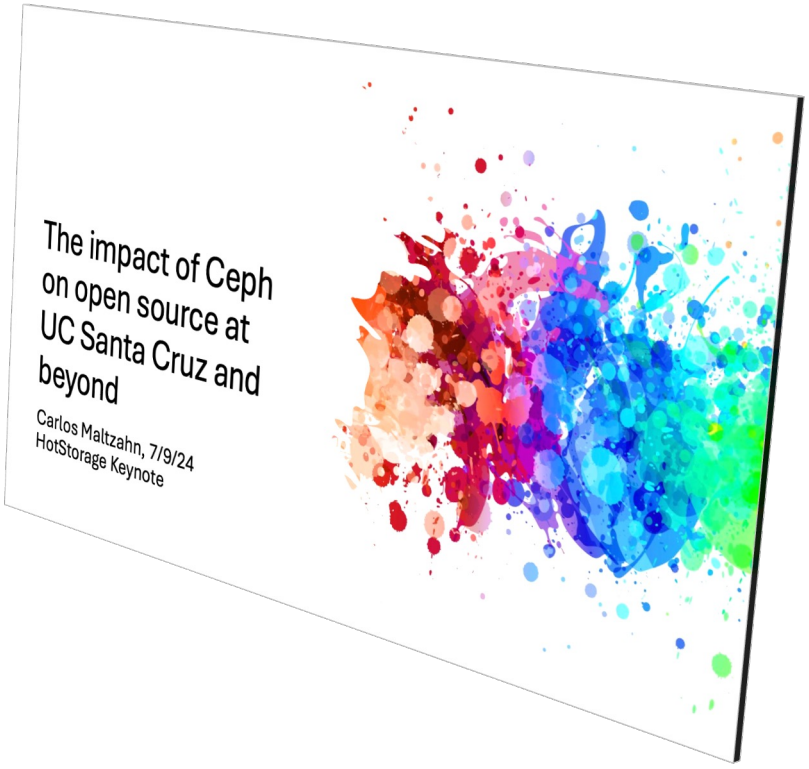


HELIOS

ASSOCIATION OF RESEARCH LIBRARIES

cni
Coalition for Networked Information

FAIROS RCN

Thank you!

Contact: carlosm@ucsc.edu