



2025

# STATE OF OPEN INFRASTRUCTURE

Trends in characteristics, funding,  
policy, and community health

May 2025



Invest in Open  
Infrastructure

# Publication notes

Title: 2025 State of Open Infrastructure: Trends in characteristics, funding, policy, and community health

Publisher: Invest in Open Infrastructure

This publication is shared under the following Creative Commons License:

Attribution 4.0 International (CC BY 4.0):

<https://creativecommons.org/licenses/by/4.0/>



Suggested citation: Skinner, K., Collister, L., Huang, C.-K. (Karl)., Lippincott, S., Neylon, C., Riordan, D., Sellanga, J., Steinhart, G., Thaney, K., Tsang, E., & Wu, C. (2025). 2025 State of Open Infrastructure: Trends in characteristics, funding, policy, and community health. Invest in Open Infrastructure. Zenodo. <https://doi.org/10.5281/zenodo.15198874>

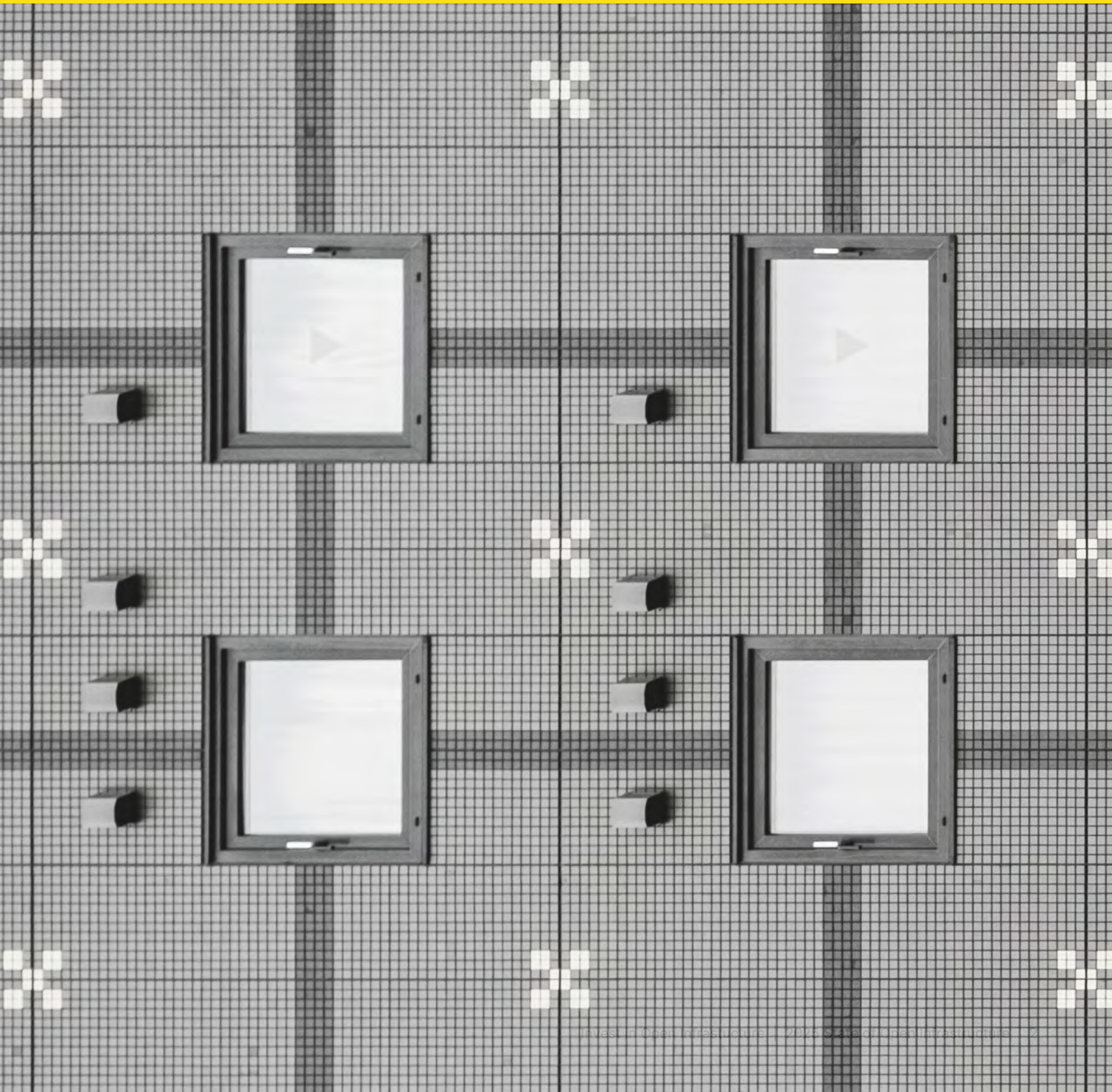
All images in this report are sourced from Unsplash.

# Table of contents

Foreword .....	2
The state of open infrastructure grant funding.....	5
Characteristics of selected open infrastructures.....	14
The cost of public access: Do we get what we pay for? .....	21
Vital open infrastructures in a volatile moment.....	34
Signals from the field .....	50
Shorthand falls short: Why open infrastructure defies simple labels .....	52
The intersection between digital public infrastructure and open infrastructure.....	57
Trust, transparency, and technology: Do community health frameworks shape open infrastructure decisions?.....	62
Contributors .....	65



# Foreword





The open tools and systems we rely upon need our immediate attention and action. We continue to reaffirm our commitment to a vision where every place of higher learning has access to the tools and infrastructure necessary to access and participate in research.

What a time to be writing a foreword on critical infrastructure for research and scholarship. When I first joined IOI, the world was rapidly going into a state of lockdown, economic collapse, and shifts to virtual learning and work. Openness was in the news as a necessary condition to spur innovation and development of life-saving vaccines and treatments; infrastructure and equity were terms and concepts inserted into global policy discussions, funding proposals, new programs, and organizations like ours.

In the five years since IOI formally launched, we have delved deep into the conditions and communities surrounding open infrastructure to power the research ecosystem and civil society.

Today, we're proud to share our second annual *State of Open Infrastructure* report and peel back on the analysis and research our team conducts to power our work to drive investment in and adoption of the tools and systems research relies on.

At IOI, we have long wrestled with what makes an open tool, protocol, platform, system, or network an “open infrastructure.” The term is particularly important for us; not only is it in our name, it also resides at the core of our mission to *increase investment in and adoption of open infrastructure*.

While we have come to no single, let alone strict, definition of the term, we have often returned to a key element, a kind of “true north” that stands out as we try to differentiate between what *is* and *is not* open infrastructure: **reliance**.

At IOI, we are focused on the research ecosystem, and we take “reliance” seriously. Researchers, curators, and technologists *rely* on open infrastructure as critical underpinnings of our information landscape. Disruption in the availability of open infrastructures in the research ecosystem can jeopardize swaths of human knowledge. Imagine just a few of the major open infrastructures serving scholarship today: arXiv, PubMed, DSpace, Open Journal Systems, or perhaps Creative Commons.

Consider how much of our memory, our knowledge trail, depends on open research infrastructures, and consider how many of these have been created on a cycle of grants and sustained on a combination of donations, membership fees, service hosting, and/or specialized development. Repositories and publishing platforms, metadata standards and persistent identifiers — as long as they all work reasonably well, most scholars, teachers, and researchers across disciplines and focal areas accept them as a given and depend on them as a reality.

But these infrastructures are not, on the whole, sustainable businesses with robust fiscal models and diversified revenues. Many do not turn profits; most operate at steady losses that are absorbed by philanthropic and government funders and a variety of research institutions including labs, universities, and colleges. They operate on systems rife with technical debt, and they depend upon volunteer labour to cover much of their human costs from governance to editorial review to code development.

As we worked on this year's *State of Open Infrastructure* report, we watched shifts reported on last year in our Policy report continue at a pace few could expect or even wrap their heads around. The global impacts we started to see in 2023-2024 across Europe and Latin America with political shifts and slashed budgets for science, research, and civil society took on a different intensity at the start of 2025 with many of the foundations underpinning research and knowledge production coming under attack in the US. Reductions in funding for open science had already become a widespread (and growing) problem, with the Netherlands, France, Argentina, Mexico, and others all serving as examples of funding fragility over the last few years. As we worked to reveal patterns and shine light on the investments made in open infrastructure in 2024, we became increasingly aware of how vulnerable our funding sources, which seemed maybe inadequate at times, but certainly stable, actually are.

While we've been writing this report, we've watched a dramatic change hit the US research ecosystem, with massive cancellations of federal grants and contracts numbering in the high billions of dollars, some due to their use of terms like "equity" or "diversity" in titles and descriptions, others because they were being used to study gender or sexuality under a new administration that is working to eradicate those terms. We've watched USAID be dismantled. We've seen NEH and IMLS virtually shuttered, and widespread termination of grants across many federal agencies. We've witnessed direct attacks on universities, including active proposals to slash the standard indirect cost rate for federal grants to fifteen percent (barely a third of what is typically negotiated), which would decimate support for facilities, infrastructure to conduct research, and other critical staffing.

We have long stated that open infrastructure requires better forms and mechanisms of investment, that open infrastructure needs more adoption. We've questioned leading myths like "open = free" and that there's not enough money to fund open infrastructure.

And today — as we launch the *State of Open Infrastructure 2025*, which of course mostly covers funding and activity through December 2024 — we are challenged to describe what we think the open infrastructure landscape might look like by this time next year.

That brings us back to reliance. The open tools and systems we *rely* upon need our immediate attention and action. We continue to reaffirm our commitment to a vision where every place of higher learning has access to the tools and infrastructure necessary to engage and participate in research. This report is a tool, one that brings attention to the current state of open infrastructure and that highlights actions we might take to radically improve that state.

So much is needed to build that better future we seek, where open infrastructure is truly supported and sustained. We're hard at work, as we know so many of you are as well — let's use this moment, together, to make *open* the default for knowledge.

In appreciation,

Kaitlin Thaney  
Executive Director, Invest in Open Infrastructure



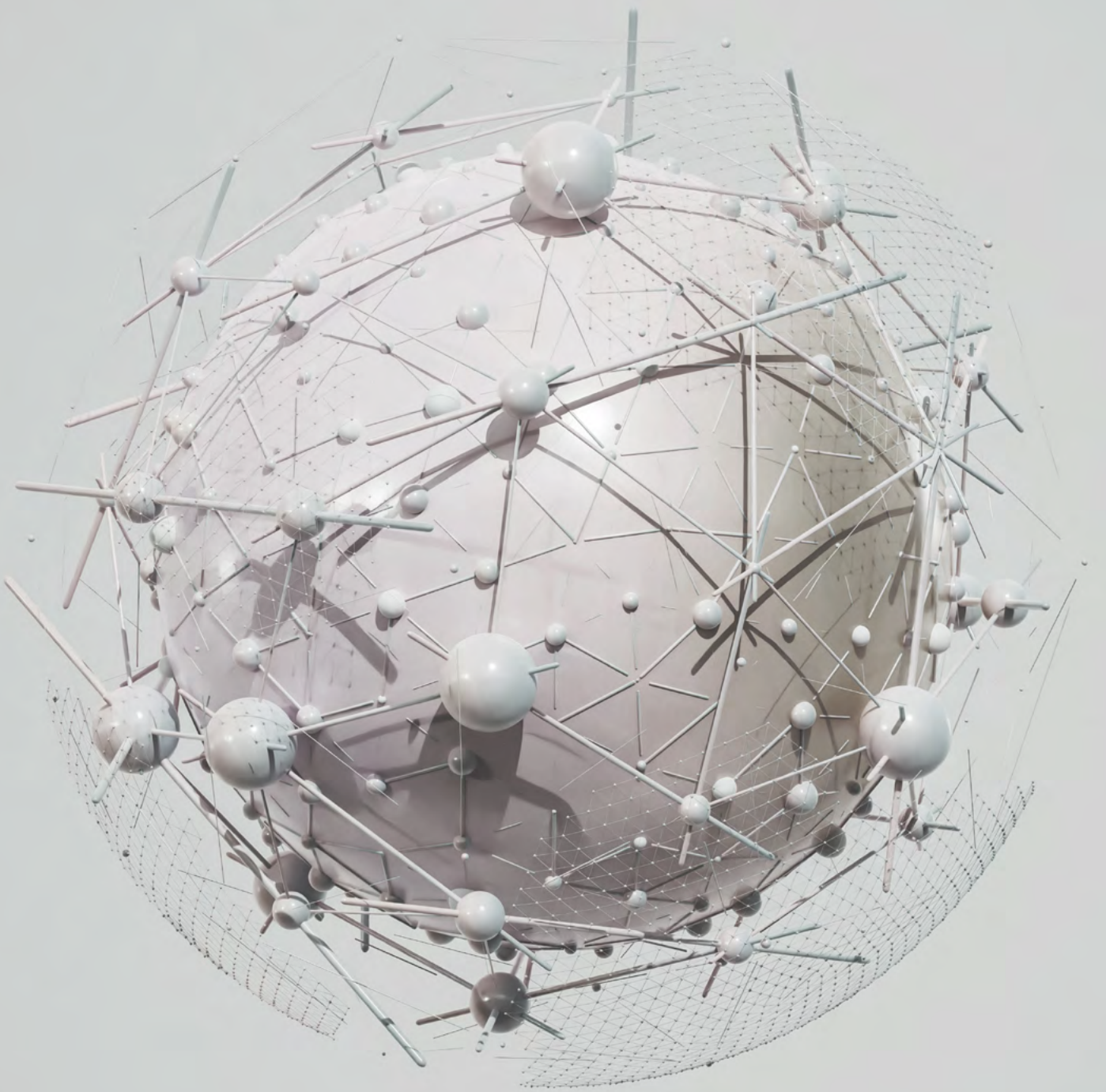
We hope this report sparks your curiosity about the tools and systems the research community relies on and helps you to think about how your decisions line up with your vision for the future of research and scholarship.

If you need additional information and support around these decisions, please reach out to us. We are experienced in providing tailored research and recommendations, including discovery, due diligence, and stakeholder feedback collection, to inform the decisions you are facing.

We're here to help and we can't wait to continue the work of advocating for the future of open infrastructure with you.



# The state of open infrastructure grant funding



**Guiding questions**

- Which funding organizations are investing the most in infrastructure and in what ways?
- What is the potential impact of federal funding instability in the United States on Open Infrastructure (OI)?

**Key insights**

- \$550M+ in funding from 2001-2024. Analysis of 641 grants shows substantial investment in OI; \$191.7M in direct support and \$339.7M in adjacent activities that open infrastructure enables.
- Comparing direct funding to OIs with the research activity they enable (measured by adjacent awards) indicates that OIs provide a strong return on investment.
- Top-funded OIs: Europe PMC, Open Science Framework, arXiv, and OpenEdition.

## Introduction

In support of IOI's mission "to drive informed, strategic, and coordinated investment in and adoption of open infrastructure," we aim to deepen our collective understanding of investment in the sector. IOI's 2024 research into the characteristics of and sources of financial support<sup>1</sup> for open infrastructure (OI) showed that direct financial contributions, including grants, are the primary source of support for many OIs. We build upon our analysis of grant funding in last year's *State of Open Infrastructure* report by including more funders and more OIs in our awards dataset, continuing to examine the importance of OIs in supporting other sponsored

research, and attempting to put OI grant funding from US federal funders into the context of current events in the US. Our updated dataset includes 641 awards totaling approximately \$550M USD, from 2001 through late 2024. As we did last year, we include awards made directly to OIs ("direct" awards) as well as awards that express a reliance on OI but are made to some other recipient ("adjacent" awards) or support the adoption of an OI ("adoption" awards). We conclude with a discussion of some of the general challenges in collecting and analysing grant funding data and what would improve matters.

## Methods

Our full data collection and preparation methods are described in detail in the documentation that accompanies the raw data — this is available for download.<sup>2</sup> We encourage readers to explore the full documentation for additional information on how we processed the data, as well as key assumptions and their likely trade offs. Also note two blog posts describing the challenges we encountered in collecting<sup>3</sup> and processing<sup>4</sup> the data.

We focused on funder-reported and centrally reported data as the sources of record, adding several new funders to our list from last year's report. We focused our data collection on the more than 70 OIs that had been invited to participate in IOI's Infra Finder by 30 September, 2024, and had complete entries by 14 January, 2025. We continue to focus on OIs included in Infra Finder in order to be able to leverage the additional data available there.

<sup>1</sup> <https://investinopen.org/state-of-open-infrastructure-2024/sooi-characteristics-2024/>

<sup>2</sup> <https://doi.org/10.5281/zenodo.15198421>

<sup>3</sup> <https://investinopen.org/blog/sooi-2025-funding-data-collection-challenges>

<sup>4</sup> <https://investinopen.org/blog/sooi-2025-funding-data-analysis-challenges>



Briefly, we:

1. Harvested data directly from funders' websites when it was available, from OpenAIRE, from [usaspending.gov](https://www.usaspending.gov), from 360giving, and in the case of one funder, reviewed their US Internal Revenue Service 990 forms.<sup>5</sup>
2. Added awards from IOI's earlier dataset<sup>6</sup> that we did not capture with our current methods.
3. Searched the data using a predefined list of search terms to search the description, title, and recipient of each award, and interpreted a match in any of these fields to indicate that an award was of plausible interest. We then manually reviewed award titles and descriptions to determine whether they were relevant and excluded those which had no clear relationship to any of the OIs of interest. Duplicate awards were also excluded from the final dataset (although deduplication is fraught with challenges<sup>7</sup>).
4. Converted all award amounts to USD; all amounts in this report are in USD.
5. Assigned each award to a category based on its title and description, as follows:
  - a. **Direct** awards flow to an OI, and are further categorized as research and development, operations, and other.
  - b. **Adjacent** awards do not flow to an OI but support an activity that makes use of a particular OI, either as an end user (depositing a research work to an open repository, for example), or extending or building upon existing infrastructure, independently of the community that supports it.
  - c. **Adoption** awards support the adoption of a particular OI by an individual user or community of users.
  - d. Awards with insufficient information to be classified are categorized as "unknown."
  - e. We identified multi-recipient awards if we could, based on the award description or information provided to us by individual OIs. We did count multi-recipient awards for each named OI, wherever total counts are reported. Because we almost never knew how the funds were distributed among collaborators, the amounts were not counted in any totals for each OI.



When we compare the direct support to OIs with the research activity they support [...], OIs appear to offer sound return on investment in terms of the activity they enable.

<sup>5</sup> Form 990 is filed annually by tax-exempt organizations in the US and lists grants made.

<sup>6</sup> <https://zenodo.org/records/7259472>

<sup>7</sup> <https://investinopen.org/blog/sooi-2025-funding-data-analysis-challenges>

# Results and discussion

## General characteristics of OI funding

The 641 grant awards in our 2025 dataset provided a total of approximately \$550M in support to OIs and activities making use of open infrastructure (Table 1). 178 awards provided direct support from 23 funders to 42 OIs, accounting for about one-third of all funding (Figure 1A). These direct support awards totaled \$191,744,420, representing about \$17M more than what we compiled in the 2024 dataset. Of the amount of funding that goes to direct support, nearly two-thirds is used for research and development, 20% is used for operations, and the rest for other activities such as events, business planning, etc. (Figure 1B). While we did simplify our categorization of awards in this year's reports, this distribution of funding is very similar to what we found last year.<sup>8</sup>

**TABLE 1.**

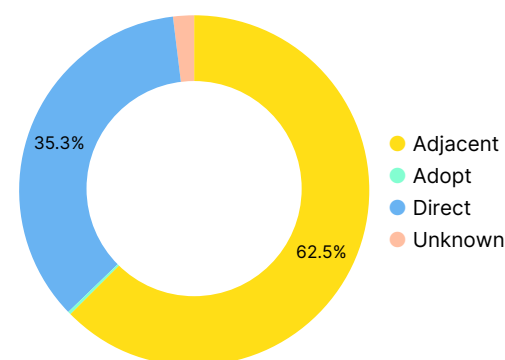
Total funding, and counts of awards, funders, and OIs for all awards and for awards categorized as direct support, adjacent support, and adoption support

*Note that not all awards had amount information, some had an amount of zero, and some we were not able to categorize.*

	All awards	Direct support	Adjacent support	Adoption support
<b>Total funding (USD)</b>	\$551,379,853	\$191,744,420	\$339,711,539	\$2,135,014
<b>Award count</b>	641	178	433	11
<b>Funder count</b>	28	23	20	7
<b>OI count</b>	54	42	41	8

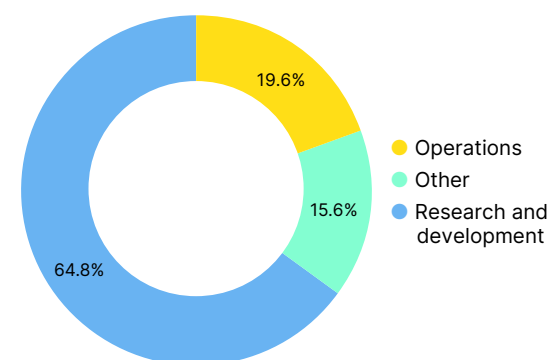
**FIGURE 1A.**

Distribution of all funding by super-category



**FIGURE 1B.**

Distribution of direct funding by category



<sup>8</sup> <https://investinopen.org/state-of-open-infrastructure-2024/sooi-grant-funding-2024/>

## Direct funding

The top recipients of direct support in our dataset are Europe PMC, Open Science Framework (OSF), OpenEdition, and arXiv (Table 2). Datacite was the receiving institution for a number of large multi-recipient awards, but we exclude multi-recipient awards from this table because in most cases we have no information about how the award is distributed among collaborators.

**TABLE 2.**

Sum and count of DIRECT support awards by open infrastructure, top 10 OIs (by total amount)

OI	Total amount (USD)	Total count
Europe PMC	\$26,431,542	13
Open Science Framework (OSF)	\$23,188,855	14
OpenEdition	\$21,415,748	2
arXiv	\$10,253,495	17
Fedora	\$8,281,139	7
OpenAlex	\$7,500,000	1
Dryad	\$7,403,895	9
Dataverse	\$6,500,240	8
Mukurtu	\$6,030,028	15
rOpenSci	\$4,575,071	3

The top funders providing direct support to the OIs we looked at are listed in Table 3. The European Commission tops the list, due primarily to the previously mentioned multi-recipient awards made to Datacite and its collaborators as well as support for Europe PMC and OAPEN Library. Across the larger funders based in the EU the top recipients of direct support are Datacite and its collaborators, Europe PMC, and OpenEdition. Among US-based funders, the top recipients are Open Science Framework (OSF), Fedora, and Dryad (Figure 2B).

**TABLE 3.**

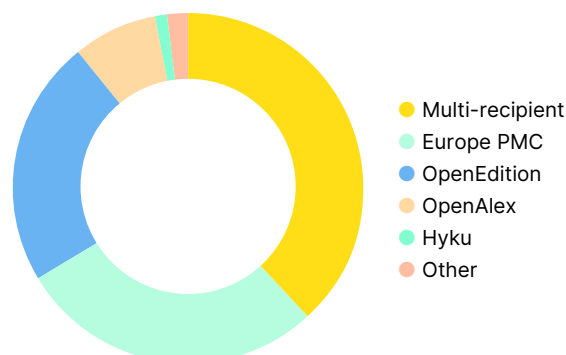
Sum and count of DIRECT support awards by funder (by total amount, funders with >\$5M)

Funder	Total amount (USD)	Total count
European Commission	\$36,070,103	12
Wellcome Trust	\$27,439,563	17
French National Research Agency (ANR)	\$21,415,748	3
National Science Foundation (NSF)	\$16,469,237	16
Arnold Ventures	\$14,310,360	6
Chan Zuckerberg Initiative	\$13,115,875	12
Arcadia Fund	\$9,101,000	4
Andrew W. Mellon Foundation	\$7,834,500	18
Institute of Museum and Library Services (IMLS)	\$7,795,093	27
Gordon and Betty Moore Foundation	\$7,300,020	4
National Institutes of Health (NIH)	\$5,856,270	4
Leona M. and Harry B. Helmsley Charitable Trust	\$5,524,700	2
Templeton World Charity Foundation	\$5,109,856	3

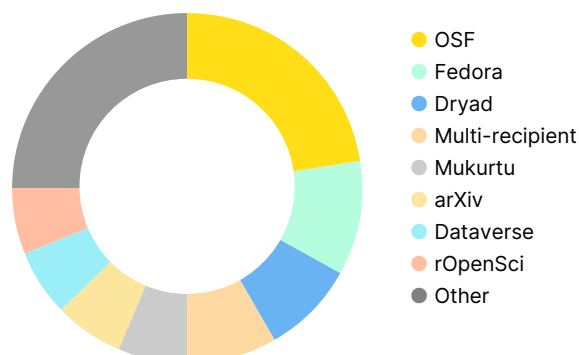


**FIGURE 2A.**

Distribution of DIRECT support awards by OI for EU-based funders

**FIGURE 2B.**

Distribution of DIRECT support awards by OI for US-based funders



## Adjacent funding

We define “Adjacent” awards as awards that are made to projects that utilize open infrastructure in some way. This could be as significant as developing new applications that make use of or interact with existing infrastructure, the use of the entire corpus hosted in an open repository for research purposes, or simply depositing a dataset or preprint to an open repository. This type of support may provide an important signal of the overall impact of open infrastructure by quantifying just how much research activity the existence (and openness) of open infrastructure makes possible. Within the bounds of our dataset, when we compare the direct support to OIs with the research activity they support (as measured by the sum of adjacent awards for that OI), OIs appear to offer sound return on investment in terms of the activity they enable (Table 4).

**TABLE 4.**

Comparison of sums of ADJACENT and DIRECT awards by OI, ranked by ratio of adjacent:direct awards

OI	Adjacent	Direct	Adjacent:Direct
IIIF	\$25,366,512	\$493,842	51.4
International Generic Sample Number (IGSN)	\$5,973,245	\$249,777	23.9
Creative Commons Licenses	\$13,672,115	\$575,000	23.8
Galaxy	\$21,293,831	\$1,780,233	12.0
DSpace	\$15,934,509	\$1,631,643	9.8
Zenodo	\$17,307,880	\$1,926,895	9.0
Dryad	\$44,390,508	\$7,403,895	6.0
Integrated Rule-Oriented Data Systems (IRODS)	\$9,869,443	\$1,685,757	5.9
Dataverse	\$23,910,231	\$6,500,240	3.7
Islandora	\$437,118	\$153,000	2.9
DataCite	\$10,516,917	\$4,220,449	2.5
Open Science Framework (OSF)	\$46,573,978	\$23,188,855	2.0
arXiv	\$15,206,796	\$10,253,495	1.5
Omeka	\$4,846,224	\$3,443,203	1.4
Fedora	\$10,977,897	\$8,281,139	1.3
Knowledge Commons	\$1,249,282	\$1,170,000	1.1



Where the total amount of “adjacent” activity far exceeds direct investment by a funder, it’s worth asking whether it is also in the funder’s interest to assume more responsibility for supporting the infrastructure their grantees depend on.

We could also examine whether funders of the users of OI also fund those OIs directly. For example, the top two OIs named in Adjacent awards are Open Science Framework (OSF) and Dryad. Yet when we compare the amount of Direct and Adjacent funding from the same source, there is a noticeable lack of alignment (Table 5, 6). Where the total amount of “adjacent” activity far exceeds direct investment by a funder, it’s worth asking whether it is also in the funder’s interest to assume more responsibility for supporting the infrastructure their grantees depend on.

**TABLE 5.**  
Source and amount of DIRECT and ADJACENT awards related to Dryad

Funder	Adjacent	Direct
Alfred P. Sloan Foundation	\$0	\$635,915
Institute of Museum and Library Services (IMLS)	\$0	\$87,408
National Aeronautics and Space Administration (NASA)	\$196,276	\$0
National Institutes of Health (NIH)	\$0	\$1,380,325
National Science Foundation (NSF)	\$44,194,232	\$5,115,797
Robert Wood Johnson Foundation	\$0	\$184,450
<b>Total</b>	<b>\$44,390,508</b>	<b>\$7,403,895</b>

**TABLE 6.**  
Source and amount of DIRECT and ADJACENT awards related to Open Science Framework (OSF)

Funder	Adjacent	Direct
Alfred P. Sloan Foundation	\$0	\$500,000
Arnold Ventures	\$385,463	\$14,310,360
French National Research Agency (ANR)	\$0	\$0
Institute of Museum and Library Services (IMLS)	\$0	\$248,247
John Templeton Foundation	\$11,120,475	\$0
National Institutes of Health (NIH)	\$14,205,064	\$1,422,178
National Science Foundation (NSF)	\$19,023,560	\$1,498,214
Templeton World Charity Foundation	\$249,999	\$5,109,856
UK Research and Innovation (UKRI)	\$1,589,417	\$0
William and Flora Hewlett Foundation	\$0	\$100,000
<b>Total</b>	<b>\$46,573,978</b>	<b>\$23,188,855</b>

## The potential impact of federal funding instability in the US

The federal funding landscape in the US is experiencing considerable upheaval and uncertainty. The situation is evolving so rapidly that virtually any summary will be out of date almost immediately, but the situation as of early to mid-March 2025 illustrates some of the challenges even as the specifics continue to evolve:

- The National Institutes of Health (NIH) has been strongly targeted and provides a useful example of the challenges to research funding. In February 2025, the NIH was blocked from publishing notices<sup>9</sup> of review activities in the *Federal Register* as is required by law before review meetings can be held, bringing review of grant proposals to a halt. The agency was also directed to limit indirect costs charged to research grants to 15%.<sup>10</sup> Indirect cost rates typically range from 30–70%<sup>11</sup> and reflect the real cost of supporting research at a university. These charges support shared infrastructure and services, including the cost of facilities and their upkeep, research administration (including regulatory compliance), information technology support, and more.
- Federal funders have also been instructed to identify and stop funding awards supporting activities deemed to be at odds with the priorities of the current administration, including “projects studying transgender populations, gender identity, diversity, equity and inclusion (DEI) in the scientific workforce, environmental justice and any other research that might be perceived to discriminate on the basis of race or ethnicity.”<sup>12</sup> Some orders have been declared invalid by judges<sup>13</sup> who have ordered that funding be released, but it remains unclear whether the executive branch has been complying.

This ongoing uncertainty presents multiple challenges to open infrastructure. First, significant direct financial support from US federal funding sources is in jeopardy. Table 7 lists the total amount of funding from US federal funders for awards in our dataset with an end date of 1 January 2025 or later. Our cut-off end date is somewhat arbitrary (but we think reasonable) given the fairly common practice of requesting a no-cost extension for award activities that continue past a grant’s

original end date, if funds remain. We found one active award for each of the OIs listed, and award duration ranges from one to seven years. To put this potential loss into perspective, arXiv reports annual operating costs of \$2.3–3.4M for fiscal years 2023 and 2024 (from their 2023 annual report<sup>14</sup>) and Dryad reports operating expenses of \$0.9–1.5M for fiscal years 2022 and 2023.<sup>15</sup> Notably, nearly all of the OIs with active federal funding play an important role in the free and open dissemination of research results.

**TABLE 7.**

Sum of direct support awards from US Federal funders with an end date of 1 January 2025 or later

OI	Funder	Direct
arXiv	National Science Foundation (NSF)	\$4,966,530
Dataverse	National Institutes of Health (NIH)	\$1,752,129
Open Science Framework (OSF)	National Science Foundation (NSF)	\$1,498,214
Dryad	National Institutes of Health (NIH)	\$1,380,325
Zenodo	National Institutes of Health (NIH)	\$1,301,638
Knowledge Commons	National Endowment for the Humanities (NEH)	\$500,000
2i2c	National Aeronautics and Space Administration (NASA)	\$362,875
<b>Total</b>		<b>\$11,761,711</b>

<sup>9</sup> <https://www.npr.org/sections/shots-health-news/2025/02/22/nx-s1-5305276/trump-nih-funding-freeze-medical-research>

<sup>10</sup> <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-068.html>

<sup>11</sup> <https://www.pbs.org/newshour/show/the-possible-long-term-impact-of-trumps-cuts-to-medical-research-funding>

<sup>12</sup> <https://www.nature.com/articles/d41586-025-00703-1>

<sup>13</sup> <https://www.statnews.com/2025/03/05/nih-indirect-costs-lawsuit-federal-judge-extends-order-blocking-trump-cuts/>

<sup>14</sup> [https://info.arxiv.org/about/reports/2023\\_arXiv\\_annual\\_report.pdf](https://info.arxiv.org/about/reports/2023_arXiv_annual_report.pdf)

<sup>15</sup> <https://github.com/datadryad/governance/blob/main/annual-reports/990s/Dryad%20FY%202023%20990.pdf>



Second, the research activity that relies on open infrastructure is experiencing the same uncertainty and potential loss of funds due to the termination of current funded projects as well as reductions in future funding. For OIs that operate with any measure of cost recovery from their users, this makes financial planning difficult and may result in a loss of revenue.

Third, the threat of dramatic decreases to “overhead” charges will exert substantial budgetary pressure on colleges and universities. They will be looking for cost savings, potentially resulting in a decrease in membership or other types of contributions to OIs. It is encouraging that some philanthropies recognize the crisis<sup>16</sup> and are increasing their giving; whether this will benefit OIs directly remains to be seen.

## The need for more (and more standardized) funding data

We think we bring significant value to conversations around funding for open infrastructure and its use. That said, we wish we could conduct a comprehensive analysis of funding across the sector. We discussed in a pair of 2025 blogposts some of the challenges in collecting<sup>17</sup> and preparing and analysing<sup>18</sup> this type of data, and advocated for widespread endorsement and adoption of the Barcelona Declaration on Open Research Information and its principles. Sharing grant award metadata in open and standards-compliant repositories and transfer systems, with permissive licensing, would support the discovery and reuse of this information. Invest in Open Infrastructure is a supporter of the Barcelona Declaration, and we encourage others to support and adopt its principles.

## Data availability statement

The grant awards data used for the analysis presented here (Riordan et al 2025) is available online: <https://doi.org/10.5281/zenodo.15198421>.

Interactive dashboards allowing direct exploration of the data are available at <https://investinopen.org/data-room/grant-funding-data-dashboard/>.

Questions or concerns? Please contact us at [research@investinopen.org](mailto:research@investinopen.org).

Sharing grant award metadata in open and standards-compliant repositories and transfer systems, with permissive licensing, would support the discovery and reuse of this information. Invest in Open Infrastructure is a supporter of the Barcelona Declaration on Open Research Information, and we encourage others to support and adopt its principles.

<sup>16</sup> <https://apnews.com/article/macarthur-foundation-endowment-payout-increase-d371dede7ca34830d4653b949e90c647>

<sup>17</sup> <https://investinopen.org/blog/sooi-2025-funding-data-collection-challenges>

<sup>18</sup> <https://investinopen.org/blog/sooi-2025-funding-data-analysis-challenges>

# Characteristics of selected open infrastructures



### Guiding questions

- Critical infrastructure hubs shape the open research ecosystem — do we have safeguards if one fails?
- Funding needs for open infrastructures (OIs) are shifting — can pooled investments strengthen long-term resilience?
- Global distribution of open infrastructure raises digital sovereignty questions — who controls access to knowledge?
- “Open” vs. “closed” isn’t a simple binary — how can we refine our understanding to better support sustainable models?

### Key insights

- OIs most commonly cite the need for funding for research and development, closely followed by operations.
- OIs based in North America were more likely than those in other regions to say that their primary source of funding comes from program service revenue, rather than contributions, or another source.
- At first glance, it might seem simple to categorize tools and platforms as either “open” or “closed.” Deeper examination reveals a spectrum of practices and characteristics that blur this binary distinction.

## Introduction

IOI’s annual snapshot of data from Infra Finder<sup>1</sup> provides an authoritative source of information about the governance, business form and finances, policies, community engagement, and technical attributes of over 100 open infrastructures (OIs).

This year, we’ve paired our data highlights with prompts to encourage deeper exploration and real-world application by readers.

As a companion to this chapter, we make the full Infra Finder dataset<sup>2</sup> (including previous years’ data) available in Zenodo under a Creative Commons Attribution 4.0 International license to support reuse. You can also browse selected data points via our interactive dashboard.<sup>3</sup>

If our findings inspire your research or you’d like to collaborate, reach out to [research@investinopen.org](mailto:research@investinopen.org) — we’d love to connect.

## Transforming data into actionable insights

### Explore integrations to reveal networks of reliance

We ask OIs to list their integrations with other tools, services, and standards that appear in Infra Finder, revealing networks of reliance and interoperability. For example, Thoth Open Metadata describes a number of integrations that leverage its ability to create, store, and distribute robust book metadata.<sup>4</sup> Thoth, in turn, relies on upstream services and standards such as Research

Organization Registry (ROR) identifiers<sup>5</sup> and digital object identifiers (DOIs) to enrich the metadata it collects.

This web of connections highlights not only the collaborative potential within the ecosystem but also its complexity and fragility. Exploring these maps of interreliance provokes a range of questions about the strengths (e.g., distributed costs and risks, modularity) and vulnerabilities (e.g., single points of failure, redundancy) relevant to the open infrastructure ecosystem.

<sup>1</sup> <https://infrafinder.investinopen.org/>

<sup>2</sup> <https://zenodo.org/records/14647337>

<sup>3</sup> <https://investinopen.org/data-room/characteristics-of-open-infrastructure-dashboard/>

<sup>4</sup> <https://copim.pubpub.org/pub/growing-network-of-open-infrastructures-federated-services-with-thoth#nrwp0dvip5u>

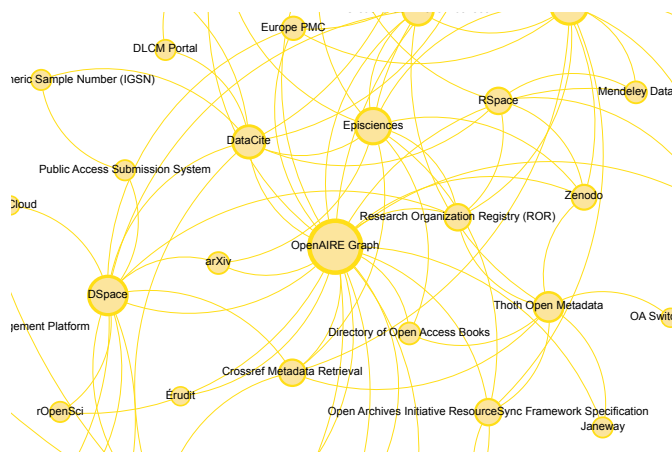
<sup>5</sup> <https://infrafinder.investinopen.org/solutions/research-organization-registry>



- Which tools and services are serving as “hubs” in the open research ecosystem due to their high level of interoperability?
- What would the ripple effects be across the ecosystem if a widely used infrastructure solution were to cease operations? Would critical services experience disruptions? Are there sufficient redundant or alternative systems in place to fill potential gaps, or would the loss of a single tool leave entire segments of the community scrambling for solutions?

FIGURE 1.

Illustration of the connections between a sample of OIs in Infra Finder



One of the notable advantages of open research tools is their flexibility. As a user, you can customize solutions to meet your needs and migrate your data to new solutions if those needs change. But this flexibility also comes with challenges. Understanding which tools to select and how to connect them to achieve your goals can be complicated. By looking at how these tools work with one another, we can ask:

- Where are the opportunities for vertical interoperability between OIs to support diverse research workflows?<sup>6</sup>
- How can we make service “stacks”, or sets of tools that work together for a specific purpose, more visible and easier to implement?<sup>7</sup>

In some cases, having too many choices can be a detriment, such as when interoperability is the goal. Multiple OIs in the same solution categories risk overly segmenting the market, leaving too few resources for any to operate sustainably. Merging overlapping infrastructure solutions or strategically sunseting redundant tools may strengthen the ecosystem, making it easier for users to make choices and infrastructure organizations to build resilient business models.<sup>8</sup>

- Where do such opportunities exist, and how can we ensure that consolidation happens in a way that preserves community values and enhances long-term sustainability?

Dig into solution categories to find opportunities to help users identify the right tools

Many of the OIs listed in Infra Finder help users manage and find information. The most common types of solutions have remained unchanged since last year and include discovery systems (22 solutions), standards, specifications, or protocols (15 solutions), repository software (11 solutions), repository services (11 solutions), and publishing systems (11 solutions).

- Which research or scholarly communication functions is Infra Finder missing?
- Do the existing categories make it easy for people to figure out which OIs will work for different needs? Are there other ways of organizing them that might make this more straightforward?

<sup>6</sup> Read more about vertical interoperability and open source research tools at <https://upstream.force11.org/the-time-is-now-vertical-interoperability-between-research-tools-an-essential-enabler-for-the-fairification-of-data/>

<sup>7</sup> Read more about the concept of service stacks at <https://mindthegap.pubpub.org/>

<sup>8</sup> [https://educopia.org/wp-content/uploads/2024/09/Census2019\\_EducopiaPublications.pdf](https://educopia.org/wp-content/uploads/2024/09/Census2019_EducopiaPublications.pdf)

**TABLE 1.**

OIs are asked to select up to three solution categories that best describe their primary purpose or function

*This table compares reported solution categories in the 2024 and 2025 Infra Finder data.*

Category	2024	2025
Discovery system	10	22
Standard, specification, or protocol	6	15
Repository software	9	11
Repository service	9	11
Publishing system	10	11
Open scholarly dataset	3	6
Digital preservation system	2	6
Digital library, collection or exhibit platform	2	6
Research profiling system	2	5
Peer review system	2	5
Digital preservation service	0	5
Digital asset management system	2	5
Archive information management system	1	5
Persistent identifier service	2	4
Index or directory	2	4
Authoring tool	2	4
Annotations system	2	4
Research software community	0	3
Data management planning tool or service	0	3
Submissions system	1	2
Open access or subscription management tool	1	2
Informal scholarly communications	2	2
Digital preservation tool	0	2
Web archiving system	0	1
Software preservation service	0	1
Open access policy information compilation	1	1
Media viewer/player	1	1
Format conversion tool or service	0	1
Federated identity or authentication management	1	1
Computing library	0	1
Computing framework	0	1

## Identify shared funding needs and opportunities for collective investment

Last year, we reported on the range of funding needs infrastructures cited in their Infra Finder entries.<sup>9</sup> Funding for *operational* and *community engagement* needs dominated the responses. These are areas where available grant opportunities often fall short. This year, we modified our classification scheme for funding needs to mirror that of our grants dataset, narrowing the categories to *research and development* (including enhancing technical features or developing new services); *operations* (such as paying or adding staff); *adoption* (such as community engagement activities and recruiting new members); and *other* (activities such as events, and strategic/governance/business planning).

This year, 62 of the 101 OIs described funding needs related to their ongoing work for which they need additional revenue. Based on IOI's analysis of funding awarded to the OIs in Infra Finder, over 64 percent of direct awards between 2016 and 2024 funded research and development work, aligning with the most commonly cited need, but falling short of the demand for operational and adoption support. The gutting of research funding in the US that is underway in 2025 will surely exacerbate revenue gaps in all categories in the coming years.

**TABLE 2.**

Count of OIs citing each category of funding need

Category	#
Research & Development	39
Operations	26
Adoption	10
Other	10

You'll find each OI's freetext description of its funding needs in our dataset.

- **Where funding needs overlap, are there opportunities to pool investment and effort?**

## Examine the geography and geopolitics of open infrastructure

The OIs in Infra Finder listed 12 different nations as their primary location of activity or incorporation. Slightly more than half of the infrastructures in Infra Finder are US-based. While research is global, OI governance, funding, and operations are often subject to national laws, politics, and funding.

- **How does the distribution and concentration of infrastructures in different regions relate to questions of digital sovereignty?**<sup>10</sup>
- **Which infrastructures might be at greatest critical risk due to geopolitical changes, politically motivated efforts to dismantle academic research, and unprecedented funding challenges in higher education?**<sup>11</sup>
- **Have open access/public access policies around the world contributed to infrastructure development? Can we tie infrastructure in different regions directly to these policies?**
- **To what extent is the information included in Infra Finder useful outside of North America and Western Europe? What could be added to make it more relevant to communities around the world?**

<sup>9</sup> <https://investinopen.org/state-of-open-infrastructure-2024/sooi-characteristics-2024/#governance-business-form-and-finances>

<sup>10</sup> For more on digital sovereignty, see <https://investinopen.org/state-of-open-infrastructure-2024/sooi-future-signals-2024/>.

<sup>11</sup> For more on the critical risks to open infrastructure provoked by the US political environment, see <https://katinamagazine.org/content/article/open-knowledge/2025/US-funding-cuts-imperil-open-infrastructure-globally>



TABLE 3.

Primary location of activity or incorporation of the OIs listed in Infra Finder

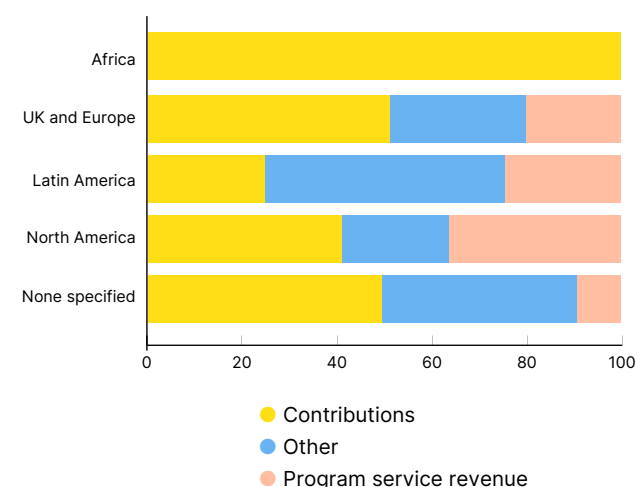
Primary location	#
United States	48
United Kingdom	11
Canada	10
Germany	5
Netherlands	4
France	4
Switzerland	4
Malawi	2
Brazil	2
Uruguay	1
Italy	1
Ecuador	1

In our dataset, OIs based in North America were most likely to say that their primary source of funding comes from program service revenue (income from selling their products and services) rather than contributions (including donations, membership fees, and grant funding) or another source.

- **Are variations in primary funding sources meaningful indicators of how infrastructures sustain themselves in different regions?**

FIGURE 2.

Primary funding source by location of activity or incorporation of the OIs listed in Infra Finder



## Complicate the open vs. closed binary

At first glance, it might seem simple to categorize tools and platforms as either “open” or “closed.” However, a deeper examination reveals a spectrum of practices and characteristics that blur this binary distinction. What do we mean by open or closed? Is it about openly licensed code, transparent governance, community engagement, no-cost usage, or something else entirely?

The profiles of OIs in Infra Finder further reveal the diversity of practices in the open infrastructure world, and shed light on what characteristics are erased in the use of simple criteria. Non-profit organizations that do not make their code openly available, despite operating with community-oriented missions and values. For-profit companies that openly license and extensively document their code, allowing public contributions and fostering collaboration, yet maintaining control over governance and strategic direction. Projects with strong community engagement — through forums, user feedback loops, and active social media presence — yet lacking formal community governance structures that define the ground rules for participatory decision-making. The false binary of open versus closed can obscure important nuances. Infra Finder data can help expose these complexities, fostering a more informed and critical understanding of openness. You can read more about exposing false binaries in a later section of this report, [“Shorthand falls short: Why open infrastructure defies simple labels.”](#)

Consider engaging with the following questions:

- **What can Infra Finder data tell us about how to describe “open”?**
- **Use data from Infra Finder to complete an evaluation of an OI using the FOREST Framework (specifically looking at the value of “Openness”).<sup>12</sup> Does looking more deeply give you a different impression or perspective on the OI’s openness?**
- **Are varied “flavors” of openness more common among OIs at certain stages of maturity, based in different regions of the world, or within different solution categories?**

<sup>12</sup> <https://www.nextgenlibpub.org/forest-framework>

## More information about the data

This year, we invited 77 new OIs to participate in Infra Finder and also accepted self-nominations via our expression of interest form. We asked our existing 57 infrastructures to update their information as needed. We received 44 new entries through our outreach and the expression of interest form, as well as updates (or confirmation that no data had changed) from about half of the initial 57 infrastructures.<sup>13</sup> The total number of OIs included in our dataset now stands at 101.

This year, we focused specifically on recruiting and adding OIs for research data sharing and analysis and digital preservation and archiving in response to trends and needs in the open science ecosystem.<sup>14</sup>

Because participation in Infra Finder is voluntary, and largely by invitation, this dataset remains modest in size and scope and is not a random sample. However, the breadth of information captured provides a robust basis for exploring research questions about the open research infrastructure ecosystem, especially when paired with other quantitative and qualitative data sources.

We made minor modifications to our data collection instrument this year, including consolidating several solution categories and introducing nine new solution categories.<sup>15</sup> The new categories are

- Digital preservation service
- Data management planning tool or service
- Research software community
- Digital preservation tool
- Web archiving system
- Software preservation service
- Format conversion tool or service
- Computing library
- Computing framework

An infrastructure was eligible for inclusion if, at the time of the invitation, it was fully operational and in active use as a service, protocol, standard, or software that the academic ecosystem needs in order to perform its functions throughout the research lifecycle. Infrastructures also had to meet one or more of the following eligibility criteria:

- Meets the definition of open source software (OSS);
- Primarily or exclusively distributes openly licensed (open access) content;
- Is free to use by anyone (free of charge or other restrictions);
- Is community governed and is transparent in its operations and finances;
- Is operated by a non-profit or non-commercial entity.

Participating OIs were sent a partially completed data form and asked to correct or add information. IOI's team validated the responses against publicly available documentation and resolved discrepancies with the respondents.

As a companion to this chapter, we make the full Infra Finder dataset<sup>16</sup> (including previous years' data) available in Zenodo under a Creative Commons Attribution 4.0 International license to support reuse. You can also browse selected data points via our interactive dashboard.<sup>17</sup>

Readers are encouraged to explore Infra Finder and the associated data, whether to further investigate some of the research prompts we have presented or to support their own work. If that describes you, please let us know by [completing our feedback form](#).

<sup>13</sup> Recognizing that changes can happen at any time, in addition to IOI's annual outreach campaign, OIs can make as-needed data updates throughout the year. These changes are reviewed by the IOI team and each individual service page of Infra Finder shows when data was last updated.

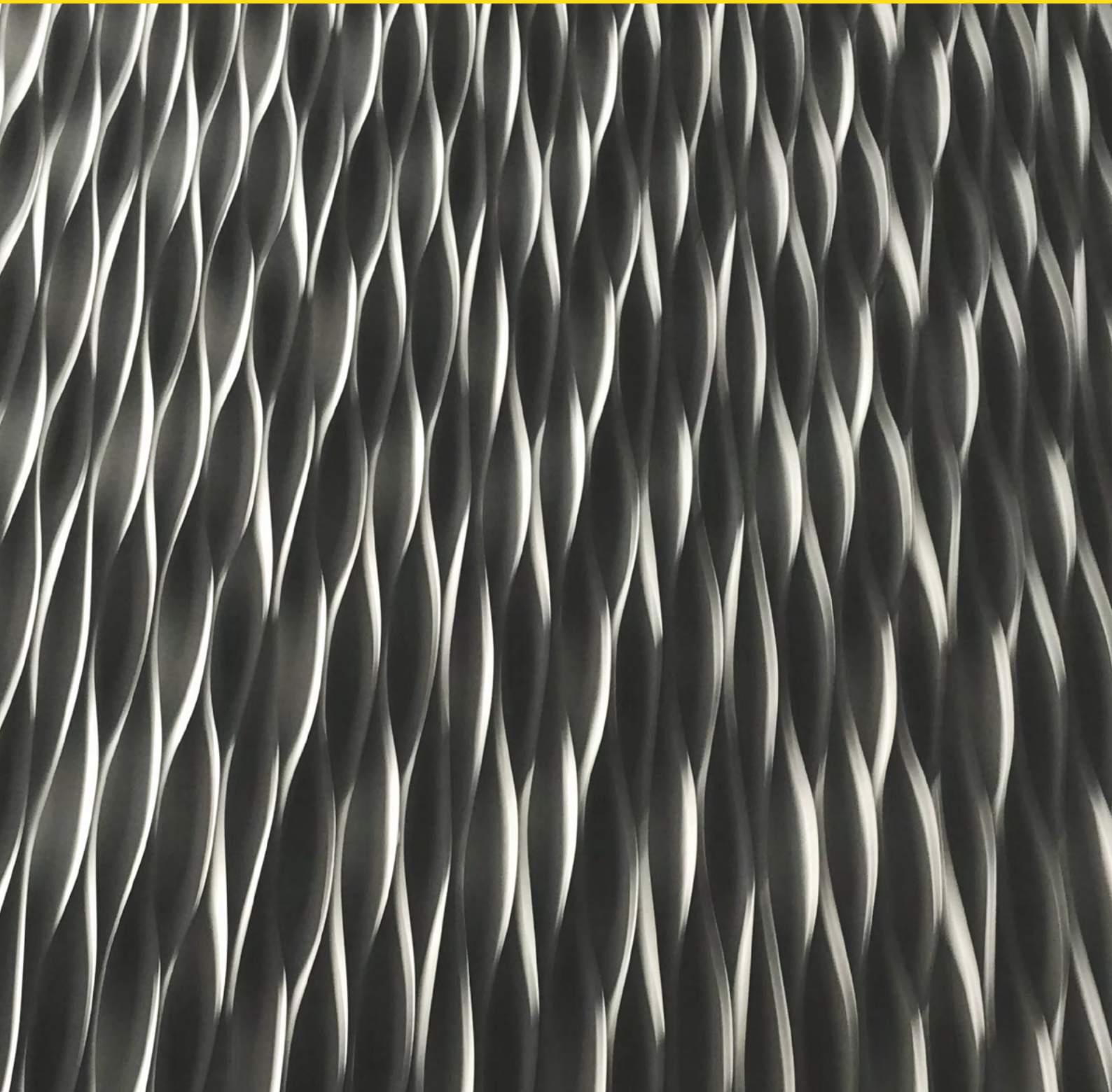
<sup>14</sup> <https://investinopen.org/blog/help-shape-the-future-of-infra-finder-explore-new-features-and-share-your-feedback/>

<sup>15</sup> Read our [announcement](#) and find definitions for each solution category in our Infra Finder documentation.

<sup>16</sup> <https://zenodo.org/records/14647337>

<sup>17</sup> <https://investinopen.org/data-room/characteristics-of-open-infrastructure-dashboard/>

# The cost of public access: Do we get what we pay for?





### Guiding questions

- What cost factors compromise public access policies as levers of change?
- Can public access policies catalyze open science practices or does a focus on *access* sideline *openness*?
- Where can funders and institutions invest to build on the momentum that public access policies create?

### Key insights

- The success of public access hinges on well-supported research infrastructure, including repositories, persistent identifiers, publishing platforms, training programs, and expert labor.
- Abrupt policy shifts, political uncertainty, and funding cuts make it difficult for institutions to plan their response to public access policies, even in the near-term. Many academic institutions are taking a wait-and-see approach to investments in boosting policy compliance.
- Public access policies do little to address broad participation in knowledge creation, focusing instead on the right to consume.

## Introduction

Free and immediate access to the results of taxpayer-funded research has gained widespread support from governments around the world over the last several decades. Public access policies, memoranda, and guidance have proliferated, promising to make science more transparent, equitable, and collaborative.

While these policies have garnered widespread praise for their potential to transform research culture, they also raise important questions about resources. Who pays for implementation? What happens when policies mandate access without providing funding mechanisms?

Drawing on perspectives from stakeholders across three global regions implementing such policies — including **Kathleen Fitzpatrick**, **Jennifer Gibson**, **Bianca Kramer**, **Jonah McAllister-Erickson**, and **Eunice Mercado-Lara** — as well as findings from IOI's Reasonable Costs project,<sup>1</sup> we categorize various costs involved and identify opportunities to invest in these areas to promote public access and advance open research principles. We reached out to a broad range of stakeholders for interviews and scheduling problems limited us to those who could accept; we note that the interviewees in this piece are more heavily library and repository focused, though we do try to account for additional perspectives that we have heard from the many society, repository, library, publishing, administrative, and research-focused stakeholders in our Reasonable Costs work.

## The cost of culture change

As more and more governments and private funders require the research outputs they fund to be publicly available, researchers themselves are leading an ever-growing movement toward “open science” — a broad set of research principles and practices that emphasize transparency and inclusiveness.

Jennifer Gibson, Executive Director at the data publishing platform Dryad, believes that while the spirit of directives like the OSTP Public Access Memo,<sup>2</sup> also known as the Nelson Memo, in the US are on the right track, “policy isn’t enough” to actually change how researchers work (personal communication, April 2, 2025). **The symbiosis of public access policies and open science practices is key to creating meaningful change. Jumping straight to mandates without laying essential groundwork can lead to resistance and failed implementation.**

Eunice Mercado-Lara, who helped create Mexico's public access policy, advocates a multi-step formula for success: raise awareness, build skills, create infrastructure, and only then introduce requirements (personal communication, March 6, 2025).

<sup>1</sup> <https://investinopen.org/data-room/reasonable-costs/>

<sup>2</sup> The text of the memo has been removed from the White House website, but is accessible via the Internet Archive's Wayback machine at <https://web.archive.org/web/20250120050644/https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>.



## Where to invest

- **Next generation research assessment.** Traditional evaluation metrics can conflict with open science practices, creating career risks for researchers (Belles et al., 2023). Assessment systems must recognize and reward open practices to drive meaningful culture change. Investment in the work of groups like CoARA,<sup>3</sup> which promote “more inclusive and effective assessment practices,” helps create the necessary conditions to promote openness.
- **Data sharing support.** Campus data services, often based in university libraries, offer critical training, guidance, and (in some cases) infrastructure for researchers. Public access requirements elevate these services to institution-level strategic priorities that deserve commensurate investment. West Virginia University offers an example of this investment in action. When Jonah McAllister-Erickson, Scholarly Communication Librarian at the WVU Libraries, reviewed recent proposals from investigators at WVU, he found that nearly half had budgeted nothing for data management. This may stem from uncertainty about allowable costs and underestimating the labor involved in proper data sharing. The lack of specific budget lines for data management could force research teams to shift project resources from data collection and analysis to accommodate data cleaning and processing (personal communication, March 21, 2025). A cross-campus working group at WVU concluded that “ensuring data sharing compliance would require new processes, guidance, training, services and roles” and is collaborating to put these systems in place (“Rising to meet research needs”, 2025).
- **Open science skills.** *Public access and open science* are not synonymous. However, public access mandates create opportunities to introduce researchers to a wider range of open science practices and motivate them to participate. Kathleen Fitzpatrick, Interim Associate Dean for Research and Graduate Studies at Michigan State University (MSU) and Director of the open repository Knowledge Commons, is seeing changes at MSU in the form of “projects to promote the idea that scientific enterprise is one giant collaboration and figuring out how we can work together in ways that are supportive rather than competitive.” Initiatives like the NASA-led 2023 Year of Open Science,<sup>4</sup> building on the momentum of the Nelson Memo, “are really attempting to think beyond making publications and data available and open up practices more generally” (personal communication, April 2, 2025). Campus-based and community-led initiatives can provide necessary training in coding, data management, and reproducibility, among other open science skills. Investment from institutions, consortia, as well as public and private funders can help make this type of training the norm across disciplines and geographic regions.

Public access and open science are not synonymous. However, public access mandates create opportunities to introduce researchers to a wider range of open science practices and motivate them to participate.

<sup>3</sup> <https://coara.eu/>

<sup>4</sup> <https://www.nature.com/articles/d41586-023-00019-y>

# The cost of equity

Public access and open science policies from around the world reference “equity” as a core principle and a desired outcome. For example, the Nelson Memo explicitly establishes a strong connection between public access and equity, emphasizing that access should benefit all segments of society; the UNESCO Recommendation on Open Science<sup>5</sup> prominently features equity as a core principle. However, when public access policies conflate “public access” with “equity,” they flatten the differences between the very narrow ability to read published research results with the broader ideal of enabling full participation in research by all people (Khisro & Fenlon, 2025).<sup>6</sup>

**Public access policies do little to address broad participation in knowledge creation, focusing instead on the right to *consume*.** In the context of India’s open science policy, for example, critics noted that “knowledge dissemination receives immediate attention with little or no contemplation of the knowledge creation process” and insufficient consideration is given to “cultural diversity in knowledge production” and knowledge systems (Koley, 2022). Achieving an inclusive and equitable research ecosystem requires addressing deeper cultural and organizational barriers (National Academies of Science, Engineering, and Medicine, 2024).

## Where to invest

- **Community ownership.** Fitzpatrick notes that, while the Nelson Memo explicitly addresses the importance of high-quality, trustworthy repositories, it doesn’t “open the possibilities for exploring non-corporate models and ensuring that those are supported.” In Europe, organizations have urged public investment in open infrastructure “to prevent open science appropriation by private multinational companies” (Bacalexi, 2022). Public funding can offer critical, stable funding for open infrastructure, though reliance on any single funding source can create risks and vulnerabilities, as demonstrated by the radical and unprecedented cuts to federal funding currently hitting the US. Open infrastructure organizations need to build diverse revenue streams — which may include public and private grants, membership and service fees, donations, and in-kind contributions — coupled with governance models that put researchers over profits and incorporate inclusive scholarship into their mission.
- **Regional initiatives.** Investment in local and regional open science infrastructure helps counter the dominance of Western models and perspectives. Regional initiatives can better account for contextual differences in research capacity, cultural knowledge systems, community needs, and norms and engage end users in the design of knowledge infrastructure that works in their context (Arthur et al., 2023; Okune et al., 2018). Funding models should acknowledge and accommodate different economic realities across regions and solutions should be co-designed with local stakeholders, include diverse representation in governance, and reflect regional differences in connectivity and hardware. A successful example of a regional initiative supported by a global collaboration is LIBSENSE,<sup>7</sup> which has defined priorities and developed concrete deliverables in support of open science policy adoption in Africa (Chiwari & Skelly, 2022).

<sup>5</sup> <https://www.unesco.org/en/open-science/about>

<sup>6</sup> Khisro and Fenlon (2025) reviewed US federal agency plans developed in response to the Nelson Memo and found that few “take equity meaningfully into account as a concept distinct from public access; and among those that do, there is little consensus on the meaning and entailments of equity in public access to research results.”

<sup>7</sup> <https://libsense.ren.africa/home>

- **Data reusability.** Proper investment in FAIR (Findable, Accessible, Interoperable, Reusable) data practices can create more equitable research opportunities. Broader implementation of the CARE Principles<sup>8</sup> helps ensure that data governance aligns with Indigenous worldviews and interests while providing collective benefit. McAllister-Erickson notes that access to underlying data breaks down barriers for validation and replication and provides a rich resource for stakeholders who may not have the resources to generate the data for themselves. Poor data management practices and gatekeeping by researchers has historically prevented much of this data from being used to its full potential. Data reusability requires standards, platforms, tools, and training that support diverse research communities. Public and private funders that care about catalysing scientific research breakthroughs need to be investing not only in research studies, but in standards-based infrastructure for data sharing and reuse.
- **Non-APC-based publishing models.** As public access requirements become the norm, publishers have increasingly adopted article processing charges (APCs) — where an author pays a fee to make their article openly readable — to replace subscription revenue. Business models built around maximizing open access fees end up “excluding people from all the same groups as the subscription model, except that they’re now excluded from participating in the conversation rather than seeing the results of the conversation” (Kathleen Fitzpatrick, personal communication, April 2, 2025). Concern about replacing paywalls with “playwalls” (i.e., researchers with more funding get higher visibility as the more prestigious journals turn to APC-based OA business models) demonstrate the need for more radical change in how we support scholarly publishing (Hampson & Steinhauer, 2023; Krauskopf, 2021; Peterson et al., 2013). The high cost of publishing forces many researchers to choose between career advancement through prestigious but expensive journals and making their work widely accessible (Belles, et al., 2023). Diamond and Green Open Access (OA) initiatives, led by the academy, can put the power back in the hands of researchers and provide opportunities for more equitable participation in scholarly communication. Funders interested in furthering research participation by marginalized groups could be investing in initiatives that advance Diamond OA, such as the Open Journals Collective (OJC),<sup>9</sup> which aims to direct collective investment into hundreds of Diamond OA journals, as well as in open source repositories and publishing platforms, such as AfricArxiv, that specifically promote scholarship from developing regions.



No single institution can or should support a repository that’s robust enough to do all the work that open scholarship requires, but with pooled resources and collective effort, we can build robust and sustainable solutions.

<sup>8</sup> <https://www.gida-global.org/care>

<sup>9</sup> <https://www.openjournalscollective.org/>

# The cost of infrastructure

Public access mandates require significant ongoing investment in technical infrastructure to host, distribute, and preserve research outputs.

The technical infrastructure that enables public access to research is built and maintained by a range of entities, including national governments and agencies, research institutions, infrastructure providers, and publishers. This distributed responsibility creates both opportunities and challenges for sustainable public access implementation.

For example, the NIH is spending millions to ensure critical generalist data repositories meet desired technical specifications while fostering culture change and policy compliance (Jennifer Gibson, personal communication, April 2, 2025). The European Commission invested over 5 million euros on the preliminary round of development and operationalization of Open Research Europe, a venue for researchers to make their outputs publicly available in compliance with the Open Access mandate of Horizon 2020 and Horizon Europe (Johnson, 2022).

Research institutions are increasingly investing in local data and publication repositories as well as subscriptions or memberships with third party vendors who help them meet the compliance needs of their researchers. In many regions, these are funded entirely at the institutional level, requiring ongoing care and funding. Mercado-Lara notes that, of the 200 Mexican universities that received funding to build institutional repositories to support public access, only about a third are still active and coordinating. Those that succeeded had buy-in from university leaders and researchers that provided momentum for ongoing support after national funding ended. US-based institutions that participated in IOI's Reasonable Costs project expressed uncertainty about whether and how much to invest in campus data storage/hosting and preservation infrastructure given the lack of specificity from federal agencies regarding designated repositories and appropriate formats and technical standards for shared data.

In the US, facilities and administrative (F&A) costs paid by funders have historically supported building and maintaining necessary campus research infrastructure. On some campuses, proposed federal caps on these rates are already resulting in job eliminations and provoking difficult choices about postponing or cancelling needed infrastructure projects (Jonah McAllister-Erickson, personal communication, March 21, 2025).

## Where to invest

- **Leveling up.** To ensure they can support high-quality, trustworthy, and compliant public access, infrastructure organizations need to invest in feature and service enhancements. In some fields, this work is being supported by direct investments from funding agencies, as in the case of the National Institutes of Health's Generalist Repository Ecosystem Initiative (GREI),<sup>10</sup> which aims to help a group of widely used data repositories provide robust and consistent service to investigators. Even in areas where guidance and intervention from funding agencies have been less specific, policies are prompting infrastructure organizations to reflect on how well they align with best practices for technology and services.
- **Building collectives.** Supporting a vast ecosystem of individual infrastructures to operate independently and with redundant functions across the board costs more money than consolidating (Jennifer Gibson, personal communication, April 2, 2025). Despite the challenges of inter-institutional collaboration, Fitzpatrick and others contend that this might be the moment to say "no single institution can or should support a repository that's robust enough to do all the work that open scholarship requires," but with pooled resources and collective effort, we can build robust and sustainable solutions.

<sup>10</sup> <https://datascience.nih.gov/data-ecosystem/generalist-repository-ecosystem-initiative>



- **Academy-owned cloud.** Can research institutions support large-scale open research and public access? Broadly available, academy-owned cloud hosting infrastructure may be one place to start. Research is highly dependent on commercial cloud providers like Amazon Web Services (AWS), another example of “siphoning resources out of the academy into corporate pockets” and generating concerns about data security, privacy, and other academic values that often misalign with corporate practices (Kathleen Fitzpatrick, personal communication, April 2, 2025). Examples of infrastructure hosted on academic servers, rather than via commercial cloud providers

are increasingly rare. Recent news that arXiv plans to move to Google Cloud Platform (GCP) prompted discussion about the risks of vendor lock-in and increased costs, while also underscoring the difficulty of a single institution maintaining its own modern, reliable hosting infrastructure (Kasanmascheff, 2025). Pooling resources to build necessary hosting infrastructure could allow libraries across entire regions or countries to support the wide array of institutional infrastructure projects, such as repositories, digital collections, publishing platforms, and research software that currently use commercial cloud hosting services.

## The cost of publishing

In response to a request for comment on the Nelson Memo in 2022, signatories from the Ivy Plus Consortium in the US wrote, “We both applaud this policy change and are aware that it may result in significant additional costs related to publication, repositories, data management, and staffing which we anticipate will be shouldered by individual researchers and institutions” (“Library Statement on the Nelson Memo,” 2023).

While policies like Plan S explicitly endorse Gold Open Access (the business model where authors pay to publish), others like the Nelson Memo remain ambiguous about the mechanisms for public access (Hampson & Steinhauer, 2023). Some have suggested that it may favour APC-based models (Hampson & Steinhauer, 2023; Crotty, 2022).

A 2023 study of spending by US universities shows that “the majority of states published between 1,000 – 7,000 Gold Open Access publications and spent up to 6 million dollars [on publication fees] in the past 10 years” (Halevi et al., 2023). The financial burden extends beyond covering the APC. Data from a study of 29 universities

in the UK shows that when taking into account both the APC and *administrative costs*, Gold OA publishing costs universities 2.5 times more than Green (repository-based) alternatives (Johnson et al., 2016). Meanwhile, well-resourced commercial publishers can quickly adapt business models, develop new revenue streams, and maintain influential advocacy positions that shape policies in their favour (Jennifer Gibson, personal communication, April 2, 2025).

Research institutions are already shifting limited resources to accommodate these changes, largely by converting portions of their library collections budget to “cover some publication costs, whether via an APC fund available to institutional authors, via Read-and-Publish contracts that cover APCs in bulk, or via direct subsidies to support open access ventures” (Sharp et al., 2023). In regions experiencing stagnating and shrinking university library budgets (such as the US, UK, and parts of Latin America,) new spending needs to be cost neutral. Supporting new publishing models means reducing investment in other initiatives.

## Where to invest

■ **Rethinking scholarly communication.** According to Fitzpatrick, the “focus of Nelson Memo on making end products of research *as we currently know them* freely and openly available” can thwart opportunities to rethink the functions and forms of scholarly communication. Similarly, in the Netherlands, investment since Plan S has largely taken the form of publisher agreements, leaving out alternative models if all the money is going to traditional publishing. Bianca Kramer, Executive Director for the Barcelona Declaration on Open Research Information, notes that great work around Diamond OA continues in the Netherlands though recent open science budget cuts risk stymying progress (personal communication, April 3, 2025). By centering traditional research outputs — namely peer-reviewed journal articles — the policy risks reinforcing entrenched models. Efforts like Plan U, which advocate for the mandatory deposition of preprints as a condition of funding (Sever et al., 2019), offer a different vision: one that enables faster, more transparent scientific exchange decoupled from the prestige factor of “high-impact” journals. Kiermer et al. (2025) similarly argue for a broader reevaluation of the scholarly publishing system, suggesting that a true shift must include reimagining peer review, credit attribution, and the role of community feedback. Embracing diverse research artifacts — such as data sets, code, protocols, and preprints — as primary scholarly outputs could promote new forms of transparency, collaboration, and impact.

■ **Establishing reasonable costs.** In the long term, the capping of grant allocations for publishing fees by large federal agencies may be an effective means of preventing further increase in the cost of publishing (Belles et al., 2023; Gorelick & Li, 2021). But what do “reasonable” caps look like? The Nelson Memo and other policies have prompted more publishers to provide cost and price transparency information, but it has remained difficult to establish fair and objective metrics for gauging reasonable costs (Kemp & Skinner, 2024). IOI’s Reasonable Costs project is establishing benchmarks for calculating costs (i.e., the direct and indirect expenses involved in publishing content) and price (i.e., the fees charged to authors or their institutions to make content public). The openCost<sup>11</sup> project has developed a schema and database to track the costs involved in scientific publishing. This critical work provides funding agencies with the information they need in order to establish allowable costs associated with public access compliance.

## The cost of ensuring compliance

Evidence shows that when there’s no clear enforcement, many researchers simply don’t comply with public access policies. In the US, compliance with National Science Foundation policies has been low, partly because there’s not enough oversight (Powell et al., 2025). Research funders incur considerable ongoing costs for paid services that help them monitor whether researchers comply with their requirements (Cobb-Lewis, et al., 2024). In the case of data, ascertaining whether data sharing has been performed in compliance with the investigator’s data management plan can require detailed and time-consuming verification and potentially even domain expertise.

A lack of specific guidelines can create administrative uncertainty and complicate compliance efforts, as institutions interpret broad mandates in various ways. What was meant to be flexible actually ends up creating a patchwork of practices that make it harder to achieve public access goals.

<sup>11</sup> <https://www.opencost.de/projekt/>

We need to know that the regulations won't continue to be a moving target. We also need better information and consistency from federal agencies.

In 2023-24, IOI surveyed 83 representatives of universities, colleges, and research labs in the US as part of our Reasonable Costs<sup>12</sup> project. Eighty percent of respondents said the Nelson Memo had resulted in few or no meaningful changes to the work of their unit. Only 14 percent identified specific ways in which their unit had increased its investment to support activities related to public access to grant-funded publications and data/other outputs as a result of the Nelson Memo. Respondents were frustrated by the lack of specificity from many agencies regarding their requirements, including where research outputs are expected to be deposited and in what formats. As one respondent noted, "We need to know that the regulations won't continue to be a moving target. We also need better information and consistency from federal agencies."

Who is ultimately charged with compliance remains unclear, further complicating the question of how policies get enforced and by whom. Do institutions or principal investigators bear primary responsibility for ensuring research outputs are shared in accordance with requirements? Most institutions participating in IOI's Reasonable Costs project do little to no monitoring of public access compliance. Will this need to change? If so, who bears the costs of such monitoring?

## Where to invest

- **Reducing friction.** "It's friction," says Gibson. "If folks don't know what to do, they're not going to do it." Researchers are being asked to divert attention from their core work to learn and implement new systems and procedures. Institutions and open infrastructure organizations need to build new tools and pathways that make it easy for researchers to understand and meet their obligations. This is especially true in the US going forward, as publishers may be less motivated to support automated deposit in funder repositories (which has been a major driver of public access compliance) now that the 12-month embargo period has been lifted (Crotty, 2022). However, IOI's research found that many academic institutions in the US are taking a wait-and-see approach on the Nelson Memo, delaying changes to workflows, outreach, and internal policies until they understand the level of enforcement and consequences of non-compliance.
- **Promoting persistent identifiers (PIDs).** Machine-readable funder acknowledgements facilitated by PIDs remain underutilized, especially for data, making it even more challenging to associate research outputs with their funders (Schaes, 2024). Funders can help by participating in Crossref's Grant Linking System (GLS)<sup>13</sup> and Open Funder Registry (OFR)<sup>14</sup> and by requiring grantees to use PIDs in the funding acknowledgements of the outputs they share. Institutions can invest in promoting the adoption and use of ORCID<sup>15</sup> and Research Organization Registry identifiers<sup>16</sup> that facilitate machine- and human-readable attribution.

<sup>12</sup> <https://investinopen.org/data-room/reasonable-costs/>

<sup>13</sup> <https://www.crossref.org/services/grant-linking-system/>

<sup>14</sup> <https://www.crossref.org/services/funder-registry/>

<sup>15</sup> <https://orcid.org/>

<sup>16</sup> <https://ror.org/>

# The cost of politics

When political administrations change, particularly toward more austerity-focused or anti-research positions, open science and public access initiatives become collateral damage.

**Policies associated with a particular administration are vulnerable to being dismantled.** In Mexico, for example, starting in 2019, “anything associated with the previous administration was condemned, including the programs around the open science policy” (Eunice Mercado-Lara, personal communication, March 6, 2025). The Nelson Memo, strongly associated with the Biden administration in the US, faces an uncertain future, which might include being retracted, delayed, unenforced, or deprioritized.

**Abrupt policy shifts and political uncertainty requires frequently adjusting even near-term plans.** The disappearance of implementation guidance, even as policies themselves remain in effect, can be equally as disruptive. In the US, for example, the NSTC’s Desirable Characteristics of Data Repositories for Federally Funded Research, an influential set of guidelines for data sharing infrastructure, appear to have been erased from the White House website.<sup>17</sup> Identifying which services, positions, and technologies are critical may depend on unpredictable policy changes, making it difficult for institutions to know where to focus or where to cut when they don’t know which requirements they will need to address (Jonah McAllister-Erickson, personal communication, March 21, 2025).

In Argentina and Chile, cuts to overall science funding have created an environment where open science becomes a lower priority as institutions struggle to maintain basic research operations (de los Ángeles Orfila, 2024). The US faces a similar challenge as proposed federal funding caps on research overhead rates imperil the ability of research institutions to perform their basic functions, let alone invest in the necessary infrastructure and support structures that support public access compliance (White, 2025). The Trump administration’s push to dismantle federal agencies, eliminate research funding, and censor research poses grave risks to public research infrastructure in the US (e.g., ERIC, PubMedCentral) and around the world (Collister et al., 2025).<sup>18</sup> Fitzpatrick cautions that we could see significant US federally funded research platforms crumble and noted the urgency of national and international attention on the implications. Conversations are happening in the US and elsewhere about how to fund and execute a plan to replicate and bring up these research platforms outside of the federal government if and when such action is needed. In Europe, institutions have begun warning researchers to avoid storing data on US-based infrastructure and researchers have launched efforts to rescue at-risk data (Schoupe, 2025; Schapp, 2025).

Even in countries with strong open science foundations, political shifts can threaten progress. Funding cuts in the Netherlands, for example, could undermine years of advancement in a country previously seen as a leader in open science and public access (“Open Science NL budget”, 2024).

When political administrations change, particularly toward more austerity-focused or anti-research positions, open science and public access initiatives become collateral damage.



<sup>17</sup> These guidelines, previously available at [whitehouse.gov](https://www.whitehouse.gov/wp-content/uploads/2022/05/05-2022-Desirable-Characteristics-of-Data-Repositories.pdf), are now available at: <https://web.archive.org/web/20250116081728/https://www.whitehouse.gov/wp-content/uploads/2022/05/05-2022-Desirable-Characteristics-of-Data-Repositories.pdf>

<sup>18</sup> <https://hechingerreport.org/proof-points-eric-under-threat/>



Supporting the development of community-owned infrastructure, built on open source software, where researchers, academic institutions, and the public have a stake in ownership and governance, can create a more resilient system and insulate against government interference or disruption.

## Where to invest

### ■ **Diversify infrastructure ownership and control.**

Kramer says that the situation in the US is reinforcing discussions in Europe about digital sovereignty, a conversation that has spread well beyond the academic sphere (Reynolds, 2025). Fitzpatrick notes that Knowledge Commons is thinking hard about what it means to be subject to the whims of a single research institution (Knowledge Commons is fiscally hosted by MSU), a single presidential administration, or a single tech CEO. For researchers outside North America and Western Europe, open source technologies “offer a viable, cost-effective alternative to the technological infrastructure provided by big tech and/or powerful, well-resourced nations” (Steinhart et al., 2024). Supporting the development of community-owned infrastructure, built on open source software, where researchers, academic institutions, and the public have a stake in ownership and governance, can create a more resilient system and insulate against government interference or disruption. This requires establishing sustainable business models for these independent organizations to flourish without being overly reliant on government funding.

■ **Support productive duplication.** Given growing concern about government censorship or manipulation of research results and public data, it is essential to collaborate on and fund projects that are preserving and safeguarding copies in trusted, distributed repositories. Internet Archive,<sup>19</sup> the Data Rescue Project,<sup>20</sup> Environmental Data & Governance Initiative (EDGI)<sup>21</sup> in the US, as well as European organizations like Pangea<sup>22</sup> are working to ensure that the public retains access to critical research data and other outputs produced by the US federal government or with federal funding (Tollefson, 2025).

<sup>19</sup> <https://archive.org/>

<sup>20</sup> <https://www.datarescueproject.org/>

<sup>21</sup> <https://envirotatagov.org/about/>

<sup>22</sup> <https://www.pangea.de/>

# References

- Arthur, P. L., Hearn, L., Ryan, J. C., Menon, N., & Khumalo, L. (2023). Making Open Scholarship More Equitable and Inclusive. *Publications*, 11(3), 41. <https://doi.org/10.3390/publications11030041>
- Bacalex, D. (2022, November). Open Science policy. *Workshop "Open Science, a Landscape under Construction with a Horizon of Possibilities"*. <https://hal.science/hal-03939950>
- Belles, A. B., Beatty, K. E., Rodman, C. H., & Connolly, C. J. (2023). Publish, Don't Perish: Recommendations for Mitigating Impacts of the New Federal Open Access Policy. *Journal of Science Policy & Governance*, 22(1). <https://doi.org/10.38126/JSPG220101>
- Chiware, E. R. T., & Skelly, L. (2022). Open Science in Africa: What policymakers should consider. *Frontiers in Research Metrics and Analytics*, 7, 950139. <https://doi.org/10.3389/frma.2022.950139>
- Cobb-Lewis, D. E., Synder, D., Dumanis, S., Thibault, R., Marebwa, B., Clark, E., Clair, L. S., Kirsch, L., Durborow, M., & Riley, E. (2024). *Investing in Open Science: Key Considerations for Funders* (p. 2024.12.09.627554). bioRxiv. <https://doi.org/10.1101/2024.12.09.627554>
- Crotty, D. (2022, October 27). *Speculation on the Most Likely OSTP Nelson Memo Implementation Scenario and the Resulting Publisher Strategies—The Scholarly Kitchen*. <https://scholarlykitchen.sspnet.org/2022/10/27/speculations-on-the-most-likely-ostp-nelson-memo-implementation-scenario-and-the-resulting-publisher-strategies/>
- de los Angeles Orfila, M. (2024, November 8). 'Scienticide': Argentina's science workforce shrinks as government pursues austerity. *Science*. <https://www.science.org/content/article/scienticide-argentina-s-science-workforce-shrinks-government-pursues-austerity>
- Gorelick, D. A., & Li, Y. (2021). Reducing open access publication costs for biomedical researchers in the U.S.A. *MIT Science Policy Review*. <https://doi.org/10.38105/spr.4nu1qfif3t>
- Halevi, G., van Leeuwen, T., Milićević, N. & Fry, R. (2023). Gold Open Access output and expenditures in the United States in the past decade [preprint]. 27th International Conference on Science, Technology and Innovation Indicators (STI 2023). <https://doi.org/10.55835/64410a4a643beb0d90fc4707>
- Hampson, G., & Steinhauer, J. (2023). *OSI Policy Perspective 6: Considering evidence-based open access policies*. Open Scholarship Initiative. <https://doi.org/10.13021/osi2023.3553>
- Johnson, R. (2022). *Operationalising Open Research Europe as a collective publishing enterprise*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2777/061886>
- Johnson, R., Pinfield, S., & Fosci, M. (2016). Business process costs of implementing "gold" and "green" open access in institutional and national contexts. *Journal of the American Society for Information Science and Technology*, 67(9). <https://doi.org/10.1002/asi.23545>
- Kasanmascheff, M. (2025, April 18). arXiv Is Swapping Cornell University Servers for Google Cloud in Modernization Push. *WinBuzzer*. <https://winbuzzer.com/2025/04/18/arxiv-is-swapping-cornell-university-servers-for-google-cloud-in-modernization-push-xcxwbn/>
- Kemp, J., & Skinner, K. (2024). *The Cost and Price of Public Access to Scholarly Publications: A Synthesis*. <https://zenodo.org/records/14013060>
- Khisro, J., & Fenlon, K. (2025, January 7). Equity in Public Access to Scientific Research Results: Insights from Federal Agency Responses to the Nelson Memorandum Policy. *Proceedings of the 58th Hawaii International Conference on System Sciences*. <https://hdl.handle.net/10125/109103>
- Kiermer, V., Mudditt, A., & O'Connor, N. (2025). Rethinking How We Publish to Support Open Science. *Learned Publishing*, 38(2), e2006. <https://doi.org/10.1002/leap.2006>
- Koley, M. (2022). Analysis of Open Science Policy Recommendations Proposed in India's 5th Science, Technology & Innovation Policy Draft. *Journal of Science Policy & Governance*, 21(02). <https://doi.org/10.38126/JSPG210208>
- Krauskopf, E. (2021). Article processing charge expenditure in Chile: The current situation. *Learned Publishing*, 34(4), 637–646. <https://doi.org/10.1002/leap.1413>
- Library Statement on the Nelson Memo—Open Access—Yale University Library Research Guides at Yale University*. (2023). Retrieved April 15, 2025, from <https://guides.library.yale.edu/openaccess/nelsonmemo>
- National Academies of Sciences, Engineering, and Medicine. (2024). Promoting Equitable and Inclusive Implementation of Open Scholarship Policies: Proceedings of a Workshop—in Brief. Washington, DC: The National Academies Press. <https://doi.org/10.17226/27801>

Okune, A., Hillyer, R., Alborno, D., Posada, A., & Chan, L. (2018, June 15). Whose Infrastructure? Towards Inclusive and Collaborative Knowledge Infrastructures in Open Science. *22nd International Conference on Electronic Publishing*. <https://doi.org/10.4000/proceedings.elpub.2018.31>

*Open Science NL budget cut by half: "The transition to open science is as important as ever"* Open Science NL. (2024, September 17). <https://www.openscience.nl/en/news/open-science-nl-budget-cut-by-half-the-transition-to-open-science-is-as-important-as-ever>

Peterson, A. T., Emmett, A., & Greenberg, M. L. (2013). Open Access and the Author-Pays Problem: Assuring Access for Readers and Authors in the Global Academic Community. *Journal of Librarianship and Scholarly Communication*, 1(3), Article 3. <https://doi.org/10.7710/2162-3309.1064>

Powell, K. R., Townes, J., & Rascoe, F. (2025). Open But Hidden: Open Access Gaps in the National Science Foundation Public Access Repository. *Journal of Librarianship and Scholarly Communication*, 13(1), Article 1. <https://doi.org/10.31274/jlsc.17767>

Rising to meet research needs. (2025, March 6). *Ex Libris Magazine*. <https://exlibris.lib.wvu.edu/news/2025/03/06/rising-to-meet-emerging-research-and-open-science-needs>

*Recommendation on Open Science*. (n.d.). Retrieved April 15, 2025, from <https://www.unesco.org/en/legal-affairs/recommendation-open-science>

Reynolds, J. (2025, March 17). Tech leaders call for EU sovereign infrastructure fund in "crisis" moment. *Tech. Eu*. <https://tech.eu/2025/03/17/tech-leaders-call-for-eu-sovereign-infrastructure-fund-in-crisis-moment/>

Schares, E. (2024, July 31). Data on Datasets: Quantifying US federally funded records in DataCite. *Invest in Open Infrastructure*. <https://investinopen.org/blog/data-on-datasets-quantifying-us-federally-funded-records-in-datacite/>

Schapp, S. (2025, April 21). Nederlandse instituten slaan handen ineen om wetenschappelijke data in VS te beschermen. *NOS Nieuws*. <https://nos.nl/artikel/2564401-nederlandse-instituten-slaan-handen-ineen-om-wetenschappelijke-data-in-vs-te-beschermen>

Sever, R., Eisen, M., & Inglis, J. (2019). Plan U: Universal access to scientific and medical research via funder preprint mandates. *PLOS Biology*, 17(6), e3000273. <https://doi.org/10.1371/journal.pbio.3000273>

Sharp, P. A., Bonvillian, W. B., Desimone, R., Imperiali, B., Karger, D. R., Chakanetsa Mavhunga, C., Brand, A., Lindsay, N., & Stebbins, M. (2023). Access to Science and Scholarship: Key Questions about the Future of Research Publishing. *Access to Science and Scholarship: Key Questions about the Future of Research Publishing*. Retrieved April 15, 2025, from <https://access-to-science.pubpub.org/>

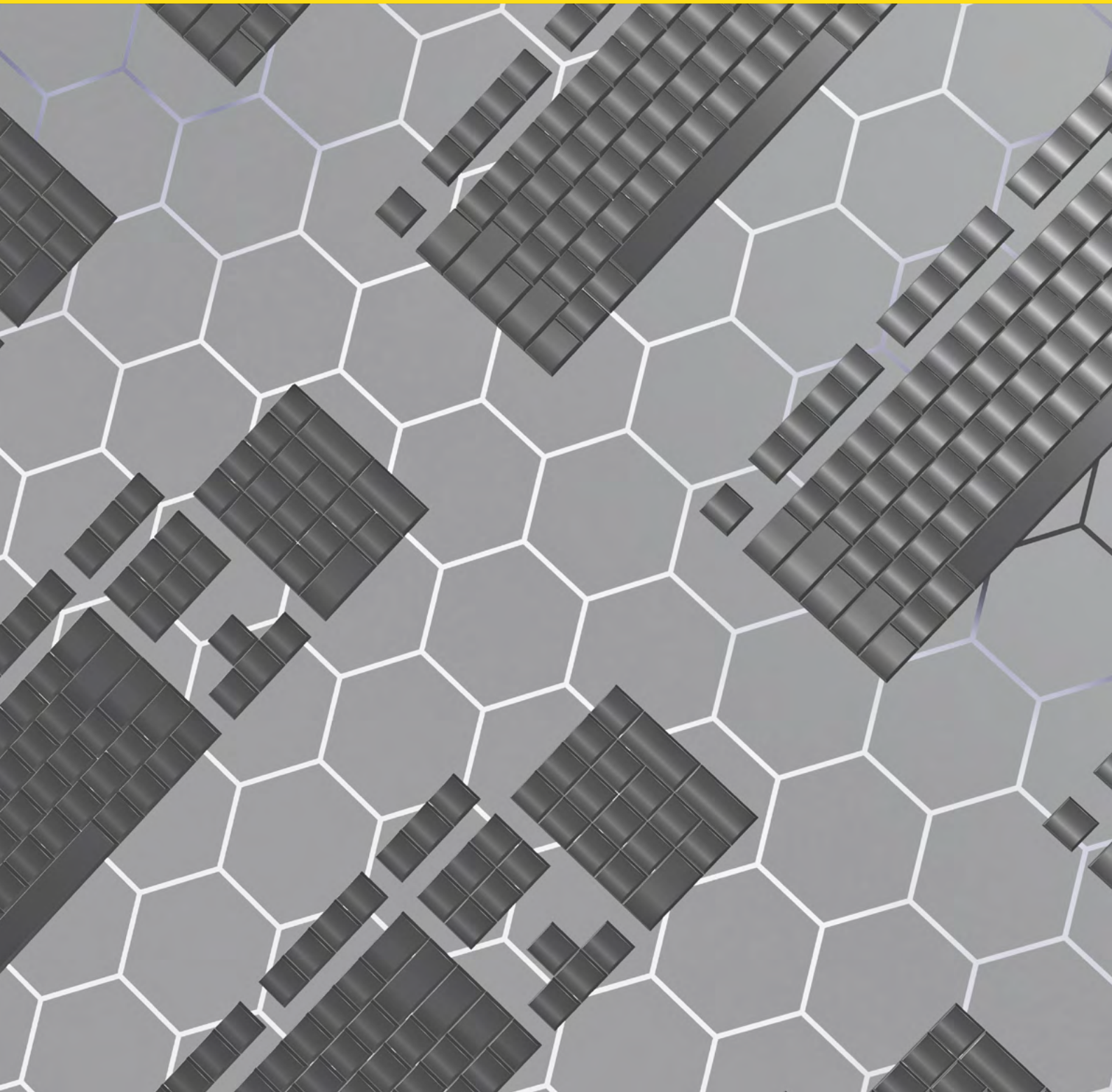
Steinhart, G., Collister, L., Huang, C.-K. (Karl), Lippincott, S., Neylon, C., Riordan, D., Sellanga, J., Skinner, K., Thaney, K., Tsang, E. (2024). *2024 State of Open Infrastructure: Trends in characteristics, funding, governance, adoption, and policy. Invest in Open Infrastructure*. <https://doi.org/10.5281/zenodo.10934089>

Tollefson, J. (2025). Major European institutes join race to save US science data. *Nature*. <https://doi.org/10.1038/d41586-025-01309-3>

Schouppe, W. (2025, April 22). UGent raadt medewerkers wereldwijd aan "Amerika-gevoelige" data op eigen servers te bewaren. *VRTNWS*. <https://www.vrt.be/vrtnws/nl/2025/04/22/ugent-universiteit-president-trump-amerika-data-bescherming-beve/>

White, S. (2025, February 15). *The Hidden Costs Of Freezing Indirect Research Funding*. *Forbes*. <https://www.forbes.com/sites/scottwhite/2025/02/15/the-hidden-cost-of-freezing-indirect-research-funding>

# Vital open infrastructures in a volatile moment





### Guiding questions

- What role can we, as OI stakeholders, play in order to prevent real disruptions and losses on a broad scale in a moment of funding volatility that is impacting so many elements of the research ecosystem?
- How numerous and diverse, or concentrated and related, are the OIs that respondents consider “vital”? How many categories of infrastructure do they include, and how do they connect to each other?
- What relationships are implicit and crucial in our open infrastructure ecosystem (e.g., between OIs, or between OIs and different supporting/using groups), and how can we surface these for action without comparing or ranking these infrastructures in the process?
- How can stakeholders come together across the ecosystem in this critical moment to better understand and bolster support for these OIs?

### Key insights

- The volatility of the current funding landscape in the US may directly and indirectly affect OI organizations and services tied to research and scholarly communication worldwide; we need ways to study, understand, and decisively act across stakeholder roles and geolocations to bolster and strengthen these critical resources.
- We suspect that data about infrastructure use could be layered with additional information about the infrastructures, including the types of data we collect in Infra Finder, to tell much more nuanced stories about the types of relationships and potential engagements possible between OIs. This could help to identify groups of infrastructures that are best suited to work closely together, something that may help to reduce cost, increase visibility, and/or achieve scale for OIs in a challenging funding climate.
- Funding is an obvious and widespread concern, but not all organizations and services in Infra Finder (and in the wider community) face the same types and levels of threats. Specificity is key.

## Introduction

**At IOI, we’re increasingly concerned about how the volatility of the current funding landscape in the US may directly and indirectly affect open infrastructure (OI) organizations and services tied to research and scholarly communication.**

As we’ve described elsewhere in this report, in the first 100 days of the new administration in the US, we’ve borne witness to extreme policies and actions, including billions in terminations of federal grants and contracts; dismantling of agencies including USAID, NEH, and IMLS; and a mounting series of direct attacks on US universities, including politically motivated pauses/threats to cancel additional billions of active grants and contracts and other federal funding sources.

The implications of these and other US-based actions on open research practices throughout the world are significant. Prior to this moment, we were already seeing cuts to open science in parts of Europe and Latin America; but now, as the world deals with the collateral damage done by US withdrawals from various defense, foreign aid, and unification environments, we expect that some of the impacted countries will be motivated to divert even more of their available resources in these directions, potentially to the detriment of their funding for research and science.

To call this “disruption” is an understatement. The landscape is quickly and radically changing in ways that seem certain to negatively impact multiple funding streams for OIs worldwide. Understanding which open infrastructures are at risk at pivotal moments will be key to preventing real disruptions and losses on a broad scale.

## How numerous and diverse are the infrastructures that respondents consider “vital”? How many categories of infrastructure do they include, and do they relate to each other in interesting ways?

Getting access to information about what is most at risk is difficult: it depends on many developing external factors, as well as how secure/embedded/resilient each OI is. And while much attention has been given to the ways this will impact higher education (especially in the US) and scientists and other researchers, far less attention has been given to the way these changes will impact the bottom lines and narrow margins of OIs serving the research community, including hosted or embedded groups and free standing organizations that rely on membership revenues from libraries.

In February we blogged about our concerns and invited infrastructures, particularly those featured in Infra Finder,<sup>1</sup> to share their stories with us. We received responses<sup>2</sup> and we heard additional stories through our personal networks. Based on both formal and informal reports, we gathered that most OIs were (painfully) aware that their own business models were likely to be challenged or compromised by the massive changes underway regarding US federal funding for research and higher education. They also were aware that they didn't know how to predict, forecast, or respond because the changes (and court actions to fight those changes) were still unfolding at dizzying speed. Most OIs are still assessing their own levels of risk, as well as a timeline of when various risks are likely to happen.

We wanted more “in the moment” information from the field to better inform our own approaches to advocating for and supporting open infrastructures. In March 2025, our team designed a poll for limited circulation, only to IOI's newsletter subscribers,<sup>3</sup> in what we knew was a very challenging moment for open infrastructures and those who depend on them. These subscribers include many different types of OI stakeholders, including those who

build and manage infrastructure, as well as those who use and/or fund infrastructure. To mitigate potential concerns, we designed the poll so that:

1. It was brief, with five questions, only two of which were required.
2. It had clear boundaries and purpose, centering explicitly on the then-104 infrastructures featured in IOI's tool, Infra Finder<sup>4</sup>
3. The email invitation, newsletter blurb, and poll intro briefly and consistently stated, *“This poll is meant to help IOI identify groupings, trends and scenarios;”* and promised respondents that results would *“not be used to rank, exclude, or create competition among the breadth of open infrastructures that support our community.”*

Even still, it was hard to push the “send” button that launched it out to 1,000+ of IOI's closest, direct contacts for a two-week period. We knew some recipients might misunderstand our intent; we knew others might be irritated by our premise that US funding volatility could unsettle open infrastructures (especially infrastructures outside of the US). But after running the poll, we're glad we took the risk of putting this quick study out into the field. We're also grateful for those that responded for providing us with data that we could use to test a few really important things in this critical moment. We haven't decided what our next steps will be, but we are seeking ways (and partners!) for further study and surfacing of pertinent, actionable details about what impact volatility is having on OIs and on how we can best support their needs so that they can continue supporting the research ecosystem. Please reach out to us at [research@investinopen.org](mailto:research@investinopen.org) if you have ideas or interest in working with us on this.

<sup>1</sup> <https://infrafinder.investinopen.org>

<sup>2</sup> We briefly chronicled the compelling responses in a blog post at <https://investinopen.org/blog/elevating-open-research-insights-from-the-responses-to-our-call-to-action/>

<sup>3</sup> We initially considered creating a survey for wider input; we worried about the scope and what stories we would be able to tell responsibly if we sought out a broader, more general respondent pool. Pending the outcome of the current, limited-scope study, we may launch a larger study in the future, likely in partnership with additional organizations.

<sup>4</sup> Infra Finder currently features more than 100 open infrastructures serving research and scholarship. These open infrastructures can be viewed as corollaries to physical infrastructure like roads and bridges: They are elements that undergird the way knowledge moves among creators, reviewers, service providers (commercial and non-profit) and users. They include digital and technical components (e.g., repositories, platforms, aggregators, and search-and-discovery systems); standards (e.g., frameworks, identifiers, metadata schemas, and a host of best practices, norms, assessment, and certification programs); and social “glue” (e.g., associations, working groups, and collaborative networks). <https://investinopen.org/data-room/about-infra-finder/>

# Methods

On March 31, 2025, we sent out a quick, anonymous poll via email just to IOI newsletter subscribers, a list of just over 1,000 email addresses.

## Poll design

The Jotform-based poll had five questions, of which only two were required (indicated below by an asterisk):

1. Which of the following infrastructures are most vital for your work? Choose from the Infra Finder list a total of up to 5 selections.\* (see [Appendix A for this list](#))
2. In which country or region/geolocation do you work (primarily)? If you prefer to list this differently, please add it under Other.\* (with checkboxes for: Africa, Asia-Pacific (APAC), Europe, Latin America, Middle East, North America (non-US), and United States, plus a write-in line for "Other")
3. Which best describes your focus/work area most of the time? (with checkboxes for: Data science and/or metadata, Funding, Libraries, Publishing (scholarly/academic/research), Research (any discipline), Research administration, Scholarly communications (non-publishing), Teaching, and Technology, plus a write-in line for "Other")
4. If you were unable (for any reason) to use the infrastructures you selected, what would the result be for your work (rate 1-5, where 1 is no bad effects and 5 is catastrophic)?
5. If you have additional comments, please use the space below. We are particularly interested in infrastructures that you find valuable to your work, your view on the current funding environment for infrastructures specifically, and/or any thoughts about your discipline or field's use of open infrastructures.

As described above, our intent was to query those with connections to IOI (as evidenced through their newsletter subscriptions) about a distinct set of important OIs that are featured in IOI's Infra Finder. Particularly given the complicated set of definitions that emerge around the term "open infrastructure," we knew that we needed clear boundaries and a concrete set of selections to ensure consistency in our poll results.<sup>5</sup>

We structured the poll so that it intentionally focused on trends, not on rankings or other markers pertaining to single infrastructures. We also chose a very specific term, "vital," meaning absolutely necessary or essential, for our main question ("Which of the following infrastructures are **most vital** for your work?") to focus respondent attention on what matters to their own work, no matter what type of work they do or how esoteric or mainstream that work might be. We did this for many reasons, including the diversity of roles in our respondent pool (librarians, publishers, researchers, technologists, funders, etc.) and to try to sidestep or avoid the meanings that other questions might evoke. For example, if we had asked about what infrastructures respondents **use the most** or what infrastructures they consider to be **most important**, either generally or by functions (e.g., repository software) or focal areas (e.g., book publishing), we would be assessing how popular or central a particular infrastructure, or even a set of infrastructures, might be. Again, we were not interested in rankings; our aim was to get at something more important to us and trickier to surface: how numerous and diverse are the infrastructures that respondents consider "vital"? How many categories of infrastructure do they include, and do they relate to each other in interesting ways?

We created the poll to surface connections, especially the breadth, spread, concentration, and focus of the OIs "most vital" to people in different geographic locations and professional roles. To protect the infrastructures from being subjected to popularity rankings in other contexts (intentional or not), we have also made a very difficult (and rare for IOI) choice to not share this data as an open dataset.

<sup>5</sup> Kaitlin Thaney. "What we talk about when we talk about 'open infrastructure,'" *Katina*, Jan 15, 2025. <https://katinamagazine.org/content/article/open-knowledge/2024/talk-about-defining-open-infrastructure>

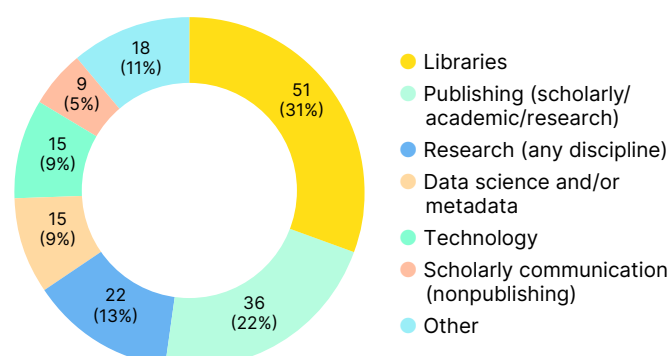
## Respondents

170 respondents (a response rate of approximately 17%) completed the poll in the two-week period during which it was made available. Poll respondents selected 80 OIs — nearly three quarters of all infrastructures in Infra Finder. Their selections demonstrate the wide and distributed usage of many infrastructures in IOI stakeholders' daily work, not just tangentially but in ways that these respondents described as "vital."

Nearly all respondents (166) answered the optional work role question; responses skewed heavily toward scholarly communications work, including librarianship and publishing. Most respondents identified their focus and/or work area as libraries (51 respondents, 31%); additional respondents cited publishing (36 respondents, 22%), research (22 respondents, 13%), data science and/or metadata (15 respondents, 9%), technology (15 respondents, 9%), and scholarly communication (9 respondents, 5%) as their focus or work area.

**FIGURE 1.**

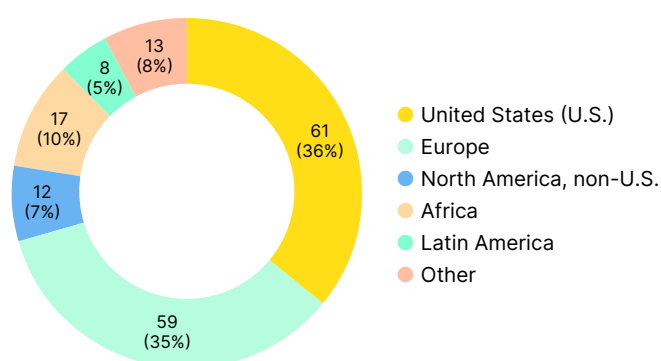
Which best describes your focus/work area most of the time?



Given that the US funding landscape is driving the upheaval in the research enterprise, we specifically listed the US as the only country-level response in the region/geolocation question. We also invited respondents to name their location by country or another designation using "Other." Over a third of respondents (61, 36%) said they work in the US (at least primarily). An almost equal amount (59, 35%) identified Europe as their work location. An additional 10% (17) selected Africa and 5% (8) selected Latin America. North America (non-US), Asia-Pacific, and the Middle East were all represented in the responses as well, and six individuals used "other" to write in Oceania, United Kingdom, and International.

**FIGURE 2.**

In which country or region/geolocation do you work (primarily)? If you prefer to list this differently, please add it under Other.





# Analysis

Wherever reasonable herein, we avoid identifying specific OIs in order to avoid fostering a ranking tone or lens on the results.

We began our work by studying the degree of concentration in the dataset overall, and then analysing these according to respondent geolocation and respondent role to see what, if any, differences were detectable. By the term **concentration**, we mean how compact or how spread out the results are. In particular, we sought distinct patterns and trends in this concentration, first in the dataset as a whole, and then sorted by geolocation and by role, in terms of the overall distribution of answers. For example, we wanted to see if respondents named a small number of OIs far more often than others, or if the distribution of respondent answers was more evenly spread, with many open infrastructures cited at a relatively average rate. We also wanted to see how long the “long tail” was, whether it included many or few infrastructures with lower counts (e.g., 1-3).

We then sought to understand the interconnectedness of the infrastructures in the set, first by how often one infrastructure was **co-named** with another, and then by **constellations** of infrastructures based on patterns of interrelation and trends in OI co-occurrence. By **co-named**, we mean how often infrastructures were listed together by single respondents (e.g., bioRxiv and Zenodo were co-named 10 times; DOAJ and OJS were co-named 21 times). By **constellations**, we mean how deep the relationships between particular sets of OIs are (in our case, using Gephi’s<sup>6</sup> modularity function to detect “communities” within the dataset).

We used Gephi to visualize co-named pairs of infrastructures in the dataset; we also used Gephi’s modularity function to detect constellations of OIs. We adjusted the modularly parameters to create visualizations of four and seven communities in order to identify meaningful constellations. We used the Forceatlas2 layout to format the visualizations.

**Concentration:** How compact or how spread out are the results.

**Co-named:** How often are infrastructures listed together by single respondents.

**Constellation:** How deep are the relationships between particular sets of institutions.

Because we remain concerned about how the poll data could be used and interpreted out of context, we have intentionally masked some of the signalling that Gephi provides through size of node (e.g., making infrastructures that are cited more often larger, and those cited less often smaller) and thickness of lines (e.g., linking infrastructures that are co-named more often with thicker lines and those that are co-named less frequently with thinner lines) in all of our images included herein. We do share infrastructure names and lines between them (standard sized), as well as color coding for some of the constellations that appear within the data, as these details reveal far less about ranking than they do about relationships.

Finally, we analysed the reported effect of being unable to use the OIs and the open response “additional comments” section. Results and discussion are included below.

<sup>6</sup> <https://gephi.org/>

# Results

## Infrastructure concentration

Our first research question related to the **concentration**<sup>7</sup> of responses to the initial poll question (*Which of the following infrastructures are most vital for your work? Choose from the Infra Finder list [below] for a total of up to 5 selections*) as measured in several ways.

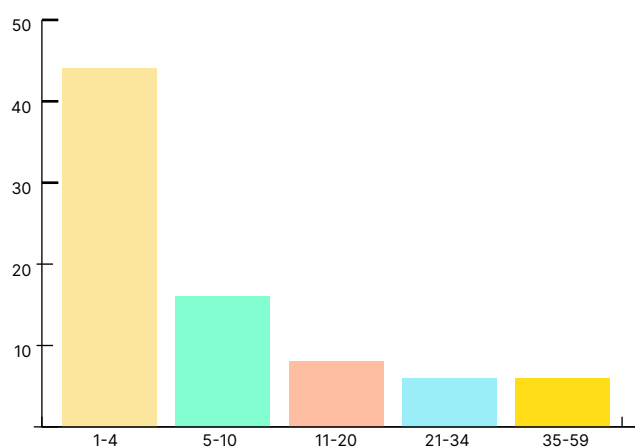
We had 170 total responses to this question. Most (145) of these included the maximum allowed (5 infrastructures); however 25 of our responses included fewer than 5 infrastructures (4 respondents included 4 infrastructures; 6 respondents included 3 infrastructures, 10 respondents included 2 infrastructures, and 5 respondents included only 1 infrastructure). The total number of times the 80 infrastructures were selected across all 170 responses was 784.

### Overall concentration of infrastructures

First, we wanted to understand the overall concentration of infrastructures.

Our 170 respondents named 80 infrastructures between 1 and 59 times. On average, each infrastructure was selected by 10 respondents. The median number of times that each infrastructure was selected was 4.

**FIGURE 3.**  
How many times a single infrastructure was selected by respondents



As depicted in Table 1, data was not highly concentrated around a small number of infrastructures; instead it was broad and included a long tail of infrastructures named 1-2 times. More than half of the 80 selected infrastructures were named between 1 and 4 times<sup>8</sup> (44, or 55%); 16 were named 5-9 times (20%), 8 were named 10-20 times (10%), and 6 were named to each of the two highest categories, 21-34 times (8%) and 35-59 times (8%). The relatively narrow set of 12 infrastructures named more than 20 times stand out as especially “vital” to many respondents. These may be worth further investigation in a future study to determine if this finding holds with a broader pool of respondents.

None of these results are major surprises, though the consistency with which many people across stakeholder groups selected the same 12 infrastructures was relatively striking. While we are not sharing the names of those 12 infrastructures, as per earlier explanations, the range of *functions* represented within them included the following: computing library (1); discovery system (3); index or directory (2); open scholarly dataset (3); publishing system (1); repository service (1); repository software (1); and standard, specification or protocol (3).<sup>9</sup> These 12 infrastructures also represented a diverse range of content types (data, software, journals, monographs, archival, preprint, PIDs, citations) and functions (creation, analysis, dissemination, deposit, discovery). The most notable function that was missing in this diverse set of most-cited infrastructures was digital preservation.

<sup>7</sup> By the term concentration, we mean how compact or how spread out the results are.

<sup>8</sup> As Infra Finder grows over time, this long tail is likely to increase as well; it is no small factor in considering what infrastructures are at risk in the current landscape. Within the poll results, some larger constellations even have their own long tails.

<sup>9</sup> Here and throughout, numerical counts of “functions” include multiple functions for single infrastructures. Our categorization of infrastructures is based on Infra Finder’s solution categories.

## Concentration of infrastructures by region/geolocation

Second, we wanted to see if there were notable differences between which infrastructures were selected by respondents from different regions/geolocations. This comparison was complicated by the large differences in the number of respondents and (as a result) the number of infrastructures selected.

As shown in Table 1, overall, we had similar numbers of respondents and named infrastructures from the US (61 respondents, 60 infrastructures) and Europe (59 respondents, 48 infrastructures), which allowed for limited comparisons across these geolocations. While the number of respondents (17) and named infrastructures (33) was much lower in Africa, which was the next-largest respondent geolocation, we also drew careful comparisons across the US, Europe, and Africa where these seemed meaningful.

We noted the following counts across the US/Europe/Africa responses:

**TABLE 1.**

Selected infrastructures by respondent region/geolocation

	United States	Europe	Africa
Respondents	61	59	17
Named infrastructures	60	48	33
Total infrastructure selections	305	270	65
Average/Median	Average 5, Median 1	Average 6, Median 2	Average 2, Median 1
Repository service	53	62	11
Repository software	66	21	11
Discovery system	59	58	17
Publishing system	18	33	10
Standard, specification or protocol; PID	77	74	21
Open scholarly dataset	35	61	15

### The biggest surprises, perhaps worth further investigation, were the differences we found in repository services, repository software, and publishing system counts across these three locations.

Open repository software and tools listed in Infra Finder were cited over three times more by US-based respondents than by European respondents and six times more than by African respondents (this is a lot even given the notably different respondent sizes in Europe and Africa). Publishing systems in Infra Finder, on the other hand, were much more often cited by respondents in Europe (33 times) than by US respondents (18) or African respondents (10); this size gap between US and European respondents is especially interesting. Explanations for these differences may be related to Infra Finder's current coverage of repository and publishing platforms and tools; they may also signal differences in reliance on open infrastructures for these functions in different regions/geolocations.

We also analysed which infrastructures were selected by respondents across either two or all three of these geolocations. The common core between the US, European, and African respondents in this poll included 23 OIs; licensing tools, publishing tools, and key metadata standards and protocols featured as the most common types of infrastructures in this core grouping. There were also 38 infrastructure selections in common between the US and Europe, 25 infrastructure selections in common between the US and Africa, and 25 infrastructure selections in common between Europe and Africa. Nearly a third of all named infrastructures in the dataset (32%) were held in common between these three regions/geolocations, demonstrating strong support for a set of backbone infrastructures within this limited study.

Although the number of named infrastructures is relatively similar between the US and Europe, concentration of these infrastructures was higher in Europe, with only 49 infrastructures named, and concentration was lower in the US, with 60 infrastructures named.



People in different roles rely on different infrastructure components.

### Infrastructure concentration by role

Third, we looked for differences based on self-identified roles. The poll provided nine options, plus a write-in option marked “other.” The highest scoring categories were Libraries (51, or 31%), Publishing (scholarly/academic/research (36, or 22%), Research (22, or 13%); Data science and/or metadata (15, or 9%), and Technology (15, or 9%); we concentrated our analysis primarily on these categories accordingly.

We found that distinctions in infrastructure selection were both clear and meaningful across these role types. The data show how important it is to include the perspectives of multiple stakeholder types/roles in analyses of OI tools, and also in any planning activities undertaken at a meta-level to understand and/or respond to OI needs. People in different roles rely on different infrastructure components.

We noted the following counts for the main five role types we analysed:<sup>10</sup>

**TABLE 2.**  
Selected infrastructures by respondent role

	Libraries	Publishing	Research	Data science / metadata	Technology
Respondents	51	36	22	15	15
Named infrastructures	55	35	32	25	32
Total infrastructure selections	244	159	102	66	65
Average/Median	Average 4, Median 2	Average 6, Median 2	Average 3, Median 1	Average 3, Median 1	Average 2, Median 2
Repository service	43	43	38	29	7
Repository software	43	4	12	4	17
Discovery system	48	29	14	13	11
Publishing system	22	32	8	3	2
Standard, specification or protocol; PID	49	26	35	28	23
Open scholarly dataset	37	41	8	18	6

<sup>10</sup> To place infrastructures into normalized categories (repository service; repository software; discovery system; publishing system; standard, specification, or protocol; or open scholarly dataset) we relied on Infra Finder’s solution categories.

We also analysed the breakdown of roles by region/geolocation (and vice versa) and found that the distribution of roles across regions/geolocations may complicate some of these findings. Specifically, we had a higher number of librarians in our United States respondents, and a higher number of publishers in our European respondents.<sup>11</sup> This could help to explain some the differences we see in reporting of “Repository software” (much higher overall in the US than in Europe, and also much higher overall in librarians than in publishers) vs. “Publishing system” (much higher overall in Europe than in the US and also much higher overall in publishers than in librarians).

Additional observations we make about this limited data include:

- Respondents from three of the main roles we analysed (Libraries, Publishing, and Research) share a clear dependance on open licensing and copyright tools. This dependence was far less pronounced in the open infrastructure selections by Data science/metadata or Technology respondents; however, Infra Finder does not yet include any open source licensing/copyright tools. If such tools had been represented in Infra Finder, we expect the selections could have risen for these stakeholder groups.<sup>12</sup>
- Respondents from Libraries and Technology roles were more likely to be invested in repository software; as were a few respondents from Research. Respondents from Publishing and Data science/metadata almost never selected Repository software, but did indicate their reliance on Repository services. This is a crucial distinction that may have important implications for what open infrastructures are supported and how. Those who partake in services may not be investing in the tools that those services rely upon, both in terms of their volunteer contributions of time and energy and also in terms of financial contributions.
- Respondents across the board rely heavily on Standards, specifications and protocols and PIDs; however, respondents from Publishing roles seem to be disproportionately lighter in their selection of these elements as “vital” for their work. Given the degree to which the current scholarly communication ecosystem, perhaps especially publishing, relies upon these standards, this may be worth deeper exploration in future studies to understand how visible these are for publishers and why they may not have been top of mind for these respondents.

- Respondents from Libraries selected larger numbers of specifications, standards and protocols, repository software, data tools, and catalogues; however, they selected publishing tools much more rarely. This group also evidenced the highest selection of preservation tools and services.
- Respondents from Research often selected the connective tissue, like standards, specifications, and protocols; they seem to be most focused on Discovery services and on Repository services.

## Infrastructure connections and constellations

After analysing the **concentration** of the data, we sought to understand how the infrastructures represented in the results were connected, both in terms of **co-naming** (e.g. how often single respondents have included the same infrastructures among their up-to-five answers) and in terms of **constellations**, or relatedness of particular sets of infrastructures based on their patterns of co-occurrence.<sup>13</sup>

### Co-named infrastructures

Part of what motivated this poll was our hypothesis that there would be meaning in the relationships between named infrastructures that we could begin to see, even in this relatively small and informal study. We have struggled with how best to share what we are learning without putting undue emphasis on particular infrastructures. We know that the weightedness of the connections between infrastructures (e.g., the difference between two infrastructures being cited together once versus two infrastructures being cited together 15 times) is a crucial marker of relationship, one that we could study on the basis of such factors as interdependence (e.g., one infrastructure is built upon another infrastructure); relatedness (e.g., named infrastructures interoperate); and other forms of connection (e.g., two named infrastructures often support the same content workflow, or two named infrastructures perform related functions, or even two named infrastructures were created by or are organizational hosted by the same group of people).

<sup>11</sup> E.g., Librarians included respondents from Africa (4), Asia-Pacific (1), Europe (13), North America, non-US (1), Other - UK (2), US (30); Publishers included respondents from Africa (3), Asia-Pacific (2), Europe (19), Latin America (3), US (6), Other - UK (1).

<sup>12</sup> We will likely prioritize these standards in our next calls for submissions to Infra Finder (note: open infrastructures can fill out an expression of interest to join at any time; we consider applications on a rolling basis).

<sup>13</sup> As previously noted, because we are concerned about how the poll data could be used and interpreted out of context, we have intentionally masked the signalling that Gephi provides through features like node size and line darkness.



We see each of these scenarios and many more in the 1391 pairs identified in this dataset. Below, we focus on a couple of specific case studies that we think illustrate interesting trends. We use pseudonyms to mask the identities of the infrastructures in these case studies to keep the focus on the relationships we see, not on how many times particular infrastructures were named.

Each pair of infrastructures (e.g., two infrastructures named by the same respondent) appeared together between 1 and 21 times in this dataset. Although the vast majority of pairings only occurred between 1 and 3 times, 20 such pairings occurred between 10 and 15 times; another two pairings occurred more than 15 times. Some of these high-match pairs were relatively predictable cases of two heavily selected infrastructures coming up together repeatedly (e.g., two of the 12-most-often-selected infrastructures). Some were not. We'll look at a couple of case studies representing different situations below.

### Case 1: Common match, one less-selected infrastructure (X and Y)

In the first case, we see a single pair that appears together 10 times. The “match” in this case is between two Repository services. One of the infrastructures (X) is among the 12 infrastructures that were selected most often by respondents; the other (Y) is not. Looking back to the master data for context, these two matched up as a pair nearly every time the less-selected infrastructure (Y) appears in the data set. That less-selected infrastructure (Y) is co-named with 24 other infrastructures in total; many of these lean more toward what we might categorise as data analysis and researcher tools, including Open scholarly dataset, Computing library, and additional Repository services (notably, not Repository software).

The co-occurrence and co-naming in this case likely relates to the slightly more specialized nature of the less-selected infrastructure (Y), which is closely tied to a disciplinary function. Respondents who select this Y infrastructure may be focused on specific disciplinary frameworks. Y was selected more often with similar, though not at all interrelated, infrastructures (e.g., the pairs do not share underlying code, nor do they interoperate or connect explicitly together. The relationship seems to be based on the substance of the research topic and type, and it is one directional. Y responses tended to include X, but the reverse was much more seldom true. X, the more often selected infrastructure, pairs with a total of 44 other infrastructures (with 150 matches).

### Case 2: Common match, more-often selected infrastructures (A and B)

Looking across the data, another match that occurred 15 times stood out in our analysis because the pair was so often related with other elements in the same small field of monograph/book publishing. These two infrastructures were both among the 12 infrastructures that were selected most often by respondents.

What becomes interesting in this case is that the other infrastructures that are co-named with A and with B are, almost without exception, part of the same “family” of book publishing and book discovery infrastructures. In this case, the familial relationships between the infrastructures extend to regular grant-funded partnerships and deeply collaborative work. In essence the pairing, when considered in context, is part of a close-knit community of tools and services that regularly interact. One member of the pair (we'll term it A) pairs only with 20 other infrastructures (with 67 total matches), despite its high selection. The other (B) pairs with 23 infrastructures (with 54 total matches), again despite its high selection. The density of any single pairing (between A and 20 infrastructures and B and 23 infrastructures) happens six or fewer times, in contrast to the A/B pair appearing together 15 times.

By contrast, most high-selection infrastructures have a much broader array of matches, e.g., one of the Open scholarly datasets that respondents selected a similar number of times to both pairs with 36 infrastructures (with 106 total matches).

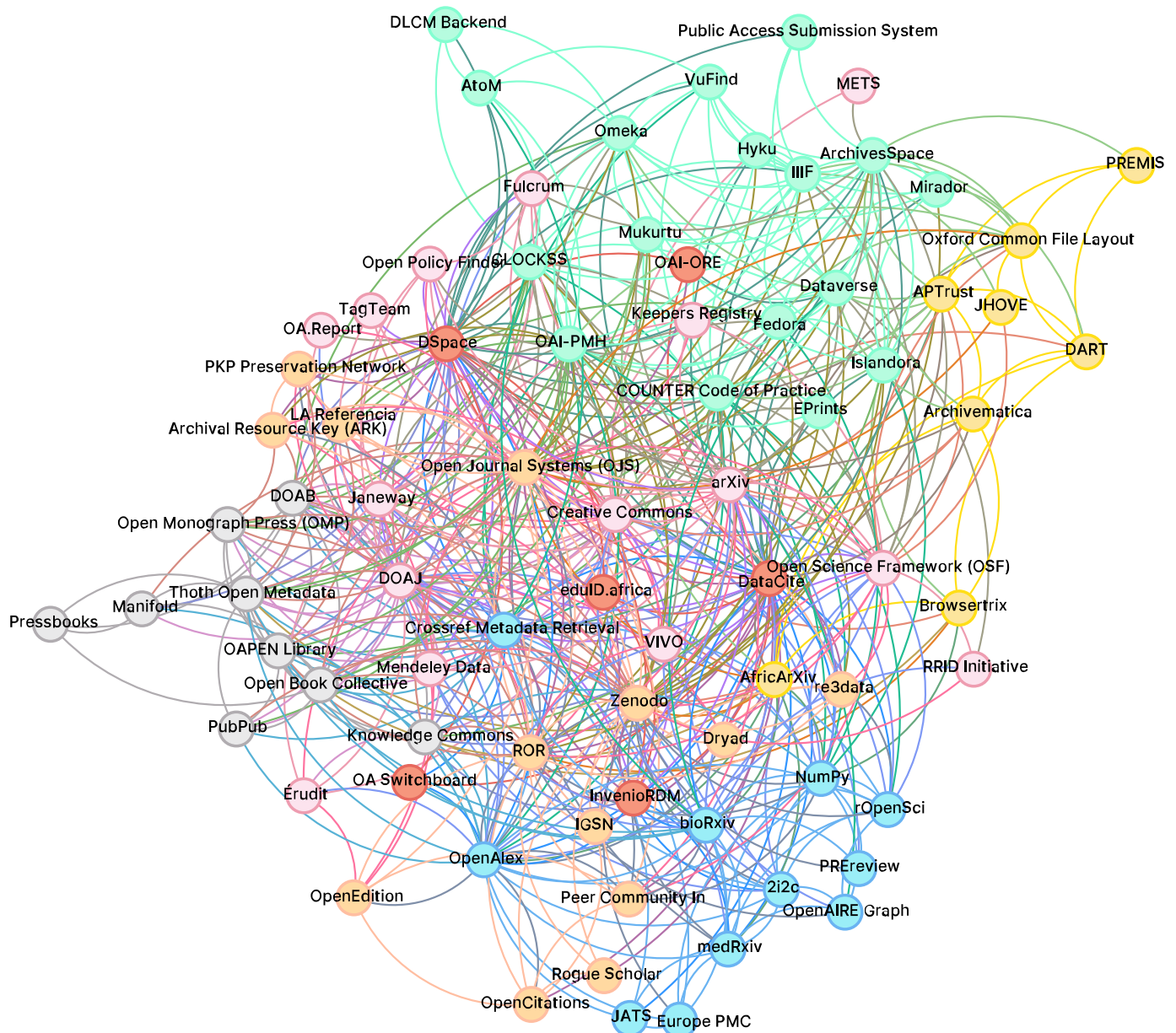
### Constellations

As we designed the poll, we hoped to be able to identify which sets or groupings of infrastructures most often appear together. We term these groupings **constellations** and we suspect this type of analysis might provide the most nuanced and interesting stories about the relationships between Ols, particularly if this data were layered with additional information about the infrastructures, including the types of data we collect in Infra Finder regarding each infrastructure's business/organizational model, governance model, location, code language(s), dependencies, and other key facets. This could identify groups that might be well suited to work more closely together, especially if that helps to reduce cost, increase visibility, and/or achieve scale.

We adjusted Gephi’s modularity parameters to create the visualization we share below. As we describe elsewhere, we have standardized the line weights and the sizes of individual infrastructures in the images we share herein in order to reduce the temptation to see the infrastructures as “larger” or “smaller” and to instead focus attention on the types of constellations we can start to see using Gephi’s color-coded “communities” as our base.

The patterns and trends that are displayed here show a couple of interesting elements that we expand on briefly below.

**FIGURE 4.**  
Infrastructure relationships



First, interconnectivity between infrastructures is dense, even with the very basic data we gathered in this poll. (“Which of the following infrastructures are most vital for your work?”) We didn’t ask respondents to actively share groupings, nor have we added other data (e.g., details from Infra Finder about business or organizational model or location) to complement the poll responses. Just based on the relationships implicit in 170 respondents’ answers that included up to five infrastructures each, we can already start to see meaningful patterns.

As one example, one set of “book” publishing components appeared together in tight accord across respondent answers. When one or two of these infrastructures were selected, others in this grouping tended to also be present. This is immediately visible in the visualization (Figure 4) — the left middle constellation includes elements like Open Monograph Press (OMP), Manifold, Thoth, Open Book Collective, OAPEN Library, PressBooks, and DOAB, as well as Knowledge Commons and PubPub, both of which lean more toward humanities and social sciences publishing and community building.

Another interesting pattern is revealed by looking at the Digital Preservation and Digital Asset Management elements. These don’t all correlate tightly with each other, e.g., CLOCKSS and PKP Preservation Network are both nestled in with journal publishing platforms and standards that each one is often paired with in workflows. One distinctive set in the upper right provides another recognizable preservation constellation, involving APTTrust, PREMIS, JHOVE, DART, Keepers Registry, and Archivematica, all of which factor into archival collection, curation, and preservation workflows.

Other constellations emerge from this picture with journal publishing relationships showing up in pink (platforms like Erudit, Janeway, and OJS, as well as services including DOAJ, Creative Commons, and Keepers’ Registry), Catalogs (DOAJ, DOAB), metadata schemas/standards/collection and reporting (OpenAlex, Crossref Metadata Retrieval, ARK). A blue cluster in the lower right focuses more on science and data tools, standards, services, and environments including repositories and review platforms (EuropePMC, bioRxiv, medRxiv, PREreview, and rOpenSci), and tools (NumPy, 2i2c).

These constellations are far from perfect — data-oriented repositories, discovery tools, frameworks, and standards tend to be scattered throughout, including Dataverse, Dryad, re3data, DataCite; metadata standards and exchange protocols likewise tend to be more dispersed in this graph. Still, this initial depiction demonstrates that even the relationships represented in lists of five “vital” infrastructures as selected by 170 respondents are strong enough to visibly appear in simple visualizations.

IOI is interested in undertaking further investigation into what we can learn through additional research, particularly coupled with the data (and business intelligence) embedded currently in the Infra Finder tool, to help point to promising “backbones” of common and/or complementary traits between existing infrastructures. Given the volatility of the current US research environment in particular, establishing how infrastructures function in relationship with each other can help them to raise their visibility and connectedness. This may be helpful in creating new fundraising mechanisms that work with groups of institutions (e.g., library consortia or disciplinary societies) that share a vested interest in specific sets of infrastructures. It may also help OIs find ways to team up to gain adoption, scale, and reinforce, rather than compete with, each others’ assets, skills, and operational environments.

## Infrastructure reliance level

One poll question asked about the effect of being unable to use the OIs respondents selected, on a scale of 1 to 5 where 5=catastrophic and 1=no bad effects. More than 90% of respondents answered this optional question; of those, more than 80% selected 5 or 4 (catastrophic or very detrimental effects). Only 11 respondents (7%) indicated that the loss of these infrastructures would have only mild or no bad effects, with answers of 2 or 1.<sup>14</sup>

<sup>14</sup> Analysis of this set of responses according to role and region/geolocation revealed that these individuals self identified across a representative range of categories, including Europe, Africa, US, and Libraries, Publishing, Research, Teaching, Data Science and metadata, and Funding.

## Open responses

Nearly a quarter of respondents (41; 24%) replied to the final, optional question in the poll asking for additional comments. The full text of the question helps provide context for some of the results:

*If you have additional comments, please use the space below. We are particularly interested in infrastructures that you find valuable to your work, your view on the current funding environment for infrastructures specifically, and/or any thoughts about your discipline or field's use of open infrastructures.*

10 respondents took the opportunity to express the difficulty in being limited to selecting only five organizations from the Infra Finder list in the poll. Nineteen also named other infrastructures on which they rely (some of those named are included in Infra Finder, others are not; those are noted below with an asterisk):

- |                        |                                   |
|------------------------|-----------------------------------|
| 1. 2i2c                | 25. ORCID*                        |
| 2. Avalon*             | 26. Peer Community In             |
| 3. Biomed News*        | 27. PID networks (general)        |
| 4. CORE                | 28. Problematic Paper Screener*   |
| 5. COUNTER             | 29. PubMed*                       |
| 6. Creative Commons    | 30. PubPeer*                      |
| 7. CRediT*             | 31. R*                            |
| 8. Crossref            | 32. RAiD*                         |
| 9. DataCite            | 33. Redalyc*                      |
| 10. DOAB               | 34. Research Data Alliance (RDA)* |
| 11. DOAJ               | 35. RePec                         |
| 12. Dryad              | 36. Retraction Watch*             |
| 13. DSpace             | 37. ROR                           |
| 14. Fedora             | 38. Samvera*                      |
| 15. Figshare*          | 39. SciELO*                       |
| 16. Hyrax*             | 40. Software Heritage*            |
| 17. JATS               | 41. Thoth Open Metadata           |
| 18. La Referencia      | 42. Unpaywall*                    |
| 19. Latindex Catálogo* | 43. VuFind                        |
| 20. NumPy              | 44. Wikimedia Foundation*         |
| 21. OAI-PMH            | 45. Zenodo                        |
| 22. OJS                |                                   |
| 23. Open Policy Finder |                                   |
| 24. OPERAS*            |                                   |

\*Names with an asterisk are not currently listed in Infra Finder.<sup>15</sup>

A few respondents called out efforts of note in their country or region and/or amplified the resource needs in still-underrepresented regions, including Africa/ sub-Saharan Africa, Argentina, Ethiopia, the Global South and Nigeria (each of which was named explicitly in responses).

Some comments discussed the layered, interrelated nature of Ols, noting that users may be one or more steps removed from an infrastructure and as a result, they may not realize they are reliant upon it. Others wondered what the appropriate granularity is for naming what they rely upon — a specific product or service or the parent organization, for example. Some respondents made clear distinctions between infrastructures they find most vital and others they use, admire or otherwise want to support. Many provided context for their use cases, for example, relying on open infrastructures they personally work on or contribute to.

It's difficult to know whether or to what extent poll responses reflect the particulars of Infra Finder entries or (as seems more likely) if these names call to mind more personal associations for respondents. For example, DataCite encompasses Make Data Count and it is part of the governing collaborative of ROR. Make Data Count does not have its own Infra Finder entry. If a respondent named DataCite as vital for their work, does that reflect reliance on Make Data Count? On ROR? PKP (Public Knowledge Project) appears in four Infra Finder entries; all of its services rely on PKP. How to count 'PKP' in poll results and what kind of follow up to consider based on the responses to each entry are two different questions. What is clear is that there is a lot of dependency among open infrastructures (and their communities).

A few respondents articulated the power and promise of what open infrastructure can do, beyond specific organizations, including supporting the need for academic and knowledge sovereignty. Others provided specific thanks to infrastructures on which they depend or that they see doing important work to make access to knowledge more equitable. One respondent noted something we found particularly resonant — that the loss of the humans behind maintenance of local instances of Ols and of the Ols' own development work would be more catastrophic than losing the infrastructures themselves.

<sup>15</sup> Note that Infra Finder is an "opt in" service to which any open infrastructure can submit an application for review; those open infrastructures that are not included have not been omitted intentionally, they just have not yet approached us. Information on Infra Finder eligibility can be found in the FAQ: [https://hackmd.io/4Lugk\\_tWSMyJBT1TXCmBxA#How-do-you-define-open-infrastructure](https://hackmd.io/4Lugk_tWSMyJBT1TXCmBxA#How-do-you-define-open-infrastructure)



# Discussion

Gaining the necessary vantage point for interdependencies in the larger landscape requires first understanding how “open infrastructure” serves research and scholarship, including key characteristics such as what they do, where they are based, how they are funded, and who owns, governs, or controls them. Looking at these factors in groupings of infrastructures — constellations — is meant to probe commonalities and differences that may be points of vulnerability and strength in networks of infrastructures surfaced by poll results.

As a network of interconnected stakeholders, we — the funders, builders, and users of open infrastructure — have choices to make. We cannot support everything that we have built to date. Doing so is not supporting “infrastructure,” it is more likely to increase clutter and distraction in the landscape. Our knowledge and research landscape is rapidly changing. What we can choose is how we want to change — but we only get to choose how by taking deliberate steps along an active path rather than passively waiting to first see what gets discontinued or what dies on the vine.

Infrastructure relies on meeting broad-based needs with community-wide investment and adoption. Especially in a moment in which major sources of funding are disappearing and long standing partners (e.g.,

government) are no longer consistent or trusted, we need to differentiate between the tools, protocols, services, and environments that are needed by many, versus those that satisfy more specific (though still crucial) needs of a smaller subsets of research communities. These two types of resources need different investment structures and strategies to succeed.

What running this poll in this historical moment has largely done for us at IOI is to convince us that these are questions we have to be asking in new ways, even though (or maybe because) they are subjective and very hard to work with and/or tally without distortion.

The poll approach provides insights that we hope can be used with other data, including sources like Infra Finder, to give a cogent sense of what people most need, care about, and value alongside other information about design, coding language, type of openness, community engagement, governance, usage and adoption levels, dependencies, interoperability, financial stability, hosting model, values (e.g., transparency, openness), and many other characteristics that matter a great deal when trying to create a plan for the future sustainability and care of our OIs. If you are interested in working with us as we seek new vantage points and perspectives, please let us know.



What we can choose is how we want to change — but we only get to choose how by taking deliberate steps along an active path rather than passively waiting to first see what gets discontinued or what dies on the vine.



# Appendix A: Infra Finder list

The first question in the poll asked:

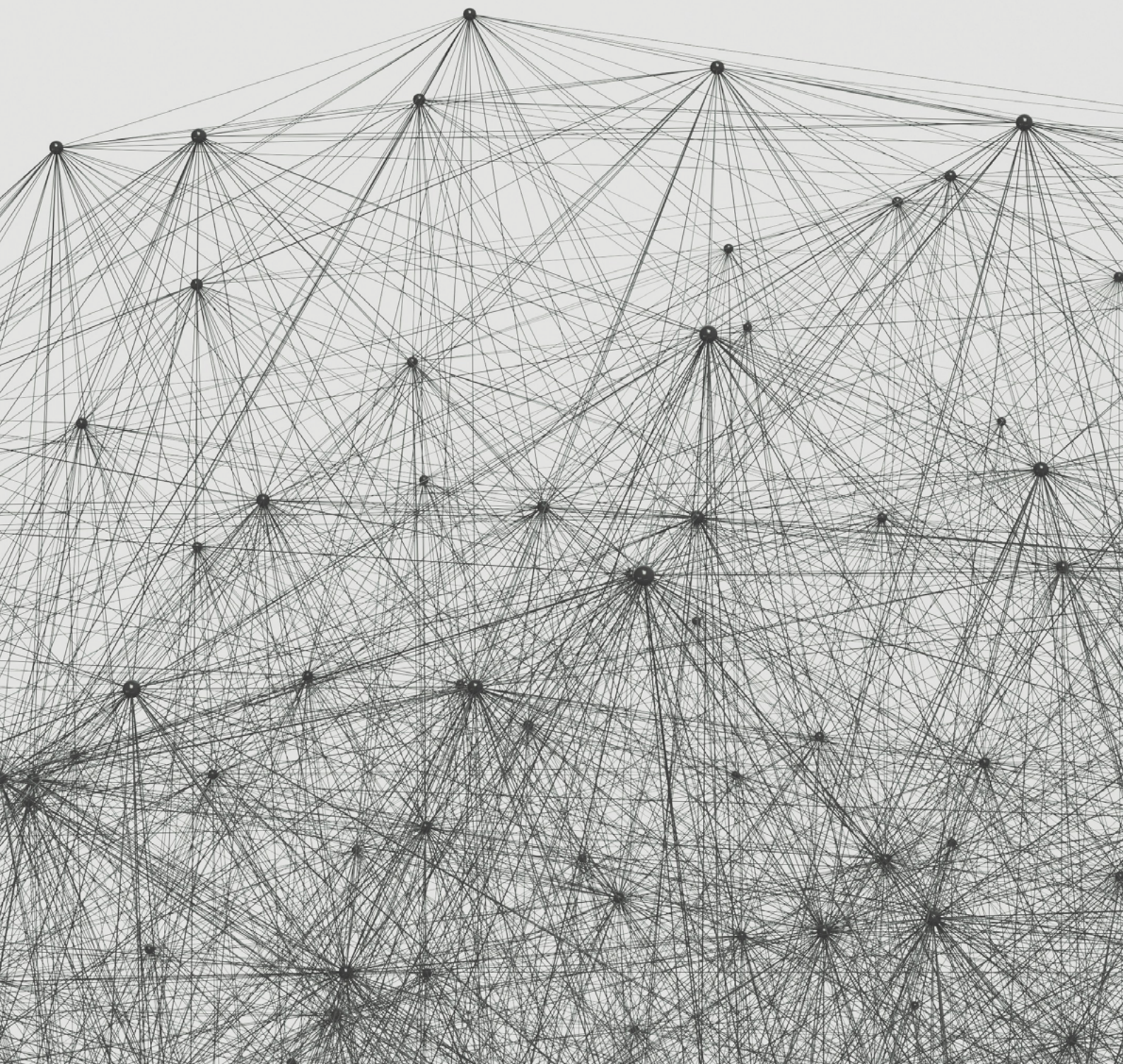
Which of the following infrastructures are most vital for your work? Choose from the Infra Finder list for a total of up to 5 selections.

The following is the list of the 104 open infrastructures from which respondents could choose, ordered alphabetically as it was in the poll:

- |  |  |   |  |
|--|--|---|--|
| 1. 2i2c                                      | 29. DSpace   | 60. NumPy   | 79. OSF (Open Science Framework)                     |
| 2. Academic Preservation Trust (APTrust)     | 30. DuraCloud  | 61. OAPEN Library   | 80. Oxford Common File Layout                        |
| 3. AfricArXiv                                | 31. eduID.africa                                       | 62. OA.Report   | 81. Peer Community In                                |
| 4. Arches Heritage Data Management Platform  | 32. Episciences  | 63. OA Switchboard  | 82. PKP Preservation Network                         |
| 5. Archipelago Commons                       | 33. EPrints  | 64. Omeka   | 83. PREMIS Data Dictionary for Preservation Metadata |
| 6. Archival Resource Key                     | 34. Érudit   | 65. OpenAIRE Graph  | 84. PREreview  |
| 7. Archivemata                               | 35. Europe PMC   | 66. OpenAlex  | 85. Pressbooks                                       |
| 8. ArchivesSpace                             | 36. FAIRiCat   | 67. Open Archives Initiative Object Reuse and Exchange (OAI-ORE)  | 86. Public Access Submission System                  |
| 9. arXiv                                     | 37. FAIR Signposting                                   | 68. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)   | 87. PubPub   |
| 10. AtoM                                     | 38. Fedora   | 69. Open Archives Initiative ResourceSync Framework Specification   | 88. pyOpenSci  |
| 11. bioRxiv                                  | 39. Fulcrum  | 70. Open Book Collective  | 89. re3data.org                                      |
| 12. BrCRIS                                   | 40. Galaxy   | 71. OpenCitations   | 90. REDI   |
| 13. Browsertrix                              | 41. Hyku   | 72. OpenEdition   | 91. Renku  |
| 14. CLOCKSS                                  | 42. IIF  | 73. Open Journal Systems (OJS)  | 92. RePEc  |
| 15. CollectiveAccess                         | 43. International Generic Sample Number (IGSN)         | 74. Open Monograph Press (OMP)  | 93. Research Organization Registry (ROR)             |
| 16. CORE                                     | 44. InvenioRDM   | 75. Open Policy Finder (formerly Sherpa Services)   | 94. Research Resource Identification Initiative      |
| 17. COUNTER Code of Practice                 | 45. iRODS  | 76. Open Preprint Systems (OPS)   | 95. Rogue Scholar                                    |
| 18. Creative Commons Licenses                | 46. Islandora  | 77. Openverse   | 96. rOpenSci   |
| 19. Crossref Metadata Retrieval              | 47. Janeway  | 78. O Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto (The Brazilian Open Access Publications and Scientific Data Portal) (Oasisbr) | 97. RSpace   |
| 20. DART (Digital Archivist's Resource Tool) | 48. JHOVE  |   | 98. Sciety   |
| 21. DataCite                                 | 49. Journal Article Tag Suite                          |   | 99. Synapse  |
| 22. Dataverse                                | 50. Keepers Registry                                   |   | 100. TagTeam   |
| 23. Directory of Open Access Books           | 51. Knowledge Commons                                  |   | 101. Thoth Open Metadata                             |
| 24. DLCM Backend                             | 52. LA Referencia                                      |   | 102. VIVO  |
| 25. DLCM Portal                              | 53. Manifold   |   | 103. VuFind®   |
| 26. DOAJ (Directory of Open Access Journals) | 54. medRxiv  |   | 104. Zenodo  |
| 27. dokiel                                   | 55. Mendeley Data                                      |   |  |
| 28. Dryad                                    | 56. Metadata Encoding and Transmission Standard (METS) |   |  |
|  | 57. Mirador  |   |  |
|  | 58. Mukurtu CMS  |   |  |
|  | 59. MyCoRe   |   |  |



# Signals from the field





# Introduction: Signals from the field

Because of the wide-ranging work we do at IOI, we often catch glimpses of emerging trends in open infrastructure and adjacent fields. In the 2024 report, we called this section *Future Signals*<sup>1</sup>; this year, we've adjusted the name slightly to reflect the fact that sometimes these signals help us to look back and reflect in order to see better ways forward. To that end, in this section now titled *Signals from the Field*, we explore three topics that have emerged repeatedly in our work this year.

These three pieces certainly tackle very different topics. The first piece is a meditation on observations related to the language used within our field, the second is an exploration of an adjacent field of digital public infrastructure, and the third covers our foray into researching measures of community health in open infrastructure. In compiling these pieces for this publication, we noted a particular throughline: a call to

reflect on the values of *open*, whatever they might be in different areas of the field, and to articulate and express those values without reservations. In our first section on the binaries that shape our language around open, we explore how our words can be shortcuts for meanings that may become increasingly obscured. In the second section, we discover that the very concept of open is under debate, and tensions exist between the large-scale adoption of open technologies and the need to ensure those technologies serve their users. Finally, we consider some of the recent signals from the field regarding the health of open source and open infrastructure communities, and how those measures reflect the underlying needs of our communities for sustainable, values-aligned tools and services.

We hope you'll enjoy reading these Signals from the Field.

... we noted a particular throughline: a call to reflect on the values of *open*, whatever they might be in different areas of the field, and to articulate and express those values without reservations.



<sup>1</sup> <https://investinopen.org/state-of-open-infrastructure-2024/sooi-future-signals-2024/>

# Shorthand falls short: Why open infrastructure defies simple labels

## Guiding questions

- What value judgments are encoded in terms like “open,” “non-profit,” and “innovation”?
- What gets lost when we use shorthand to describe the spectrum of practices and characteristics of open infrastructure?
- Does binary thinking impact the ways we choose to adopt and invest in open infrastructure?

## Key insights

- **Binary labels like “open vs. closed” or “non-profit vs. commercial” often obscure more than they reveal.** This section interrogates how these categories can oversimplify the nuanced realities of open infrastructure and proposes a view grounded in core values like community governance.
- **Values-driven work in open infrastructure transcends organizational labels.** Alignment with community needs, sustainable practices, and collaborative practices can matter more than the names of business models.
- **Language is infrastructure.** The words we use, like “sustainability,” “mission-driven,” and “open,” carry assumptions that can shape not only perceptions but also opportunities for funding and partnerships.

In our work at Invest in Open Infrastructure (IOI), we often encounter assumptions that, on the surface, seem obvious or useful but quickly unravel under closer inspection. One of the most persistent of these is what we’ve come to call a binary fallacy: the assumption that complex, nuanced realities can be neatly divided into either/or categories.

This section explores some of these binary descriptions we’ve encountered in the open infrastructure ecosystem, particularly in our work on Infra Finder and through conversations across our network. Our aim is twofold: to illuminate these binaries for what they are — shortcuts — and to unpack the deeper meanings and assumptions that often accompany them and that have surfaced in important ways during our work this past year. To help with this section, we draw not only from our own experiences, but also from interviews with four colleagues who shared their perspectives on this matter: **Geoffrey Bilder**<sup>2</sup> and **Joshua Neds-Fox**<sup>3</sup> shared their expertise and observations from navigating the field, and **Jennifer Roberts**<sup>4</sup> and **Katie Punia**<sup>5</sup> provided their experiences working on and enacting some of these concepts in their time at Artefactual Systems, Inc.

We hope this reflection shows how language can both reveal and obscure the meanings and values encoded in open infrastructure work. We provide this initial reflection as a foundation for exploring two additional signals from the field that showcase the richness and complexity of open infrastructure work.

<sup>2</sup> <https://gbilder.com/>

<sup>3</sup> <https://library.wayne.edu/info/staff-directory/dp5745>

<sup>4</sup> <https://www.artefactual.com/our-team/jennifer-roberts>

<sup>5</sup> <https://www.artefactual.com/our-team/katie-punia>



## Beyond open vs. closed

Many readers may be familiar with one particular well-trodden example in the open movement: open vs. closed. This binary has long been used as a shorthand to describe the nature of access, participation, or licensing.

Many in the community have pointed out that openness exists on a spectrum. For example, the 2014 “How Open Is It?” guide for evaluating the openness of journals from SPARC maps out gradations of openness across six key aspects of journal publishing, reminding us that the term “open” is not a monolith. Guides like these and other spectra can help us describe what “open” means in a way that holds more value than a rigid definition. Furthermore, they can serve as aspirational frameworks to help guide development, instead of shutting out potential participants in our open infrastructure work. In this way, we view “open” as a gradable adjective, rather than an absolute one — something can be more or less “open” according to a set of descriptive parameters.

Following this assertion, we know that “open” can be expressed in various ways, but that there are some key shared values underlying the use of that term. Some of those are context-specific, such as the SPARC journal evaluation tool, which includes these parameters: reader rights, reuse rights, copyrights, author posting rights, automatic posting, and machine readability. Geoffrey Bilder shared a core value that underlies many of the context-specific parameters: that there is “an opportunity for users of a tool, service, or resource to exit”. If, for example, an open source software tool removes a feature that some of its users found valuable, those users should be able to go back to a previous version or fork the software to restore that feature. Suppose a researcher found a helpful article and the publisher removed it from their online archive; in that case, the researcher should be able to keep the copy they downloaded and continue to use it. If an organization provides a service to a group of members, those members should be able to retrieve their data and information in a useful way to move to another service.

Joshua Neds-Fox shared a related core value of “open” in his view: that a tool, service, or resource is open to “operating collaboratively with its community of users in a way that’s not focused simply on growth, extraction, or capture of the commons”. According to Neds-Fox’s view, this focus on community governance and input from users is a key mechanism to longevity and trust in open infrastructure. We see this as a complement to Bilder’s exit strategy assertion: open, community governance is a mechanism to avoid users’ need to exit, but the exit strategy can operate as a guarantee of real and meaningful community engagement.

As we see here, “open” can connote a wide field of meanings, including but not limited to access, participation, and licensing. At IOI, we see this first-hand with our entries in Infra Finder and the many different ways they express the various aspects of “open”. As such, we designed our eligibility criteria in a way that tries to capture some of these parameters, remaining flexible to try to capture what is most useful to our communities.<sup>6</sup>

## The many ways of signalling values

One particular binary description has come up often since our last State of Open Infrastructure Report in 2024: commercial vs. non-profit. When we were designing Infra Finder initially, we used “commercial” and “non-profit” as basic business-type categories for infrastructure services. It seemed straightforward — until it wasn’t. Conversations with our global colleagues quickly surfaced the complexity within these broad terms. Some commercial vendors operate with deep commitments to mission and values. Some non-profits pursue market-based strategies and can lack transparency or community alignment. Furthermore, the term “non-profit” was not specific enough for some of our colleagues who desired more details to inform their understandings of how organizations operate. In response, we expanded our categories (to seven different general business types, and 20 different specific kinds of non-profit status, plus an “other” option).



Open, community governance is a mechanism to avoid users’ need to exit, but the exit strategy can operate as a guarantee of real and meaningful community engagement.

<sup>6</sup> <https://investinopen.org/add-an-infrastructure-to-infra-finder/>

As we did so, we started recognizing ways this binary is used in other contexts. Geoffrey Bilder shared an anecdote that aligns with our experience: at a conference, a new organization was slated to present a plan that promised to disrupt the regular operations of many others in the room. Everyone was tense, wondering about the motivations behind this new organization and its plan. The representative from the group began their presentation by saying, “We’re a non-profit.” Instantly, everyone in the room relaxed. The label triggered a sense of trust, community, and mission even before any further context. We’ve experienced the same when introducing IOI. Saying “non-profit” carries a cultural shorthand that suggests being mission-driven, people-centred, and values-aligned.

But as Bilder and others in our network have pointed out, this shorthand can be misleading. Some non-profit organizations don’t act in alignment with community values, and some commercial organizations do. That’s why watchdog organisations like Charity Navigator<sup>7</sup> and CharityWatch<sup>8</sup> exist; “non-profit” isn’t synonymous with “trustworthy.”

We wondered: what do we really mean when we say “non-profit” in this context? When our team talked to Joshua Neds-Fox for a recent project, he described conducting a “smell test” on organizations he was considering for partnership on a service. He aimed to determine “whether the organization showed signs of being driven by a mission and values” that aligned with his and his library’s, and “demonstrated a commitment to open governance and community contributions”. For this, an organization operating under a non-profit business model might be one factor that contributes to the signals of being a potentially good partner to work with for him and his organization, but it was not the only factor.

“Mission-driven” and “community-led” are other terms that we have seen used similarly as a signal of that alignment, but which also fail to capture the potential complexity. If the mission itself isn’t in harmony with the broader community or if that community is not the one leading the initiative — say, if a mission prioritizes the interests of shareholders above all else — that alignment may be counterproductive in a values-based ecosystem like open infrastructure. We see this prioritization of values in documents like the World-Historical Gazetteer’s Digital Initiative Sustainability Report (Mostern & Straub, 2025), which lays out the need for open infrastructure organizations not only to develop, but frequently return to, “a clearly articulated and consistent set of values” and showcases this practice across several case studies of digital humanities and allied initiatives.

<sup>7</sup> <https://www.charitynavigator.org/>

<sup>8</sup> <https://www.charitywatch.org/>

## Innovation and maintenance

Another binary that has surfaced repeatedly in the past year is a distinction between “research” and “infrastructure” and, relatedly, between “creators” and “maintainers.” These terms and their related constellations of meaning shape funding landscapes, perceived value, and even career trajectories.

Much of the funding in our ecosystem is geared toward innovation: pilot projects, new tools, and experimental platforms. We see this in our funding analysis in the earlier section of this report entitled “The state of open infrastructure grant funding.” This new development work is certainly exciting and vital. But what happens once those platforms and experiments are completed and the final report is sent to the funder? Maintenance — the ongoing work of keeping those tools running, updating documentation, and fixing bugs — is too often underfunded or invisible, and leads to a depreciation of projects and a long-term sustainability issue.

This isn’t a new problem. It’s well-documented that maintainers are overburdened, underappreciated, and ageing out (see, e.g. Nagle et al., 2020; Tidelift, 2024). This is in tension with the need expressed in our community for proof of longevity when selecting infrastructure. Joshua Neds-Fox shared that when he is choosing an infrastructure to partner with, he thinks about a system’s utility decades down the line, and whether it will continue to evolve with its community of users or if it will be captured by a corporate entity and incorporated into proprietary structures. Open infrastructure’s long-term sustainability and longevity can establish trust in this space. That longevity doesn’t happen without maintenance, and maintenance suffers without funding and attention.

So why does the “creator” or “innovator” role get more attention and reward in the funding environment? There is a lot of semantic baggage tied up in these words, and funding mechanisms often define sustainability narrowly as something that can outlive a grant cycle without ongoing financial support. But what if we described sustainability instead as more aligned with maintenance — as a network of relationships, practices, and shared commitments that evolve over time? We explored this concept in a case study through a conversation with Jennifer Roberts and Katie Punia from Artefactual Systems, Inc.

## A case study from the field: Artefactual Systems Inc.

We first met the team from Artefactual Systems, Inc., (hereafter referred to as Artefactual) when we recruited entries for Infra Finder from their open source digital archiving tools Archivematica and AToM. Artefactual was founded by Peter Van Garderen in 2001 and is incorporated as a private for-profit company in Canada. The goal of Artefactual was to create alternatives to commercial digital preservation software that were both technically robust and financially accessible for archivists working in a variety of organizations and conditions. This vision materialized in a model where the software developed by the group is free and open source, and they generate revenue through professional services such as consulting, support, training, and development (Van Garderen, 2009). The software and services are all directed under Artefactual's four values: openness, collaboration, sustainability, and trustworthiness.

We asked about Artefactual's business model and how it helps them do the work in their space. As a private, for-profit company, Artefactual is not externally funded, and the six shareholders are all current or former employees who work within the company's shared values. This business model allows the company to both provide open source tools that are valuable in the digital preservation field, and provide fair compensation and benefits for the people who work for the company. Their operations push against the idea that "Free and Open Source Software" must be free in every aspect. Jennifer Roberts, Systems Archivist, summarized their approach to their business: "The code is free. The ability to use it is free. Our time isn't free." This clarity helps manage expectations around community support while reinforcing the importance of sustainable labour practices in the open ecosystem.

One example of an Artefactual initiative that employees see as aligning practice with values is the creation of "contributor gatherings", designed to facilitate collaboration between the company's maintainers and active members of the user community.<sup>9</sup> These meetings help build continuity in development, foster a distributed support ecosystem, and ensure that the software's ongoing development responds to the community's needs. Jennifer Roberts said about this work that, "Colleagues describe us as a node on a network. Rather than controlling everything about the project, it's about distributing the work, understanding the community's needs and practices, and trying to facilitate a way to make sure our tools will last into the future." At the same time, they acknowledge the reality for many software projects that there will come a time when features and entire software packages may become obsolete. Echoing Geoffrey Bilder's comments in the previous section, Roberts noted that they take a data-focused approach where users can easily take their metadata and digital objects, or use the connectivity built into the tools to send their data elsewhere. This "exit strategy" and intentional counter to the possibility of lock-in is key to how Artefactual operates as an organization, and maintaining those connectors and dependencies has been a focus of their recent development work, sometimes in tension with requests from the community for new features. According to Roberts and her colleague Katie Punia, this intentional prioritization of foundational upkeep over innovation reflects the company values of sustainability and trustworthiness, as well as their understanding of digital preservation as a "practice of care" for the material being preserved, the users who depend on the software, and for the developers doing the work (Simpson, 2020).

This short case study shows one organization's approach to operating beyond the binary distinctions of commercial vs. non-profit and open vs. closed. It shows one way that an organization can show alignment and shared values with its community of users beyond presuppositions encoded in company types.



**"The code is free. The ability to use it is free. Our time isn't free."**

<sup>9</sup> <https://www.artefactual.com/post/gathering-the-contributor-community>

## Looking forward: Language to empower infrastructure

The message of this section is not intended to call out specific uses of language. Binaries can be tempting; they're easy, fast, and sometimes rhetorically powerful. They can be useful shorthand in various situations, yet they also run the risk of flattening complexity and obscuring nuance. And when we're not all on the same page — or when we presume we are but don't articulate our presuppositions — these binary distinctions can carry unintended meanings that erase valuable information and shape decisions in ways we don't yet fully understand.

At IOI, we believe that language is part of the infrastructure. The words we use to describe our work shape how it's understood, valued, and supported. So when we talk about being open, sustainable, or values-aligned, we want to be precise and generous with our descriptions.

We hope that you will keep this discussion in mind when reading the next two sections of our Signals from the Field. When looking into the alignment between digital public infrastructure (DPI) and open infrastructure, for example, we discovered a basic disagreement about what “open” means in the context of DPI and in articulating the goals and values of DPI work. And when it comes to articulating goals and values, this approach is embedded in some community health frameworks such as the Principles of Open Scholarly Infrastructure, CHAOSS Metrics, and the FOREST framework; even though uptake of these frameworks has been limited to specific communities, some of their components are consistently on the minds of the people who work in open infrastructure.



At IOI, we believe that language is part of the infrastructure.

## References

Collister, L. (2024, November 20). Describing open. Katina Magazine. <https://doi.org/10.1146/katina-112024-1>

Mostern, R., & Straub, A. (2025, February). Digital initiative sustainability report. University of Pittsburgh: World Historical Gazetteer. <https://heyzine.com/flip-book/270892f95d.html>

Nagle, F., Wheeler, D. A., Lifshitz-Assaf, H., Ham, H., & Hoffman, J. L. (2020, December 10). Report on the 2020 FOSS contributor survey. The Linux Foundation & The Laboratory for Innovation Science at Harvard. <https://www.linuxfoundation.org/resources/publications/foss-contributor-2020?hsLang=en>

Simpson, J. (2020, November 9). The business of preservation is the practice of care. Digital Preservation Coalition Blog. <https://www.dpconline.org/blog/wdpd/blog-justin-simpson-wdpd>

SPARC, PLOS, & OASPA. (2014). HowOpenIsIt? A guide for evaluating the openness of journals. <https://sparcopen.org/our-work/howopenisit/>

Thaney, K. (2025, January 15). What we talk about when we talk about open infrastructure. Katina Magazine. <https://doi.org/10.1146/katina-01152025-1>

Tidelift. (2024, September). The 2024 Tidelift state of the open source maintainer report. <https://tidelift.com/open-source-maintainer-survey-2024/>

Van Garderen, P. (2009). The ICA-AtoM project and technology. Association of Brazilian Archivists. [https://ica-atom.org/download/VanGarderen\\_TheICA-AtoMProjectAndTechnology\\_AAB\\_RioDeJaniero\\_16-17March2009.pdf](https://ica-atom.org/download/VanGarderen_TheICA-AtoMProjectAndTechnology_AAB_RioDeJaniero_16-17March2009.pdf)



# The intersection between digital public infrastructure and open infrastructure

## Guiding questions

- What can open infrastructure and digital public infrastructure learn from each other?
- How does digital public infrastructure (DPI) perceive what is “open?”
- What are some key similarities and differences between DPI and open infrastructure?

## Key insights

- Participatory governance is key to ensuring that dominant players do not gain undue influence at the expense of the common good.
- As DPI as a concept gains more traction globally, there is a clamour for more decentralized DPIs that are community governed and not beholden to private sector interests.
- There is a need to ensure that as DPIs continue to be adopted, stakeholders actively work to narrow the digital gap by supporting DPI development in developing economies.

Over the past 10 years, the concept of digital public infrastructure (DPI) has gained traction globally. As a field, DPI is still evolving and as a result, many things about how to implement, govern, and contextualize DPI are still in flux. What is clear though is that DPI and open infrastructure are fields that already have a lot in common and could even in future both learn from each other. Some of the tensions in DPI, ranging from issues with the definition of the concept of sustainability, to the role of community in governance, are also all challenges that open infrastructures struggle with, even to date.

For this chapter, we spoke to two experts in the field, **Mila Samdub**, Visiting Fellow at the Yale Information Society Project studying digital government infrastructure in India and **Emmanuel Oloo-Khisa**, Africa Director at the Centre for Digital Public Infrastructure (CDPI). They shared their perspectives on some of the interplays between DPI and open infrastructure. This article summarizes the key thematic areas that surfaced through our conversations.

## Demystifying DPI

Digital public infrastructure is a concept that has gained traction over the past decade or so, but what does it really mean? DPI can be broadly defined as platforms that facilitate the provision of digital products and services, built to enable access to a range of “essential society-wide functions,” such as services in the areas of identification, payments, and data exchange.

Two other key terms are also used interchangeably in the DPI space. According to the Digital Public Goods Alliance,<sup>1</sup> Digital Public Goods (DPG) refers to technologies such as open-source software, open standards, open data, open AI systems, and open content collections that adhere to privacy and other applicable laws and best practices, do no harm, and help attain the United Nations (UN) Sustainable Development Goals (SDGs).<sup>2</sup> The Digital Public Goods Alliance maintains a registry of DPGs, vetting them against baseline requirements that must be met and fulfilled to earn recognition as a DPG. In theory, DPGs provide the open-source building blocks for DPI, allowing governments and communities to adopt interoperable tools without vendor lock-in.

<sup>1</sup> <https://www.digitalpublicgoods.net/>

<sup>2</sup> <https://www.digitalpublicgoods.net/digital-public-goods>

Public Digital Infrastructure (PDI) is a relatively new concept that refers to digital infrastructures designed to maximize public value by combining public attributes with public functions and various forms of public ownership (Krewer & Warso, 2024). Unlike DPI, which can lean heavily on centralized, state-led models, PDI often reflects a more grassroots, decentralized approach.

While DPI enables access to and transfer of information, and often relies on open standards and interoperability between systems, it is not necessarily built on open source solutions. Payment apps like Unified Payments Interface (UPI) in India and Pix in Brazil use proprietary technology and operate as public-private partnerships. These types of partnerships can bring needed capacity, resources, and expertise to bear on societal needs, but can also present ethical and logistical dilemmas. In India, for example, Samdub notes that public-private partnerships may circumvent government guidelines that recommend the use of open source software. Industry involvement in the public sector also provokes questions about how “DPIs are using public data to support the private capture of that value.”<sup>3</sup>

The conflicting definitions of “open” in DPI are similar to the debates in the open infrastructure space, where there have been numerous discussions about defining open infrastructures. Some have tried to define open infrastructures holistically, and others lean toward describing open infrastructures based on specific characteristics (SCOSS, n.d.; Collister, 2024).

## DPI principles and practice

Some key concepts distinguishing DPI from other systems are human centred design philosophy, robust governance, and market/private sector participation (Sang et al., 2025).

Within the human-centered design philosophy there are several key criteria that DPI has to incorporate. One is that DPI needs to be people-centric to ensure that the public’s rights to privacy and autonomy of their data are respected. A key element of the common design philosophy is that DPI needs to be **interoperable** to allow the sharing of data across different systems. Closely

related to this interoperability is the fact that DPI needs to be **modular (extensible)** to allow components to be built on top of each other over time. Another design philosophy is that DPI in principle favours **data federation** (sharing data in a secure and controlled manner over different components) over data centralization. The final design philosophy is that DPI should be **protocol based** — using open protocols and standards to allow for diverse innovation within the system, rather than relying on a single platform.

Another key feature of DPI is **participatory governance**. This is critical because, given the sensitive data domiciled in DPI systems, it is essential that no one singular entity gains undue control over them. These safeguards include approaches like *Privacy by Design* (where security is built in the code) or *Privacy by Policy* (where after the infrastructure is built, there are intentional policy frameworks implemented to enhance data security).

**Market/private sector participation** is meant to ensure that the private sector is an active participant in the increased adoption and scaling of DPI, that there is enough latitude to ensure a competitive market that allows businesses of all sizes to participate, and that incentives are provided to avoid market capture by powerful players.

Oloo-Khisa commented, “Other tech concepts differ here because, for the first time, we are mainstreaming the whole idea of safeguards and then market play. DPI is a technology for the good of society but with a clear thought process around market enablement.”

Samdub commented, “While several of the principles invoked in relation to DPI are admirable, in practice several DPI deployments, including model examples like India’s Aadhaar identity system, don’t follow these principles. Aadhaar, for example, centralizes data storage, leading to data security concerns. Its development excluded civil society participation. And it exhibits several characteristics of a platform, leading some scholars to argue that it has become a form of alt big tech (see Parsheera, 2024).”

<sup>3</sup> <https://www.internetgovernance.org/research/india-stack-public-private-roads-to-data-sovereignty/>



“Could you have a digital public infrastructure that has ownership and governance at a more local or regional scale?” Samdub asks. “These are questions we need to be thinking about.”

## Contextualizing DPI in developing economies

In the formative stages of DPI, a number of non-profit organizations and philanthropies were at the forefront of promoting DPI use globally. While, in principle, this is a good thing, it poses a challenge for countries to apply DPI infrastructures that were built in different countries with different realities and, therefore, not aligned with the needs of local contexts. One example of exported DPI is Modular Open Source Identity Platform (MOSIP), which was built off of India’s Aadhar system, where 17 out of the 26 installations of the system globally are in Africa (Santhanam et al., n.d.). The widespread adoption of MOSIP can be attributed to its interoperability, cost effectiveness but also a lack of African alternatives.

“One way DPI can be more aligned to the local realities in the developing economies is by setting up supporting ecosystems in the creator and developer economies,” said Oloo-Khisa. “There is also a need to work on strengthening governance via platforms like governance roundtables or circles and increasing the amount of funding allocated to support the implementation and maintenance of DPI. Equally important is developing a belief or a can-do attitude that developing nations can build their own DPI systems and not just buy and customize from other contexts.”

Some of the ground realities that make developing DPI in developing economies challenging are as follows:

- An inherent friction between need and speed of rollout: the need for the DPI is urgent for people, but there are often challenges in speed of implementation;
- The local capacity to architect and build minimalist technology infrastructure solutions may be low or vary at different levels of government;
- There may be funding or budget gaps for execution (not just for the new software systems but also for compute/hardware); and
- Procurement cycles may be long and tedious to kickstart progress. There are multiple hoops to jump through before a country can even test out DPI through a pilot.

One of the parallels that can be drawn between DPI and open infrastructure is the fact that development in both sectors is undergoing rapid change. In open infrastructures, it is commonplace that in developing economies, practitioners get tools that have been developed elsewhere and customize them to their own needs. In the DPI space, similar patterns are evident.

## Centralization vs. decentralization: Who builds and who governs?

One of the central tensions in digital public infrastructure today is between centralized, state-led deployment and decentralized, community-driven initiatives. Over the years, DPI implemented at scale has often emphasized adoption metrics (number of registered users and number of people using a particular system) as an indicator of success (Dolan & Satapathy, 2024). To achieve rapid, large-scale adoption, governments may opt for centralized, top-down models due to reduced complexity in interoperability across ministries, consistency in service delivery, and control over implementation.

However, this scale and speed can come at a cost: less transparency, reduced flexibility, and minimal space for civil society or grassroots actors to shape the system. As Mila Samdub observes, “You get a bias from building out a system as fast and as big as possible... it’s not clear that’s the best way to deliver public value.”

By contrast, decentralized or federated approaches — where local or regional actors co-create and govern digital systems — may better reflect community needs and values. These models align more naturally with open-source principles like transparency and collective ownership. Yet they face serious challenges: coordination, technical capacity, and lack of funding can all impede implementation at scale. Currently, the dominant approach is the centralized model. Still, given the increasing awareness of the need for more localized control of people’s data, decentralized DPIs could gain ground in the future.

“Could you have a digital public infrastructure that has ownership and governance at a more local or regional scale?” Samdub asks. “These are questions we need to be thinking about.”

## Balancing public and private interests

In DPI, one of the important aspects is the partnership between the private and public sectors to facilitate the provision of essential services to the citizenry. The public and private sectors have a symbiotic relationship — governments need the capital investment and technical expertise of the private sector, which in turn needs the government to provide them with a conducive business environment.

Transparent and participatory governance is one key pillar of the DPI approach and is crucial in ensuring that dominant players do not gain undue influence over DPI implementation at the expense of the common good.

“The questions around forking of code, source code, and how we contract for DPI are evolving questions. CDPI attempted to address some of this by creating what we call DPI as a package Solution (DaaS) which refers to the rapid deployment of DPI, through upgrades of existing infrastructure (instead of greenfield implementations) through non-procurement routes.<sup>4</sup> In this case, DPI software developers can build a DPI and then make it readily deployable on popular cloud solutions such as Google Cloud or Amazon Web Services, while a separate institution manages the source code, contracting, hosting of different versions, upgrades, and maintenance. There are several instances of this dual-structure model — one that separates community-based development from institutional-level oversight and management. Some of the organizations that have deployed this approach are Red Hat,<sup>5</sup> Drupal,<sup>6</sup> and OpenCVRS.<sup>7</sup> DaaS is a more straightforward way to implement DPIs and can help circumvent some of the complexities in DPI implementation that we are already seeing in the market now. For example, with the development and implementation of e-Citizen in Kenya, where details on the ownership of the platform and how data is utilized has caused some controversy,” remarked Oloo-Khisa. (For more on e-Citizen, see Daily Nation Editorial Board, 2025.)

We can also see a corollary between DPI and open infrastructures in the tension between public and private interests (Samdub, 2025). In open infrastructures, there is a big tension when publicly funded research is then domiciled in repositories and journals owned by private entities.

## Innovations and future outlook for DPI

DPI, as with many other technologies, is rapidly evolving. Advances in blockchain and artificial intelligence, the need to ensure alignment with community needs and decentralization of DPI are some of the emergent topics that may influence how DPI evolves in the future. We asked Samdub and Oloo-Khisa to share what they are excited about in the DPI space going forward.

“There are things that I find inspiring that I wish would be part of the DPI conversation,” said Samdub. “Community networks are forms of commons-based ownership and sovereignty over connectivity. I wish there could be more conversations between community networks and DPI, which has unfortunately largely ignored the question of connectivity. In India, we have Social Audits, a transparency mechanism, especially in public service delivery, which are built in by law to certain programs. This is a collaborative process where the public creates an open, horizontal space to hold the government accountable. Both these community-centric approaches are characterized by a sort of popular sovereignty and accountability that’s entirely missing from DPI as we know it.”

On the other hand, Oloo-Khisa is interested in developing the capacities and a conducive environment for DPI to thrive in developing economies. “As DPI uptake in Africa continues, there are a number of issues that we are trying to address like the shortage of developers, creation of an enabling environment for DPI, and also mobilizing resourcing to support DPI implementation and adoption,” said Oloo-Khisa. “The Africa Digital Economy Lab (ADEL) is a virtual secretariat at the moment that aims at building an environment that facilitates policy conversation on DPI, aggregating funds to support development of the ecosystem, and finally, creating a support structure that helps countries to implement via developing much needed skills like project manager, project quality assurance, and project technical architects that help countries develop DPI solutions.”

<sup>4</sup> <https://docs.cdpi.dev/initiatives/dpi-as-a-packaged-solution-daas/daas-in-a-nutshell>

<sup>5</sup> <https://www.redhat.com/en>

<sup>6</sup> <https://www.drupal.org/association>

<sup>7</sup> <https://www.opencrvs.org/>

## Conclusion

This section was meant to provide an initial exposition of the intersection between two fields that are gathering traction globally. DPI and open infrastructures share a lot in common ranging from how in principle they approach issues like responsible data management, to governance and participation, among others. However, there are also issues that both DPI and open infrastructures are grappling with like how to ensure the interest of the public is always centered despite government and private sector interests, sustainability, as well as impact assessment. While this section provides valuable foundational context, there are still plenty of additional insights that we hope to highlight in the future.

## References

Center for Digital Public Infrastructure. (n.d.). *DaaS in a nutshell. DaaS in a Nutshell*. <https://docs.cdpi.dev/initiatives/dpi-as-a-packaged-solution-daas/daas-in-a-nutshell>

Collister, L. (2024, November 20). *Describing Open*. Describing Open. <https://katinamagazine.org/content/article/open-knowledge/2024/describing-open>

Daily Nation Editorial Board. (2025, March 10). Private ownership of eCitizen a grave risk. Daily Nation. <https://nation.africa/kenya/blogs-opinion/editorials/private-ownership-of-ecitizen-a-grave-risk-4959540>

Desai et al., n.d.. How digital public infrastructure supports empowerment, inclusion, and resilience. <https://blogs.wrldbank.org/en/digital-development/how-digital-public-infrastructure-supports-empowerment-inclusion-and-resilience>

Digital Public Goods Alliance. (n.d.). *Digital Public Goods*. Digital Public Goods Alliance. <https://www.digitalpublicgoods.net/digital-public-goods>

Dolan, J., & Satapathy, S. (2024, March 27). As investment in DPI increases, how can we meaningfully measure its impact? <https://dial.global/how-to-measure-dpi-impact/>

Krewer, J., & Warso, Z. (2024). *Digital Commons as Providers of Public Digital Infrastructure*. Open Future. [https://openfuture.eu/wp-content/uploads/2024/11/241113\\_Digital-Commons-as-Providers-of-Public-Digital-Infrastructures.pdf](https://openfuture.eu/wp-content/uploads/2024/11/241113_Digital-Commons-as-Providers-of-Public-Digital-Infrastructures.pdf)

Lim, Q. (2024, February 1). *NEXT-GENERATION DIGITAL PUBLIC INFRASTRUCTURE*. <https://www.elevandi.io/hubfs/Next%20Generation%20Digital%20Public%20Infrastructure.pdf>

Parsheera, S. (2024, July 10). Digital Public Infrastructure and the Jeopardy of “Alt Big Tech” in India. *Digital Public Infrastructure and the Jeopardy of “Alt Big Tech” in India*. <https://casi.sas.upenn.edu/iit/smrity-parsheera-2024>

Samdub, M. (2025, March 14). What Would it Take for Digital Public Infrastructures to Actually Serve the Public? *The Wire*. <https://thewire.in/tech/what-would-it-take-for-digital-public-infrastructures-to-actually-serve-the-public>

Sang, D., Sambuli, N., & Munga, J. (2025). *Digital Public Infrastructure: A Practical Approach for Africa*. Carnegie Endowment for International Peace. <https://carnegieendowment.org/research/2025/02/digital-public-infrastructure-a-practical-approach-for-africa?lang=en>

Santhanam, N., Narayanan, R., & Rajagopalan, S. (n.d.). *Journey Towards a Billion IDs*. Modular Open Source Identity Platform (MOSIP). [https://connect.mosip.io/resources\\_pdf\\_2025/Journey%20Towards%20a%20Billion%20IDs.pdf](https://connect.mosip.io/resources_pdf_2025/Journey%20Towards%20a%20Billion%20IDs.pdf)

SCOSS. (n.d.). *Defining open infrastructures*. <https://scoss.org/what-is-scoss/defining-open-infrastructure/>



# Trust, transparency, and technology: Do community health frameworks shape open infrastructure decisions?

## Guiding questions

- Which factors count most when people are deciding which research tools to use?
- How relevant are assessment frameworks like CHAOSS, FOREST, and POSI to adopters, maintainers, and funders of open infrastructure?
- Can these frameworks help maintainers improve their sustainability and stay accountable to their user community?

## Key insights

- Practical needs drive decisions: adoption and support of infrastructure tools are primarily driven by practical needs such as cost and fit for purpose. Community health attributes are seen as value-added factors.
- Low framework awareness, but values matter: awareness and use of community health frameworks like POSI, CHAOSS, and FOREST are generally low, but the underlying values of transparency, sustainability, and community governance align with the frameworks' goals.
- Apprehension about framework use: concerns exist about exposing shortcomings when using community health frameworks, particularly when encouraged by funders.

Open infrastructure is a socio-technical construct that needs people, culture, technology, and funding to survive and thrive. The collaborative approach that addresses these needs to ensure the ongoing success and sustainability of the open infrastructure is called “community health”.<sup>1</sup> In recent years, various frameworks — collections of principles, ideals, and metrics — have been designed to help foster understanding of the community health of open infrastructure in the research ecosystem. These frameworks are useful as a tool for orienting people and organizations toward a collective intention, particularly when they are reinforced through incentives and accountability mechanisms.

With the support of a grant from the Digital Infrastructure Insights Fund, Invest in Open Infrastructure (IOI) conducted discovery interviews with a cross-section of members of the open infrastructure ecosystem, including academic libraries, research institutions, funders, technology creators, vendors, and collectives.<sup>2</sup> We wanted to understand what impact, if any, community health frameworks have had on the use, operation, and funding of open infrastructure. To that end, we sought to interview people from organizations that were not signatories to the frameworks in order to get a clearer sense of overall priorities and how widely known the frameworks are.

## Overview of community health frameworks

In our interviews, we focused on three community health assessments that have garnered attention within the open research infrastructure ecosystem:

- 1. The Principles of Open Scholarly Infrastructure (POSI)**<sup>3</sup> — Introduced in 2015, POSI is designed to “guide the operation and maintenance of open scholarly infrastructure” through 16 principles that cover governance, financial sustainability, and “insurance” that code, data, and patents will not lock-in users.
- 2. Community Health Analytics in Open Source Software (CHAOSS) Metrics**<sup>4</sup> — Launched in 2019, this framework focuses on open source software and its contributors, communities, companies, and foundations, helping them understand the health of the open source projects they engage with.

<sup>1</sup> Our initial analysis revealed confusion between “community health” as it relates to open source software and its meaning in public health. We’ll further clarify this distinction in future reports.

<sup>2</sup> <https://infrastructureinsights.fund/>

<sup>3</sup> <https://openscholarlyinfrastructure.org>

<sup>4</sup> <https://chaoss.community/kb-metrics-and-metrics-models/>

**3. The FOREST Framework for Values-Driven Scholarly Communication<sup>5</sup>** — Released in 2022 by the Next Generation Library Publishing project, FOREST is intended to help scholarly communication organizations and communities demonstrate, evaluate, and ultimately improve their alignment with values including: financial and organizational sustainability, openness, representative governance; equity, accessibility, and anti-oppression; sharing of knowledge, and transparency.

This section contains preliminary findings and a sample of the themes we've observed so far. Additional reporting on this project will be released following this report's completion in late 2025.

## Framework awareness and underlying values

A notable finding from our interviews was that **awareness and use of these frameworks was generally low**.

Most interviewees were unfamiliar with POSI, CHAOSS, or FOREST, or had encountered them only in passing. Despite this, when asked about how they view a healthy open infrastructure community, participants consistently emphasized priorities such as transparency, sustainability, and community governance, aligning closely with the frameworks' goals. This finding suggests that the **values embedded in these frameworks are present in the ecosystem**, even if the specific framework tools themselves are not widely referenced.

Another notable finding is that **community health is rarely the deciding factor when organizations select infrastructures to use**. Participants described decisions about whether to adopt or support infrastructure tools as driven primarily by practical needs: does the tool do what we need it to, and does it meet our budget constraints? In some cases, interviewees also wanted to find out who else was using the infrastructure to better understand if it was an appropriate selection. This basic question of "fit" for the needs of an organization indicates to us that open infrastructures and vendors could improve their chances of being selected if they prioritize making it easy to understand what features and interoperability they offer, and list of who is using their solution, or what their ideal users look like.

While most decisions were based on practical considerations such as cost and fit for purpose, **community health attributes were seen as value-added factors** which could contribute to comfort and trust in the decision.

## Relationships matter: The many facets of trust

Given the community-and-collective nature of open infrastructure, trust plays an outsized role. Interviewees described how difficult it is to know what's best-suited for any single organization's needs, constraints, and intentions. Interviewees shared that gathering as much relevant information as possible was crucial to their decision making.

Most frequently, they talked about how important it was to gather information by talking with people involved in the infrastructure. They named that it was also useful to review public-facing communication such as engagement messaging, technical discussions, feature lists and user testimonials to get a sense of technical and social-cultural fit. However, these same interviewees shared that when it came to assessments, the information presented might not, and in some cases can't, expose uncomfortable realities.

Some interviewees said they would be more comfortable with their final choices if they could easily know about more alternatives and have deeper trust in solutions or partnerships, but it was difficult to find information and develop sufficient interpersonal trust ahead of decision time. **We see an opportunity for OIs** to share more information about their user base and capabilities, conducting outreach/engagement to build awareness and trust amongst existing and new audiences.

Some organizations had the capacity and organizational desire to contribute labour or funds back to the community, or saw community governance as a proxy promise to prioritize the institution's needs. Members of these organizations said that governance, community contribution, and financial transparency heavily factored into their evaluations. With this intention, interviewees repeatedly emphasized the importance of **trust** in their decision-making processes. They wanted to know who maintains and governs infrastructure projects, how finances are handled, and under what circumstances labour contributions are integrated.

Interviewees who were, or who represented, OI users wanted to know how welcome they would be in the community, how much influence they would actually have, how long and by what mechanisms would they gain influence, and whether there were political or personality frictions that might cause difficulties.

For those who intended to fund an OI, it was important to be in touch with people responsible for finances, leadership, and work because readily available and accurate financial data was hard to come by, and funders wanted confidence that financial contributions would advance their larger goals.

<sup>5</sup> <https://doi.org/10.5281/zenodo.6557302>

## Apprehension about framework use

Among the few participants who were familiar with community health frameworks, most had **not applied them in practice**. We heard a theme of concern about exposing shortcomings which could result in negative outcomes. As one participant put it, very frankly: “No one likes being judged.”

We also heard a contrasting perspective: that transparency about shortcomings could be an opportunity for values alignment, which could both foster trust and create opportunities to offer support.

Participants who had considered using assessment frameworks said they felt more comfortable using them as a self-assessment to help improve their own organization or project. There was also a general acknowledgement that a third-party evaluation would carry more weight, but would require prioritizing evaluation above other community needs and desires.

Fear of negative consequences surfaced specifically when funders encouraged the use of community health frameworks as part of grant activities, even when these activities were framed as growth and development activities rather than requirements or criteria for qualifying for grants. An interviewee from a funding organization noted: “We’re used to measurement being associated with competition. We’re not accustomed to measuring in service of cooperative behaviors.”

A pattern we saw in the interviews was that organizations needed or wanted to reach a certain level of organizational and operational maturity before even considering a self-assessment. One interviewee noted: “Health frameworks are useful but often feel aspirational. They should be more explicitly tied to solving real problems.”

Additionally, interpretation of key framework values varied widely. For example, the idea of “community governance” meant different things in grassroots, collective-led groups than it did in institutions with hierarchical decision-making structures. Some interviewees said they modified or included individual framework bullets as guidelines or hard criteria in their own documents or evaluation rubrics. POSI and the It Takes a Village Guidebook<sup>6</sup> were the most commonly referenced in these modified uses.

## Moving forward

Despite a general lack of familiarity with the community health frameworks, there is a shared desire across the ecosystem for **clearer signals of functional capabilities, transparency, collective trust, and alignment** from open infrastructure projects. Interviewees expressed that better communication, more visibility into how the organization is governed, who its users are, and what opportunities exist to support capacity building would all help infrastructure users make more confident, values-aligned decisions. Ols themselves expressed appreciation for the aspirational nature of the frameworks, and the opportunity to use them for development of their organizations as they mature.

As we finalize this phase of our research, we welcome your insights and ideas for how to make it easier to select, contribute labour to, or fund open infrastructures, whether by making improvements to Infra Finder or by some other means.

If you have thoughts or suggestions, please send them to us at [research@investinopen.org](mailto:research@investinopen.org).



“We’re used to measurement being associated with competition. We’re not accustomed to measuring in service of cooperative behaviors.”


<sup>6</sup> <https://itav.lyrasis.org/>





# Contributors





We use the Contributor Roles Taxonomy (CRediT) to specify collaborators' contributions: <https://credit.niso.org/>


Lauren Collister,  <https://orcid.org/0000-0001-5767-8486>: investigation, methodology, writing — original draft, writing — review & editing


Chun-Kai Huang,  <https://orcid.org/0000-0002-9656-5932>: data curation, software, visualization


Jennifer Kemp,  <https://orcid.org/0000-0003-4086-3196>: data curation, formal analysis, investigation, methodology, validation, visualization, writing — original draft, writing — review & editing

Sarah Lippincott,  <https://orcid.org/0000-0002-5700-5844>: data curation, formal analysis, investigation, methodology, validation, visualization, writing — original draft, writing — review & editing


Cameron Neylon,  <https://orcid.org/0000-0002-0068-716X>: data curation, software, visualization


David Riordan,  <https://orcid.org/0000-0002-6257-1859>: data curation, software


Jerry Sellanga,  <https://orcid.org/0009-0006-1224-0548>: investigation, writing — original draft, writing — review & editing

Katherine Skinner,  <https://orcid.org/0000-0003-0139-7524>: data curation, formal analysis, investigation, methodology, supervision, writing — original draft, writing — review & editing


Julia Smith, visualization

Gail Steinhart,  <https://orcid.org/0000-0002-2441-1651>: data curation, formal analysis, investigation, methodology, validation, visualization, writing — original draft, writing — review & editing

Kaitlin Thaney,  <https://orcid.org/0000-0002-7217-4494>: conceptualization, funding acquisition, writing — original draft, writing — review & editing

Emmy Tsang,  <https://orcid.org/0000-0002-9248-1280>: writing — review & editing

Kathleen Turaski, visualization

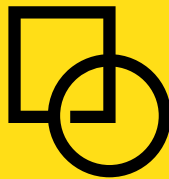
Chrys Wu,  <https://orcid.org/0000-0002-8431-1580>: conceptualization, investigation, methodology, writing — original draft, writing — review & editing

## Funding acknowledgement

This work was generously supported by the Andrew W. Mellon Foundation, the Arcadia Fund, the Digital Infrastructure Insights Fund, and the National Science Foundation (EAGER award, Grant No. 2330827).

We are also grateful for our Sustaining Circle members, whose support is critical to bolstering IOI's capacity to conduct targeted research, provide strategic tools for decision makers, pilot funding mechanisms, and foster a shared agenda for the investment in and adoption of open infrastructure. Find out more about how you can join to advance the use and sustainability of open infrastructure at [investinopen.org/about/funding/#Sustaining-Circle](https://investinopen.org/about/funding/#Sustaining-Circle).





**Invest in Open  
Infrastructure**

## About Invest in Open Infrastructure

---

Invest in Open Infrastructure (IOI) works to increase the investment in and adoption of open infrastructure to further equitable access to and participation in research. We do this by providing actionable, evidence-based tools and recommendations for decision-makers, offering tailored strategic support to infrastructure services and funders, and catalysing investment in open infrastructure. For more information, please visit [www.investinopen.org](http://www.investinopen.org).

### Connect with us

**in** [linkedin.com/company/invest-in-open](https://linkedin.com/company/invest-in-open)

**@** [indieweb.social/@investinopen](https://indieweb.social/@investinopen)

**✖** [bsky.app/profile/investinopen.bsky.social](https://bsky.app/profile/investinopen.bsky.social)

Invest in Open Infrastructure  
c/o Code for Science and Society  
3439 SE Hawthorne Blvd, #247  
Portland, OR 97214-5048  
US

[info@investinopen.org](mailto:info@investinopen.org)  
[www.investinopen.org](http://www.investinopen.org)