2i2c CZI final grant report

Contents

Chapter 1

1. 2i2c grant report overview

This MyST Document is a final grant report generated for the Chan Zuckerberg Initiative for the seed grant that founded 2i2c. It covers the major goals, operations, and impact of 2i2c over its first three years of existence.

Holdgraf (2024)

2i2c is now three years old, having been initially founded with this grant from CZI in late 2020. Over this time, 2i2c has grown and matured as an organization, has served thousands of users doing open science and education, and has made extensive contributions to the open source ecosystem for interactive computing.

This document contains an overview of the impact and journey of these three years. It is broken down into three key areas of organizational impact.

A note on measurements

Over the last three years, we've learned the importance of having measures that reflect progress towards your strategic objectives and goals. Measures like this give us a more objective way to decide we are making progress (or not), and if we need to modify our approach.

We've tried to include a few key measures for each of the major areas of progress below. We know they aren't perfect, and in some cases need an overhaul or better mechanisms for measuring. You'll see them in pull-quotes like the following:

> Here's an example pull quote!

We'll include a brief explanation of why we use each of these measures, and what we might want to improve about them. You'll find those in dropdowns similar to the one you're reading now.

See our KPIs documentation for more information about the metrics we track and aim to track.

This document was created with MyST

We created this website with the MyST Document Engine, a new system for authoring scientific narratives and documentation. Try hovering over this link for a cool feature preview.

MyST is similar to the Jupyter Book project, though uses more modern web technologies. We're exploring more ways to improve MyST for documents like these, and are using this report as a way to learn from it.

1.1 Impact through our cloud platform and service

2i2c provides a community-centric platform that enables the lifecycle of knowledge creation and sharing. At the heart of this model is a community JupyterHub that provides a digital space where

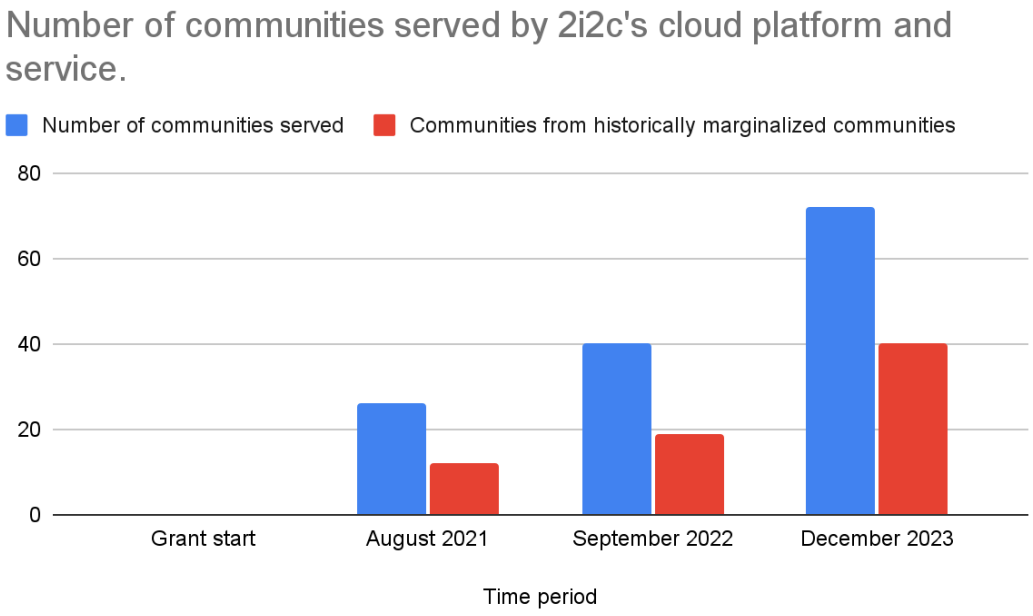

Figure 1.1. *The total number of communities served on our platform each year. We show both the total number of communities, as well as communities designated as "historically marginalized" (for example, community colleges or communities from the global south).*

many community members come together to access shared resources and work together[1][2].

More than 70 Active Communities and 7200 Monthly Active Users.

The most direct measure of impact for 2i2c is in the number and diversity of communities that utilize our service. At the time of writing, our open science cloud platform serves more than **70 different communities**, at a total of around **7200 monthly active users**.

Below we show the total number of communities served on our platform each year, with a focus on historically marginalized communities.

In addition we track **Weekly Active Users** across all communities to understand how frequently community members make use of the platform[3]. Below we show an estimate of Weekly Active Users at the end of each year of this grant.

These communities are intentionally diverse in their scope and focus, and range from university-wide educational services (University of Toronto), to multi-organization research collaborations (the LEAP project at Columbia University), to major federal projects fighting climate change (the US Greenhouse Gas Center), to networks of community colleges in California (the CloudBank and UC Berkeley educational hubs collaboration), to international networks of historically under-resourced communities in the Global South (the CZI-funded Catalyst Project).

[1] "Communities" are a combination of research groups, classrooms, communities of practice, and institutions that were actively using the service at the end of each date period.seelsocommunities Historically marginalized communities are defined as educational institutions that are not R1 universities (for example, community colleges) as well as communities that are primarily not based in North America or Western Europe.

[2] For a more in-depth picture of community hubs see our Community Hubs Usage Dashboard.

[3] This is defined as a user that starts their own Jupyter server within a week-long window. Note that this varies heavily based on the academic calendar, and spikes around the end of each academic semester.

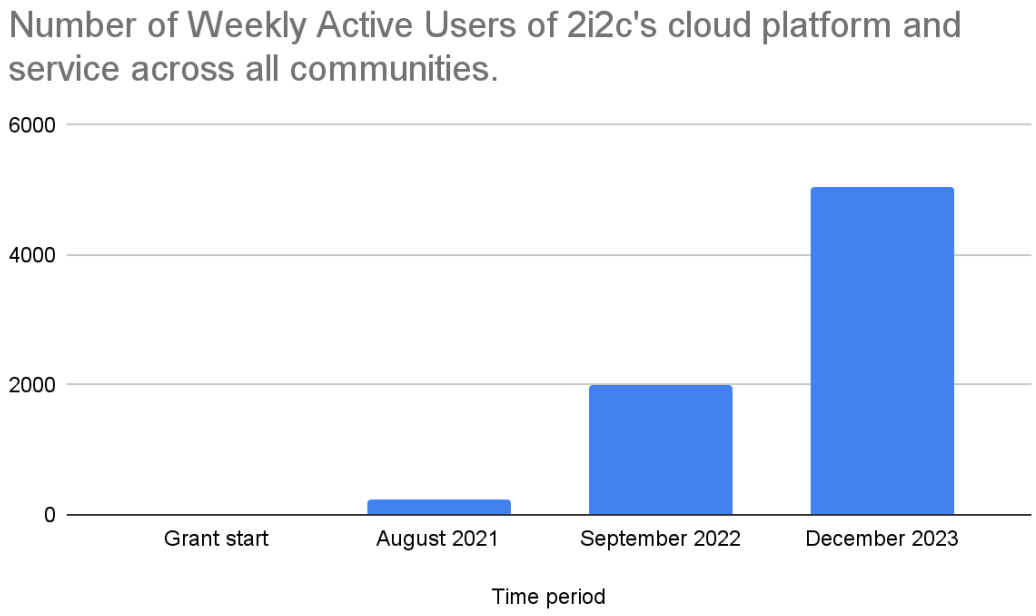

Figure 1.2. *Weekly Active Users (WAUs) at the start of this grant, and at one-year checkpoints throughout this grant.*

These numbers are beyond the expectations that we had for 2i2c when we first began this service, and speak to the value of open science platforms rooted in community-centric design and open infrastructure.

How we chose measures for community impact

We chose to measure impact both at the community level (number of community hubs), as well as the amount of usage across all communities (monthly active users). This is because 2i2c aims to balance *breadth* and *depth* of service, serving both many different kinds of workflows, but also providing enough flexibility that communities can tailor their hub to their needs and grow their total users.

We additionally track communities we define as "marginalized" because we believe that equity and democratizing access to infrastructure is a core part of 2i2c's mission. Historically marginalized communities often lack the resources or expertise to manage and access infrastructure, and we want to ensure 2i2c's service benefits these communities.

In the future, we hope to perform more in-depth research to understand how these communities use our platform, as well as the diversity of user archetypes and workflows that are contained within a single community.

1.2 Capacity building for open source projects

Another key outcome of this grant was using 2i2c as a mechanism to grow capacity in open source ecosystems. 2i2c builds its open science platform by integrating community-driven open source projects. Providing upstream contributions and support to the communities that underlie these projects is a secondary goal that underlies 2i2c's open practices as an organization[4].

2,000 merged pull requests across more than 600 repositories.

[4]For a more in-depth picture of our upstream contributions, see our Upstream Contributions Dashboard.

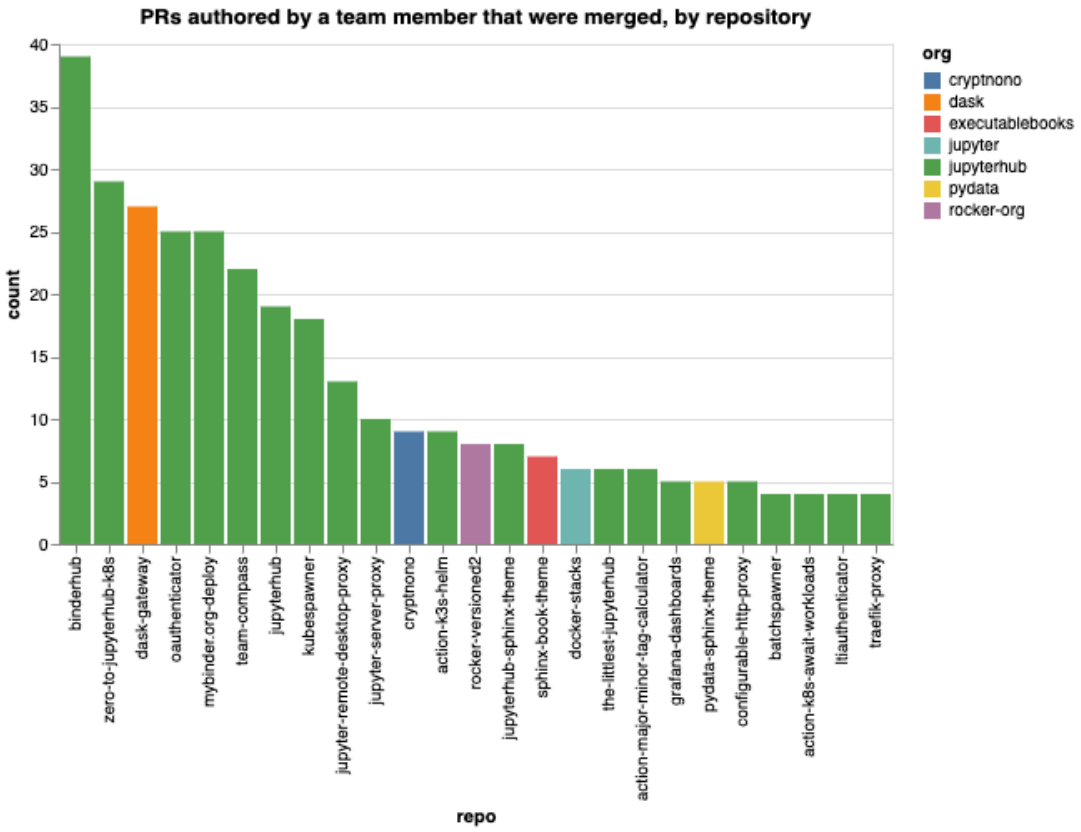

Figure 1.3. *PRs authored by a 2i2c team member that were merged in the last two quarters, for key upstream communities. See latest data at 2i2c.org/kpis.*

Since January, 2021, 2i2c's team members have authored over **2,000 merged pull requests across more than 600 repositories** in "key upstream communities" that underlie our infrastructure[5]. For example, here is a summary of merged Pull Requests in key upstream repositories that were authored by a 2i2c team member over the last two quarters:

2i2c now **employs 6 "distinguished contributors" to the Jupyter Community**, as well as several team members that create training and guidance programs to help communities use these open source tools to their full potential. We have driven major enhancements to technology in the stack, such as bringing Binder-like environment building to JupyterHub, or creating a more flexible and powerful user environment selector, or defining a new documentation back-end for the Jupyter Book ecosystem.

We believe that no other organization exists that has served such a diverse collection of community partners in open science workflows while providing entirely open source technology that is driven by multi-stakeholder communities of practice. This highlights our commitment to ensuring these open source communities are healthy.

How we chose measures for upstream impact

Measuring "impact" in open source communities is difficult without additional resources to interview community members and understand the true impact of a person's participation in a community. Many projects track "volume-based" metrics like "number of comments", "issues

[5] A Key Upstream Community is an open source community that builds infrastructure considered strategically critical for 2i2c's services and impact.

opened", etc, but these are not necessarily correlated with "useful contributions" (and in some cases may be anti-correlated).

We decided to use "Pull Requests authored by 2i2c team members that were merged" as a reflection of both "work done by 2i2c team members" in addition to "work that was deemed useful by the community and thus merged into its codebase". This almost certainly under-counts the many non-development ways that our team contributes (for example, providing guidance in issues, reviewing the work of others, and providing broad leadership and strategic guidance). We hope to find better ways to measure the breadth of our upstream contributions in the coming years.

1.3 Organizational capacity and sustainability

Finally, 2i2c aimed to achieve this impact in a sustainable way that grew the capacity of the open science and open source ecosystems. This ensures that we have a model for sustainably serving our communities and the open source ecosystem into the future.

More than $5.5 million in funding brought in over three years.

Over the last three years, 2i2c has brought in **over $5.6 million** (including this grant) to fund the capacity needed to carry out its operations. The majority of our funds have come from direct grants to 2i2c, or grant-funded collaborations with other communities. Over time, we have increased the percentage of our revenue that comes from service contracts, and aim to continue this trend to further diversify our sources of funding.

Below is a graph of 2i2c's total funding broken down by major category. Note that for several grants, 2i2c was not the sole recipient of funds, but served as the Principal Awardee for the grant.

And the same data broken down by quarter since the beginning of this grant:

Here are the new service contracts in each quarter since the start of this grant. This is a reflection of the revenue-generating partnerships that 2i2c has created.

Finally, over the past three years 2i2c has also grown and matured as an organization. We began with a team of **two** (an Executive Director and one Engineer). We are now a team of **twelve**, with organizational capacity split across engineering, community partnerships, product management, and organizational operations and strategy.

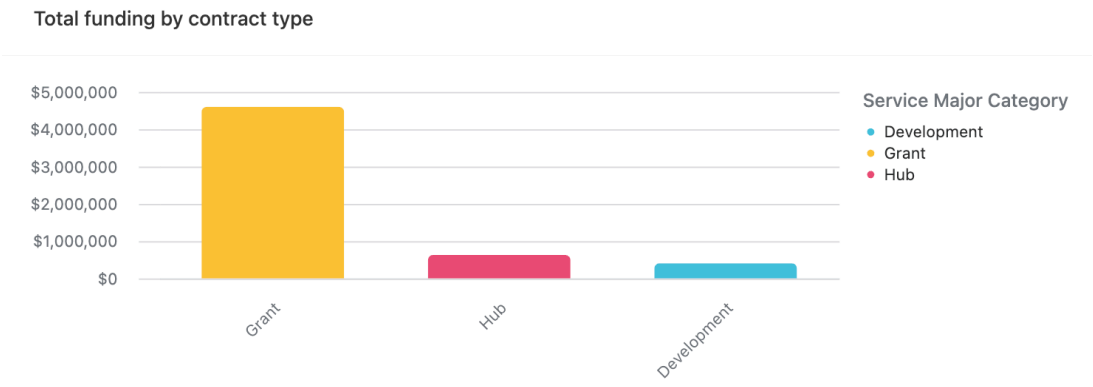

Figure 1.4. *Total funding brought in by 2i2c throughout this grant (the first three years of its existence), broken down by major category. Note that not all funding went to 2i2c as in many cases we distributed funding to other collaborators and partners via grant sub-contracts.*

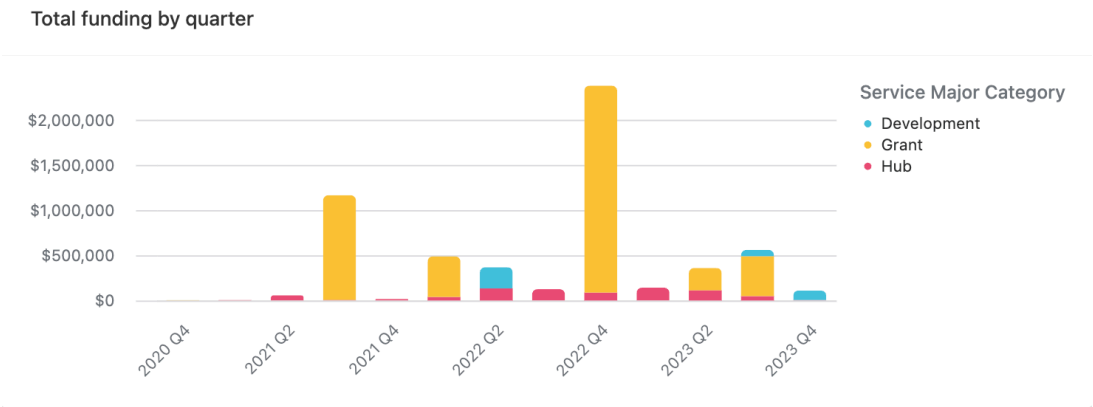

Figure 1.5. *Total funding brought in by 2i2c for each quarter of this grant, broken down by major category. Note that not all funding went to 2i2c as in many cases we distributed funding to other collaborators and partners via grant sub-contracts.*

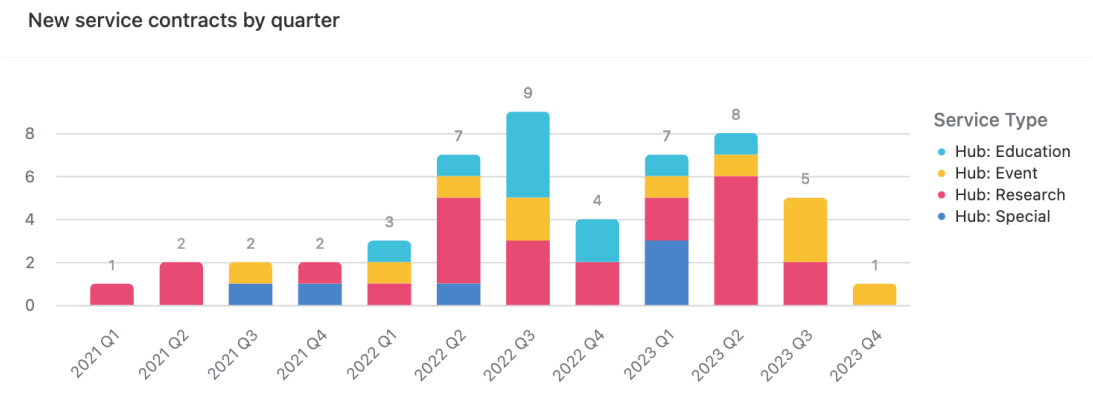

Figure 1.6. *Service contracts each quarter during the lifecycle of this grant, broken down by major category.*

This more complex structure reflects a more mature organization that understands the roles that are needed in order to deliver excellent open science cloud services to research and education

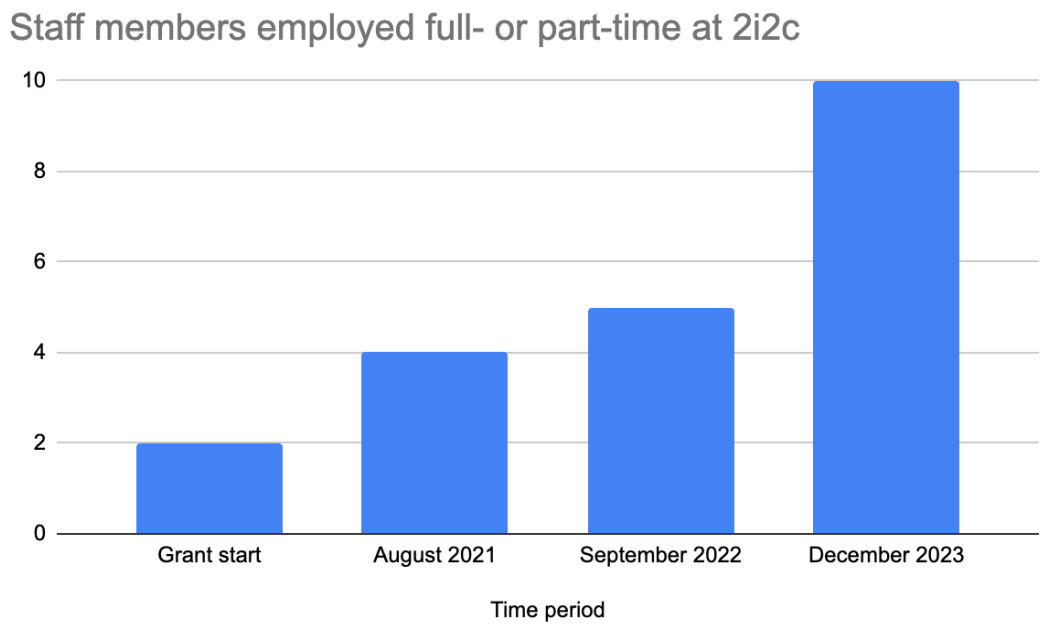

Figure 1.7. *Total number of paid staff (full or part time) at 2i2c. This reflects our organizational capacity and an increasing complexity and scope of operations. Note that two additional staff members have been hired since the end-date of this grant, and 2i2c's current team is 12 paid staff.*

communities. We also believe this model can serve as a guide for mission-driven organizations that facilitate collaboration between the open science and open source communities.

However, the additional staff in our team also has a significant impact on our organizational costs. As a result, one of our main strategic objectives heading into the next year is to raise funds that allow us to continue providing 2i2c's community-centric service at our current staff levels.

> **How we chose measures for organizational capacity**
>
> Measuring organizational capacity is also difficult because it is an inherently multi-faceted quality that is not easily captured by a single number. In addition, 2i2c is still relatively young, and thus is still growing its capacity to measure organization-wide traits like this. We believe that our recent Delivery Manager / Chief of Staff role will help us refine our understanding and practice around organizational capacity.
>
> In the meantime, we choose "financial support" as well as "team size" as a reflection of overall capacity. This is because at a fundamental level, any organization is both balancing the cost of its operations against its ability to bring in revenue to fund those operations. Growing an organization is a constant balancing act between matching your organizational capacity (your costs) to your organizational commitments (which bring in revenues). It is also a way for us to track improvements in team efficiency and skills (by being able to grow our operations without significantly growing our costs).
>
> In the future, we aim to define measures of team velocity and efficiency, in order to more directly measure our ability to deliver work and enhance our service. We hope to balance this against the numbers-based "cost" and "revenue" metrics in order to ensure our team becomes more efficient and effective over time.

1.4 Areas for growth and next steps

After three years, 2i2c's service has demonstrated a clear demand for community-centric science platforms built on open infrastructure, and continues to grow the percentage of costs it recovers from service contracts versus grants. We've also identified challenges to overcome in order to carry the organization forward into its next phase. Below is a brief description of each.

1.4.1 Refine our operating and service model

In 2023, our team felt like we hit a ceiling in how quickly we could make strategic progress. The growth of our service, paired with a lack of evolution of our team's planning mechanisms, led to a lot of reactive work and a lack of alignment on next steps. The organic growth of our service is an excellent indicator that we are delivering value and having an impact. However we also believe that we need a sharper definition of our value proposition, service offering, and pricing/funding model.

To address this, we partnered with an organizational consultant to audit 2i2c's structure and practices and make recommendations for adjustments. We Sirisena (2023) and provide more detail in this blog post. We next turn to implementing the recommendations made in the report, and measuring whether this has had a positive impact on our team, with the goal of **refining our organizational value proposition and operating model**.

To begin, we've created new roles dedicated to overseeing our team's system of work as well as product strategy and user research. We've also invested in People Operations capacity in order to design support and growth plans for our team. We believe this will allow us to make more strategic improvements to our platform and to the upstream communities we support.

1.4.2 Grow our service and runway through fundraising

As of this writing, we have approximately 11 months of runway remaining. This is within the lower end of our "comfort range", and the cost of our new hires create a sense of urgency to raise funds for the next phase of 2i2c's operations. The organizational refinement described above sets up 2i2c for a subsequent phase of growth in the communities that we serve, in order to cover these new costs.

We are currently exploring a combination of grants, donations, and service contracts, with a goal of extending our runway by another three years. Once we make the necessary adjustments to improve our operating model described above, we will next focus on this goal to ensure that 2i2c may continue to provide value to open science communities across the world. By improving our value proposition, product, and service operations, we will improve our chances of bringing in funding to extend our organizational runway and move us closer towards long-term sustainability.

1.5 Acknowledgements

2i2c's success would not be possible without the critical support it received from CZI at its founding. The capacity provided by this grant was a crucial moment in crystallizing the vision of 2i2c and taking our first steps as an organization. We are honored to have had this chance to build an impactful, open, and ambitious organization like 2i2c, and we're excited about the next step in our organization's journey.

Chapter 2

2. Year 3 report (final year)

The following project assessment covers the third and final year of our operations. It covers the dates from 10/2022 through 11/2023.

2.1 Brief summary of year 3

In year 3 of 2i2c's operations, we expanded 2i2c's operations with several new community partnerships, development partnerships, and grant-funded projects to provide guidance to communities that wish to use open source infrastructure for open science. In addition, we underwent significant organizational changes as the team began to feel strain brought on by our increasing commitments. We spent several months introspecting and identifying where we needed to make improvements to our structure and team capacity, and finished the year by hiring critical new roles. As we head into 2024, 2i2c aims to incorporate these roles into our organization and improve our operational model as a result, we then aim to fundraise for the next phase of 2i2c's organizational growth.

> **Note**
>
> This section focuses on recording and linking to deliverables and milestones. For a narrative about 2i2c's current situation and next steps, see the grant-wide summary.

2.2 Key outcomes

2.2.1 Strategic planning and capacity building for Jupyter in research and education

In the final year of this grant, 2i2c significantly expanded its partnerships both in its open science cloud platform, as well as its strategic and development efforts. Below is a table of funds that we raised for a combination of services and grants:

Type of funds	**Funds raised in Year 3**
Development contracts	$175,000
Open science cloud service contracts	$330,745
Grants awarded	$1,083,541
Total	**$1,589,286.00**

Notable milestones in several major categories are described and linked below.

2.2.1.1 Notable community partnerships

- We began a partnership with the US Greenhouse Gas Center to run Jupyter infrastructure for their data teams.
- We partnered with the CryoCloud project for Cryosphere research. This led to a project leader Tasha Snow receiving a prestigious open science award with AGU.
- We began two new major educational collaborations with the Canadian Callysto project for high school learning as well as running Jupyter infrastructure for UC Merced.

2.2.1.2 Development efforts

- We began a collaboration with GESIS to bring environment building to BinderHub, and documented the outcome of this collaboration in this blog post.
- We began a grant-funded project from CZI that provides interactive computing services for historically marginalized global communities.

- We began a partnership with CILogon for institutional log-ins.
- We brought QGIS to a JupyterHub workflow for earth science communities.
- We began a collaboration pilot with HHMI to provide JupyterHubs for biomedical workflows.

2.2.1.3 Leadership in open science technology and cloud

- We were the co-authors on Rose et al. (2023).
- We were awarded a Munroe et al. (2023), a collaboration between 2i2c and MetaDocencia.
- Our Partnerships Lead, Jim Colliander, authored a response to a NASA Call for Proposals on the value of hosted cloud environments for open science. Colliander & 2i2c Community (2023).
- The 2i2c team authored a response to a NASA Call for Proposals on the value of ephemeral interactive computing environments. Colliander et al. (2023).
- The OpenScapes community gave 2i2c's open science cloud model a strong endorsement. OpenScapes on their cloud environment with JupyterHub.
- OpenScapes also authored a response to the same NASA Call for Proposals that included a pitch for Nickles et al. (2023).

2.2.1.4 Organizational growth and maturation

- We Sirisena (2023).
- We created and hired for five new roles on our team (links to descriptions of each below):
    - Partnerships Lead
    - Community Success Lead
    - Technical Content Specialist
    - Product Lead
    - Chief of Staff and Delivery Manager

2.2.2 Facilitating communications and connections within the Jupyter developer community and its stakeholders in research/education

In our final year, 2i2c's team authored more than 700 merged pull requests in key upstream repositories. We also engaged in several major community initiatives listed below.

Jupyter-specific conversations and open source engagement/leadership…

- We released a blog post on Principles for contributing to Open Source and encoded this in our Open Source Strategy section of 2i2c's team compass.
- Our engineer Georgiana Dolocan published a blog post about her professional journey from Outreachy intern to 2i2c engineer on the Jupyter blog.
- Our engineer Sarah Gibson served as JupyterHub's Community Strategic Lead and led a cohort of Outreachy interns.
    - We partnered with The Turing Institute on a second phase of JupyterHub's Community Strategic Lead work, which we will begin later this year.
- We piloted a new form of meeting with the JupyterHub team called the Collaboration Cafe
- We helped the Binder project navigate and adjust to a loss in our cloud infrastructure resources.
- We expanded the Executable Books steering council and led efforts to incorporate many of the project's repositories into the Jupyter project.

2.2.3 Enabling the execution of projects and collaborations for Jupyter in research and education

We did not focus on this goal in the final year of the grant, because we had team capacity from other funds to lead these efforts.

- We created a new role to help improve 2i2c's system of delivery and planning. This role will lead efforts around this goal in the future.

2.3 Artifacts, publications, and software code

Blog posts

- Bringing QGIS to Jupyter and JupyterHub via 2i2c's service
- 2i2c running cloud infrastructure for the US Greenhouse Gas Center
- An organizational report describing 2i2c's growth bottlenecks and challenges
- Georgiana Dolocan's journey from Outreachy intern to 2i2c engineer
- Describing our partnership and integration of CILogon for institutional authentication
- Exploring our strategy around ethically engaging with open source communities.
- Our annual update in year 2 of 2i2c's operations
- Announcing our collaboration around user-driven image building in JupyterHub with GESIS.
- Describing the completion of the GESIS project and the improvements we made
- Introducing JupyterHub's Outreachy interns
- A new format for JupyterHub's team meetings: the Collaboration Cafe

Blog posts and videos from community partners about 2i2c

- OpenScapes on their cloud environment with JupyterHub
- CryoCloud's cloud platform for Cryosphere research

Publications and grants about or by 2i2c

- Analyzing wildfire data with NASA EarthData cloud, a collaboration between 2i2c and MetaDocencia.
- The value of hosted JupyterHub instances for community learning, by the OpenScapes team, in response to a NASA CfP on the value of hosted cloud environments for open science.
- Building the open source science stack by Jim Colliander, in response to a NASA CfP on the value of hosted cloud environments for open science.
- The value of ephemeral interactive computing for open science communities by the 2i2c team, in response to a NASA CfP on the value of hosted cloud environments for open science.

Videos and presentations

- No Magic Added Deploying Multiple JupyterHubs to Multiple Clouds from one Repository, by Sarah Gibson
- Accelerating Discovery for NASA Cryosphere Communities with JupyterHub, James Colliander
- Building a Cloud Computing Infrastructure for Research and Education, Chris Holdgraf
- On Building open source infrastructure for research, education, and local communities, Jim Colliander
- Sarah (2023), by Sarah Gibson
- Gibson (2023), by Sarah Gibson

Chapter 3

3. Year 2 report

The following project assessment is from year 2 of 2i2c's operation (08/2021 through 09/2022). It is copied from our previous report to CZI, as a historical record of the project's progress.

3.1 Brief summary

In year 2 of 2i2c's operations, we gradually **scaled the communities that we served in order to test and refine our new service model**. This helped us understand the bottlenecks that existed in our previous model, and the ways in which it did not match the communities that we work with. As a result we have come to a better understanding of the roles and collaborative model to use around this service, which we will define, implement, and scale in the final year of this grant.

3.2 Key outcomes

3.2.1 Strategic planning and capacity building for Jupyter in research and education

Below are several major accomplishments over the last year that grew our capacity to partner with research and education communities:

- We assessed our "alpha" service offerings and pricing, in order to set a foundation to use for partnering with new communities.
- We overhauled our JupyterHub infrastructure to be more reliable and scalable.
- We defined a shared responsibility model to provide agency to our partner communities.
- We defined a new Product and Community Lead role and hired James Munroe to serve in this role for 2i2c.
- In anticipation of scaling up our partner communities in 2023, we also defined a Partnerships Lead role to oversee our efforts in making new connections with partner communities, and to develop sustainability models out of these efforts.
- We partnered with several organizations around an open science cloud service for Latin America and Africa. For full details, see 2i2c et al. (2022).
- At the end of this period, we audited our strategy and service model thus far, and identified several key areas where we needed to invest our time moving forward.

3.2.2 Facilitating communications and connections within the Jupyter developer community and its stakeholders in research/education

In addition to our strategic and capacity building efforts above, 2i2c's Executive Director also led a number of key efforts at growing connections and capacity within the Jupyter ecosystem.

- Stewarded the Notebook v7 enhancement proposal.
- Led efforts to refactor and update the JupyterHub governance and team structure.
- Oversaw a major overhaul of the PyData Sphinx Theme and assisted with several efforts to update Jupyter documentation to build on this theme for its own documentation.
- Set a foundation for the Binder project to grow its organizational support.
- Secured another year of cloud funding for mybinder.org via a grant of cloud credits from Google Cloud.
- Co-led efforts to define a formal governance structure for the Jupyter Book and MyST projects, and began laying the groundwork for incorporation into the Jupyter project.
- Served on the Steering Council for NotebooksNow!. An effort to bring notebooks into modern publishing workflows for scientific communities.
- Began serving as an advisor for JupyterHub's first Community Strategic Lead role, a role that was funded from a previous year's EOSS grant from CZI.

3.2.3 Enabling the execution of projects and collaborations for Jupyter in research and education

While most effort on this grant was spent on major strategic and capacity building efforts within 2i2c, we wish to highlight some of our operational accomplishments that required extra management effort.

- We designed a new Project Manager role that is dedicated to managing our deliverables and planning for major projects within 2i2c.
- Doubled the number of simultaneous partner communities in our Managed JupyterHubs Service from about 20 to 41 community hubs with clusters in each major cloud provider.
- Partnered with several strategically-important communities, including a new NSF-funded center called the Alabama Water Institute, the NASA Cryo in the Cloud project, the LEAP-Pangeo project, and the Jack Eddy Symposium.
- We began a development partnership with GESIS to bring Binder's dynamic image building capabilities directly into the JupyterHub technology stack.

3.3 Areas for growth

We identified a number of ways we must improve our service moving forward. Much of these were described in our annual strategic update blog post. Below is a brief summary:

- Engineering roles are not enough to achieve the impact we wanted via our Managed JupyterHub Service. Communities also need support in *using* the infrastructure. This led to the creation of the Product and Community Lead role described above.
- Our pricing and service offerings must be further-refined to reflect the diversity of communities we wish to serve. We must explore models that are inclusive to *a global population*, particularly those with fewer resources (we describe some of these challenges in our recent CZI proposal for communities in Latin America and Africa).
- *Billing and invoicing* quickly became a bottleneck that limited the number of communities we can partner with at our current capacity (and cost recovery model). We are exploring ways to more sustainably invoice the communities that we work with - both for 2i2c and for their administrative teams.

3.4 Artifacts, publications, and software code

Below are links to public reports and blog posts that 2i2c produced over the past year. Many of these posts are linked above with more context as well.

- **[Blog post]** Q3 Community Update
- **[Blog post]** Q1 Community Update
- **[Blog post]** One year later: an update of 2i2c's mission, strategy, and impact
- **[Grant application]** 2i2c et al. (2022) (CZI-OS invited application)

In addition, 2i2c Executive Director co-authored several pre-prints and published articles that mentioned 2i2c's model as a part of domain-specific visions for cloud-enabled collaborative interactive computing. Below are links to notable publications, blog posts, pre-prints, and presentations given by personnel funded on this grant.

- **[Article]** DuPre et al. (2022) (PLoS Computational Biology)
- **[Article]** Uchida et al. (2022) (European Geosciences Union)
- **[Preprint]** Rokem et al. (2021)

Finally, these presentations were given on behalf of the Jupyter community at various workshops and events.

- **[Presentation]** Scientific Communication and Reproducible Publishing in the Jupyter Ecosystem and Beyond (AGU 2021)
- **[Presentation]** Jupyter Book 101: Create Beautiful, Publication-quality Books and Documents from Jupyter Notebooks and Computational Content (AGU 2021)
- **[Presentation]** Holdgraf (2022) (Jupyter Community workshop in Maastricht)
- **[Presentation]** Holdgraf & Community (2022) (NotebooksNow! 2022 workshop)

This is a brief update of major developments with 2i2c over the first nine months of its existence (and of this grant). Much of this information also exists in various places in the 2i2c Team Compass, which is an open resource about the 2i2c team, our major projects, and our organizational structure. For a high level view of our strategy and goals, see the Team Compass strategy page.

Chapter 4

4. Year 1 report

The following project assessment is from year 2 of 2i2c's operation (12/2020 through 08/2021). It is copied from our previous report to CZI, as a historical record of the project's progress.

4.1 Major developments

4.1.1 Funding and grants

As this is a capacity building award, we list below funding that we have facilitated or applied for over the past 9 months. A subset of these grants are directly for 2i2c, while many of them are for other communities and open source projects that align with 2i2c's mission. All of them are related to infrastructure and services that improve Jupyter's impact in research and education.

Item	Date	Awardee	Source	Description	Amount	Status
Pangeo Infrastructure Support	2020 - Fall	2i2c	Moore Foundation (sub-award from Columbia)	Building and running JupyterHub and cloud infrastructure for the Pangeo Project	$479,295 / 2 years.	Awarded
JupyterHub Community Strategic Lead	2021 - Spring	JupyterHub	CZI EOSS	Funding a strategic lead role for the JupyterHub Community to improve diversity and inclusion dynamics.	~$400,000 / 2 years	Awarded
PyData Sphinx Theme development	2021 - Spring	PyData Sphinx Theme	NumFocus development grant	Development and maintenance for the PyData Sphinx Theme, the parent theme used by Jupyter Book	$25,000	Awarded

Item	Date	Awardee	Source	Description	Amount	Status
Infrastructure for Scalable Science Institute	2021 - Spring	2i2c	NSF	Development and hosted infrastructure in collaboration with UW, UCB, JH universities for a scalable science institute	~$1.5m / 5 years	Waiting
Educational Hub Infrastructure Pilot	2021 - Spring	2i2c	JROST Rapid Response Fund	Funding cloud infrastructure for educational JupyterHub pilots	$5,000	Awarded

4.1.2 Managed Hub Service revenue

In addition to funding from these grants, we have also been developing potential revenue and sustainability models for the Managed JupyterHub Service. We are still revising these models and using our JupyterHub pilots (see below) to gain experience and arrive at the best pricing and sales structure. We anticipate finalizing these models and running prototypes of revenue-generating Managed Hub Contracts in the second year of this grant.

4.1.3 New team members

We've welcomed two new members to the 2i2c core team. These individuals will both work towards 2i2c's major projects, and collaborate together on running our 2i2c Pilot Hub infrastructure. Here's a bit about each new team member.

- Damián Avila. Damian joined 2i2c in May. He has been a Jupyter core team member for many years now, and has done work across many different parts of the PyData stack (in particular, Jupyter, Bokeh, RISE, and Nikola). Damián will focus his efforts on supporting JupyterHub infrastructure for the Pangeo project, as well as development across the Executable Books Project
- Sarah Gibson. Sarah joined 2i2c in June, after spending several years as a Research Software Engineer at the Turing Institute. She has also been involved with the Turing Way for many years. Sarah will focus her efforts on JupyterHub development and operations for the Pangeo community.

4.1.4 Governance and a code of conduct

2i2c has made important steps towards defining a stable and transparent organizational model moving forward. 2i2c now has a Steering Council and an early organizational structure. In addition, we defined a one-year bootstrap strategy that we'll use to guide our path in the first year of 2i2c's existence.

One of the first acts of the Steering Council has been to adopt a Code of Conduct. This is a set of guidelines, and a process for resolving incidents, that makes our community more inclusive, equitable, and enjoyable for all.

4.1.5 Early pilot JupyterHub infrastructure

In addition to organizational structure and foundations, we have also made progress along the JupyterHub deployment infrastructure we wish to provide and support. One of our major organizational goals is to build a sustainable service managing open source cloud infrastructure for interactive computing. This service will provide hosted, customized JupyterHubs for communities of practice in research and education. They'll be built entirely with open source tools that are community-driven, and that respect the customer's Right to Replicate.

In order to accomplish this, 2i2c is running several pilots with partners and interested organizations. These pilots are meant to be learning opportunities to understand what kind of infrastructure and service it needs to build moving forward.

The documentation for our pilot hubs infrastructure contains information about our deployments and infrastructure. It is served from this 2i2c-org/pilot-hubs repository, a centralized location for configuring and deploying a federated network of JupyterHubs. Each JupyterHub is independent of one another, and could be spun out from the centralized repository with minimal extra work, giving hub users the ability to replicate their infrastructure, with or without 2i2c. We will continue refining the code in this repository as we learn more from our hub infrastructure pilots.

This page has a list of our currently running JupyterHubs. Each one roughly corresponds to a community that we are serving (e.g., a classroom, a research group, or an institution). As of this report there are around 30 JupyterHubs that we manage.

4.1.6 JupyterHub for geospatial analytics - A collaboration with Pangeo

As originally announced on the Pangeo blog, 2i2c is forging a collaboration with the Pangeo project around operating and developing cloud infrastructure for large-scale geospatial analytics. This collaboration is funded through a grant from the Moore Foundation (via Pangeo investigator Ryan Abernathey).

2i2c will assume operation of infrastructure underlying the Pangeo project, allowing the Pangeo team to focus their efforts on their core scientific and development missions. Once this is complete, we'll next shift our attention to some new areas of development that support use-cases in the Pangeo community (and in the scientific community more broadly).

4.1.7 JupyterHub for education - A collaboration with CloudBank and UC Berkeley

We've begun a partnership with the UC Berkeley Data Science in Undergraduate Studies program, as well as CloudBank. This collaboration aims to provide hosted JupyterHub infrastructure for community colleges across the state of California. It is an attempt at providing vendor-agnostic and open-source infrastructure to several institutions who would otherwise not be able to deploy this infrastructure on their own.

2i2c will provide the deployment and configuration architecture for this collaboration, working with Sean Morris in operating this educational infrastructure. All of the cloud infrastructure for this pilot will be funded via CloudBank. We will begin by offering environments that are modeled after the Data 8 course at UC Berkeley. This is part of an effort to build a community of practice around Data Science education using open source tools.

4.2 Areas for growth

In the next phase of this grant, we aim to continue executing on our major strategic priorities, briefly described below:

- **Managed JupyterHub Service:** after gaining experience running infrastructure with our Managed JupyterHub Pilot, we plan to crystallize and prototype a service model for managed infrastructure. This includes an infrastructure plan, a support plan, a sales and billing plan, and a pricing plan. There are still some major unknowns to resolve before this is complete - particularly, what kind of pricing structure will be the most effective at making this service accessible, scalable, and sustainable for us and for the communities we serve. This will require consultation with those familiar with non-profit / open source business and service models.
- **Organizational Foundations:** as 2i2c is still a young organization, there is much work to be done in laying the basic policies and practices of our organization. We are defining more distributed-friendly team coordination practices, and improving our project planning and milestones process. We also plan to revisit our governance and steering council structures to ensure that they are well-positioned to guide 2i2c's strategy moving forward. This will require consultation with other leaders in the non-profit sector to understand how we can improve our organizational practices.
- **Major collaborations and development:** we have launched several major collaborations in our first year, most notably with the Pangeo Project, and continuing a collaboration with the Jupyter Book project. In our next phase, we wish to develop an infrastructure strategy that allows development with these communities to complement one another, and feed into a broader infrastructure system that we can re-use and customize for other communities as well. This will require more technical design in partnership with key stakeholders across the Jupyter and open source community.

4.3 Unexpected changes

2i2c has experienced one major unexpected change, which was transitioning our fiscal sponsorship from the International Computer Science Institute to Code for Science and Society. The decision to change fiscal sponsors came after several months of discussion with the ICSI team, and both organizations ultimately decided that it was in our best interests to shift our organizational relationship to one of collaboration rather than fiscal sponsorship. This has delayed some of 2i2c's planned operations by several months, but we do not anticipate that it will have a major impact on our ability to execute on the long-term mission of this award. ICSI has done an excellent job of facilitating the transition to CSS, and we expect to have this completed within a month or two.

Chapter 4

References

2i2c, Carpentries, T., Center For Scientific Collaboration And Community Engagement, Invest In Open Infrastructure, MetaDocencia, & Open Life Science. (2022,). *A Collaborative Interactive Computing Service Model for Global Communities*. Zenodo. https://doi.org/10.5281/ZENODO.7025288

Colliander, J., & 2i2c Community. (2023,). *Building the Open Source Science Stack*. Zenodo. https://doi.org/10.5281/ZENODO.7662828

Colliander, J., Avila, D., Dolocan, G., Gibson, S., Holdgraf, C., Munroe, J., & Panda, Y. (2023,). *Ephemeral Interactive Computing for NASA Communities*. Zenodo. https://doi.org/10.5281/ZENODO.7892224

DuPre, E., Holdgraf, C., Karakuzu, A., Tetrel, L., Bellec, P., Stikov, N., & Poline, J.-B. (2022). Beyond advertising: New infrastructures for publishing integrated research objects. *PLOS Computational Biology*, *18*(1), e1009651. https://doi.org/10.1371/journal.pcbi.1009651

Gibson, S. (2023). *Embedding a Right to Replicate in RSE practices*. https://doi.org/10.5281/ZENODO.8228653

Holdgraf, C. (2022). *Open Infrastructure for Open Science*. https://doi.org/10.5281/ZENODO.7233586

Holdgraf, C. (2024,). *2i2c three year report and retrospective for our founding grant from CZI*. Zenodo. https://doi.org/10.5281/ZENODO.10790818

Holdgraf, C., & Community, T. E. B. (2022). *Jupyter Book and MyST: A community-led, extensible, modular ecosystem for creating computational narratives*. https://doi.org/10.5281/ZENODO.7287626

Munroe, J., Palopoli, N., & Acion, L. (2023). *Reproducibly Analyzing Wildfire, Drought, and Flood Risk with NASA Earthdata Cloud*. https://doi.org/10.5281/ZENODO.8212073

Nickles, C., Friesz, A., Hunzinger, A., Barrett, A. P., Lind, B., Welch, J., Lopez, L., Jami, M., Thornton, M., Robinson, E., Lowndes, J. S., & Community, T. N. O. M. (2023,). *The Value of Hosted JupyterHubs in enabling Open NASA Earth Science in the Cloud*. Zenodo. https://doi.org/10.5281/ZENODO.7667299

Rokem, A., Dichter, B., Holdgraf, C., & Ghosh, S. S. (2021). *Pan-neuro: interactive computing at scale with BRAIN datasets*. https://doi.org/10.31219/osf.io/mwh2b

Rose, B. E. J., Clyne, J., May, R., Munroe, J., Snyder, A., Eroglu, O., & Tyle, K. (2023). *Collaborative Research: GEO OSE TRACK 2: Project Pythia and Pangeo: Building an inclusive geoscience community through accessible, reusable, and reproducible workflows*. https://doi.org/10.5281/ZENODO.8184298

Sarah, G. (2023). *Reproducible Computational Environments with Binder*. https://doi.org/10.5281/ZENODO.10075621

Sirisena, D. (2023,). *Organizational review of 2i2c: Opportunities and challenges for our structure and operations*. Zenodo. https://doi.org/10.5281/ZENODO.10081003

Uchida, T., Le Sommer, J., Stern, C., Abernathey, R. P., Holdgraf, C., Albert, A., Brodeau, L., Chassignet, E. P., Xu, X., Gula, J., Roullet, G., Koldunov, N., Danilov, S., Wang, Q., Menemenlis, D., Bricaud, C., Arbic, B. K., Shriver, J. F., Qiao, F., … Wallcraft, A. (2022). Cloud-based framework for inter-comparing submesoscale-permitting realistic ocean models. *Geoscientific Model Development*, *15*(14), 5829–5856. https://doi.org/10.5194/gmd-15-5829-2022