Invest in Open Infrastructure: *Exploring Hidden Costs of Open Infrastructure (Mellon Foundation)*

Kaitlin Thaney (Executive Director, Invest in Open Infrastructure)

Proposal Information

Project Title: Exploring hidden costs of open infrastructure

Amount Requested: $135,125 USD

Grant Start Date: February 1, 2021

Duration (in months): 15

Program: Public Knowledge

Description of Proposed Work (1-2 paragraphs):

Invest in Open Infrastructure (IOI)'s work is centered around providing targeted, evidence-based guidance, analysis, and strategic support to institutions and funders to improve the resourcing, viability, and sustainability for open technologies that underpin research and scholarship. IOI is a fiscally sponsored project of Code for Science & Society, a leading 501(c)3 that supports a portfolio of projects in the public interest technology and research sectors.

In the attached proposal, *"Exploring hidden costs of open infrastructure"*, we request $135,125 USD for a 15-month initial exploration of the costs associated with ongoing maintenance, sustainability, and resourcing of the technology, tools, and systems that scholarship and research depend on. Funding will be dedicated in large part to bringing on a Research Data Analyst, whose work will be centered around building out a funding data prototype for the sector and further interrogating and analyzing the costs associated with running and maintaining key open infrastructures, building on initial findings from our Future of Open Scholarship research.

Proposal Narrative

Proposed Activities and Rationale:

Invest in Open Infrastructure (IOI) was launched to help institutions and funders better invest in the open technology and systems that research and scholarship rely on. This effort, arising from conversations and shared frustration about the perceived scarcity of funding and support for open infrastructure services to serve research and scholarship in a sustainable way, was founded in August 2018. Initially a coalition committed to acting collaboratively to sustain the open infrastructure critical to the future of scholarship and open knowledge, IOI is now actively working to fulfill its mission through research and community engagement.

Between August 2018 and March 2020, IOI existed solely as a volunteer effort, led by a 20 person Steering Committee of infrastructure providers, institutional leaders, and advocacy organizations. This early stage laid the groundwork for the collaborative spirit and open communication across organizations that is central to this effort. In late 2019, the effort secured initial funding from Schmidt Futures and the Alfred P. Sloan Foundation, leading to the hiring of IOI's inaugural Executive Director (Kaitlin Thaney) in March 2020. The effort is a fiscally sponsored project of Code for Science & Society (CS&S), a leading 501(c)3 that supports open collaboration in the public interest technology and research sectors through fiscal sponsorship and other programs supporting sustainable open source. CS&S's project portfolio includes initiatives such as Measurement Lab, Open Refine, PREreview, Reproducibility for Everyone, and the Research Software Alliance. (CS&S Executive Director Danielle Robinson is also an active member of IOI's Project Advisory and governance.) CS&S has a founding mission that centers on advancing the *"charitable purpose of helping people, as well as the scientific, academic, and public institutions they serve, to more effectively use data to address social and economic problems, and to enhance our lives"* and robust open infrastructure across all scholarly domains is core to this work. CS&S and IOI believe that in order to build and sustain open infrastructure for scholarship we must collaborate with a broad scholarly community, including the humanities, arts, economics, sciences, and social sciences. While domain-specific challenges exist and should not be minimized, digital object and/or data management are now critical in nearly all fields. As such, the common issues of digital infrastructure management, resourcing, and maintenance are dominant across fields, and evidenced by the increasing commercialization of publishing workflow tools, content and data sharing platforms, and more.

IOI's efforts to date have focused on working in concert with institutional decision makers, infrastructure providers, philanthropic funders, and other supporting organizations to explore how open infrastructure is developed and sustained today. This includes documenting the decision points, funding and governance models available, and costs associated with maintaining, sustaining, and scaling open infrastructure projects. We define open infrastructure as the sets of open services, protocols, standards and software that the academic ecosystem needs in order to perform its functions throughout the research lifecycle — from the earliest phases of research, collaboration and experimentation through data collection and storage, data organization, data analysis and computation, authorship, submission, review and annotation, copyediting, publishing, archiving, citation, discovery and more. *Open* infrastructure represents the narrower sets of services, protocols, standards and software that can empower communities to collectively build the systems and infrastructures that deliver new improved collective benefits without restrictions to participation, engagement, or usage.

The need for shared, open infrastructure transcends disciplinary boundaries through, as examples, shared content repositories, digital preservation services, knowledge discovery tools and platforms. The

call for open infrastructure also empowers specific tools and platforms needed for more specialized research and scholarship (e.g., computational notebooks, capture and archiving of digital art and images, publishing workflow tools, etc.). The The European Science Foundation's Standing Committee on the Humanities in 2011 issued a policy brief *"Research Infrastructure in the Digital Humanities"* that stated for some humanists "...an infrastructure is the technical and operational framework that allows them to collaborate and share data and results; for some it is the content to which access is offered rather than the facilities around it; and for some it is both." In the last decade, the challenges to maintaining a robust scholarly open digital infrastructure have become clear. At the core of our work around open infrastructure are platforms, tools and services that enable knowledge creation and sharing, as well as the places for scholarship and learning to occur to understand the societal context in which open knowledge sits.

Since hiring Kaitlin Thaney in March 2020, IOI has been conducting research on open infrastructure and its changing landscape, gathering and analyzing data, and crafting resources and reports to share those findings with decision makers, tool providers, and members of the research community. IOI's work is centered on employing an evidence-based approach to guiding decision makers on where to invest, at the university administrator and director level, at the level of infrastructure project leadership, and in partnership with other funders to support their visions. To accomplish this, IOI looks outside traditional open research silos to collaborate with experts in management, business, finance to shed light on costs, risks, and models to employ and produce open reusable data and resources. Unlike major consultancies, however, our work is community governed and deeply rooted in open knowledge, ensuring what is produced for the community is approachable, actionable, and reusable. In the first nine months IOI has spearheaded a major research effort on the future of open scholarship and its infrastructure, convened over 500 members of the open source, research, and scholarly communities for the Joint Roadmap of Open Science Tools (JROST) conference, and modeled a Rapid Response Fund to support maintenance costs for the open infrastructure community, distributing $50,000 USD to projects in need. In 2021, we will add capacity in key areas (including the research data analyst position described in this proposal) to launch research that will build a strong case for continued investment in open infrastructure.

The problems we seek to solve (and why)

1. *Open and community-owned infrastructure projects remain resource-constrained, or perceive resource constraint, despite healthy profits for commercial entities*

IOI was founded out of shared frustration in the limitation of funding and sustainability of open source and non-profit infrastructure to support research and scholarship. This is in comparison to the commercialization and acquisition of both for and non-profit services such as bepress, Social Science Research Network, Mendeley and more by an ever concentrated series of commercial players like Elsevier and Clarivate. As Katherine Skinner, Executive Director of Educopia Institute (and IOI Steering Committee member) wrote in July 2019 after surveying 45 key infrastructure programs and organizations as part of the Mellon-funded "Mapping Scholarly Communication Infrastructure Project", "Academy-owned" and "academy-governed" (a connotation used by Skinner in her analysis) tools, platforms, and services have been well trained to run on as little funding as possible, are rewarded for building new tools/platforms/services (rather than maintaining a solid base for existing work), and are set up to compete with each other for increasingly scarce resources. She further states that we are missing key information to explore the power of the scholarly communication sector. Without knowing how much money is currently spent on scholarly communication as a field, the community is unable to measure, grow, or leverage its own market power.

2. *Shifts in philanthropic giving have the potential to shore up open infrastructure, but critical data on the true costs of long term sustainability are missing*

At the same time, philanthropies have also been more actively exploring shifts towards understanding and investing in maintenance and development of shared infrastructure. The Siegel Family Endowment recently released a white paper declaring a new strategic focus for grantmaking dedicated to infrastructure, which will lead to over $20M in grantmaking this year. The Ford Foundation, Alfred P. Sloan Foundation, Omidyar Network, Open Society Foundations, and the Mozilla Open Source Support awards program recently concluded their second request for proposals for their Digital Infrastructure funding program (the first round supported by the Sloan and Ford Foundations in 2018).

The Sloan Foundation, who are also founding supporters of IOI and ex-officio advisors of CS&S, have expressed interest in exploring grant making opportunities around maintaining open infrastructure more explicitly, as well. These foundations, despite their size, typically lack the internal team of analysts that you'd find in an impact investing fund or other commercial investment shop to outline opportunities, market needs, and risks. Additionally, without the data on the initial and long term costs, philanthropic investments may fail to set up infrastructure for long term impact.

Exploring funder data answers an explicit call from members of the Open Research Funders Group, a partnership of philanthropic organizations committed to open sharing of research outputs, and representative of many of the key funders of open infrastructure in research and scholarship. Their 2018 survey on open infrastructure identified significant interest in coordinating funding across member foundations, but called for additional insight to direct that support.

3. *COVID-19 has illustrated the increased urgency for investment in stable digital infrastructure while institutions face unprecedented budget and staffing shortages*

In addition to these shifts, the past year further highlighted the need to better understand the costs and maintenance needs of shared research infrastructure as teaching, learning, and research moved quickly online. The global pandemic and response to COVID-19 increased demand for and a call for more readily available, equitable and open access to research with an economic crisis that brought into even more stark relief the need for effective and strategic investment in the core systems that support open research and scholarship. Suddenly, institutions were facing staffing cuts and hiring freezes as well as mandates to quickly rebalance and cut spending, the consequences of which help significant implications for the future of open research and scholarship. The sudden shift to online learning and pace of research also led to a need to make decisions about solutions and platforms that could be rolled out instantly across entire faculty and student populations, in some cases, at the expense of privacy, data sovereignty, and academic freedom.

IOI's approach over the last 9 months

To understand the sector and define the problems articulated above, IOI launched a call in June 2020 for participation in a research effort called the "Future of Open Scholarship" project designed to provide support in outlining costs, benefits, risks, collective action opportunities, and assist in scenario planning to support open scholarship and infrastructure amidst the confluence of global health, economic, and racial justice crises of 2020. This 6-month project included over 75 hours of user interviews with institutional staff and decision makers, press and scholarly society directors, infrastructure providers and more, to better understand their dependencies and histories with supporting open infrastructure projects, gain a sense of the budget implications they were facing, and work collectively on modeling costs to enable faster, more informed decision making in support of open scholarship. Participation in

this project was entirely self-selected, with 18 countries represented, 76 institutions and organizations, and 112 participants ranging from core developers on infrastructure projects to Deans of Libraries and heads of publishing outfits like eLife. (A full participant list can be found here. Additional reporting will be shared in early 2021 as IOI concludes this project.)

This work was intentionally scoped to support communities invested in furthering open scholarship, with heavy representation from those in or adjacent to institutional libraries, and does not represent the entirety of the audiences IOI views central to its work or to "open infrastructure" work more broadly.

In the interviews and workshops conducted for the Future of Open Scholarship project by IOI Executive Director Thaney, she heard repeatedly from library budget owners the need for more information to guide their investment decisions — ranging from ways to understand the efficacy of their current spend as they weigh waves of budget cuts to understanding on an ecosystem-wide level how their level of investment lines up to other institutions and how their funding is being allocated to maximize their return on investment based on their institutional needs. IOI has also heard frustration at lack of information regarding the "invisible" costs of maintenance, staffing, and resourcing, which can hinder appropriate budgeting for support to keep key services afloat to provide access to knowledge and scholarship. We have begun to create the means for institutional leads to discuss and analyse elements of their investments through the costs and benefits modeling work conducted over the past few months by Kate Pugh as part of the Future of Open Scholarship work. That work includes tools and resources for institutional leaders to employ in their own work and budgeting, indexing core costs across institutions and partners where collective investment in open infrastructure is currently difficult to quantify (e.g., open repository services, investment in open library services platforms such as FOLIO, etc.) The focus on funding data represents a first step in building out a foundational understanding of investment in open technology investment, so that we can begin modeling additional revenue streams and systems of support (often augmenting institutional investment) for use by institutional leaders, members of the funding community (to compare their investments across the sector, identify gaps and potential concentrations in support, etc), and for infrastructure providers.

The demand and need for this level of insight, analysis, and data is currently being met by for-profit consultancies (such as McKinsey & Company, Accenture, and Bain & Company) detached from the work and needs of the community, who are designing investment strategies for top-level university administration and budget holders (ie, Vice Provosts and Deans of Research and Libraries). In addition, commercial players such as Elsevier and Clarivate (formerly Thomson Reuters) spend significant amounts on lobbying industry and university leaders, as well as lobby against government policy, in ways that serve their profit margin and business motives. The analyses being produced are often expensive, opaque, and cursory, as evidenced in the University of Arizona's $14M contract with McKinsey & Company in 2019, the largest contract of its kind in the university's history, with many details of the work elusive to the public due to the firm's demand for secrecy, claiming their methods and strategies are trade secrets. IOI has heard from Vice Provosts and Deans of Research directly that these strategic efforts often fail to represent the complexity of current investments and adoption of core infrastructure services and more often cater to commercial offerings over open technology solutions the institution may already be investing in. As Brandon Butler, the Director of Information Policy at the University of Virginia shared from a budget conversation with the school's head of Information Technology "you wouldn't question Microsoft or Google suite licenses for students, so why are you questioning a membership fee to support our institution's repository at a fraction of the cost?" Butler's comment represents one dimension of the challenges in making the case for open infrastructure investments to take priority, with branding, stability, and scale just some of the additional components, even for an institution with a history of being an early adopter and supporter of open source software solutions and open infrastructure.

Discussions over cost, value of service, and return on investment are complicated. In the Future of Open Scholarship project, Jennifer Vinopal, the Associate Dean for Distinct Collections and Digital Programs at Ohio State University stated that for her, it was less about showing that choosing to invest in open solutions were cheaper in the near-term, but more so an opportunity to better outline (and not shortchange) the investment needed to make open infrastructure the better, more competitive choice, even if that university expenditure would increase. IOI heard repeatedly over the past 6-months of research in the Future of Open Scholarship project the challenges from institutional leaders, heads of scholarly societies like the American Geophysical Union, and university press directors in the tradeoffs and tensions that lie in needing a robust product in the near term to serve the needs of their faculty, students, and researchers, while still prioritizing open, community-governed and developed solutions that are more in line with their values and mission. The reasons cited for the lack of competitiveness in open solutions range from inadequate funding to scale and grow offerings as a for-profit entity would, lack of product or business expertise, and staffing constraints. As our work progresses, IOI is exploring ways to provide avenues for conversations among institutional leaders who are grappling with those realities of efficiency, levels of service, and a desire to choose more values and mission aligned software and infrastructure solutions, to help facilitate more collective problem solving and peer-to-peer support.

The data needed to understand current investment in open infrastructure for research and scholarship is disaggregated and difficult to analyse given its distributed nature and complex accounting. Core to building out a better understanding of how open infrastructure offerings and projects are currently supported is data from leading governmental and philanthropic funders, data from institutions about their spend and in-kind support for open infrastructure offerings they employ and contribute to, and other external investment and cost-sharing support provided by consortia, national funding programs, and commercial vendors like EBSCO and dedicated software development vendors like Atmire who contribute to open source code bases and support initiatives through funding directly.

IOI is currently working with a renowned data scientist to do a feasibility study and initial analysis of philanthropic investment across a subset of foundations known for their investment in open infrastructure.[1] IOI intentionally chose to start with funder data due to its availability and perceived influence on the sustainability of open infrastructure projects. The funders included extend beyond the scholarship/library scope of IOI's Future of Open Scholarship research to also account for investments in research tooling and infrastructure (such as computational frameworks and tools for data analysis). The decision to begin with funding data also helps us model a foundational dataset for use by institutional leaders and technologists that represents a core financial driver and support mechanism for open infrastructure projects employed within institutional contexts today. While we endeavor to work towards approaches that elicit more data about institutional investments (and will through the targeted use case portion of this proposed work), we also recognize that for the staffing and timeline of this work the hurdles that have been identified in the findings of the "Mapping the Scholarly Communication Infrastructure" project in quantifying institutional investment. Project lead Katherine Skinner, the Executive Director of the Educopia Institute, recently said, " The single biggest finding that they could document across the board from interviews with library deans, was that [institutional leaders] have no way to measure investment, given the degree to [that data] being intertwined with other data that makes it impossible to extricate". We also have heard similar in our Future of Open Scholarship research, where the breakdown of investment across full time employees, library collections budgets, departmental budgets (often multiple, across the institution), in-kind support, other services that are accounted for under "overhead" and not detailed, make drawing a clear picture of institutional spend a challenge that

[1] This work is being led by Q Ethan McCallum, an industry data scientist and strategist out of Chicago who has worked extensively with nonprofits such as DataKind, and clients in the corporate and financial sector.

would warrant additional staffing and time to execute, which we view as beyond our capacity currently. We are continually looking for models and creative ways to approach gaining additional insight into institutional investment that minimize burden and can be systematized, looking to examples such as the University of Michigan's Institute for Research on Innovation in Science, a consortium that uses data architecture to map internal data from procurement systems, human resources, and sponsored projects to track things like job creation, research impact, and outputs related to government funding. Such approaches take years of development and testing. We are currently tracking and assessing potential approaches and collaborations more broadly for IOI's work in this space to creatively tackle the challenges identified above more systematically and effectively.

The work in this proposal, "Exploring hidden costs of open infrastructure" seeks to outline a means to begin to shed light on costs and current levels of investment via increasing our analytic capacity and further building on a funding data prototype we began this fall.

Previous and related work

Prior to the pandemic, there were efforts underway for universities to re-negotiate and even cancel "Big Deal" journal subscription bundles with publishers like Elsevier. There were conversations bubbling to the surface in philanthropy about infrastructure as an essential service (as noted above in the Digital Infrastructure funding calls and the Siegel Family Endowment's recent grantmaking strategy shift), and with groups like The Maintainers forming to advocate for a shift in focus to maintenance, rather than building from scratch. Efforts such as the Open Platforms Group, a coalition of top US research libraries, were planning ways to identify and pool resources to better sustain a few key pieces of infrastructure across their 30+ member institutional network. Community led efforts, like the JROST conference, centered on exploring ways open tools and services could better integrate with one another, rather than compete.

With the recent emergence of non-profit decision making analysis tools and resources such as Unsub, SPARC's Big Deal tracker, and the Academic Observatory from the Curtin Open Knowledge Initiative, we've seen a shift occur away from for-profit research/business intelligence tools to offerings curated, grown, and provided by the community. These tools and resources have fostered an increase in institutional leaders' confidence in making more informed decisions about their research priorities and investments, in some cases leading to cancelling expensive contracts leading to millions of dollars of cost savings for institutions like the State University of New York (SUNY) system, University of North Carolina, Chapel Hill, and Iowa State University, saving them anywhere from high hundreds-of-thousands to several million dollars in one go. There has been growing momentum over the past few years towards cancellations, aided by data on pricing and usage (evidenced here and here) as well as models set by groups such as the University of California system and Massachusetts Institute of Technology (MIT).

What's missing today is clear guidance on where to best re-invest those savings - be it from a Big Deal cancellation, a collections budget re-allocated towards open technology, consortia dues, or a pooling of funds from private philanthropy and institutions to provide a more sustainable and effective means of supporting open infrastructure in higher education. Our goal with this work - and more broadly for IOI - is to meet that need in a way that not only supports top-line budget owners and decision makers, but democratizes that analysis and information for the broader research and technology communities, as well.

Through our work to shed light on the costs and economic factors associated with the maintenance of open infrastructure as well as our efforts to further investigate data about the current funding landscape, we aim to add an additional dimension to support institutional leaders in understanding how to adequately resource the core systems and technologies they rely on for research and scholarship, provide a lens for funders to ask questions about the current flow of capital to the sector, and to help infrastructure providers better model their costs upfront to enable more sustainable resourcing and support for their important work.

IOI's work is also informed and shaped by research done to examine the challenges and opportunity costs of digital public infrastructure (outside of the research sector) and open source software development, more broadly. That includes Nadia Egbhal's report "Roads and Bridges: The Unseen Labor Behind our Digital Infrastructure" and her recent book, "Working in Public: The Making and Maintenance of Open Source Software". Both of these works interrogate a number of key assumptions about open source development, especially those around the economics of open source development, maintenance, and participation. Eghbal challenges the assumption that open equals free (or cheaper), and proposes additional frames for thinking about the costs of producing open source software, separate from the cost of consuming open source software, positing that they should be viewed as two separate economic goods. Both of these works also dig into participation in open source development and the relation to product maintenance, and have been useful in elevating IOI's thinking this past year as we began to explore effects of budget cuts and staffing reallocations and furloughs on the health and stability of core infrastructure services, as many of those staffing costs for project development, maintenance, and governance are poorly accounted for but viscerally felt when they are missing.

IOI's work is also influenced by the research of Frank Nagle, an assistant professor in the Strategy Unit at Harvard Business School, and Shane Greenstein, a professor at Harvard Business School, co-director of the HBS Digital Initiative, and co-director of the program on the economics of digitization at the National Bureau of Economic Research. Together, their work explores the economics of IT, Internet infrastructure collaboration, and crowdsourcing, specializing in that of the open source software movement. Their 2013 working paper in the National Bureau of Economic Research on "Digital Dark Matter" in particular sought to assess the rate of return on open source software solutions like Apache and to quantify the value of technology and commerce built on top of open solutions. Nagle and Greenstein's work present an inspiration and a model for IOI to explore for open source solutions in research and scholarship, to assist in our work to better quantify and articulate to budget holders what the return on investment is for choosing open source solutions over proprietary offerings that may not offer the same flexibility, transparency in pricing, and/or data security and privacy.

We aim with this effort (and our research more broadly, moving forward) to explore ways to further incorporate this level of economic thinking about value adds, associated and hidden costs, and rates of return for adopting, developing, and maintaining open infrastructure tools and services for research and scholarship.

The work outlined in this proposal is informed by and designed to build on our preliminary research and prototyping efforts to explore the problem space around hidden costs and funding for open infrastructure in research and scholarship. That work takes two forms, both of which are relevant to this proposal, and especially to the research data analyst we seek to hire, as that role is scoped to further the modeling, analysis, and investigative work conducted over the past six months at IOI.

One body of work is our Future of Open Scholarship research, a six-month project with 100+ decision makers from institutions and supporting research and scholarship organizations, outlining their needs,

constraints, opportunities for collective action, and first cuts at costs and benefits models to bolster institutional decision making (in particular to argue for more investment in shared infrastructure). Over the course of that project, we've collected over 75 hours of interviews with participants on their budget constraints, dependencies on and attitudes towards open infrastructure, and needs regarding cost modeling and tools to support their own budgeting within their institutions and organizations.

Through that project, we've been working with former industry analyst Kate Pugh, from AlignConsulting and Columbia University, who is modeling the costs and benefits of supporting open infrastructure within institutions as a collective, and has led a series of collaborative design workshops with participants of the project (consisting of Vice Provosts and Deans of Libraries, institutional staff, infrastructure providers such as members of the Public Knowledge Project and Our Research, and other scholarly communication initiatives such as preprint offerings housed outside of the institution) to explore further questions around return on investment, reliance on commercial off the shelf software, relationships between vendors and open infrastructure support, and how best to approach quantifying participation in open projects at the institutional level. That work is ongoing through January 2021, and will be written up and shared with project participants as well as the broader community.

We are also actively working on the initial phases of data exploration to examine available funding data from the philanthropic sector and a sampling of government agencies. That work is also slated to continue through January 2021, and has given us a baseline understanding of what's available to ground the work articulated in this proposal. To date, with the help of industry data scientist and consultant Q Ethan McCallum, we have pulled the past 1-3 years of funding data from a variety of philanthropic funders.

In our research to date, information about the costs for open research infrastructure development, adoption, and maintenance are difficult to come by, often contributing to challenges at the institutional level in making a more robust investment case for better resourcing for critical open infrastructure relied on by an institution.

What needs to be done

The work outlined in this proposal will build off of initial explorations of funding data to scope an initial expanded dataset and prototype, as well as initial costs benefits modeling efforts from our Future of Open Scholarship project. With the help of a data scientist on short-term contract as well as a former industry analyst, we've begun work on pulling and parsing available grants data to begin asking questions while also indexing models and variables to start exploring ways to quantify collective investment in open infrastructure.

Key phases of this work include:

(1) Recruitment and onboarding to build our research capacity

In the Schedule of Activities, we have noted a 2-3 month timescale to account for a competitive and equitable recruitment process for the Research Data Analyst position. This is an important role with noted complexity, as it may necessitate recruiting and onboarding an analyst from outside of the traditional scholarly communication and academic research analyst networks.

(2) Funding data exploration and modeling:

IOI's initial funding data prototype currently has data from 18 funding sources, pulled together with the aim to model an extensible framework and dashboard for interrogating philanthropic investment in open

technology and systems for research and scholarship. This work initially focuses on available philanthropic and government funding data intentionally to help establish a foundational dataset to build upon, beginning with data that is inherently more readily available through grant databases and listings than tracking institutional investment across departments, institutions, and discrete budget lines (not to mention in-kind contributions of staff, maintenance support, and other dimensions that often are accounted for under "overhead" and not easily extractable).

IOI aims in this work to craft a foundational base to incorporate additional data into as it is made available, including institutional data. Currently, there is no systematized way to access budgetary information from institutions without significant investments in building partnerships with procurement offices and top-level administration across institutions, as well as investments in staffing to handle data collection and harmonization, security, parsing, and analysis. The recently concluded "Mapping the Scholarly Communication Infrastructure" project funded by the Andrew W. Mellon Foundation ran a survey to assess library investments in Scholarly Communications Resources, building on David Lewis' 2.5% Commitment provocation, calling on library budget holders to invest 2.5% of their budget in supporting the open scholarly commons, of which open infrastructure investment is a part.

In following up with the project team as well as participants in that survey, concerns were voiced about how to best account for shared costs across the institution, staffing time and percentages viewed as "in-kind" or calculated as part of an overhead cost, and other "hidden costs" that made gaining insight into true investment into open infrastructure and other services challenging. While we aim to build towards incorporating more institutional data in the future, we recognize the limitations that exist in expanding our scope without additional capacity or funding beyond our request in this grant, and will endeavor to incorporate institutional data as it is made available over the duration of this funded work.

In addition to identifying and incorporating funding data sources, there is also a need to create and continually refine a shared data model to harmonize data sources for easier querying across. The level of data available from the available data sources varies greatly, with few data fields in common, especially when it comes to shared terminology for describing grants. Creating a data model that fosters a shared vocabulary across data sources will benefit the field and improve the landscape for future work in this area.

We currently have initial data from the following funding sources:

- Alfred P. Sloan Foundation
- Arcadia Fund
- Arnold Ventures
- Gordon and Betty Moore Foundation
- Howard Hughes Medical Institute
- James S. McDonnell Foundation
- John Templeton Foundation
- Lumina Foundation
- Rita Allen Foundation
- Robert Wood Johnson Foundation
- Templeton World Charity Foundation
- Andrew W. Mellon Foundation
- Chan Zuckerberg Initiative
- Institute for Museum and Library Services
- National Endowment for the Humanities
- National Science Foundation
- National Institutes of Health

- Department of Energy

The selection of initial funding agencies and organizations to explore began with members of the Open Research Funders Group, a partnership of philanthropic organizations committed to the open sharing of research outputs, holding assets in excess of $100 billion and representing many of the leading philanthropic funders of open technology projects in research and scholarship.

In addition to that group, we also identified a series of other key players in funding scholarly communications, open scholarship, and research. It's worth noting that funding organizations such as the Bill & Melinda Gates Foundation, Open Society Foundations, Siegel Family Endowment, Schmidt Futures / Eric & Wendy Schmidt Fund for Strategic Innovation, the Wellcome Trust, Leona M. and Harry B. Helmsley Charitable Trust, JISC, and the European Union are excluded from the initial dataset curated due to a lack of data available for parsing (e.g., not available on their website, no funding data/amounts listed, etc.).

The data gathered so far provides us with an ability to ask some high-level questions regarding spend and for whom — which institution/organization, categories of expenditure such as which funding portfolio a grant falls under, and/or year.

Additional data exploration and modeling would increase our ability to ask more targeted, filtered questions, such as how much funding has gone to projects based in western institutions versus other parts of the world and what does that tell us about possible inequities in the sector, where are there repeated or redundant investments and what can we learn from that, what trends emerge when we compare investments across major funders over a series of years (declining/increasing spend, shifts in portfolios, drop-off times after repeated investment cycles, favoring feature based development versus ongoing support, etc). Dedicated research staff would also enable us to request more detailed information from various funding sources and identify external datasets to join with our data, such as demographic and descriptive data (e.g., categorization of institutions, such as historically black colleges and universities, community colleges, etc.) As this exploration evolves, we will also continue our efforts to prototype an extensible framework and dashboard for interrogating philanthropic investment in open technology and systems for research and scholarship.

(3) Scoping open infrastructure use cases for further cost analysis:

A key part of our work in 2021 is to outline an approach and begin investigating the costs associated with supporting key infrastructures, hosting and maintenance costs, staffing support, margins, vendor relationships / outsourcing, in-kind support, and breakdown of past and current resourcing. (Examples include Open Library Environment / FOLIO projects, DSpace / Fedora) By framing the impact and benefits of open infrastructure with real costs, we can better understand the scale of the commitments required to sustain open infrastructure.

We know that many of the costs we're looking to investigate are difficult to assess and/or poorly documented, making the need for a creative, strategic approach to identify key use cases where data may be available in some form to help us get started an imperative. We also recognize that while we look to examples of this sort of analysis in other non-academic/research sectors, that this sort of analysis is new, and will warrant flexibility in our approach as we explore ways to reveal a fuller picture of project costs. We will work with the (to be hired) Research Data Analyst and our networks of stakeholders to choose 2-3 open infrastructure use cases to do a deeper analysis on.

In selecting open infrastructure use cases to focus on, we would look for the following:

- Projects that are open source and not-for-profit;

- Projects that serve the research and scholarly communication communities;
- Projects that are stakeholder governed;
- Accessibility of documentation of financial support of the project ideally over a 1-3 year timespan;
- Transparency and diversity of business models (e.g., membership-based support vs national subsidy);
- Adoption levels and usage as a proxy for influence and impact;
- Diversity in services represented;
- Other environmental considerations include:
    - Transformative influence (e.g., does this shift away from a dated, closed, less ideal model?)
    - Degrees (if any) of external investment (e.g., EBSCO investment in FOLIO and OLE)
    - Affordability and accessibility by underserved audiences

A primary factor in our decision making is the accessibility of adequate financial data for analysis. Given the distributed support nature of many of the open infrastructure services in use today, this may be the biggest challenge in ensuring we have the necessary level of detail to gain a broader picture of costs to model our analysis, while also ensuring we're effectively managing time in gathering that data.

As our Future of Open Scholarship research is actively being synthesized and written up, we have outlined a few areas of exploration to be further honed in partnership with our new hire once they are in place. Those proposed research questions include, but are not limited to:

- Costs associated with operating and maintaining open infrastructure:
    - Modeling transition costs (off or away from commercial, off the shelf software)
    - Costs associated with building interoperability among existing open source systems utilized in higher education (e.g., content and data repositories)
    - Costs and benefits associated with crafting, adopting, and employing open standards to facilitate system interoperability?
    - Exploration of the "hidden" or "invisible" costs associated with open, distributed infrastructure? (e.g., direct vs indirect costs, staffing and governance, ongoing maintenance)
- Current investments in the sector
    - How much is currently being invested in open infrastructure, tools and systems to support research and scholarship? Where is capital flowing?
    - How do funders compare in terms of investment level? What does that tell us?
    - What sort of themes emerge in that funding? Are there any obvious gaps or concentrations?
    - What's missing? (Funding data that's not available, demographics and/or projects that are less or not represented)

(4) Synthesizing and making data accessible to decision makers

We aim to share our process, analysis, and findings on an ongoing basis via publications, discussion sessions, and events geared towards institutions and funders.

What we need to accomplish these tasks

Our main need moving forward is dedicated research and data analysis capacity on the team, with a runway to explore the questions surfaced in our preliminary work in more depth. IOI as a formal project with dedicated staff began in March 2020 at the start of the global pandemic, and has been working to build research capacity while also serving over 100 members of the research community as they grappled with economic volatility, heightened demand for services and software to bolster research and scholarship, as well as reckoned with the inequities that pervade our system and limit access to knowledge globally.

We have been fortunate to receive smaller donations, gifts, and grants to provide short-term capacity (via a series of contractors) to bring in experts to accelerate our work and level of service in this time of need. To the extent possible, we have strategically worked to begin exploring key issue areas articulated by our stakeholders through the funding data prototype and modeling efforts. Dedicated analyst support will provide us with the means to build on those findings, expand our capacity to investigate and interrogate our existing models of supporting open infrastructure, and provide that analysis and data back to the community to inform their work and the sustainability of the sector.

In addition to staffing, we also are thinking creatively about how to gain access to data on costs for running, maintaining and staffing open infrastructure projects. That data, as evidenced in the Mapping the Scholarly Communication Landscape survey, can be difficult to track down given the complexities in how universities account for overhead and investments in shared infrastructure. We also have insights now into the data that is and is not readily available from funding organizations invested in supporting this sector. Additional data from those funders as well as external datasets to provide additional descriptive and demographic information will be essential to that body of analysis.

Means by which we'll make these results available to the community

We strive at IOI to ensure our work balances between trustworthy and evidence-based with a design that makes the topics we care about actionable, approachable, reusable, and accessible to our community of stakeholders.

We are committed to openness and transparency, and endeavor to share our process, findings, and work in formats that enable participation and reuse, and aim to openly license our materials under a CC-BY-4.0 license wherever possible.

We believe in iterative development and sharing early and often, and are committed to soliciting and incorporating feedback from the community into our work products and process via facilitated workshops, community calls, and in written form on our blog.

Description of the Project Organization

This project will be supported by the following staff:

- Dr. Danielle Robinson, Co-Executive Director, Code for Science & Society (Co-Principal Investigator)
- Kaitlin Thaney, Executive Director, Invest in Open Infrastructure (Co-Principal Investigator)
- Jessica Hardwicke, Sponsored Projects Service Manager, Code for Science & Society

Code for Science & Society team members (Robinson and Hardwicke) will support Thaney in administering this work, in providing services and support to assist with hiring, recruitment, strategic guidance, and financial reporting and administration.

Thaney will oversee the programmatic work outlined in this grant, including the direct day-to-day management of contractors and hired staff, onboarding new staff and contractors, scoping work, facilitating engagement with the open infrastructure community, and assisting with documenting and sharing learnings and findings.

Curriculum Vitae for both Dr. Robinson and Ms. Thaney can be found in the attached Appendices.

Invest in Open Infrastructure

Invest in Open Infrastructure (IOI) is governed by a (3) person Project Advisory, a (20) person Steering Committee, and is supported by over 200 institutions, organizations, and individuals worldwide. The initiative is supported by (1) full-time staff member, Kaitlin Thaney, IOI's Executive Director.

Invest in Open Infrastructure is a fiscally sponsored project of Code for Science & Society, a leading US-based 501(c)(3) public charity supporting the public interest technology and research sectors.

Code for Science & Society

Code for Science & Society (CS&S) is a US-based 501(c)(3) nonprofit supporting open source and collaboration in public interest technology through fiscal sponsorship and programs that support sustainable open source. As a top fiscal sponsor for public interest technology, CS&S serves projects from research-driven open data science to open civic data with strategic support and mentorship as well as financial and administrative services. Through years of this work, we know first hand the common gaps and oversights that projects struggle with and work to add capacity to our fiscally sponsored projects.

IOI is supported by CS&S President and Co-Executive Director, Dr. Danielle Robinson. She is a proven community leader and strategic advisor to the open source digital infrastructure community (CV attached). Dr. Robinson oversees a team of excellent Fiscally Sponsored Project (FSP) program staff and administrators to manage day to day project needs.

Consultants

We have budgeted for additional technical support to aid in building out the technical backend of the funder data dashboard and prototype, data pipelines for consolidating multiple data sources for easier

analysis and visualization, and additional software development support and expertise to augment the work of the staff Research Data Analyst.

We envision this work taking a form similar to the preliminary work on the prototype started this past October with data scientist and consultant Q Ethan McCallum, where we contracted with him for a set number of hours to complete an agreed upon series of deliverables. We chose to approach the contract in this blended way (of both hours-based as well as anchored in deliverables) due to the exploratory nature of the initial work and to be responsive in our design as we investigated the availability and usability of existing funder data online.

Technology

The work outlined in this proposal will build off of initial explorations of funding data to scope an initial expanded dataset and prototype. With the help of a data scientist on a short-term contract, we've begun work on pulling and parsing available grants data to begin asking questions of, and testing initial ways to display and query the data that will provide a jumpstart to the work in developing a dashboard for use by the broader community, as outlined in this grant proposal.

We're building our tools on well-known, well-documented, actively-maintained open source libraries for pulling and parsing the grants data. This solid foundation ensures that the prototype (and it's more production-ready, official progeny) will last, and will shorten the learning curve for any new developers we bring in for future maintenance and enhancements. IOI aims to make these tools as openly available as possible for use by the wider community, so they too can ask questions of the data to fuel their analysis.

Specifically, we're using the following Python libraries:

- Scrapy - for pulling raw content
- BeautifulSoup - for extracting data fields from the raw content

For data storage, display, and analysis, we have yet to finalize our decision on which tools to use for the funding data dashboard and ad-hoc data exploration to power IOI's internal analyses. We're looking into offerings from Amazon Web Services (QuickSight, ElasticSearch) as well as Google Cloud (BigQuery, Data Studio, Looker) for internal use. We recognize the complexity of using commercial offerings from Amazon and Google for this work (as an initiative designed to support and advocate for open, values-aligned infrastructure), and will be doing a further assessment of business intelligence and data exploration tools as the work evolves.

Any additional modeling would utilize well-known, open source libraries such as scikit-learn, a free and open source software machine learning library that enables predictive data analysis in Python, and Keras/TensorFlow, open source machine learning platforms and tools (Keras is a deep learning API that runs on TensorFlow, which is an end-to-end open source machine learning platform).

Initial recommendations for the tools and services mentioned above to display and analyze data come from a leading data scientist and consultant, Q Ethan McCallum, who has worked with and advised both for profit and not-for-profit / charity efforts to explore similar data analysis problems.

Schedule of Major Activities

Below is a proposed schedule for major activities associated with this proposal. We have requested 15 months for this work to account for the hiring of our Research Data Analyst (for which we're requesting one year's worth of salary).

Key activities include:

- **Recruitment & Hiring of Research Data Analyst**
  - **Estimated**: February - April 2021
    - Job posted: February 2021
    - Recruitment and hiring: February - mid-April 2021
    - Anticipated start date: May 2021
  - **Staff lead:** Hiring search and recruitment will be driven by IOI Executive Director, Kaitlin Thaney. Additional Human Resources support to run the hiring process will be provided by Code for Science & Society staff and leadership.

- **Onboarding & initial data exploration**
  - **Estimated:** May 2021
  - **Staffing:** Led by IOI Executive Director with new Research Data Analyst.
  - **Additional detail**: This phase will focus on building the foundational data set including an assessment of data sources, articulation of gaps in available data, and initial descriptive exploratory data analysis and visualization. The goal of this is to describe the available data and its limitations, as well as how data sources or standardization could be developed. This work will lay the foundation for the Research Data Analyst's work, set a benchmark for data availability and condition at the start of the work, and articulate key areas of work on funding data moving forward (e.g., articulate the need for open data in key areas, data or metadata standardization)

- **Dataset development, standardization, and analysis**
  - **Estimated start**: May-June 2021.
  - **Staffing**: Led by Research Data Analyst (to be hired) in collaboration with IOI Executive Director.
  - **Additional detail**: **Data collection & modeling**. For the funder data dashboard and initial dataset, we will work to solicit additional data and context about the grants made from funding sources outlined in our initial exploration. We will also work to identify other publicly available datasets to augment the funder data, such as classifications of institutions and organizations, institutional funding data (as it is made available/discoverable), in addition to the data collection and modeling work that is involved with the individual open infrastructure use cases.

- **Scoping open infrastructure use cases for further cost analysis**
  - **Estimated start**: May-June 2021.
  - **Staffing**: Led by Research Data Analyst (to be hired) in collaboration with IOI Executive Director.
  - **Additional detail:** The hired Research Data Analyst will work with IOI steering committee and project leadership to research key open infrastructure offerings (building on the "Mapping the Scholarly Communication Infrastructure" project's research) and evaluate for participation against outlined criteria, resulting in the finalized selection of projects.

- **Communication of findings with community / key stakeholders**
  - **Estimated start**: August - September 2021, ongoing through end of grant

    - **Staffing**: Led by Research Data Analyst (to be hired) in collaboration with IOI Executive Director
    - **Additional detail:** The Research Data Analyst will share out their work through a series of whitepapers, community feedback sessions, and the creation of other resources (e.g., slide decks, figures, interactive data visualizations).

Ongoing activities (duration of the grant):
The following are activities that will increase the scope of the core dataset, analysis, and modeling to further our understanding of maintenance costs and flows of capital (philanthropic and government). We envision these activities as ongoing tasks of the Research Data Analyst, with the support of IOI leadership.

- **Data collection & modeling** (estimated 30%). For the funder data dashboard and initial dataset, we will work to solicit additional data and context about the grants made from funding sources outlined in our initial exploration. We will also work to identify other publicly available datasets to augment the funder data, such as classifications of institutions and organizations, institutional funding data (as it is made available/discoverable), in addition to the data collection and modeling work that is involved with the individual open infrastructure use cases.
- **Ongoing analysis and refinement** (estimated 40%). For both the funder data dashboard and the work specified to investigate project maintenance costs, we expect the explorative nature of this work to involve continuous analysis as we investigate the datasets and questions outlined above, as well as ongoing refinement. By "refinement", we mean the work to standardize, tag, categorize, and summarize data sources before they're used for advanced analytics. This work is noted as ongoing as in choosing to bring a data source online (for example, for the funder data dashboard), we are committing to make that data source available in a certain normalized format, consistent with our data model and usable for our analyses and that of the broader community.
- **Soliciting and incorporating community feedback** (estimated 15%). We aim to share this work out over the duration of this grant via monthly IOI community calls, as well as through more focused community listening and feedback sessions to share findings, explore needs and challenges, and further iterate on our analysis and prototype. These stakeholders will be defined as part of the process following the selection of open infrastructure use cases and discussion of next steps for developing the funder data dashboard.
- **Sharing out results** (estimated 15%) on a regular basis via community calls, virtual workshops/feedback sessions, working papers, blog.

Expected Outcomes

This work helps us shine a light on so-called "invisible" or "hidden" costs of open technology and infrastructure for research and scholarship, from the staffing and resourcing of participating in open source development and maintenance, to the costs associated with maintaining system integrations at scale.

Outcomes and outputs of this work include:

- **Investment guidance strategy for funders to improve coordination and maximize impact in OI:** Our work looking at funding data provides an opportunity to begin asking questions at a portfolio level about funding allocations, equity in where capital is flowing over a series of years, and also

call for broader transparency across the funding landscape to enable more robust analysis by organizations like ours.
    - **Outputs:**
        - Investment guidance strategy for funders (report)
        - Funding data dashboard, for funders to compare their investment and analyse gaps in support; for infrastructure
- **Increased investment in open infrastructure by institutions:** We will produce modeling tools and resources for key stakeholders to use in their work to advocate for budget allocation, staff, and demonstrate return on investment and the net present value of investing in open infrastructure.
    - **Outputs**:
        - Modeling and budgeting resources and toolkits
        - Community listening and feedback sessions
- **Infrastructure projects are able to resource maintenance effectively:**
    - **Outputs:**
        - Usable data on project sustainability and maintenance costs
        - Case studies on 2-3 OI project business and sustainability models
        - Funding data dashboard (see above)

Success for this grant will be measured by increased availability of resources and use by key stakeholders (institutional decision makers, funders, and infrastructure providers) to better coordinate funding and resource core open infrastructure work. Tangible outcomes include reports and analysis probing at the research questions outlined in this work, This work also builds our capacity as an organization to build our shared evidence-base to heighten awareness and convene a bigger conversation in the sector about sustainability, while also providing tactical, tangible assessments, pointed analysis, and resources for the community to utilize.

Long-term Sustainability of Project Results

We want our work to be as replicable and reusable as possible to ensure that resources are usable and sustainable over the long-term. To do that, we will ensure that documentation and reporting of process, findings, and prototypes is made available openly. Results will be freely usable or downloadable from our site, and published open access when appropriate. We will also make code and data available in repositories and backed clear documentation and appropriate metadata for others to access, reuse, and build upon.

In addition to those efforts, we are governed by a Project Advisory and Steering Committee who are responsible for stewarding the work of IOI should there be a change in leadership or change in project status.

Proposal Details

Collaborators

Please comment in this section on any organizations collaborating with the applicant organization that would be significantly involved in the conception, development, and/or execution of the proposed

activities. Consultants and contractors are not considered collaborating organizations for the purposes of this section and should be listed in the section titled Consultants and Contractors.

Does the proposal involve collaborating institutions and individuals?

No.

Collaborators Details:
If yes, please list all collaborating organizations; for each organization, briefly describe the nature of the collaboration. If a collaborating organization would be receiving grant funds, please list the amount in the Grant Budget and in the Budget Narrative section.

Contractors and Consultants

The Foundation's guidelines for grants involving consultants and/or contractors apply to proposals where a significant portion of the grant funds would be paid to a third party in exchange for services. Note that the Foundation does not ordinarily consider universities that partner with a grantee to be consultants or contractors for the purposes of these guidelines.

Does the proposal designate a significant portion of the grant funds to be paid to consultants, and/or subcontractors?

No.

Consultants, Subcontractors, and/or Vendors Details:
If yes, please briefly describe the proposed grant activities for which a significant portion of the grant budget would be paid to a third party in exchange for services. Amounts allocated to consultants and contractors should be provided in the Grant Budget and Budget Narrative sections.

Diversity and Inclusion and Anti-Discrimination

The Foundation is committed to diversity and inclusion in its grantmaking programs and to ensuring that its funds are deployed in workplaces and educational environments that maintain and enforce policies committed to safety, dignity, ethical conduct, and freedom from discrimination. Please complete the questions below.

Diversity and Inclusion

Please describe how your organization defines and approaches diversity and inclusion in relation to its mission and operations. We ask that you include one or more examples of challenges and successes the organization has experienced with respect to diversity and inclusion. (Note: if the proposed grant is intended specifically to address diversity and inclusion, this should be described in the Proposed Activities and Rationale document).

Attention to Diversity

Code for Science & Society (CS&S), Invest in Open Infrastructure (IOI), and all CS&S' Sponsored Projects are committed to meaningful equity and inclusive organizational practices.

We define diversity along multiple axes including race, class, historical representation or exclusion in a profession or field, geographic location, gender, sexuality, and people with access or communication differences.

Scholarship, science, and technology are human endeavors that are subject to influence by the biases of people. In the pursuit of knowledge, people have often replicated power structures that reinforce the disenfranchisement and exclusion communities. In the American context, this has historically centered on the exclusion of marginalized communities from work in scholarly and technical communities. CS&S strives to be a community leader on issues around meaningfully inclusive public interest technology across domains. CS&S does this through our Sponsored Projects and Collaborative Communities Programs. We believe that it is core to our mission that we share our process to establish meaningful equity occurs day to day at our organization

When working with Sponsored Projects, CS&S focuses on helping projects grow sustainable, inclusive culture through evolving governance, working transparently, encouraging open dialogue around building leadership skills. Our Collaborative Communities program focuses on building capacity in the open source ecosystem. The Open Source Alliance for Open Scholarship Handbook Project, includes a frequently referenced definition of Open Scholarship, where equity and inclusion are central. Ongoing work with inclusion professionals DeEtta Jones & Associates (DJA) will focus on the challenging conversations about inclusion that are happening (and often not happening) in the open source and scholarly space. Our 2019 work with DeEtta Jones is summarized here. In 2020 we deepened this relationship by engaging DJA's team to help guide development of anti-racist nominating and governing body onboarding processes at both CS&S and IOI.

As the nonprofit home of multiple sponsored projects, CS&S will oversee project leaders and teams, who will also collaborate with staff at outside organizations. By centering transparency and governance with our projects and leaning in to organizational and community growth, we hope to continue to lead as a voice for meaningful inclusive practices in open scholarship and open source.

Our approach and examples of how we work for meaningful equity and inclusion:

1. **Diversity in our Collaborators:** When collaborating with partners, we recruit diverse perspectives to participate, be it through user-centered design processes, governance, or to speak to our community.
2. **Equity and Representation in Staffing:** We offer structure and benefits to support our project's to recruit and retain talented people from all backgrounds.
3. **Centering Inclusion in the Project Governance:** CS&S has been actively engaged in supporting open source community to question, iterate, and mature governance models. As discussed above, in 2020 we engaged equity experts DeEtta Jones & Associates to help us to develop and implement governance processes for ourselves and our sponsored projects that meaningfully centers equity and anti-racist values. We will share the results of this process transparently in 2021.
4. **Proactively Seeking Opportunities to Engage with the Global Community:** It can be hard for small projects on limited budgets to adjust to the needs of a growing community. There are unique funding opportunities for open projects to get expert diversity support. As an example, PREreview was selected to convene a working group of international experts at 2019 TriangleSci to focus on bringing equity and diversity to peer review, and a Wellcome Trust grant specifically focused on Diversity and Inclusion.

Our current challenges:

1. **Decentering whiteness:** In majority-white and white-lead spaces the default mode of operating centers the white experience. As a white-led nonprofit working in scholarship, science, and technology our white staff are challenged to de-center their experience to build projects and programs with room for diverse voices. This work is challenging, personal work.
2. **Traditional nonprofit governance is inherently white-supremacist:** Nonprofit governance structures are based on corporate structures that are not designed to be collaborative, equitable, or inclusive. We leverage the transparency required of nonprofits to share our process in developing governance.
3. **Structures/practices meant to enhance diversity, equity and inclusion are not always implemented in ways that are meaningful:** Good intentions of project or program leaders may fail to be meaningful to the communities they are supposed to support (see also, decentering whiteness). Majority white (male, abled, etc) spaces may, for example, add diverse people to a governing body but then fail to give that governing body power or convene them infrequently.

In addition to the support outlined above by IOI's fiscal sponsor, CS&S, IOI is committed to openness and transparency, and endeavors to share our process, findings, and work in formats that enable participation and reuse. Transparency and access are two critical components of the healthy, inclusive ecosystem we are building towards, and we aim to reflect that in our work.

Underlying all of our work is a framework anchored in ensuring the work we do and recommendations we provide foster a healthier, more equitable and inclusive research ecosystem. That weaves through every aspect of our work - from our governance bodies to our decision making tools and the research we produce.

We are, in collaboration with CS&S leadership, working with inclusion professionals DeEtta Jones & Associates, on an anti-racist governance structure, to evolve our current 20-person Steering Committee and Project Advisory into a governance body that is more actively representative of the communities affected, communities served, and aware of the systemic inequities that exist at the cross-section of open source technology, higher education, funding and capitalism, and long-term sustainability/time horizons. That work builds on a series of focus groups with existing Steering Committee members, IOI supports, and funders (including representatives from the Sloan Foundation), and has led to a Nominating and Governance group to support IOI ED Kaitlin Thaney in carrying that work forward. For a full list of our existing Steering Committee, see the Conflict of Interest statement in the attached Appendices.

Anti-Discrimination:
Please briefly describe your organization's equal opportunity, anti-discrimination, and/or anti-harassment policies.

Anti-Discrimination

CS&S is committed to providing a work environment free of unlawful discrimination and harassment, including sexual harassment.

CS&S policy prohibits unlawful discrimination, sexual harassment, and/or harassment based on race, religious creed (including religious dress and grooming practices), color, national origin (includes language use and possession of a driver's license issued to persons unable to prove their presence in the United States is authorized under federal law), ancestry, physical disability, mental disability, medical

condition, genetic information, registered domestic partner status, marital status, sex (including pregnancy), gender, gender identity (including transgender identification), gender expression, age for individuals over forty years of age, sexual orientation, military and veteran status of any person, or any other consideration made unlawful by federal, state or local laws ("protected classification"). It also prohibits unlawful discrimination and/or harassment based on the perception that anyone has any of those characteristics, or is associated with a person who has or is perceived as having any of those characteristics. All such discrimination, sexual harassment, and/or harassment is unlawful and prohibited by the CS&S.

CS&S's anti-discrimination/anti-harassment policy applies to all persons involved in the operation of the CS&S, including all CS&S employees, supervisors and those in management, as well as all persons doing business with or for the CS&S including vendors, customers, independent contractors, and others who enter the workplace (e.g.,. "third parties").

CS&S's anti-discrimination/anti-harassment policy prohibits unlawful harassment by any employee of CS&S (including supervisors, managers, and co-workers of the above-listed persons) or by any third party. Applicants, employees, unpaid interns, volunteers and independent contractors are all protected from discrimination, sexual harassment, and/or harassment under this policy.

Discrimination and harassment based on a job applicant or employee's protected classification (defined above) is against state and federal law.

Sexual harassment is a form of gender discrimination. Both state and federal law prohibit discrimination and harassment based on a job applicant or employee's gender.

There are two recognized types of sexual harassment under state and federal law: Quid pro quo and hostile work environment. The definitions of both forms of sexual harassment are as follows:

- "Quid Pro Quo" Sexual Harassment. The essential elements of this type of harassment are unwelcome sexual advances, requests for sexual favors or other verbal, visual or physical conduct of a sexual nature when:
    - Submission to the conduct is made either explicitly or implicitly a term or condition of an employee's employment, or
    - Submission to or rejection of the conduct by an employee is used as the basis for employment decisions affecting that employee.

- "Hostile Work Environment" Sexual Harassment. The essential elements of this type of harassment are:
    - The employee affected was subjected to harassing conduct directed toward him or her, or the employee personally witnessed the harassing conduct and it took place in their immediate work environment;
    - The employee's gender was a motivating factor for the harassment;
    - The conduct is unwelcome and sufficiently severe or pervasive that it has the purpose or effect of altering the conditions of employment and creating an intimidating, hostile, abusive, or offensive working environment;

- The environment created by the conduct would have been perceived as intimidating, hostile, abusive, or offensive by a reasonable person in the same position as the affected employee; and
- The environment created was perceived by the affected employee as intimidating, hostile, abusive, or offensive.

Prohibited unlawful harassment based upon sex (gender or pregnancy), or other protected characteristics (age, race, national origin, etc.) includes, but is not limited to, the following behavior:

- Verbal conduct such as epithets, derogatory jokes or comments, slurs or unwanted sexual advances, invitations or comments;
- Visual conduct such as derogatory and/or sexually oriented posters, photography, cartoons, drawings or gestures;
- Physical conduct such as assault, unwanted touching, blocking normal movement or interfering with work because of sex, race or any other protected basis;
- Threats and demands to submit to sexual requests as a condition of continued employment, or to avoid some other loss, and offers of employment benefits in return for sexual favors; and
- Retaliation for having reported or threatened to report harassment

Sexual harassment does not need to be motivated by sexual desire to be unlawful or to violate this policy. For example, hostile acts toward an employee because of his/her gender can amount to sexual harassment, regardless of whether the treatment is motivated by any sexual desire.

CS&S needs, expects and encourages you to come forward, without delay, should you suspect that any form of discrimination, sexual harassment, and/or harassment has occurred in the workplace. CS&S takes all complaints regarding discrimination, sexual harassment, and/or harassment in the workplace seriously. Additionally, all CS&S events are subject to the CS&S Code of Conduct, and all CS&S Sponsored Projects are expected to maintain community-appropriate conduct guidelines. If you feel you have been subject to discrimination, sexual harassment, and/or harassment, please notify CS&S Executive Director or your manager immediately, or use the anonymous conduct reporting form here, which notifies CS&S Executive Director immediately. Discrimination, sexual harassment, and/or harassment in the workplace will not be tolerated.

Any employee, regardless of position or title, whom the Company determines has engaged in discrimination, sexual harassment, and/or harassment in violation of this policy, will be subject to discipline, up to and including unpaid suspension and/or termination of employment.

Financial Health

Has your organization experienced any financial difficulties and/or deficits in the last three fiscal years?
No.

If yes, please address any financial difficulties and/or deficits your organization has experienced in the last three fiscal years.

Code for Science and Society, Inc. has reported positive financial results and increasing net assets since its inception in 2016. There have been no deficits, and each of the organization's Form 990 filings since inception have reported revenues in excess of expenses.

Leadership Changes

Has your organization recently experienced, or does it anticipate, any leadership changes and/or significant staff turnover?

If yes, please describe any recent or anticipated leadership changes and/or significant staff turnover relevant to the proposal grant activities.

Invest in Open Infrastructure

History of Leadership: IOI was formed in 2018 and stewarded by a Steering Committee of leaders and advocates for open infrastructure in research and scholarship. In late 2019, IOI raised funds to hire an inaugural Executive Director to develop IOI from a volunteer-led coalition into an established non-profit initiative. IOI was established as a fiscally sponsored project of CS&S in late 2019, and a search resulted in hiring Kaitlin Thaney as IOI's inaugural ED in March of 2020.

Planned Leadership Changes: None

Code for Science & Society

History of Leadership: CS&S was founded in 2016 by Max Ogden who operated the nonprofit until 2017. Dr. Danielle Robinson and Joe Hand were appointed Co-Executive Directors by the Board, each with a specific programmatic focus. CS&S has operated with a Co-Executive Director structure for three years.

Planned Leadership Changes: In January 2021, due to growth in the organization requiring a shift in leadership strategy and with the support of the CS&S Board of Directors, CS&S will transition Dr. Danielle Robinson to sole Executive Director and Joe Hand to Operations Director.

Intellectual Property

Do proposed grant activities include the digitization of works or the creation of digital technology and/or digital products, such as software, databases, audio or video recordings, podcasts, and websites?

If yes, please provide a detailed account of the intellectual property to be created, any rights or permissions that your organization would need to secure, the means by which the technologies and/or content would be distributed, including the type of license that your institution would issue to users, and how your organization would ensure the long-term sustainability of any digital or software products.

Invest in Open Infrastructure (IOI)'s work involves producing documents, databases, surveys and other landscape analysis tools, and conduct workshops and interviews with a wide range of stakeholders. All content will be made available under a CC-BY license with the exception of sensitive personnel and/or funding information. It is the stated goal of the IOI project that these outputs are participatory in nature and of use to the wider community.

We anticipate this work to lead to the production of blog posts to share out research progress, working papers/preprints (where appropriate), and supporting models and/or frameworks to supplement analysis. We also endeavor to share out updates on this work openly with the community via our IOI Monthly Community Calls, which will be launched in the new year.

Written work, reports, analysis, and any videos will be made openly available under a CC-BY-4.0 license, and be made available on our website at investinopen.org. We will also explore venues to share this work that align with our values to provide openly available and accessible research, including but not limited to open repositories and preprint services such as Zenodo and the National Bureau of Economic Research.

Code that may be generated as a part of this work will be made openly available via IOI's Github repository under an MIT or BSD license. Data and models generated for this work will be stored securely and made available to the fullest extent possible under the terms noted above.

As information products are created, we will also work with Code for Science & Society to track those items to help further long-term sustainability and stewardship of any intellectual property created or held by IOI.

We anticipate the following code, software, and data products to be developed in the duration of this grant:

- Funding data foundational dataset;
- Funder data dashboard to support exploration and visualization of funding data;
- Data model(s) and taxonomies to help create a common vocabulary across disparate datasets from funding agencies and organizations to enable more robust search and analyses;
- Scraping code to support data collection of funder data;
- Cost and benefits models and other analyses to interrogate support costs for core infrastructure use cases.

Our aim is to openly license and share these works with the community in a responsible and transparent way, while also ensuring we are compliant with data security and privacy rules that may apply.

Investment Income

Please describe how grant funds would be managed, including the overall investment strategy and asset allocation, and how income would be calculated and allocated to the grant. If the organization cannot by law invest grant funds in interest- or income-generating instruments, please explain why.

Code for Science and Society, Inc. currently maintains commercial checking accounts to administer the grant funds that it receives on behalf of its fiscally sponsored projects. The organization's policy is to invest all pooled funds in cash and cash equivalents, which it defines as cash and highly liquid investments with maturities of three months or less at the date of acquisition. This conservative policy is necessary in order for us to maintain liquidity on funds held for the benefit of our fiscally sponsored projects, as required under our agreements with them. The organization has not earned any investment income to date.

Strategies for Successful Grant Execution

Are there any factors that could potentially impede the timely and successful execution of grant activities and goals? (Consider, for example, the achievability of project timeline, availability of qualified personnel, ability to secure agreements, and capacity to obtain any remaining project funding).

Please outline these factors and briefly explain possible strategies for mitigating their effects on grant activities.

There are a few key factors that could affect the timely execution of grant activities and goals:

Recruitment and hiring of a Research Data Analyst. This will be our first full-time hire for IOI, and while we have onboarded a number of contractors, this will be our first time managing a recruitment process for staff. This role is one that we've modelled off of examples found in industry and for organizations like GiveWell, which are non-academic. To mitigate any delay or disruption, we plan to lean on the broader data science networks of the ED Kaitlin Thaney as well as the IOI Steering Committee, and strategically post and advertise the job ad in a variety of places to maximize exposure across networks. These include, but are not limited to, the IOI mailing list (which consists of over 550 members), the Harvard Business School network, listservs such as Code4lib, SPARC's member list, Coalition for Networked Information (CNI), as well as through LinkedIn, and via Twitter/social media. We will also work with inclusion experts DeEtta Jones & Associates (who both IOI and CS&S have been working with this past year) to identify additional job boards and networks to share this post with to ensure we are actively reaching communities that are traditionally underrepresented in this work. We will also work with our current data scientist on contract to help with the recruitment and onboarding process. In the event we need to restart or retailor our recruitment process, we will do so with expediency and keep Mellon program staff informed of any shifts in timeline.

Access to data on funding and budget spend for open infrastructure. Our initial work to gather and model available funder data has given us a base for which we can build from. There is likely still funder data that will not be made available in the duration of this grant, or may be kept back for fear of public scrutiny. The same goes for data on operating and ongoing costs for open infrastructure projects. In some cases, we know the accounting spreads across institutions, consortia, and other vendors/commercial partners.

For example, for a project such as DSpace, a leading open source repository offering, one would need to examine membership support across almost 100 institutions and partners, in-kind development support from institutions, outsourced support from institutional partners and dedicated DSpace software development vendors like Atmire, grant-based investments and other ongoing costs shouldered by LYRASIS, DSpace's consortial home. Those costs are not readily accessible for analysis, especially given the distributed nature to the work and development.

To mitigate this hurdle, we will work with our network of institutional leads, funders, and infrastructure providers to outline approaches to solicit financial and resourcing data through collaborations with project leads and supporting organizations to the best extent possible so as to not hinder our research and analysis. We will also ensure that accessibility of data is part of our selection criteria for open infrastructure use cases.

Grant Payment to Third Party

If the proposed grant is approved, will you request that the Foundation make payment to another organization (e.g., a university foundation) to administer grant funds on your organization's behalf? No.

If yes, please provide the name of the organization that will administer the grant funds and a brief description of the relationship with your organization

Matching Requirements

Would the grant be subject to a matching requirement? No.

If yes, *please provide a description of your organization's strategies and timetable for meeting the matching requirement, including details of the prospective donor base and project uses for the matching funds.*

Please note: only newly received gifts of cash or securities - in hand valued at the time of transfer, and designated for the Foundation-supported project - can qualify as matching contributions. Pledges do not qualify.