2024

STATE OF OPEN INFRASTRUCTURE

Trends in characteristics, funding, governance, adoption, and policy

May 2024

Invest in Open Infrastructure

Publication notes

Title: 2024 State of Open Infrastructure: Trends in characteristics, funding, governance, adoption, and policy

Publisher: Invest in Open Infrastructure

Suggested citation: 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

All images in this report are sourced from https://unsplash.com/.

Table of contents

Foreword

We believe that open infrastructure has an indispensable and irreplaceable role in advancing a research ecosystem that prioritizes access and participation by and for all — and that anchoring our technical systems in community, interdependence, and openness are competitive advantages.

Infrastructure, in its purest sense, speaks to the underlying systems and structures we rely on. In research, these “roads and bridges” that make scientific discovery and scholarship possible and often go unnoticed until there’s a disruption — a service gets shut down or bought, prices increase.

Our work at Invest in Open Infrastructure (IOI) is grounded in the belief that research functions best when the underlying technology and systems — the infrastructure — is open, accessible to all, and maintained in a way that minimizes disruption and empowers the research community at large.

We’ve spent the past two and a half years obsessively studying various dimensions of the sector to identify ways to further the adoption of open infrastructure providers globally, and expand the pool of funding available to sustain their operations. We’ve gathered thousands of data points, interviewed hundreds of experts and practitioners, studied and engaged hundreds of infrastructures, and deepened partnerships across four continents to build a more comprehensive and representative understanding of the open infrastructure ecosystem — and we’ve just scratched the surface.

Today, I am proud to share with you IOI’s first State of Open Infrastructure Report, a snapshot of the deep research and analysis that underpins our work to drive more informed, strategic, and equitable investments in open infrastructure. We designed this report to shine light on the richness, complexity, and opportunity that exist when values and community needs are prioritized in our infrastructure decisions in higher education.

In this inaugural report, we dive deep into the characteristics of open infrastructure powering research and scholarship and what (we believe) sets them apart from their competitors. We take a closer look at the governing bodies and decision-makers behind the technologies your community relies on. We share the latest data and analysis of over US$415M in grant funding powering open infrastructures and research surrounding them, and bring you the latest infrastructure and policy developments in regions such as Latin America, Africa, and the European Union. We highlight success stories and the key trends in the adoption of open infrastructure. We share the latest on trends we’re tracking, such as the global movement towards Diamond Open Access and the underlying infrastructure needs, Artificial Intelligence (AI) and the intersection with open research, and “digital sovereignty” and its impact on research across borders.

There is so much to digest in this report, but we hope you take away the richness and opportunity that open infrastructure provides as an alternative to many of the models and tools we rely on today. We believe that open infrastructure has an indispensable and irreplaceable role in advancing a research ecosystem that prioritizes access and participation by and for all — and that anchoring our technical systems in community, interdependence, and openness are competitive advantages. The funding, policy, and technological landscape in which open infrastructure is embedded is ever-changing and varied, and it is more important than ever that we look deep into the evidence and trends to make more informed, strategic, and coordinated investments to increase the resilience and health of this invaluable ecosystem of infrastructure.

We hope this report sparks your curiosity about the tools and systems we rely on and helps you to think about how your decisions line up with your vision for the future of research and scholarship. We can’t wait to continue the work of advocating for the future of open infrastructure with you.

In appreciation,

Kaitlin Thaney
Executive Director, Invest in Open Infrastructure

Characteristics of selected open infrastructures

Introduction

The *State of Open Infrastructure* report provides an annual snapshot of general characteristics for open infrastructures (OIs) listed in Invest in Open Infrastructure's (IOI) open infrastructure discovery tool, Infra Finder.[1] Infra Finder was developed to inform and support decisions around the adoption of open infrastructures to support research and scholarship (Collister et al., 2024).

Briefly, 84 OIs were invited to participate in the first round of data collection for Infra Finder; 57 of which accepted the invitation and provided information. 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.

This first look at the characteristics of open infrastructures includes the 57 infrastructures included in the initial release of Infra Finder. While our ultimate objectives are to surface patterns in the sector and to illustrate Infra Finder's potential as an analytical tool, the initial dataset is modest in size and scope. As the dataset grows in diversity, size, and geographic coverage, we will be able to draw more varied and robust conclusions from it, and continue to build an evidence base that demonstrates the distinctive characteristics of open infrastructures that make them sound strategic choices for supporting research and scholarship.

Summary of findings

The 57 infrastructures represented in this dataset meet at least one of our basic criteria (see introduction) for openness. We would expect the results to tilt favourably towards open practices and characteristics as well as community accountability, and they do.

In terms of organizational or business form, non-profit organizations (independent, fiscally sponsored, or hosted by an academic institution) dominate the dataset (82%, n=47). Contributions and programme service revenue are the dominant sources of revenue. Open infrastructures perform well in terms of governance, with 86% (n=49) reporting formal or ad hoc governance structures, and somewhat fewer reporting defined activities that sustain governance (47%, n=27). Open infrastructures demonstrate a strong and widespread (80%) commitment to community engagement, with multiple paths for participation. Engagement with values frameworks (such as the Principles of Open Scholarly Infrastructure (POSI; Bilder et al., 2020), the FOREST Framework (Lippincott & Skinner, 2022), the FAIR Data Principles (Wilkinson et al., 2016), and the CARE Principles for Indigenous Data Governance (Carroll et al., 2020) is also widespread (66% or n=38). The Principles of Open Scholarly Infrastructure (POSI) is the most commonly mentioned values framework in use, with ten OIs (18%) having completed a self assessment

[1] https://infrafinder.investinopen.org

As the dataset grows in diversity, size, and geographic coverage, we will (...) continue to build an evidence base that demonstrates the distinctive characteristics of open infrastructures that make them sound strategic choices for supporting research and scholarship.

and eight more (14%) considering one. Policies that demonstrate community accountability and inclusion (e.g. a code of conduct, diversity, equity and inclusion statement, web accessibility policy, privacy policy) are widespread, particularly codes of conduct (70%, n=40) and privacy policies (77%, n=44).

Open data statements, among the OIs for which such a statement is applicable, are common (72% or 26 of 36). Among OIs where pricing information is applicable, 31 of 36 OIs (86%) report transparent pricing, and all but one support this assertion with a URL for pricing information.

Open code repositories (86%, n=49) and open technical documentation (95%, n=54) are widespread, and slightly more than half (56%, n=32) provide an open product roadmap. Seventy-four percent (n=42) of all OIs specify one or more code licenses. Most OIs included in Infra Finder are fully operational and all are actively maintained (with one unreported). For the 22 OIs with the potential to be supported by a service provider (OIs that did not respond "Not applicable" to this question), 86% (n=17) report the availability of one or more hosting options.

Readers are invited to explore the data for themselves by visiting the dashboards configured in Looker Studio and linked within the report sections below, and by downloading the dataset used for this report for independent analysis. Open infrastructure information in Infra Finder[2] is continuously updated, new OIs are being added, and the data are available for direct download.

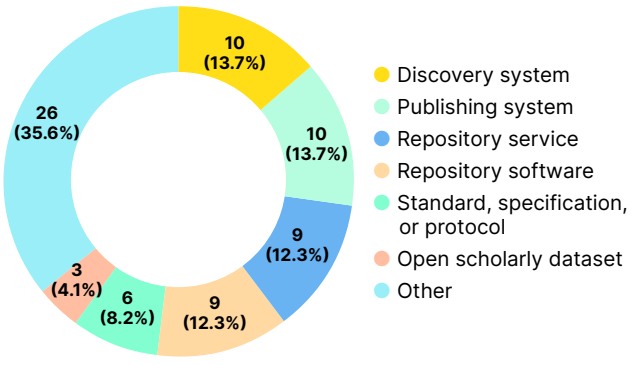
Basic characteristics of open infrastructures

In this section, we summarize the basic characteristics of the OIs in Infra Finder.

Solution categories

The first round of data collection for Infra Finder focused primarily on repository-related infrastructures, and that is reflected in the types of OIs represented in Figure 1. "Other" includes solution categories with only one or two instances in Infra Finder; see Table 1 for the complete list of categories and counts for each. We also note that an infrastructure can be assigned to more than one solution category: 20 OIs reported two solution categories, and nine reported three.[3]

FIGURE 1.
Open infrastructures by solution category



[2] https://infrafinder.investinopen.org

[3] Discovery system is the solution category most often associated with additional categories (five of 10 occurrences, associated with Persistent identifier service, Preservation system, Research profiling system, and Repository software), followed by Repository software (four of 10 occurrences, associated with Discovery system, Repository service, Preservation system, Publishing system, and Digital library, collection or exhibit platform).

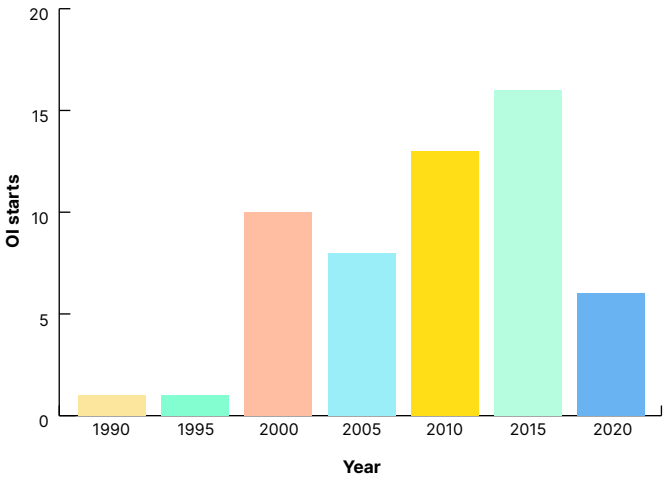
TABLE 1.
Complete list of open infrastructure solution categories

Solution category	Count
Annotations system	2
Archive information management system	1
Authoring tool	2
Digital asset management system	2
Digital library, collection or exhibit platform	2
Discovery system	10
Federated identity or authentication management	1
Index or directory	2
Informal scholarly communications	2
Media viewer/player	1
Open access or subscription management tool	1
Open access policy information compilation	1
Open scholarly dataset	3
Peer review system	2
Persistent identifier service	2
Preservation system	2
Publishing system	10
Repository service	9
Repository software	9
Research profiling system	2
Standard, specification, or protocol	6
Submission system	1

Year of incorporation

OI start dates[4] over time are shown in Figure 2. arXiv (1991) and Érudit (1998) are the longest-lived OIs in the dataset. The timing of important open access declarations (e.g. Budapest Open Access Initiative (2002), Berlin Declaration on Open Access (2003), Bethesda Statement on Open Access Publishing (2003)) and the recognition of repositories as essential scholarly infrastructure (e.g. Lynch, 2003) likely fueled many of the OI developments of the early 2000s. The ten infrastructures started between 2000–2005 include two major repository platforms (DSpace and Fedora) as well as the publishing platform Open Journal Systems. Other key OIs related to open access that were started in this time frame include the Directory of Open Access Journals and Sherpa Services (tools for navigating publisher and funder open access policies and finding open access repositories). Also dating to the early 2000s, Creative Commons sought to simplify and standardize the process of granting permission to reuse creative works, and COUNTER, Crossref, Archival Resource Key, and Journal Article Tag Suite represent important advances in the development of standards and persistent identifiers for use in publishing and repositories. The longevity of these OIs is noteworthy and evidence that they can evolve and endure.

FIGURE 2.
Open infrastructure starts by five-year period



[4] Start dates are self reported, not independently verified, and it is possible respondents interpret start date differently (i.e. formal incorporation of an organization, vs the initial development of a project).

Location of incorporation

The OIs reported in Infra Finder are globally distributed, but the majority are located in North America (32) and Europe (17). This is consistent with findings of Bezuidenhout and Havemann (2020), whose exploration of 242 digital tools for open science found a similar distribution. Four OIs reported no location; these are community-managed OIs that are not formally incorporated or part of a formally incorporated organization.

TABLE 2.
Global distribution of open infrastructures in Infra Finder

Location	Count
Belgium	1
Brazil	2
Canada	6
France	2
Germany	2
Italy	1
Malawi	1
Netherlands	4
Switzerland	1
United Kingdom	6
United States	26
Uruguay	1
Not reported	4

Special certifications or statuses

Open infrastructures were asked to report certification or validation through a recognized process or community group. This free text question garnered a very wide range of responses, but it is noteworthy that 12 OIs report participation in the Global Sustainability Coalition for Open Science Services (SCOSS).[5] Four OIs report being ORCID certified, and the three Public Knowledge Project OIs (Open Journal Systems, Open Monograph Press, and Open Preprint System) are Canadian government "Major Science Initiatives". Other mentions include Guidestar Platinum rating for financial transparency, 501(c)3 equivalency certification, and the security certification SOC 2.

> The timing of important open access declarations and the recognition of repositories as essential scholarly infrastructure likely fueled many of the OI developments of the early 2000s. (...) The longevity of these OIs is noteworthy and evidence that they can evolve and endure.

[5] https://scoss.org/

Governance, business form, and finances

The past decade has seen a flurry of interest in community governance, driven by questions about the sustainability and scalability of open infrastructure projects and high-profile examples of OIs being bought out by large commercial entities. The acquisition of such infrastructure by commercial interests raises concerns about the loss of community control over these essential resources (Hart & Adema, 2022). Best practices for community governance of open infrastructure are difficult to define. At a minimum, from the earliest stages of building an OI, community members should be able to understand how (and by whom) decisions are made (Skinner, 2018). As infrastructures mature, they may look to the abundant resources on building and cultivating healthy communities as well as to evaluative frameworks such as the Principles of Open Scholarly Infrastructure (POSI) and the FOREST Framework for guidance. The existence of governance structures and processes should not be taken as a definitive indication that an OI works exclusively in the community interest, follows best practices in its decision-making processes, or is immune to acquisition. However, having a board or other governance body and being able to point to public bylaws and evidence of stakeholder representation in decision making are important prerequisites for community governance. The Infra Finder data show that, while the majority of OIs do have some form of governance structure in place, there is significant room for growth in making governance processes and activities more formal and more transparent.

Community governance

Open infrastructures were asked to indicate whether they have implemented formal or informal governance structures and processes.[6] The overwhelming majority of respondents reported that they have some form of governance structure. Over 50% (n=32) report having formal governance, while around 30% (n=17) describe their governance as ad hoc. The remaining 14% report that they have no form of governance (Figure 3). As evidence of formal governance, 31 respondents provided links to their bylaws. Forty-seven percent (n=27) of OIs responded that they support meaningful governance via actions and events such as meetings, consultations, elections, audits, etc. (Figure 4); all but two of these provided a URL for information on governance activities.

FIGURE 3.
Percentage of respondents by reported form of governance

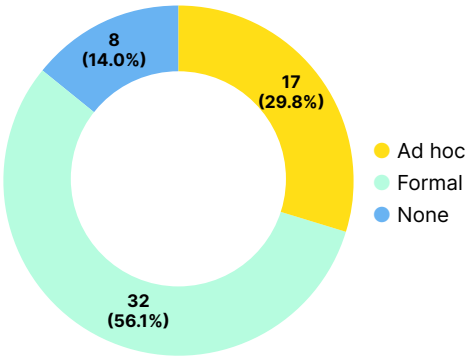
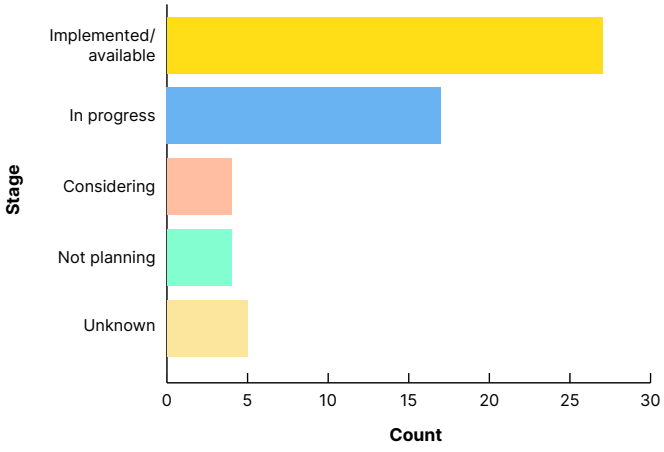


FIGURE 4.
Implementation of governance activities



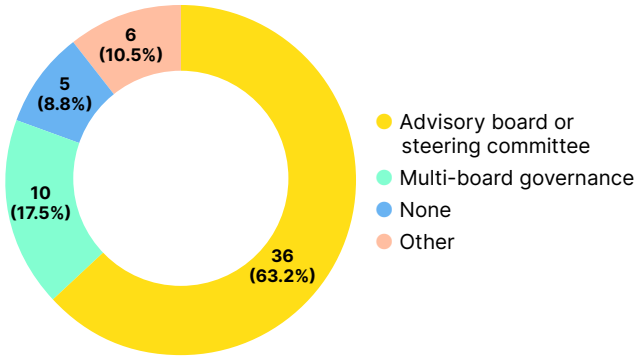
[6] Note that In the case of a fiscally sponsored OI, the answer may apply to the umbrella organization rather than the OI itself.

Board structure

Open infrastructures were asked to indicate if they have a governing board. Over 90% (n=52) report some form of governing board. Of those, over half (63%, n=36) reported they have an advisory board or steering committee. Another 17.5% (n=10) reported they use a multi-board governance style, and the remaining 10.5% (n=6) reported they use another governing board structure.

FIGURE 5.

Percentage of respondents by reported board structure



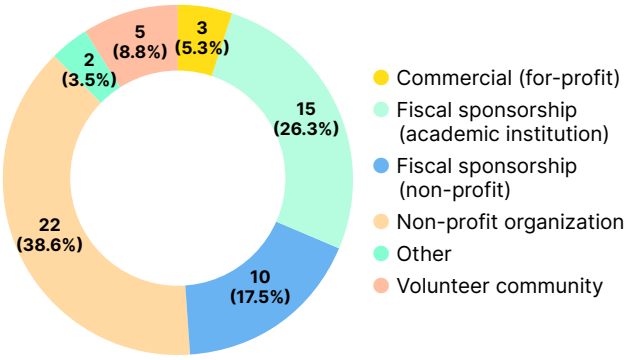
Business form and finances

Business form

Nonprofits of one form or another dominate the dataset; 56% (n=32) are independent or fiscally-sponsored nonprofits, and 26% (n=15) are hosted by an academic institution. As noted in the introduction, our criteria for inclusion in Infra Finder were fairly expansive, allowing for the full range of business forms as long as each OI met at least one of our previously defined criteria for openness.

FIGURE 6.

Business form of open infrastructures

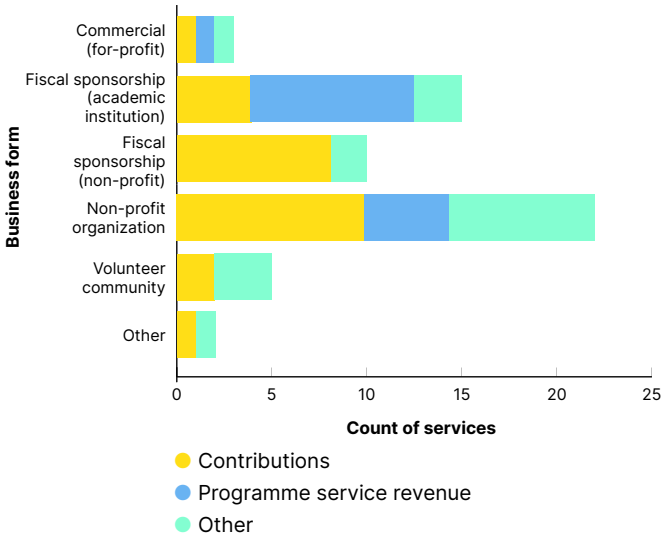


Finances

Nearly half (46%) of OIs report contributions as their primary funding source, followed by other (32%) and programme service revenue (23%). Contributions[7] are the most important source of revenue for independent and fiscally sponsored nonprofits, while programme service revenue[8] is the most important funding source for OIs with an academic institution as their fiscal sponsor. Each of the three commercial OIs reports a different primary source of funding (Figure 7).

FIGURE 7.

Primary funding source by business form



[7] The amount of money that the organization received from donors and grantmakers during the year. For US-based non-profit organizations, IRS Form 990, Part I, line 8.

[8] The amount of money that the organization received from charging fees or selling goods or services related to its mission or programmes during the year. For US-based non-profit organizations, IRS Form 990, Part I, line 9.

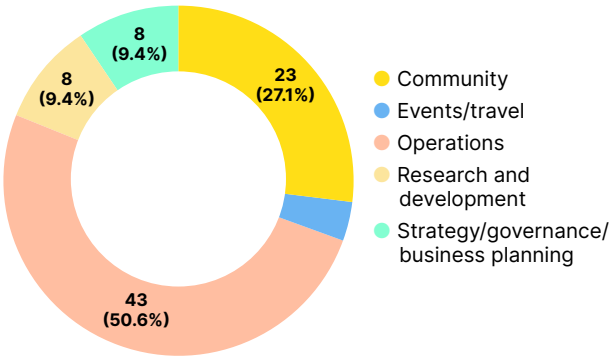
Funding needs

OIs had the opportunity to respond to a question asking them to describe their funding needs. Forty-four OIs provided a response, and the responses were coded for general themes. Some responses contained multiple funding needs that connected to more than one theme, but each theme was counted only once per response; the total count of funding needs is 85.

We mapped this coding of articulated funding needs onto the grant categories described in the "State of open infrastructure grant funding" section of this report (Figure 8). These categories are broad and we found additional richness in the OIs' responses. For example, all of the mentions of DEI work (including translation, fee waivers, and accessibility) are included alongside marketing and communications mentions in the broader grant category "Community", but it seems important to make visible the importance of DEI in community work. The most frequently mentioned funding need was to develop or maintain necessary service functionality, and was categorized as "operations" (which can also include other aspects of operations, such as user support, administration, etc.). This ongoing maintenance dominated the operations category; the other articulated operations needs were related to hiring, training, and budget.

FIGURE 8.

Funding needs by grant award category



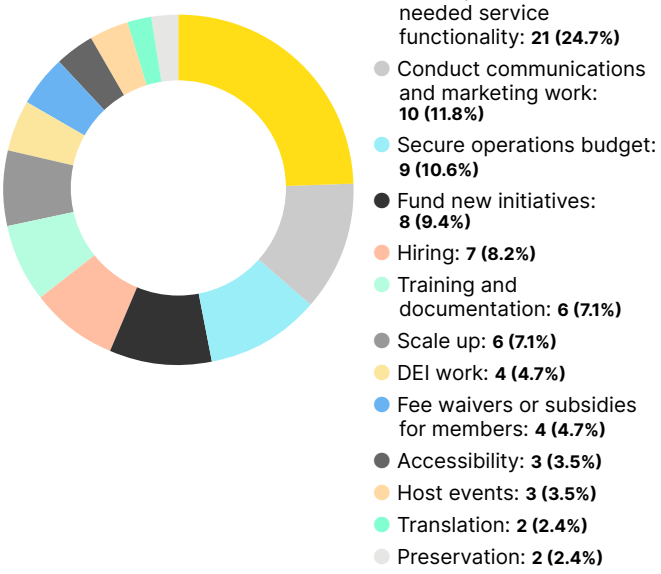
We added an additional layer of categorization to understand the nuances of the funding needs mentioned by providers (Figure 9). Notably, since nearly half (n=21) of responses included mention of critical maintenance or development of functionality, we wanted to clearly distinguish this maintenance need from the work of developing new services or engaging in new projects (n=8, or 18% of OIs). The next largest category was various responses indicating the need to engage with their community of users. Some replies in this category were general (n=10), while others more specifically listed efforts towards diversity, equity, and inclusion (4), and some even more specifically mentioning translation (2), accessibility (3), and fee waivers (4). Operational needs also surfaced, including securing a budget (9), hiring (7), and training (6).

The reported funding needs are diverse, but most of the specific needs articulated in these data involve the sustainability and day to day operations of the OI.

FIGURE 9.

Funding needs

Categories of funding needs. 85 distinct funding needs were reported by 44 of the responding services; percentages here reflect the proportion of the total count of funding needs and not proportion of responding services.



Policy and transparency

Transparency, expressed and communicated through written policies and other documents, encompasses the ways in which an OI represents itself and its business practices to its stakeholders. This may include being upfront about its official business identity and sources of funding, employing consistent and clear pricing terms, and revealing its approach to capturing and protecting user data (e.g. Lippincott & Skinner, 2022). Non-profit organizations that provide more information to their stakeholders perform better across a range of metrics, from fundraising to organizational efficiency (Harris & Neely, 2021). Implementing transparent practices increases an organization's accountability to stakeholders, helping to build trust and providing opportunities for course correction if an organization appears to be out of alignment with its community.

Open infrastructures were asked to report on the availability of several types of policies or documents, and their responses are summarized in Figure 10. Maintaining these resources can demonstrate that the community takes seriously its responsibilities towards fostering productive engagement, supporting users from diverse backgrounds and lived experiences, being transparent with respect to privacy and product pricing, and sharing data as openly as possible.

Code of conduct

Seventy percent of respondents (n=40) reported that they have a code of conduct, which is a document that describes the expectations, norms, rules, and responsibilities for individuals participating in physical and digital communities.

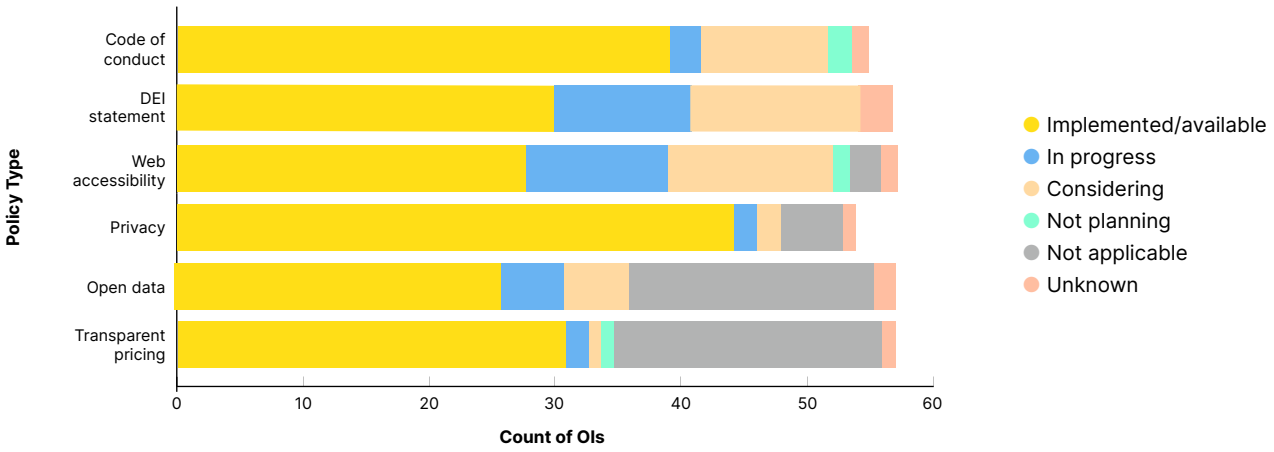
Diversity, equity, and inclusion

Fifty-three percent of respondents (n=30) reported having a stated commitment to diversity, equity, and inclusion, supported by clear initiatives, programmes, or other activities aimed at measuring and increasing the diversity of users, community members, and other key stakeholders.

Web accessibility

Slightly under half of respondents (47%, n=27) reported that they have a web accessibility statement, which is defined as a public document with information about how they make their web-based tools as usable as possible for anyone. Of the OIs reporting such a policy, ten respond that report that the policy applies to the OI or application itself; ten report that the policy applies to the OI's website; and six report that it applies to both.

FIGURE 10.
Summary of availability of policy and other documents



Implementing transparent practices increases an organization's accountability to stakeholders, helping to build trust and providing opportunities for course correction if an organization appears to be out of alignment with its community.

Privacy

Open infrastructures were asked about the availability of a policy regarding the collection and use of user data or personal information. Privacy policies should indicate whether personal information is ever sold, how users can opt out of data collection, and how the OI complies with applicable laws (e.g. GDPR[9]). The majority, 77% (n=44), report an implemented and available privacy policy.

Open data

An open data statement describes an organization's commitment to the use and availability of their data for users of the OI that allows the data to be accessed to the extent possible. An open data statement is not applicable or unknown for 37% (n=21) of OIs. Among the remaining 36 OIs, 72% (n=26) of OIs report an open data statement, and none report that they are not planning one.

Pricing

Over 50% (n=31) reported that they make their costs and fees transparent and provided a link to their pricing information. Another third of OIs (n=21) indicated that transparent pricing was not applicable to their context. Generally, this response means the tool is completely free to use and therefore does not have pricing information. Looking only at OIs where pricing information is applicable, 86% (n=31 of 36) of OIs report transparent pricing, and all but one support this assertion with a URL for pricing information. Tools that describe themselves as a "publishing system" are most likely to have transparent pricing (90%, n=9), probably because they are more likely than other solution categories to be paid services.

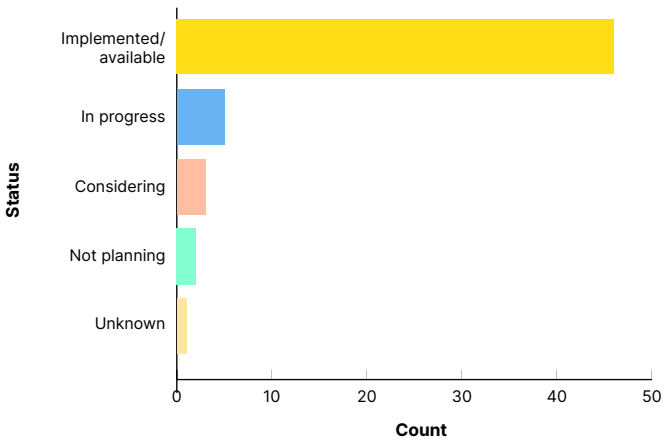
Community engagement

Commitment to community engagement

Community engagement keeps infrastructures accountable to their stakeholders and makes room for their contributions. Governance and policies play an important role and are discussed in previous sections; here we focus on support and avenues for direct, non-governance participation. Respondents were asked how they concretely demonstrate a commitment to community engagement, including indicators such as staff explicitly tasked with community engagement, publicized opportunities for stakeholder involvement, and evidence of community involvement in major decisions or initiatives.

A clear majority (80%, n=46) of respondents reported a commitment to community engagement (Figure 11).

FIGURE 11.
Implementation status of a demonstrated commitment to community engagement



[9] https://gdpr-info.eu/

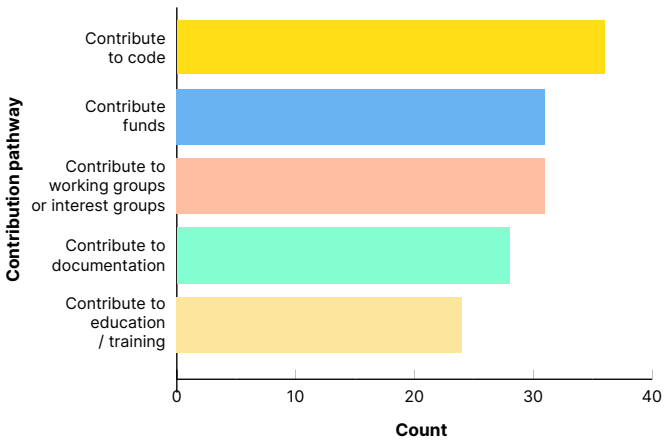
Respondents described a range of specific community engagement activities and strategies, including:

- maintaining a mailing list, discussion forum, or Slack channel;
- having staff roles that include community engagement responsibilities;
- hosting webinars or trainings;
- coordinating user, interest, working, and/or advisory groups;
- hosting regular community calls and/or an annual meeting;
- coordinating a volunteer programme or ambassador network;
- maintaining a presence on social media and blogs;
- conducting user research;
- hosting development sprints or accepting code contributions;
- and participating in conferences.

As in many of the examples above, community engagement often involves an organization pushing out content to keep stakeholders informed, but it should also include ways for stakeholders to meaningfully contribute back into the development of an OI. Contributing code was marginally the most popular response (n=36), followed by contributing funds (n=31) and contributing to working or interest groups (n=31) (Figure 12).

FIGURE 12.

Count of OIs by user contribution pathways

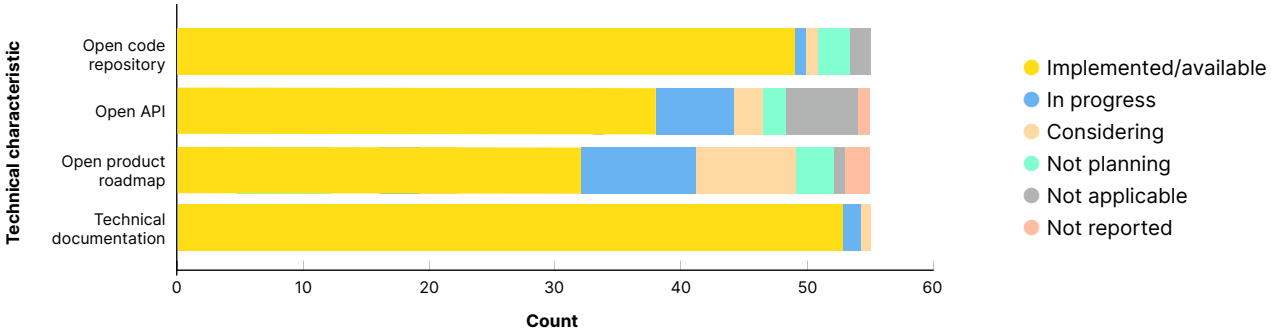


Engagement with values frameworks

Open infrastructures were asked to identify values frameworks (such as Principles for Open Scholarly Infrastructure, the FOREST Framework, etc.) that they have adopted or engaged with. Responses were free text and quite varied, but it is noteworthy that 67% (n=38) of all OIs responded that they had or were planning to engage with at least one values framework. Ten respondents indicated that they have completed a self audit against the Principles of Open Scholarly Infrastructure (POSI) and a further 8 indicated they have been inspired by POSI or are planning a self audit. Respondents also mentioned the FOREST Framework (n=2) and Lyrasis's It Takes a Village toolkit (Gemill Arp & Forbes, 2018)(n=2). Several respondents referenced the FAIR Data Principles (n=4) and the CARE Principles for Indigenous Data Governance (n=3). Additional but less frequent mentions (one or two occurrences) include the Contributor Covenant (Ehmke, 2014), the Declaration on Research Assessment (DORA, n.d.), FAST principles for preprint peer review (Iborra et al., 2022), Helsinki Initiative on Multilingualism (Federation of Finnish Learned Societies et al., 2019), Joint Statement of Principles of the Coalition for Diversity and Inclusion in Scholarly Communication (C4DISC, 2023), COAR Next Generation Repositories Principles (n.d.), HumetricsHSS values framework (Humane Metrics Initiative (HuMetricsHSS), n.d.), the UN Sustainable Development Goals (United Nations, 2015), and others.

Technical characteristics

FIGURE 13.
Status of technical attributes



Open code repository, open API, and technical documentation

Openness of code and other technical resources is a hallmark of the responding OIs. The majority (86%, n=49) of OIs maintain an open code repository (Figure 13). Sixty-five percent (n=37) of OIs provide an open application programming interface (API). The vast majority (95%, n=54) provide technical documentation.

Open product roadmap

Slightly over half (56%, n=32) of respondents reported that they maintain an open product roadmap, a publicly available document outlining the priorities and progress for product development and in some cases allowing users to submit their own ideas and feedback or vote on proposed features (Figure 13).

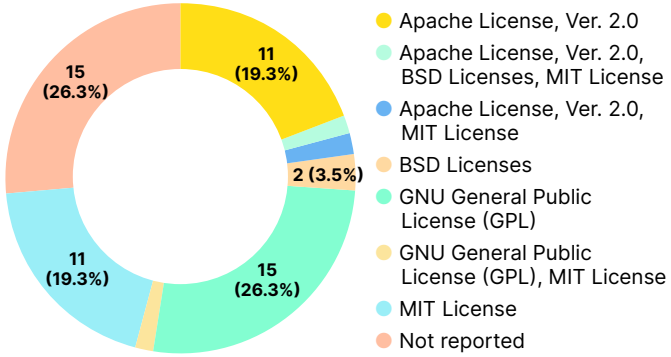
Technical documentation

Technical documentation is widely available (95%, n=54) (Figure 13).

Code licensing

Open licensing is common across the OIs, although we lack specific license information for 15 OIs in Infra Finder. GNU General Public License is the most commonly reported. Thirty-nine OIs assign a single license to their code while two assign two licenses and one assigns three (Figure 14).

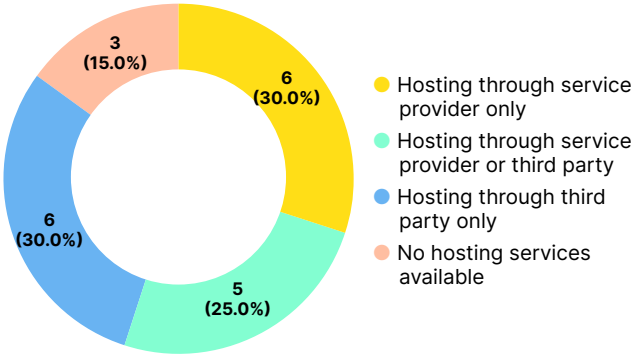
FIGURE 14.
Code licenses in use by OIs



Software as a Service (SaaS) service provider availability

For OIs that might be installed and supported by institutions, the availability of vendor-provided services for open infrastructure can be key in selecting an open solution if the institution does not have dedicated technical staff or infrastructure available locally. Excluding OIs for which hosting is either not applicable or not reported, we found that for the majority (86%, n=17 of 22) some kind of hosting option is available (Figure 15).

FIGURE 15.
Availability of hosting services



Data availability

The data used for the analyses presented here was exported from Infra Finder on 30 January 2024. Some error corrections have been made to the dataset, but subsequent updates from OIs have not been incorporated. The data (Lippincott et al., 2024) may be downloaded from: https://doi.org/10.5281/zenodo.10835677. Data may also be explored using interactive dashboards: https://lookerstudio.google.com/s/vg1ZZ13YEmQ.

The Infra Finder intake form used to collect the data reported here (Lippincott et al., 2024) is available online: https://doi.org/10.5281/zenodo.10835682.

References

Berlin Declaration. (2003). Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. https://openaccess.mpg.de/Berlin-Declaration

Bezuidenhout, L., & Havemann, J. (2021). The varying openness of digital open science tools (9:1292). F1000Research. https://doi.org/10.12688/f1000research.26615.2

Bilder, G., Lin, J., & Neylon, C. (2020). The Principles of Open Scholarly Infrastructure (v1.1, 2023). The Principles of Open Scholarly Infrastructure. https://doi.org/10.24343/C34W2H

Brown, P. O., Cabell, D., Chakravarti, A., Cohen, B., Delamothe, T., Eisen, M., Grivell, L., Guédon, J.-C., Hawley, R. S., Johnson, R. K., & others. (2003). Bethesda statement on open access publishing. http://nrs.harvard.edu/urn-3:HUL.InstRepos:4725199

Budapest Open Access Initiative. (2002). https://www.budapestopenaccessinitiative.org/read/

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. Data Science Journal, 19, 43. https://doi.org/10.5334/dsj-2020-043

Coalition for Diversity and Inclusion in Scholarly Communications (C4DISC). (2023). Joint Statement of Principles. Coalition for Diversity and Inclusion in Scholarly Communications. https://c4disc.org/joint-statement-of-principles/

Collister, L., Tsang, E., & Wu, C. (2024). Infra Finder: A new tool to enhance transparency, discoverability, and trust in open infrastructure. International Digital Curation Conference (IDCC), Edinburgh, Scotland. [Preprint] https://doi.org/10.5281/zenodo.10913249

Confederation of Open Access Repositories (COAR). (n.d.). COAR Next Generation Repositories: Principles. https://ngr.coar-repositories.org/principles/

DORA. (n.d.). San Francisco Declaration on Research Assessment. DORA. https://sfdora.org/read/

Ehmke, C. A. (2014). Contributor Covenant: A Code of Conduct for Open Source and Other Digital Commons Communities. https://www.contributor-covenant.org/

Federation of Finnish Learned Societies, Information, T. C. F. P., Publishing, T. F. A. F. S., Universities Norway, & European Network For Research Evaluation In The Social Sciences And The Humanities. (2019). Helsinki Initiative on Multilingualism in Scholarly Communication. 621757 Bytes. https://doi.org/10.6084/M9.FIGSHARE.7887059

Gemill Arp, L., & Forbes, M. (2018). It Takes a Village: Open Source Software Sustainability—A Guidebook for Programs Serving Cultural and Scientific Heritage. Lyrasis. https://itav.lyrasis.org/guidebook/

Harris, E. E., & Neely, D. (2021). Determinants and Consequences of Nonprofit Transparency. Journal of Accounting, Auditing & Finance, 36(1), 195-220. https://doi.org/10.1177/0148558X18814134

Hart, P., & Adema, J. (2022). Introduction. In P. Hart, J. Adema, & COPIM (Eds.), Towards Better Practices for the Community Governance of Open Infrastructures (1st ed.). Copim. https://doi.org/10.21428/785a6451.57d211e7

Humane Metrics Initiative (HuMetricsHSS). (n.d.). HuMetricsHSS Values Framework. https://humetricshss.org/our-work/values/

Iborra, Sandra F., et al. “FAST Principles for Preprint Feedback.” (2022). OSF Preprints, 13 Jan. 2022. https://doi.org/10.31219/osf.io/9wdcq

Lippincott, S., & Skinner, K. (2022). FOREST Framework for Values-Driven Scholarly Communication. https://doi.org/10.5281/zenodo.6557301

Lippincott, S., Skinner, K., Steinhart, G., & Wu, C. (2024). Intake form for Infra Finder - 2023 (Version 1). Zenodo. https://doi.org/10.5281/zenodo.10835682

Lippincott, S., Collister, L., Wu, C., & Steinhart, G. (2024). Data for: Characteristics of Selected Open Infrastructures, 2024 State of Open Infrastructure Report [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10835677

Lynch, C. A. (2003). Institutional Repositories: Essential Infrastructure For Scholarship In The Digital Age. Portal: Libraries and the Academy, 3(2), 327–336. https://doi.org/10.1353/pla.2003.0039

Skinner, Katherine. (2018). Community Cultivation: A Field Guide. Educopia. https://educopia.org/wp-content/uploads/2018/11/CommunityCultivationFieldGuide.pdf

United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development. United Nations: New York, NY, USA. https://sdgs.un.org/publications/transforming-our-world-2030-agenda-sustainable-development-17981

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., & others. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 1–9. https://doi.org/10.1038/sdata.2016.18

Glossary of terms

Board structure: The way in which a service or provider's governing board is structured.

- **Advisory board or steering committee.** Usually composed of a small number of experts and senior figures from allied stakeholder organizations, advisory boards are convened to provide input on decision making and strategic direction, either at the individual, ad hoc level, or through regular board meetings. This means that advisory boards are often external to the project, or at least not composed by active participants within the project itself, and so are called upon for direction and accountability.[10]
- **Multi-board governance.** Distributed model involving multiple boards with differing areas and levels of responsibility [allowing] oversight of discrete areas within a community-driven organization.[11]
- **None.** The OI does not have a formal governing or advisory board.
- **Other.** The OI has an alternate form of formal governing board.

Business form: The service's legal business structure or formal aims.

- **Commercial (for-profit):** The OI or provider operates with the primary goal of making a profit.
- **Fiscal sponsorship (academic institution):** The OI is hosted by an academic institution that provides administrative, financial, and legal support.
- **Fiscal sponsorship (non-profit):** The OI is hosted by a non-profit organization that provides administrative, financial, and legal support.
- **Non-profit organization:** The OI or provider is an independently incorporated nonprofit.
- **Volunteer community:** The OI is run by a group of uncompensated individuals and is not formally incorporated.
- **Other:** Business form other than one of the above.

Service: A product or infrastructure component. In some cases, the service may be the same as a provider. Examples of services include Open Journal Systems (OJS), DSpace, Figshare, Kotahi, and arXiv.

Service provider: A provider is a formal or informal organization responsible for running a service. Examples of providers include The Public Knowledge Project (PKP), Lyrasis, Digital Science, Coko Foundation, and arXiv.

Solution category: The function(s) that a service is primarily designed to serve.

- **Annotations system:** A tool that allows scholars to mark up and comment on digital works, either individually or collaboratively.
- **Archive information management system:** An application used to organize, control, and manage information about archival collections.
- **Authoring tool:** A tool or platform that facilitates the creation of documents destined for print or digital publication. These tools may provide real-time editing, built-in templates, version control, and support for formats like LaTeX, making them useful for academic and scientific writing.
- **Digital asset management system:** An application used to store, organize, manage, and share digital files and assets.
- **Digital library, collection, or exhibit platform:** An application used to share collections of digital objects and create media rich online exhibits.
- **Discovery system:** A tool that facilitates search and retrieval of scholarly content.
- **Federated identity or authentication management:** A secure system that enables user authorization, authentication, and digital identity management.
- **Format conversion tool or service:** A tool that supports the transformation of digital files or metadata from one format to another.
- **Index or directory:** A curated list of resources relevant to a certain audience.
- **Informal scholarly communications:** Digital tools that facilitate activities that occur outside of formal journal publication, such as communication on social media, listservs, or through other digital platforms.
- **Media viewer/player:** A tool that allows users to play digital audiovisual content.
- **Open access policy information compilation:** A curated list of open access policy information.

[10] https://copim.pubpub.org/pub/wp4-report-exploring-models-for-community-governance#advisory-board-steering-committee

[11] https://copim.pubpub.org/pub/wp4-report-exploring-models-for-community-governance#multi-board-governance-structures

- **Open scholarly dataset:** An open dataset intended to support scholarly research.
- **Peer review system:** A tool that facilitates the pre- or post-publication review of manuscripts.
- **Persistent identifier service:** A service that maintains a persistent identifier registry and associated tools and services.
- **Personal information management system:** A platform or software that allows individuals to collect, organize, and store digital objects.
- **Preservation system:** A system that facilitates long-term, secure storage of digital objects.
- **Publishing system:** A platform that facilitates digital publishing workflows.
- **Repository service:** A hosted digital platform that makes information resources available online.
- **Repository software:** A software that can be used by a community to ingest, manage, preserve, and share digital content.
- **Research profiling system:** Research profiling systems, also known as Research Information Management Systems (RIMS) or Current Research Information Systems (CRIS), "collect and store structured data about faculty research and scholarly activities for one institution, with the intention of repurposing the information in a variety of ways".
- **Standard, specification, or protocol:** A set of community-accepted rules used by two or more parties to interact or exchange information.
- **Submissions system:** A digital platform or software used by academic journals and publishers to streamline the process of authors submitting their research papers for peer review and publication consideration. They facilitate the efficient handling of manuscripts, including submission, peer review management, and editorial workflow tracking.
- **Other:** OI does not align with any of the categories above.

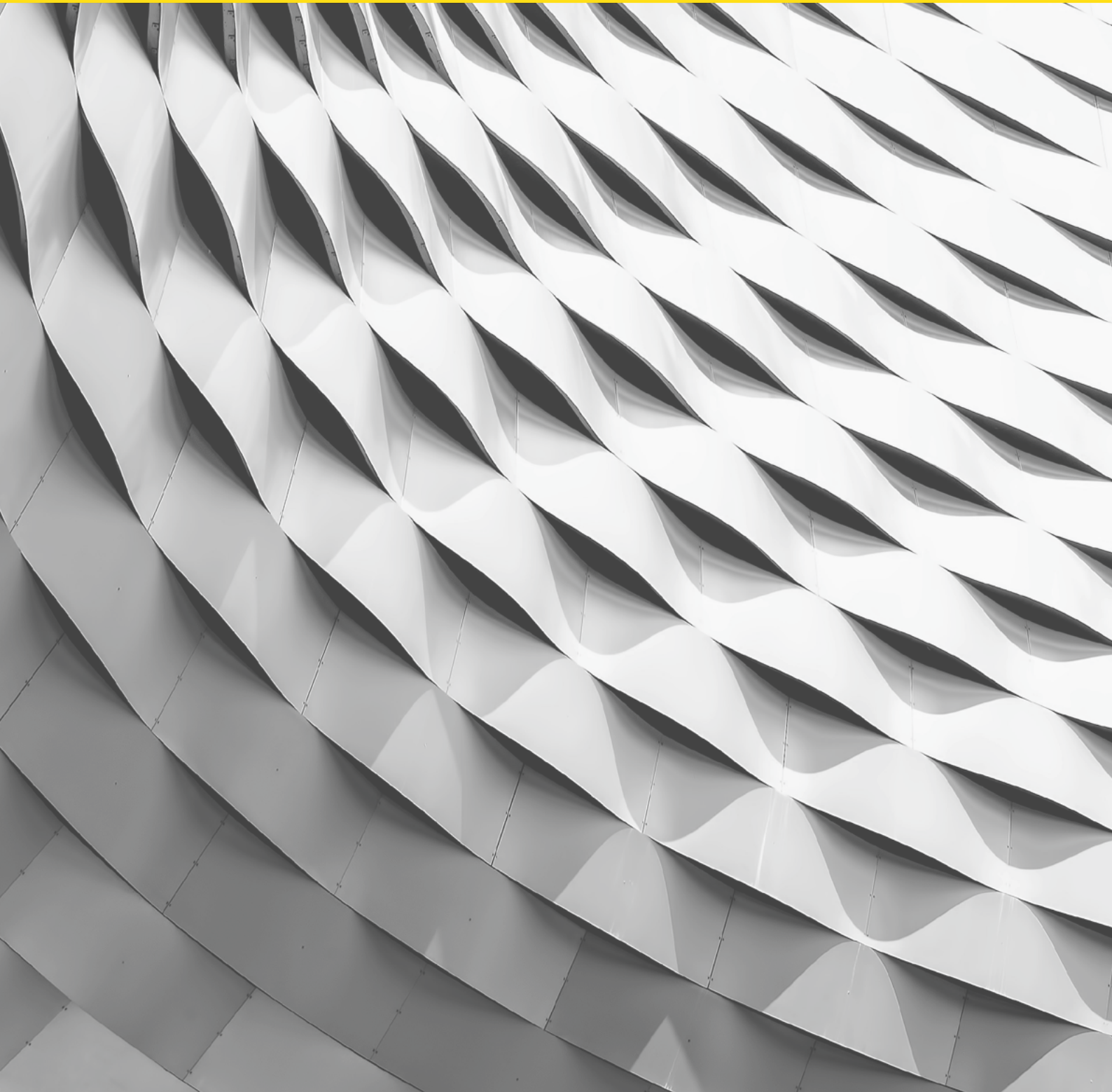
Technical documentation: Technical user documentation meaningfully describes the process for users to install, configure, maintain, and troubleshoot the code for running a service. In the case of those OIs not reliant on client installs, documentation should meaningfully describe how users perform tasks on the platform to utilize the service. Documentation may include installation guides, tutorials, manuals, FAQs, etc.

Technology readiness level:[12] The service's fitness for broad adoption.

- **Experimental proof of concept:** The OI is under development and not yet available to users.
- **Technology validated in relevant environment:** The OI has a minimum viable product but is not yet implemented in production.
- **System prototype demonstration in operational environment:** The OI is implemented in a single instance.
- **System complete and qualified:** The OI is production ready, fully featured, and in limited use within the scholarly community.
- **Actual system proven in operational environment:** The OI is production ready, fully featured, and widely in use within the scholarly community

[12] Technology readiness levels used in Infra Finder are adapted from the framework developed by the Horizon 2020 project: https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf

The state of open infrastructure grant funding



Introduction

At IOI, our work to increase investment in open infrastructure (OI) relies on a deep understanding of how infrastructure is funded and by whom. Since 2020, we've studied, analysed, and published our findings about the funding landscape for OI for research and scholarship. We recognize that grant funding is just one revenue stream that supports open infrastructure, but as our research has shown, financial contributions (which include grants) are the primary source of revenue for many OIs.[1] Here, we take our deepest dive yet into the available data in order to better understand the amount, impact, and limitations of grant funding to OIs. To the limited extent possible, we also try to position this analysis within an overall picture of the financial operations of open infrastructures.

The dataset we have assembled includes 514 awards made by 23 distinct funders to 36 open infrastructures (OIs), totaling US$415,845,753, and dating back to 2000. Of these awards, we estimate 149 of them (totaling $174,491,754) represent direct support to OIs, and the remainder support activities which depend on that infrastructure in some measure. This work greatly expands IOI's 2022 analysis of 137 funder-reported grant awards (for a total of $124,972,660) made to 28 key infrastructures and data from 22 funding organizations (Dunks, 2022) including ten members of the Open Research Funders Group. In updating our original dataset, we elected to keep the focus squarely on funder-reported grants, but we also tried to identify what the funding actually supported — e.g., did it provide direct support to the infrastructure itself, or did it support work that used an OI, potentially without actual connection to the OI in question.

These are not trivial sums, particularly given the limitations of the current dataset. We scoped our data collection efforts to focus on the OIs included in IOI's Infra Finder,[2] which are related to repository infrastructure. As such, they tilt strongly towards tools and services in current use by libraries, and were selected based on Infra Finder's criteria for inclusion. We were also limited by the availability of data that we could harvest and add to our dataset using our current methods. Even with these limitations, we think these numbers reflect something of the importance of the sector, and demonstrate the great potential a larger and more complete and varied dataset would have for understanding and informing investment decisions. We look forward to continuing to build out and make use of such a resource.

Methods

Data collection

We focused on funder-reported and centrally reported data as the sources of record. We compiled a list of funders of interest from IOI's earlier exploration of funding for open infrastructures (Dunks, 2022) and funding sources reported by the 57 infrastructures listed in IOI's initial launch of Infra Finder. We chose to focus on open infrastructures (OIs) in Infra Finder in order to be able to tie our analysis back to additional attributes of those OIs that are included in the tool, and potentially leverage the data available there.

We employed multiple methods for collecting award data associated with 20 funders, which we describe more fully in the accompanying dataset (Riordan et al., 2024). Briefly, we harvested available award data directly from the websites of 12 funders.[3]

We also obtained data from OpenAIRE, collating it from the OpenAIRE Research Graph data dump of 16 January 2024 (Manghi et al., 2024) into the COKI Academic Observatory system (Hosking et al., 2023) on 13 February 2024. Data on National Science Foundation (USA, NSF) grant awards was collected from the NSF website[4] as

[1] See "Characteristics of selected open infrastructures" in this report.

[2] See "Characteristics of selected open infrastructures" in this report for a brief introduction to Infra Finder, and Collister et al. (2024) for a fuller account. Infra Finder is available at https://infrafinder.investinopen.org.

[3] We harvested grant award data directly from the websites of the following funders: Alfred P. Sloan Foundation, Andrew W. Mellon Foundation, Arcadia Fund, Bill & Melinda Gates Foundation, Chan Zuckerberg Initiative, Gordon and Betty Moore Foundation, Institute of Museum and Library Services, Leona M. and Harry B. Helmsley Charitable Trust, National Endowment for the Humanities, Robert Wood Johnson Foundation, Social Sciences and Humanities Research Council, The Wellcome Trust.

[4] https://www.nsf.gov/awardsearch/download.jsp

XML files for the years 2010–2024. Data for all funders was manually uploaded to BigQuery.

Finally, we reviewed the funder-reported data in IOI's earlier dataset (Dunks, 2022) for awards that we did not capture with our current methods.[5] We added missing information if we could find it (most often title, description, and funder's award ID). If we could not verify that an award was to an OI on our list, we did not include it in our updated dataset.

Selecting and deduplicating awards

We used 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.

Data manipulation and enhancement

Currency conversions

For awards made in currencies other than US dollars (USD), we used the European Central Bank's currency converter[6] to convert the amount to USD, using the start date of the grant or 1 January of the award year if a specific date is not available. If no date information is available at all, no conversion is made and the award does not factor into any analysis of award amounts, but is included in award counts. Award amounts in the original IOI dataset were converted to USD using the 2010–2020 average exchange rates reported in Exchange Rates UK.[7]

Award categories

We assigned each award to a category based on its title and description (see Appendix for complete list and definitions for each category). We also group the awards into several categories to differentiate between those that constitute direct support to an OI, and those that do not, but that demonstrate the impact these infrastructures have on research and scholarship.

These assignments are somewhat subjective. For example, it may not be completely clear from an award description whether a named an existing repository infrastructure is enabling a new project, in which case the award might be categorized as "Adjacent", or whether a completely new instance of the repository infrastructure is being created, in which case the appropriate category would be "Adoption". Similarly, we attempt to distinguish between new feature development and routine code maintenance and updating ("Research and development" for the former, and "Operations" for the latter), but this is not always completely clear.

Data and data dashboards

A copy of the file we used as the basis for the analysis presented here is available for download on Zenodo (Riordan et al., 2024). We also used Looker Studio to create dashboards to allow for a more dynamic exploration of the underlying data. At the time of publication, the data driving the dashboard was identical to what we used for analysis, and readers are invited to explore it.[8] We hope to update and extend this dataset over time.

Caveats and assumptions

Please see the full dataset documentation for additional information on how we processed the data, as well as assumptions we made and their likely trade offs.

[5] Two of the 22 funders with data available from IOI's 2022 analysis, Arnold Ventures and the Simons Foundation, no longer offer straightforward access to award data in bulk. Arnold Ventures has been an important funder of the Center for Open Science (home of Open Science Framework, OSF), and many of its earlier grants were to that organization. We did scan more recent 990 forms for Arnold Ventures and did not encounter any additional awards to OSF in 2021 or 2022, although it is possible there were awards to other OIs of interest (we did not search for them). Awards from the National Institutes of Health (NIH) were also reported in IOI's 2022 dataset. In this iteration we were unsuccessful in applying our harvesting methods to the NIH's award database due to its sheer size, but we were able to retrieve some data via OpenAIRE. This almost certainly results in missing some relevant grant awards from NIH.

[6] https://data.ecb.europa.eu/currency-converter

[7] https://www.exchangerates.org.uk/

[8] State of OI funding data dashboards: https://lookerstudio.google.com/s/oStqguBdU7E

Results and discussion

Observations about the dataset

We acknowledge again up front that the scope of the data we collected is limited by our choices of OIs and of funders, and also by the lack of availability of data for some of the funders we were interested in (or the limitations of our methods for accessing it). We think this is offset to some degree by having the ability to do a deeper and more nuanced analysis of how funding is distributed, on the basis of additional information and attributes of OIs that we've collected in Infra Finder.

General characteristics of OI funding

The grant awards in our dataset provided a total of US$415,845,753 from 23 funders to 36 OIs, via 514 awards made over the time period 2000–2024 (Table 1). Award amounts ranged from $442 to $15,000,000 and the mean and median award amounts were $833,358 and $279,636, respectively. Many of the OIs in the dataset were launched over this time period, and we see an increase in overall funding and number of awards made as well (Figure 1). Because it is difficult to disentangle the overall growth in the number of OIs over time from trends in funding, we did not analyse the data for temporal trends.

TABLE 1.

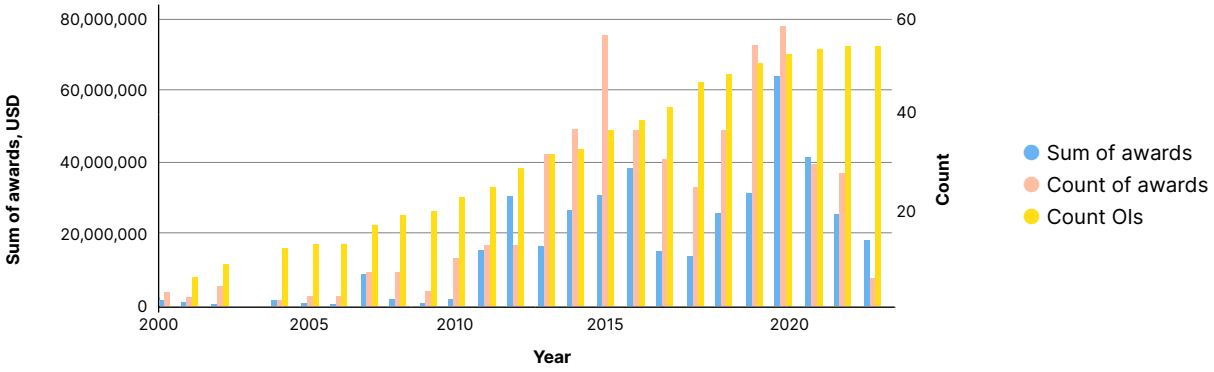
Total funding, and counts of awards, funders, and OIs for all awards and for awards categorized as direct support, indirect support, and adoption support

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	Indirect support	Adoption support
Total funding (USD)	$415,845,753	$174,491,754	$218,290,115	$10,807,041
Award count	514	149	284	40
Funder count	23	20	14	8
OI count	36	30	26	9

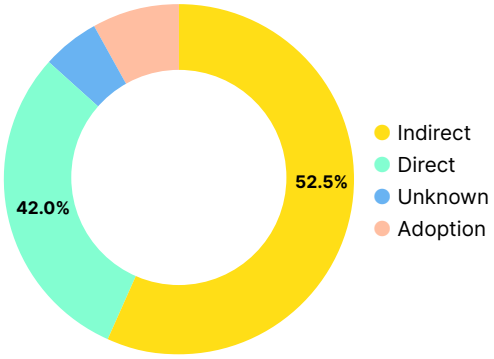
FIGURE 1.

Sum and count of awards made by year, and count of OIs available for funding each year



We categorized awards to reflect whether they provide direct support to an OI, indirect support (meaning the OI is referenced in the award title or abstract, but the funding does not directly support the OI though it may provide some indication of on OI's broader impact), adoption support (funding that supports the implementation of an instance of an OI at a local or community scale), and grants we were unable to classify (unknown). While a significant amount (42%) of funding goes to direct support, the majority of the funding (52%) goes to indirect support (Figure 2). We further analyse the direct, indirect, and adoption categories later in this report.

FIGURE 2.
Sum of all awards by super category



Direct funding to OIs

Top funders

We wanted to tease out trends in direct and indirect support to OIs in order to better understand how much funding is made available to support OIs directly, as opposed to how much additional funding relies on the use of open infrastructure without necessarily supporting it directly. The European Commission (EC) tops the list by total amount of direct support to OIs (Table 2, Figure 3). The EC's open infrastructure portfolio consists primarily of three multi-million euro awards to DataCite and multiple awards to Europe PMC and OAPEN Library. We note that the EC's awards to DataCite total nearly US$30M (one of which we found difficult to attribute with absolute certainty as direct support for DataCite, but that is how we chose to classify it) and may disproportionately affect the overall funding picture when we look at total award amounts. The Wellcome Trust is also a top funder and an important and ongoing supporter of Europe PMC, providing 11 awards totaling nearly $23M since 2013. Direct support to OIs is evenly distributed across public and private funding bodies (Figure 5).

The Institute of Museum and Library Services (IMLS) issued the largest number of awards of any funder (29), with multiple awards to Omeka (9), Mukurtu (8) and Fedora (6), and additional awards to DSpace, Dryad, Hyku, Open Science Framework (OSF), and the Research Organization Registry (ROR) (Table 2, Figure 4).

TABLE 2.
Top funders providing direct support to OIs, on the basis of award amount

Funder	Award total (USD)	Award count
European Commission	$36,070,103	12
The Wellcome Trust	$27,439,563	15
Agence Nationale de la Recherche (ANR)	$21,415,748	2
National Science Foundation	$19,583,613	18
Arnold Ventures	$14,310,360	6
Chan Zuckerberg Initiative	$11,280,875	6
Arcadia Fund	$8,500,000	2
Institute of Museum and Library Services	$7,631,406	29
Gordon and Betty Moore Foundation	$7,362,795	4
Other	$20,897,291	55
Total	**$174,491,754**	**149**

FIGURE 3.
Sum of direct support awards by funder

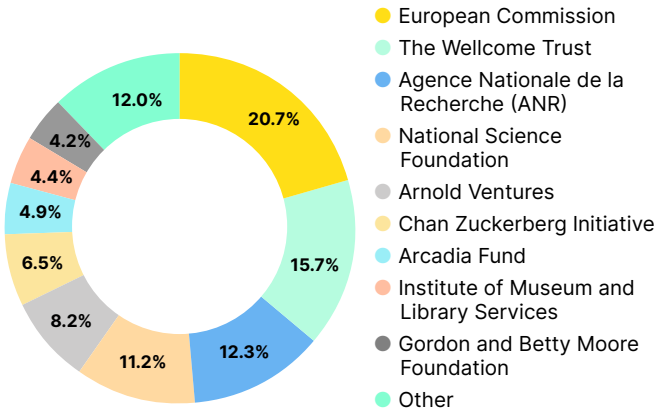


FIGURE 4.
Count of direct support awards by funder

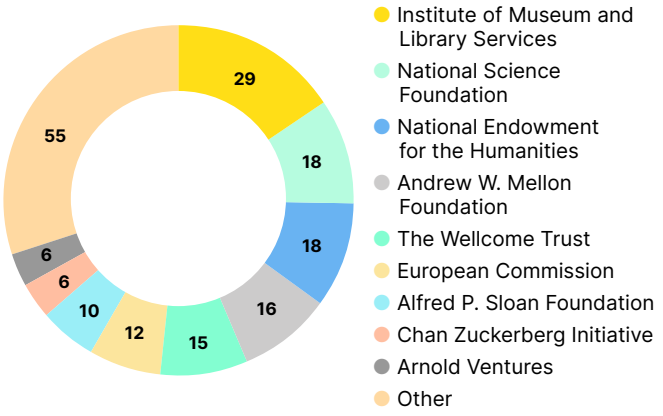
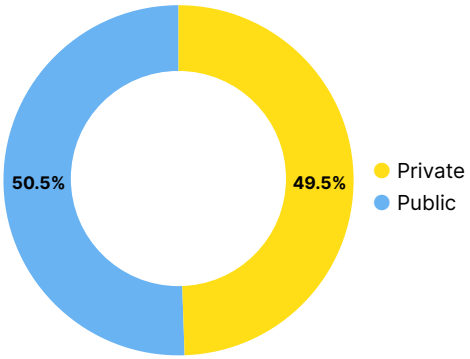


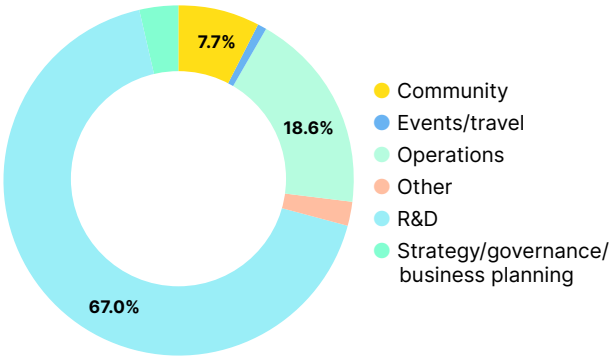
FIGURE 5.
Sum of direct support awards by funder type (public or private)



Funding by award and infrastructure categories

We categorized each award according to the definitions in the Appendix. Research and development (R&D) dominates the picture, whether we look at the award total (67%) or award counts (71%, not illustrated here), followed by operations (18.6% by total amount). Community building, events and training, strategy/governance/ business planning, and awards classified as "other" all earned less than 10% of the funding total (Figure 6). These results support the claim that philanthropy tends to favour innovation over sustaining existing endeavours (e.g. Skinner, 2019), but support for the latter is not insubstantial.

FIGURE 6.
Sum of direct support awards by award category

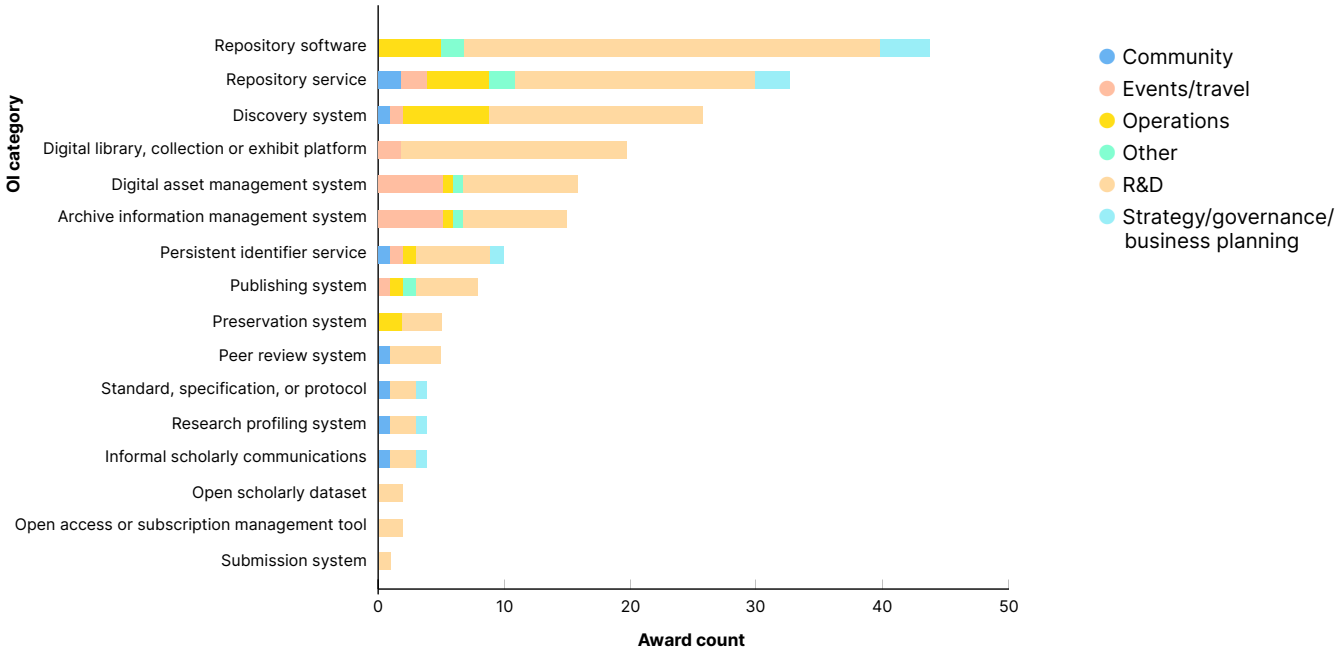


We also looked at the distribution of funding across OI category and grant category. Each OI is assigned to at least one (and usually just one) category, although six placed themselves in two categories, and two placed themselves in three categories. Because an OI may be assigned to more than one category, we consider here only the counts of awards rather than looking at totals, which would amplify the problem of counting some awards more than once. With that caveat, the constellation of open infrastructures that are digital content distribution and/or management platforms — repository software, repository service, digital library, collection, or exhibit platform, and digital asset management system — are the kinds of applications we focused our early data collection efforts on for Infra Finder. Unsurprisingly, these receive more awards than most other categories (Figure 7). Perhaps more interesting here is that the pattern of R&D as the main vehicle for investment holds across the full range of OI categories, and that there is little investment in strategy, governance, and business planning — important activities for ensuring the ongoing sustainability of OIs.

FIGURE 7.

Count of direct support awards by OI category and grant category

Note that an OI can be assigned to more than one infrastructure category, while grant awards are assigned to only one category.



Top funding recipients

DataCite, Europe PMC, OpenEdition, OSF and Fedora are the top five funding recipients by award amount (Table 3, Figure 8). Again, this is likely skewed a bit by a few very large awards to DataCite. Looking at award counts, Omeka, Mukurtu, OSF, Fedora, and Europe PMC are the top recipients. IMLS, National Endowment for the Humanities (NEH), and to a lesser extent the Mellon Foundation have made sustained investments in the Omeka (20 awards) and Mukurtu (15 awards) platforms (Figure 9), and as we will see later, IMLS and NEH have also made numerous awards to support the adoption of these same platforms.

TABLE 3.

Top funding recipients (DIRECT support), on the basis of award amount

OI	Award total (USD)	Award count
DataCite	$35,143,158	7
Europe PMC	$26,431,542	13
OpenEdition	$21,415,748	2
OSF (Open Science Framework)	$20,794,818	14
Fedora	$11,715,877	14
arXiv	$10,378,304	3
bioRxiv	$8,382,354	4
OpenAlex	$7,500,000	1
Dryad	$5,734,388	9
Other	$26,995,565	82
Total	**$174,491,754**	**149**

FIGURE 8.
Sum of direct support awards by recipient

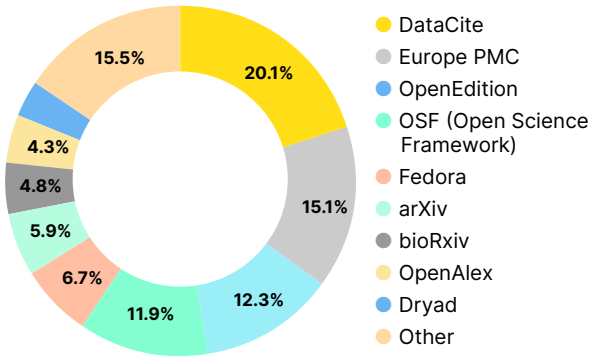
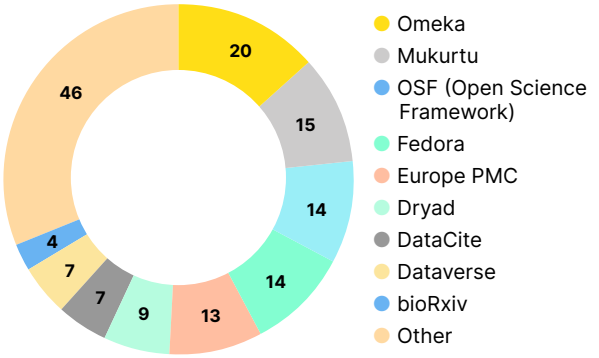


FIGURE 9.
Count of direct support awards by recipient



Indirect funding to OIs

It was not our original intention to explore indirect funding for OIs, but in the process of identifying relevant awards, we noticed that a large number of awards reference the use of open infrastructure without apparent direct support for it. These are largely references to depositing preprints, other publications or datasets into repositories (categorized as "use"), but there are also references to more substantive uses, which we categorized as "adjacent". Examples of uses that we categorized as "adjacent" include building new infrastructure that leverages existing OI, expanding or adding to digital collections that already utilize OI for access, or the development of additional features or customizations for local use (i.e. new development that is not contributed back to a community code base). Categorizing awards that fall along a continuum of uses into the two distinct categories we used was not always obvious or straightforward. Nevertheless, we posit that these uses, along with awards that support adoption of an OI, have the potential to provide a useful indication of an infrastructure's impact.

We found that US NSF awards make up the majority of *indirect* support awards, by total amount and count (Table 4, Figures 10–11). With the European Commission and the Agence Nationale de la Recherche (ANR) as the second and third most significant funders respectively (by amount), it is public funders that are the source of most of the indirect support awards (Figures 10–11). The majority of indirect support awards fall into the "use" category, and the majority of uses are related to the use of (deposit to) repository services, followed by use of repository software (Figure 12). The most used repository services are Dryad, OSF, and arXiv (Figure 13).

That open infrastructure should be such an important catalyst for the research enterprise is both exciting and a potential cause for concern. Recall that we identified relevant awards on the basis of the appearance of the name of OIs of interest in the award title, description, or recipient. We don't think it is unreasonable to suggest that mentioning a repository a researcher intends to use in one of those fields is significantly less common than it is in an award's project description or data management and sharing plan, hence our estimate of indirect support is surely a significant underaccounting of this phenomenon. Where these uses place direct demands on infrastructure, in the cases where usage or other direct fees that scale with use are not charged, these uses may place OIs under increasing strain and potentially threaten their sustainability (Steinhart & Skinner, 2024).

TABLE 4.

Top funders associated with INDIRECT support, on the basis of award amount

Funder	Sum of awards	Award count
National Science Foundation	$124,080,506	212
European Commission	$38,389,231	8
Agence Nationale de la Recherche (ANR)	$34,950,628	3
Fundação para a Ciência e a Tecnologia, I.P.	$5,518,159	11
UK Research and Innovation	$4,778,703	9
The Wellcome Trust	$3,177,695	4
National Endowment for the Humanities	$3,082,906	19
National Institutes of Health	$1,422,178	6
Institute of Museum and Library Services	$1,033,745	4
Other	$1,856,364	8
Total	**$218,290,115**	**284**

FIGURE 10.

Sum of awards (INDIRECT support) by funder

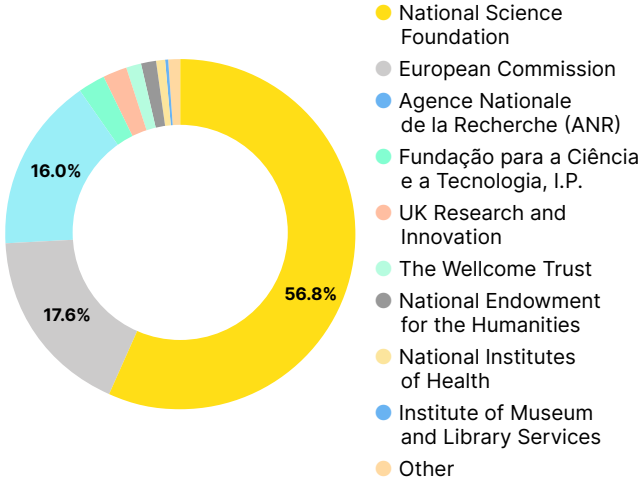


FIGURE 11.

Count of awards (INDIRECT support) by funder

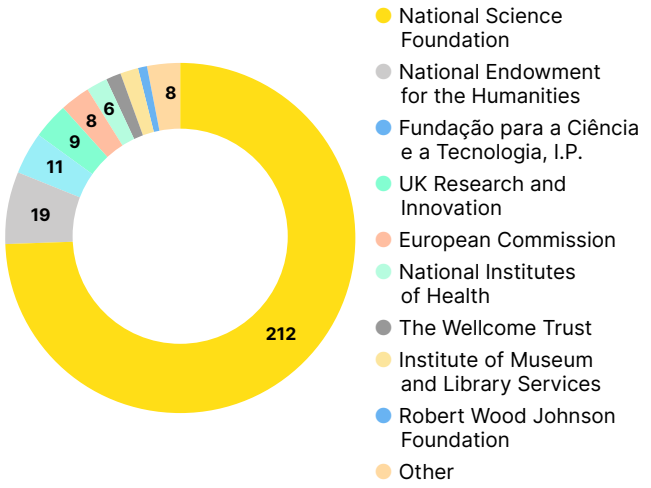
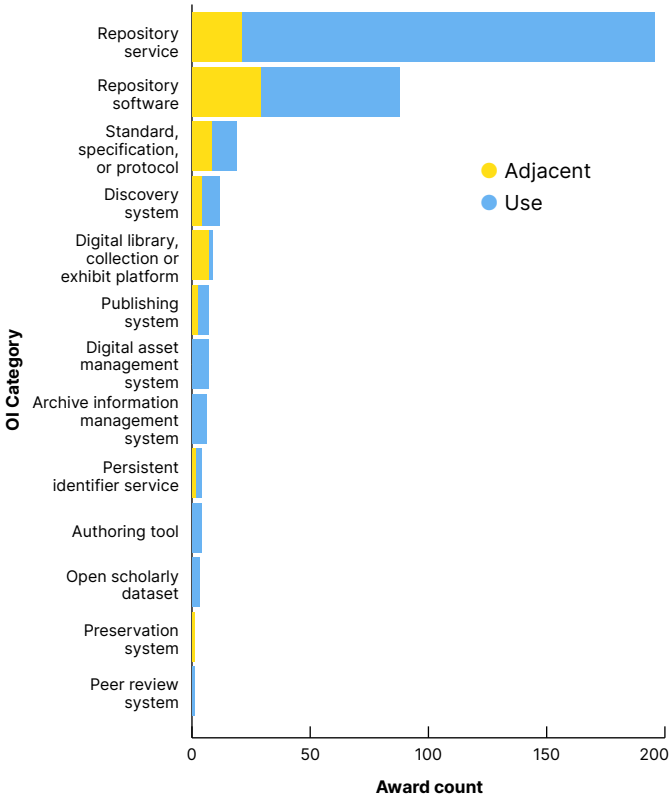


FIGURE 12.

Count of (INDIRECT support) awards by OI category and grant category

Note that an OI can be assigned to more than one infrastructure category, while grant awards are assigned to only one category.



As important as grant funding is, we think it is imperative to consider grant funding in the context of the complete financial picture for open infrastructures.

FIGURE 13.
Count of awards (INDIRECT support) by OI

OIs with less than 10 such awards are excluded for clarity.

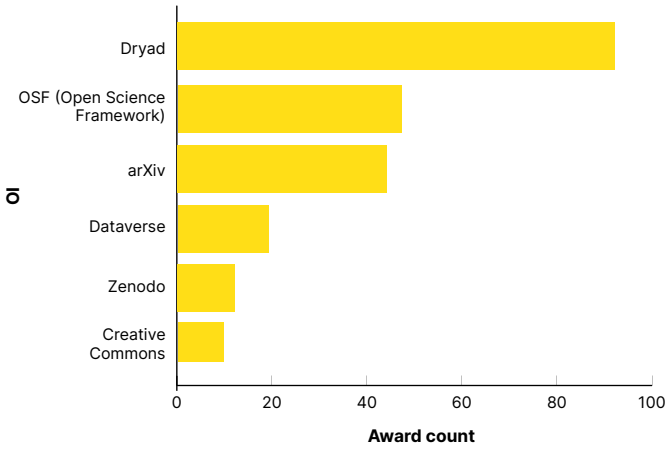
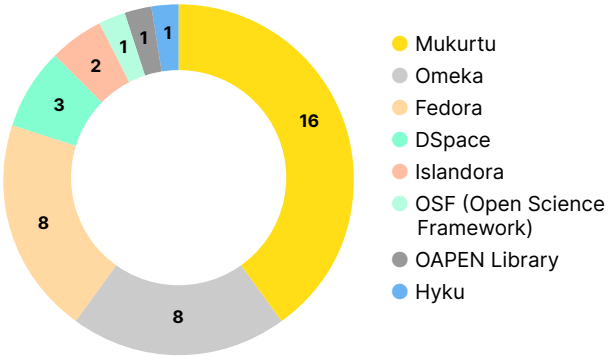


FIGURE 14.
Count of awards (ADOPTION support) by OI



Support for OI adoption

We also encountered a number of grant awards that support adoption of a particular OI to support a larger project, without an indication of funding being directed towards the OI itself. A significant amount of this funding is provided by IMLS and NEH to support the use of Omeka (8 awards) and Mukurtu (for indigenous communities, 16 awards) to build and deliver digital collections. Fedora is also named in eight awards from five different funders (Figure 14). All of these infrastructures are repository software. As with the indirect support awards we describe above, we think these awards may serve as a meaningful signal of uptake and impact that an OI has. And as with other kinds of indirect support, the possibility exists that these uses may impose an additional uncompensated burden on the underlying infrastructure, or alternatively, benefit it, as new adopters may choose to also support that infrastructure directly. Of course adoption is the *point* for open infrastructure — but it is worth examining the interplay between adoption and use, financial support, and demand on an infrastructure community's resources.

The role of grant awards in the overall financial status of open infrastructure

As important as grant funding is, we think it is imperative to consider grant funding in the context of the complete financial picture for open infrastructures. Grant awards represent just one of many potential sources of revenue (Figure 15). Other sources of revenue include membership fees and donations (paid directly by individual organizations, or by supporting consortia), client fees (again from organizations or consortia, in circumstances where the OI might provide a vended option that they support for a fee), in-kind contributions from organizations that contribute to the development of an OI or provide some other non-monetary form of support, and donations and in-kind contributions from vendors whose business utilizes an OI.

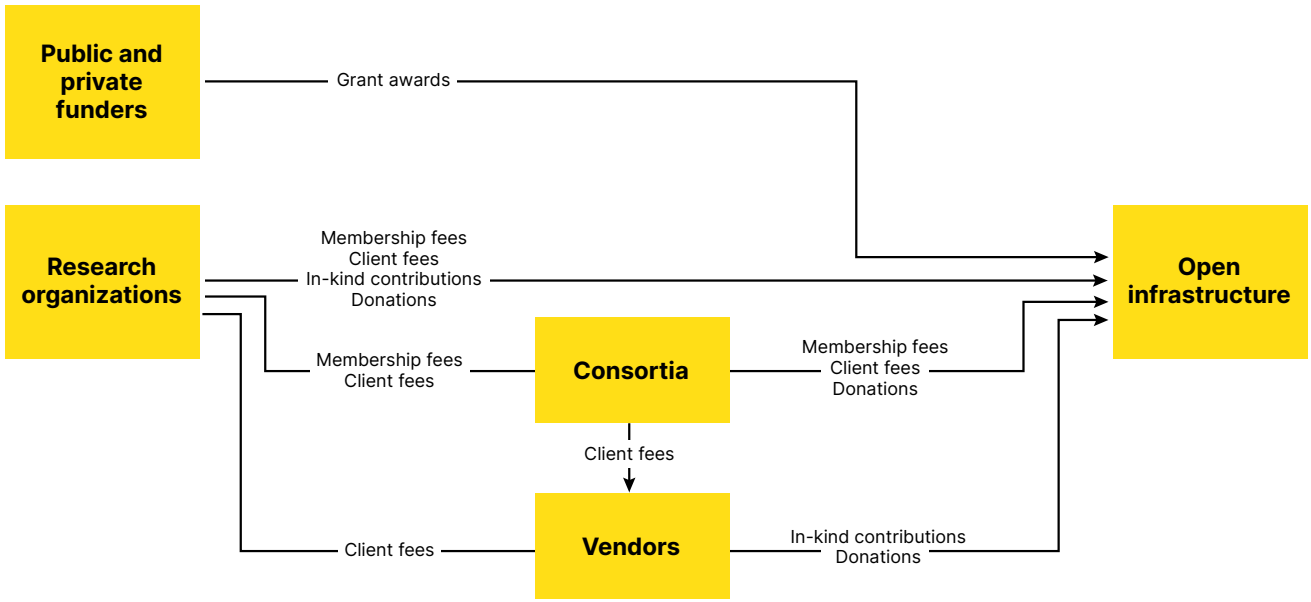
What we *can* say is that OIs overwhelmingly have unmet needs for operational support.

We would very much have liked to put this analysis of grant award data into this larger context, bringing in publicly available financial information for as many OIs as possible. We started to attempt this, and quickly encountered a couple of significant issues. First, financial reporting requirements vary by jurisdiction, and information is not universally or freely available. The most readily and freely available source of this data are the IRS 990 forms filed by non-profit organizations in the US, which we were able to obtain for four OIs that are independent nonprofits. Required reports are available for OIs incorporated in other countries, but for a fee in the case of the jurisdiction we were most interested in (Netherlands, for three OIs). Second, the placement of an OI within its host organization varies, and reported financial information may or may not be particularly meaningful at the level of the individual OI. At one extreme, an organization's entire *raison d'être* may be to manage and sustain an OI. In this case, the organization's financial information gives us meaningful insight into the OI's overall financial picture. At the other extreme, an OI may be embedded in a very large organization, with financials that are inextricably intertwined with those of the host (for example, an OI that is hosted by a large university), and the financial performance of the host tells us little or nothing about the financial status of the OI itself. Somewhere in the middle are independent organizations that sustain more than one OI (Code for Science & Society, IOI's host as well as the host of the International Interactive Computing Collaboration (2i2c), PREreview, and others). Annual reports are another potential source and often contain some self-reported financial information, potentially at the level of an individual OI. This information is not reported in any standardized way, but at least it generally reflects more closely the operations of the OI than its host organization, and we were able to locate annual reports for an additional 13 infrastructures. The best we can say from these varied and limited sources of information (and taking annual reports at face value) for 12 of the OIs that report total revenue and expenses, whether in a formal statement such as a 990 or self reported in an annual report, is that none report operating at a loss (total revenue less total expenses was always greater than or equal to zero), and that organizations who indicated to us that their primary source of revenue is contributions did so accurately where we were able to verify this independently.

What we *can* say is that OIs overwhelmingly have unmet needs for operational support. Nearly half of the OIs we canvassed as part of our data collection effort for Infra Finder reported contributions (which include grant awards) as their primary source of funding, and contributions were the single most important source of funding for OIs that operate as independent or fiscally-hosted nonprofits. We also know from that same population of OIs that 21 of the 44 who provided statements of their funding needs mention developing or maintaining basic service functionality as an explicit need (we counted these mentions as a need for operations support).[9] This is not news (see Skinner (2019), for example), but is evidence of a potential misalignment between funding OIs are able to bring in (grant awards for innovation) with the operational needs they are striving to meet.

[9] See "Characteristics of selected open infrastructures" in this report.

FIGURE 15.
Sources and flows of revenues that support open infrastructure



Conclusions

We have assembled the beginnings of a potentially useful resource for understanding the funding landscape for open infrastructure, and intend to continue to expand and update it. Our key findings in this phase of work are:

- Even in a bounded analysis such as this, the amount of funding that flows not only directly to open infrastructure, but also to *users* of open infrastructure, is substantial. This indirect support may represent both a useful measure of the impact of open infrastructure on the research ecosystem, as well as the potential strain upon it.
- Funders continue to favour making awards for innovation rather than ongoing support, with some important exceptions.
- Grant awards play an important role in the overall finances of open infrastructures where we are able to make that assessment, but consistent data at the appropriate level of granularity is hard to come by.

Data availability

The grant awards data used for the analysis presented here (Riordan et al., 2024) is available online: https://doi.org/10.5281/zenodo.10934085.

Interactive dashboards allowing direct exploration of the data are available at https://lookerstudio.google.com/s/oStqguBdU7E.

References

Collister, L., Tsang, E., & Wu, C. (2024). Infra Finder: A new tool to enhance transparency, discoverability, and trust in open infrastructure. International Digital Curation Conference (IDCC), Edinburgh, Scotland. (Preprint) https://doi.org/10.5281/zenodo.10913249

Dunks, R. (2022). Reported funding data for open infrastructure. https://doi.org/10.5281/zenodo.7259472

Hosking, R., Diprose, J. P., Roelofs, A., Chien, T.-Y., Massen-Hane, A., Smith, K. R., Handcock, R. N., Kramer, B., Napier, K. R., Tonti-Filippini, J., Montgomery, L., & Neylon, C. (2023). Academic Observatory Workflows. Zenodo. https://doi.org/10.5281/zenodo.10278531

Manghi, P., Atzori, C., Bardi, A., Baglioni, M., Schirrwagen, J., Dimitropoulos, H., La Bruzzo, S., Foufoulas, I., Mannocci, A., Horst, M., Czerniak, A., Iatropoulou, K., Kokogiannaki, A., De Bonis, M., Artini, M., Lempesis, A., Ioannidis, A., Manola, N., Principe, P., … Pierrakos, D. (2022). OpenAIRE Graph Dump (7.0.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10488385

Riordan, D., Huang, C.-K., Neylon, C., & Steinhart, G. (2024). Data for: The State of Open Infrastructure Grant Funding, 2024 State of Open Infrastructure Report [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10934085

Skinner, K. (2019). Why Are So Many Scholarly Communication Infrastructure Providers Running a Red Queen's Race? | Educopia Institute. https://educopia.org/red-queens-race/

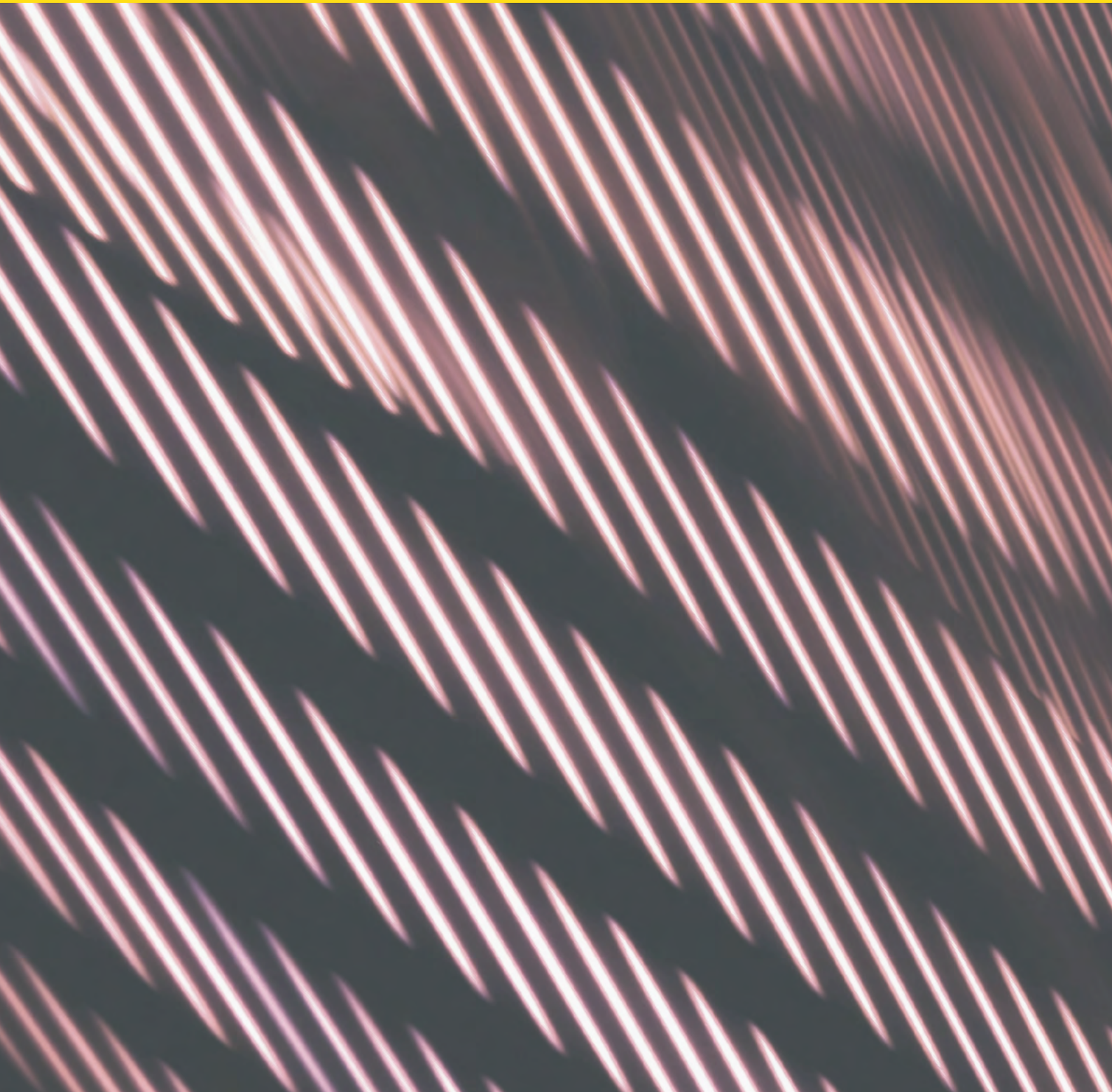
Steinhart, G., & Skinner, K. (2024). The Cost and Price of Public Access to Research Data: A Synthesis. https://doi.org/10.5281/zenodo.10729575

Appendix. Award category definitions

Award categories and the codes used in the awards data, and their definitions. We further recognize three super categories of grants: those that provide direct support to an OI, those that support adoption of an OI, and those that do not provide direct support but reflect other uses that demonstrate the broad impact of an OI.

Category (code)	Definition	Super category
Adjacent (ADJ)	Award is not directly to the OI but supports activities adjacent to it in some way. Examples: ■ Research use of an OI's data or metadata ■ Use in training or educational activities unrelated to the OI community ■ Extension of an OI to support a specific, local use, without contributing development back to the community ■ Other use that does not directly benefit or advance an OI, except those that qualify as USE, defined below	Indirect
Adoption - community (ADOPT_C)	Award supports adoption, upgrades to existing instances, or similiar direct adoption support broadly in a community.	Adoption
Adoption - local (ADOPT_L)	Award supports adoption in a single institution or local context.	Adoption
Community (COMM)	Award supports community building initiatives.	Direct
Events/travel (EV_TR)	Award supports events and/or travel.	Direct
Operations (OPS)	Award supports basic operations, including code maintenance and updates that would not normally be considered new development.	Direct
Other (OTHER)	Direct support for other activities not defined here, or multiple activities with an unclear primary activity.	Direct
Research and development (RD)	Award supports research and development, including software development. Research and development work may be performed by organizations other than the OI's host or home organization.	Direct
Strategy/governance/ business planning (STRAT)	Award supports strategic, governance, or business planning.	Direct
Use (USE)	Award references intention to use an existing OI for dissemination of content (e.g. deposit to a repository or preprint service), or other direct use (e.g. use of Creative Commons licenses). Passing mentions are ignored.	Indirect
Unknown (UNK)	Not enough information to classify	Unknown

Open infrastructure governance: Current structures, nomenclature, composition, and trends



Introduction

Many open infrastructures that support research and scholarship deploy some form of community governance, a relatively loose term that is used to describe a wide-ranging spectrum of practices (Dana et al., 2021; Hart et al., 2022; Moore, 2021). At its plainest level, governance simply means making and enforcing decisions, and within that framework, "community governance" often signifies that a community's members are in some way deliberately involved in decision-making processes. Usually, but not always, the work of community governance is unpaid and is provided by volunteers, not by staff members or those who are positioned to gain direct financial benefit from a programme or service.

The concept of community governance is championed throughout the open space, and many open infrastructures explicitly claim to be "community governed" or "community led".[1] This community involvement is often invoked as a point of differentiation and as a marker of trust, both within and beyond the scholarly communication ecosystem. It implies inclusivity and voice, but these terms are imprecise at best. What are the characteristics of community governance, and what are the models for engaging community in decision-making or advisory capacities in today's open infrastructures that support research and scholarship?

Below, we take a close look at the range of community processes employed in 54 open infrastructures to begin to answer a few key questions:

1. What types of governance models are deployed in/for these open infrastructures?
2. How are open infrastructure governance bodies named? Do these names align with specific definitions/meanings?
3. Who participates in open infrastructure community governance, and what can we know about the group of individuals or institutions that perform these roles?
4. Is there overlap in governance participation (e.g., where one participant serves on multiple governance bodies)?

A total of 82 open infrastructures that support research and scholarship were considered for this analysis, all of which were initially invited to be featured in our Infra Finder tool.[2] For a full list of these open infrastructure service providers, please see the open dataset that we have published in Zenodo (Skinner, 2024).

Methods

Between January and April 2024, we conducted web-based research to identify and record any publicly available evidence of community governance bodies and roles for 82 open infrastructures that support research and scholarship. These services were initially identified for and invited to participate in our Infra Finder tool in 2023 (Collister et al., 2024).

Our first step was to investigate which of these 82 infrastructures had community governance documentation available publicly on the web. We recorded the names and current affiliations of all currently listed members of the body; we also captured information about the governance body name(s) and any specified roles (e.g., officers). In addition, we captured information about each infrastructure's operational structure, and we recorded whether each governance body focused exclusively on a specific open infrastructure or if it focused more broadly (e.g., on a group of services or on the service provider's host organization).

Our analysis primarily focuses on the 54 open infrastructures for whom we found such public documentation on the open web. Once the dataset was created and documented (including Wayback Machine captures of all evidence), we asked a range of questions, starting with general details, such as what the 54 identified "community governance" bodies actually govern or advise, what they are called, and what size they are. We then analysed data about the unique *individuals* and *institutions* serving in these roles, and sought to understand which of those institutions and individuals are represented on more than one community governance body.

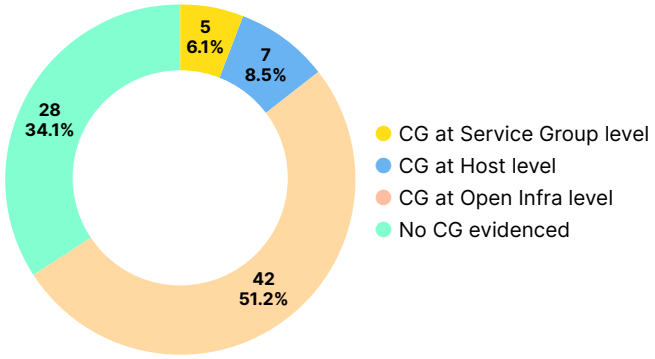
[1] For example, some description of "good" or "community" and/or "stakeholder" governance runs through many of the values and principles models that are gaining traction in the field of scholarly communication, including Principles of Scholarly Infrastructure (POSI, Bilder et al., 2020), COAR/SPARC Good Practice Principles (COAR/SPARC, n.d.) and the FOREST Framework (Lippincott and Skinner, 2022).

[2] https://infrafinder.investinopen.org/

Findings

The open infrastructures in this study included a mix of models, including those with and without community governance frameworks publicly available (see Figure 1. Community governance presence in open infrastructures). Of the 82 infrastructures we researched, 28 appeared not to have a documented community governance group.[3] More than half (42) of the open infrastructures have at least one documented, infrastructure-specific community governance body (seven had more than one).[4] Another five open infrastructures had at least one community governance body connected to what we have termed a "service group" that included several infrastructures under a single umbrella of governance (e.g., PKP's Advisory Committee, which works with Open Journal Systems, Open Monograph Press, and Open Preprint Systems). The final seven open infrastructures in this study referenced only their host institution's community governance body or bodies (e.g., ContentDM is governed by OCLC's Board; Archipelago Commons is governed by Metro's Board).

FIGURE 1.
Community governance (CG) presence in open infrastructures



All of the 19 freestanding/independent programmes have their own, open infrastructure-specific community governance bodies, while only 23 of the 35 hosted infrastructures had evidence of community governance groups dedicated specifically to the underlying open infrastructure that services were being provided for (as contrasted with service group focused, like PKP or host focused, like Metro).

Who legally/fiscally owns or bears responsibility for these infrastructures?

Open infrastructures exist in a broad array of organizational forms (e.g., university-hosted, incorporated, fiscally hosted, multi-institutional, and informal) and sectors (academic, government, commercial, nonprofit), each of which carries its own set of rules and conditions that may or may not work with particular governance frameworks.[5]

Of the 54 open infrastructures with documented community governance structures, 35 appear to be owned/operated by a host institution that provides the legal and fiscal framework under which they officially operate, while 19 appear to be freestanding or independent.[6]

The 35 hosted infrastructures represent an array of forms and relationship types. Some of the host institutions are universities (e.g., University of Bologna, Villanova, Simon Fraser, Harvard, Cornell); others are non-profit or for-profit companies that host multiple units or services (e.g., OAPEN Foundation, Coko Foundation), and still others serve as non-profit fiscal hosts that specialize in providing operational support services — including not-for-profit fiscal/legal identity — to programmes (e.g., Code for Science & Society, OpenAIRE, NumFocus).

Some of the 19 freestanding/independent entities are nonprofits, public companies, or stichtings (Dutch foundations) depending on their national contexts (e.g., Islandora Foundation, COUNTER, Vivli, PeerCommunityIn, OA Switchboard). Others are unincorporated and represent formal or informal community efforts, or partnerships between other institutions (e.g., Oxford Common File Layout).

[3] This is a point of differentiation, not judgment. These 28 groups may have different reasons for not having a governance body, including because the service is still emerging/forming, because the service is distributed by design and does not desire centralized governance processes, or because its business form makes community governance hard to implement.

[4] arXiv, DOAB, DOAJ, Dryad, Europe PMC, Islandora, and OAPEN Library each have two or more governance groups.

[5] For example, many infrastructures that support research and scholarship are based in colleges and universities, and they ultimately answer to and are controlled by their home institution's Board of Governors or Trustees; in these environments, community governance usually is constrained to advisory capacity.

[6] Our sources were publicly available documentation, including self reporting. While we have made our best attempt to categorize these accurately, hosting relationships are notoriously hard to establish without direct contact with principals.

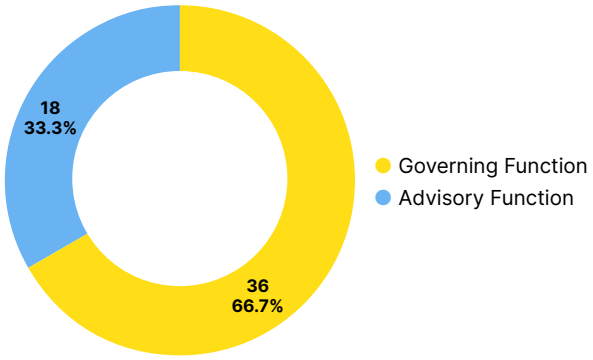
What are the infrastructure community governance bodies called?

The 54 open infrastructures we studied used 26 different terms to describe their 64 governance bodies (as previously noted: seven infrastructures have more than one governance body, hence the disparity in the numbers above). The names of these groups included many variations on common themes, including "Steering Committee," "Steering Council," and Steering Group," as well as "Board of Trustees," "Board of Directors," "Governing Board," "Supervisory Board," "Founding Board, "Executive Board," "Advisory Board," and just plain "Board." 33 open infrastructures had at least one officer position; 21 had none.

Looking across the selected 54 open infrastructures, their community governance approaches run the gamut from advisory bodies that provide input and guidance (e.g., arXiv, OpenCitations) to highly formalized groups that bear significant decision-making power and fiduciary and legal responsibility for a service (e.g., CrossRef, DOI Foundation).

Based on the terminology, we can infer how many of these community governance groups likely have an official governing (decision-making) function vs. those that are likely to be advisory-only bodies. As seen in Figure 2, of the 64 governance groups, more than half (40) have names that seem to indicate governing function.[7] An additional 18 are explicitly termed "advisory,"[8] and six (four overlapping with "advisory") reference a topical specialty.[9] A final four governance groups have names that seem too vague to warrant speculative classification as official governance bodies.[10]

FIGURE 2.
Community group function by infrastructure (ratio of full governance vs. advisory capacity)



Looking at this data from the starting point of the open infrastructure, rather than of the governance group, and again inferring from the group names, of the 54 open infrastructures, 36 (67%) have at least one group with a name that indicates governing function. Based on the names used, it appears the other 18 infrastructures (33%) have community groups with advisory responsibilities rather than full governance (decision-making) roles. Coupled with the often-confusing nomenclature, this raises questions about whether community members fully understand what role(s) their community groups do and do not play in decision making.

For example, an infrastructure may have community governance structures, including member representation (for users or contributors), and its community members may believe this "community governance" structure means the community has an active role in decision-making processes. They may be caught off guard by discovering that it actually only has an advisory role when that infrastructure makes a major decision (e.g., moving from an independent hosting arrangement to being an acquisition of a major conglomerate). The existence of community governance, in other words, may signal levels of power and involvement that in actuality do not exist; advisory groups may play important roles, but they do not bear fiscal and legal responsibility for the infrastructure, nor do they have the legal standing to contest major changes that happen without their involvement.

[7] Board (8), Board of Directors (14), Board of Trustees (1), Executive Board (5), Executive Committee (2), Executive Council (1), Governing Board (1), Leadership Group (2), Steering Group (1), Steering Committee (3), and Supervisory Board (2)

[8] Advisory Board (8), Advisory Committee (4), Advisory Group (1), Editorial Advisory Council (1), Institutions Advisory Council (1), Scientific Advisory Board (1), Scientific Advisory Committee (1), Scientific Advisory Council (1)

[9] Editorial Group (1), Editorial Advisory Council (1), Scientific Advisory Board (1), Scientific Advisory Committee (1), Scientific Advisory Council (1), and Scientific Committee (1)

[10] Funder Committee (1), Operations Team (1), Participating Organization Council (1), and Project Management Committee (1)

What is the distribution of individuals across these 54 infrastructures' community governance bodies?

These 54 open infrastructures' community governance environments include 567 total seats (averaging 11 seats per infrastructure); 496 individuals serve in these roles in 2024, with 48 individuals (10%) holding more than one seat or serving in more than one governance group and the other 448 (90%) only holding one seat each. Of those 119 seats occupied by the 48 individuals who serve on multiple groups, 30 people serve on two, 14 people serve on three, three people serve on four, and one person serves on five infrastructure governance bodies (Table 1).

This suggests a relatively wide distribution of seats to different individuals in these 54 infrastructure governance bodies, though there is some board interlock or concentration of individuals indicated by the 48 individuals occupying 119 of the total 567 seats, most particularly in the 18 people who serve on 3-5 infrastructure governance bodies in this set of 54 open infrastructures.

What is the distribution of institutions across these 54 infrastructures' community governance bodies?

The institutional distribution in community governance seats is not as wide as the individual distribution, though it still shows strong diversity overall. The 567 total seats are held by 383 institutions.[11] Of those, 91 institutions (24%) hold more than one of these 567 total seats, while 292 (76%) hold only one (Table 1).

Of the 91 institutions that were represented in multiple governance groups, 56 institutions held seats in two, 17 institutions held seats in three, six institutions held seats in four, four institutions held seats in five, and eight institutions held seats in six or more open infrastructure governance groups.

TABLE 1.
Number of community governance seats held by individuals and institutions across all 54 infrastructures

	Individuals	Institutions
One seat	448	292
Two seats	30	56
Three seats	14	17
Four seats	3	6
Five seats	1	4
Six or more seats	0	8
Total	496	383

With a total of 69 seats (12%) of the total 567 seats, the eight institutions[12] that hold seats in six or more open infrastructure governance groups strongly influence the open infrastructure ecosystem. They include core contributors, founders, and longstanding supporters of a range of tools, including OJS, arXiv, and other cornerstone technologies. Four of these eight institutions are Canadian, and some of the density of their representation can be attributed to four infrastructures they have helped to found and support over several decades: OJS, OMP, OPS, and Érudit. The remaining four institutions include one US-based not-for-profit corporation (CrossRef/ Publishers International Linking Association, Inc.), two universities (Harvard, Cornell), and the California Digital Library.

[11] Ten of these "institutions" were individuals that identified as consultants, artists, independent, or for whom we found no formal institutional affiliation.

[12] Crossref, University of Alberta, Cornell, Harvard, Université de Montréal, Simon Fraser, Public Knowledge Project.

Conclusions

There is no "right" governance framework or model that can or should be applied across these open infrastructures. We agree wholeheartedly with the COPIM team, who stated in 2022, "good governance is *situated*, i.e., it is highly specific to the resource and community in question" (Hart et al., 2022). However, after looking at the 82 infrastructures in this study, we would add that good governance is also well documented in language that the community and its extended ecosystem can readily interpret and understand.

Based on this set of 82 open infrastructures, community governance structures and terminology are prevalent, with well over half (54, or 66%) adopting and documenting such structures in some public format. That's good news for those who seek more inclusion of users and contributors (fiscal, technical, and administrative) within open infrastructure decision-making structures. The slightly more complicated news is that these structures use fuzzy terminology that makes it difficult to tell what the community groups actually are empowered to *do*.

In the 54 open infrastructures that had some type of community governance structure documented, we found 26 different names and a broad mix of solo- and multi-group structures at play. Inferring from these governance group names, 36 (67%) of the 54 open infrastructures with such groups do seem to grant them some level of active decision-making power and steering or leadership function. Others seem to be limited to advisory and topical roles, though, with 18 (33%) explicitly using only "Advisory" or a topical specialty ("Editorial" or "Scientific") in the group name(s).

The wide-ranging nomenclature in a small field hints that there are likely a lot of unique snowflake approaches to creating and naming these community roles. If that is because the governance model is carefully situated and specific to the community's needs, that might be read as a positive feature of these infrastructures. Where this instead becomes potentially harmful or dangerous is when a community develops a false sense of security, believing that its community groups possess an official decision-making authority that they do not have. For example, we could recount myriad open infrastructures that had visible, even vibrant, community governance groups that were sold or acquired *without* the consultation or involvement that many community members thought was guaranteed by the presence of community governance groups and processes. The wide range of practices (from advisory and topical to actual decision making) and the unclear language used across infrastructures can lead to confusion and misalignment. Community members' perceptions of their power and voice need to be checked and understood, ideally against official incorporation and bylaws documentation.

While the group nomenclature might be wobbly, our research did also surface solid names and employment affiliations of the 496 individuals occupying 567 community group seats for these 54 infrastructures in 2024. That data gives us several lenses to explore, both now and in the future. First, looking at the individuals who are serving on community groups in 2024 shows that, rather than having a handful of individuals holding lots of governance seats and power across infrastructures, the field has a lot of diversity in these service roles. Similarly, 383 institutions are represented in those 567 community group seats, again demonstrating strong overall diversity. We do mark small pockets of concentration, including those human and institutional outliers who serve on multiple community governance groups.

The concept of community governance is championed throughout the open space, and many open infrastructures explicitly claim to be community governed or community led. This community involvement is often invoked as a point of differentiation and as a marker of trust.

"Good governance is situated, i.e., it is highly specific to the resource and community in question" (Hart et al., 2022). However, after looking at the 82 infrastructures in this study, we would add that good governance is also well documented in language that the community and its extended ecosystem can readily interpret and understand.

The data does show that a small group of individuals (48) and (especially) institutions (91) are represented in multiple governance groups, but it also shows a high relative distribution of individuals and institutions serving overall. In other words, instead of seeing the same set of people and institutions represented in, and representing, these infrastructures, we have found a relatively wide set of stakeholders investing their time (including institutional staff time) to support open infrastructures in this field with a few exceptions.

It's important to note that this distribution of people and organizations only tells one part of the story. We can assume from the current numbers that we have a lot of different voices engaging in governance. But what does that actually mean? A few possibilities could include:

- We have too many independent entities (people and institutions) designing in silos and not enough interconnection between open infrastructures.
- We have healthy diversity and low concentration of power in community leadership.
- We have an open playing field with room for many voices, perspectives, and approaches.
- We have a small number of people and institutions wearing multiple open infrastructure hats who are encouraging interdependence, exchange, and collaborative alignment between infrastructures.
- We have a small number of power brokers who are playing an outsized role in shaping open infrastructure as both institutions and individuals.

In other words, without further research, the claims we can make based on this first year of data are still limited. Over time, we hope to be able to read and understand the health of the open infrastructure ecosystem through these types of investigations. In future years, we would like to begin comparing the community group composition across time to see if the current distribution holds or if more consolidation or diversity of roles becomes visible in multi-year analysis. We will also be able to see how much change occurs over time in the composition of specific boards. Pairing such data with additional information about funding, adoption, and use of these infrastructures should reveal much that we do not yet know about how open infrastructures are developing and maturing, not just individually, but as an interdependent system serving research and scholarship.

Data availability

The data used for the analyses presented here (Skinner 2024) may be downloaded from: http://doi.org/10.5281/zenodo.10934091

References

Bilder, G., Lin, J., & Neylon, C. (2020). The Principles of Open Scholarly Infrastructure (v1.1, 2023). https://doi.org/10.24343/C34W2H

COAR/SPARC. (n.d.). Good Practice Principles for Scholarly Communication Services. https://sparcopen.org/wp-content/uploads/2019/01/Sparc-Good-Practice-Principles-v4.pdf

Collister, L., Tsang, E., & Wu, C. (2024). Infra Finder: A new tool to enhance transparency, discoverability, and trust in open infrastructure. International Digital Curation Conference (IDCC), Edinburgh, Scotland. [Preprint] https://doi.org/10.5281/zenodo.10913249

Dana, C., Hornbein, D., Russell, V., & Schneider, N. (2021). Community Rules: Simple Templates for Great Communities. https://communityrule.info/assets/book/gov-booklet-MASTER.pdf

Hart, Adema, & COPIM (2022). Introduction, Towards Better Practices for the Community Governance of Open Infrastructures. https://doi.org/10.21428/785a6451.0304a2a8.

Lippincott, S., & Skinner, K. (2022). FOREST Framework for Values-Driven Scholarly Communication. https://doi.org/10.5281/zenodo.6557301

Moore, S. (2021). Exploring Models for Community Governance. COPIM. https://doi.org/10.21428/785a6451.0304a2a8

Skinner, K. (2024). Data for: Open Infrastructure Governance: Current structures, nomenclature, composition, and service trends, 2024 State of Open Infrastructure Report. Zenodo. https://doi.org/10.5281/zenodo.10934091

Trends in open infrastructure performance and adoption

Introduction

Our mission at Invest in Open Infrastructure (IOI) is to advance the investment in and adoption of open infrastructure for research and scholarship. From our stakeholder engagement and research work in the past years, one point that we have heard repeatedly is that those who are advocating for and making the case for the adoption of open infrastructure have faced continued resistance and questions from their peers and management. Questions were raised around the sustainability and stability of open infrastructure, and adopters are often looking for "social proof" — evidence that there is an increase in the use of open infrastructure, to support their case-making and de-risk their open infrastructure investments.

Simultaneously, from our observations and monitoring of various open infrastructure services and initiatives, the data we collected through Infra Finder,[1] and our conversations with them, we've witnessed and documented many clear signals of growth, development, and transition towards resilience in open infrastructures. These include a range of important trends, from new partnerships and funding, to technical upgrades and usage milestones. Specific examples evidencing these trends include a partnership between Dryad and Zenodo funded by the Sloan Foundation (Ioannidis, 2020), a Mellon-Funded initiative where the Nonprofit Finance Fund convened a cohort of six organizations to advance financial resiliency in the Digital Humanities (Nonprofit Finance Fund, 2020), a full rebuild of the Public Access Submission System (PASS),[2] and the use of the Open Journal System (OJS) by an estimated 60% of the world's diamond open access journals. In 2022, IOI also announced the launch of the IOI Fund for Network Adoption to catalyse a new form of investment to foster the implementation, growth, and usability of open infrastructures to advance scientific research and data sharing.[3]

We've witnessed and documented many clear signals of growth, development, and transition towards resilience in open infrastructures.

In our strategic support work,[4] we work with our partners (both infrastructure services and their funders) to improve the infrastructure's governance and co-map a strategic blueprint forward to increase their sustainability. This has led us to question our assumptions and examine the question of *what success and performance look like today* for open infrastructures, both at the level of individual infrastructure services and at the ecosystem level.

In this section, we want to celebrate the successes of open infrastructure and highlight some of the trends we see regarding the adoption of, dependence on, and shifts towards open infrastructure. We aim to monitor, track, and tell stories to highlight what we see as pivotal developments and trends in the adoption and success of open infrastructure services and to invite the broader community into conversations and co-creation of experiments around how we can better invest (financially and otherwise) to build a more resilient ecosystem of open infrastructures.

We hope that this section will, over time (with the annual release of this State of Open Infrastructure report), become a valuable resource both for those who are looking to better understand trends and transitions in the open infrastructure space, and those advocating for further adoption and investment into open infrastructure for research and scholarship.

[1] Launched in April 2024, Infra Finder is IOI's new open infrastructure discovery tool. In building the first release of Infra Finder, we worked with infrastructure service providers to collect data about 57 infrastructure services enabling open research and scholarship. More at https://infrafinder.investinopen.org, and also in other chapters of this report, including Characteristics of selected open infrastructures.

[2] PASS is also currently in the Incubation phase as an Eclipse Foundation project. Read more at https://projects.eclipse.org/projects/technology.pass

[3] https://investinopen.org/funding-pilots/ioi-fund/

[4] https://investinopen.org/strategic-support/

Stories of (inter)dependence

We started this investigation with data collected from 57 infrastructure service providers as part of our effort to build Infra Finder.[5] The intake instrument for Infra Finder asked participating infrastructure service providers to list their key achievements. From this data, we identified stories and cases of adoption of and dependence on open infrastructure services and standards beyond growth in user or member numbers. These include:

- **The inclusion of open infrastructure services in national policies and strategies:** the Research Organization Registry (ROR) is the recommended institutional identifier in a growing set of national open science policies and persistent identifier (PID) strategies worldwide. Other open infrastructures, such as ORCID, DataCite, Zenodo, and Dataverse, have also been mentioned and included in national policies and roadmaps.[6]
- **Stories of interdependence between open infrastructure services and standards:** the Open Edition of the CWTS Leiden Ranking of worldwide institutions uses OpenAlex data; the Internet Archive officially adopts the International Image Interoperability Framework (IIIF); the Directory of Open Access Books (DOAB) migrates to DSpace 6 in 2021; Fedora introduces Oxford Common File Layout (OCFL) as a preservation standard within the persistence layer in Fedora 6.0.

Taken together, these achievements demonstrate the value that open infrastructure brings to research and scholarship. It is a crucial characteristic of many open infrastructures — technological transparency and interoperability — that makes these success stories possible.

Moving towards open

Here, we want to highlight two cases where infrastructure service adopters and users are shifting away from closed solutions towards more open ones.

Beprexit

Perhaps one of the most notable cases of a collective shift away from closed infrastructure in the scholarly communication space was the migration away from Digital Commons after its provider bepress was acquired by Elsevier in 2017. Digital Commons, a turnkey institutional repository (IR) platform, was a popular solution, especially for small library presses, because it is a hosted solution (no hosting and related technical capacity and related resources required on campus) and because of its ability to provide both repository and journal publishing services.

After its acquisition, the University of Pennsylvania (Penn) Libraries became the first institution to announce their termination of the partnership with bepress and plan to look for alternatives (Penn Libraries, 2017). Since then, many institutions and organizations, including Cornell University Library, Iowa State University, Pacific University, the University of North Texas Health Science Center, the Journal of eScience Librarianship, and others, have migrated from Digital Commons to open alternatives, such as Hyku, DSpace, and Janeway (Baird & Meetz, 2022; Corrice et al., 2021; Goldman, 2022; Shelley, n.d.; Woodward, 2019).

It is worth noting that the acquisition was only part of the reason Penn Libraries chose to migrate away from bepress — many institutions migrate due to dissatisfaction with key functionalities and inability to meet their organization and community's evolving needs (Baird & Meetz, 2022). Indeed, even before bepress's acquisition, library publishers have voiced a desire for a better hosted turnkey IR and publishing solution than Digital Commons, one that is interoperable with and adaptable to the institution's existing tooling and has enhanced workflow features akin to those offered by the Open Journal System (OJS) and Janeway.

[5] https://infrafinder.investinopen.org/

[6] For example, Ireland's National Action Plan for Open Research 2023-2030 has the action point "Support the Irish ORCID Consortium and encourage further development and adoption of ORCID …", and Latvia's Open Science Strategy 2021-27 mentions "… with the support of the Ministry of Education and Science should create a network of general research data repositories, DataverseLV, …"

The Next Generation Library Publishing project (NGLP) was established in part to fill this gap and create a robust, flexible, and interoperable toolchain that both meets the needs of library publishers and aligns with academic values.[7] One of library publishers' most vital needs is visibility for its published content. In response to that, the NGLP team and their partner Cast Iron Coding developed Meru, a discovery and display platform that can ingest content from various upstream submission and curation systems, e.g. Janeway, OJS and DSpace. Building on Meru, NGLP aims to pilot a Software-as-a-Service offering. However, unlike bepress, the interoperability and modularity of their design lay solid foundations for the offering to better meet the needs of diverse users and encourage a healthy level of coopetition between various open infrastructure service providers. With its simultaneous consortial publishing pilot (more in the section on shared development and adoption), we look forward to seeing whether the maturation of NGLP's technologies will encourage more institutions to move away from bepress.

Open research information and bibliometrics

We've also been paying close attention to the bibliometric data space. The use of proprietary data and tools for research discovery, assessment, and evaluation has long been criticized because these so often lack both transparency and inclusivity. The UNESCO Open Science Recommendation includes "open bibliometrics and scientometrics systems for assessing and analysing scientific domain" as part of the open science infrastructure to further invest in (UNESCO, 2021). In January 2024, the French National Centre for Scientific Research (CNRS) announced its decision to unsubscribe from Elsevier's Scopus bibliographic database and its intention to "stop using commercial bibliographic databases altogether as soon as open solutions are sufficiently mature (CNRS, 2024)." At around the same time, the French Ministry of Higher Education and Research announced the establishment of a multi-year partnership with OpenAlex to develop a fully open bibliographic tool (Badolato, 2024). The Leiden Ranking, a robust, high-quality approach to comparing universities (distinct from but comparable to the Times Higher Education World University Rankings and the QS Universities Rankings), also introduced an Open Edition using data from OpenAlex in January 2024 (Waltman et al., 2024).

These recent movements are not only the result of the rapid development and maturation of open infrastructure and data in the bibliometrics and research information space (thanks to the work of groups including the Curtin Open Knowledge Initiative,[8] the Centre of Science and Technology Studies at Leiden University,[9] and SIRIS Academic),[10] but also coordinated effort to reform research assessment (e.g. the formation of the Coalition for Advancing Research Assessment (CoARA)[11] in 2022) and an increase in investment going into these infrastructures (e.g. Arcadia's US$7.5M grant to OurResearch to establish OpenAlex, in March 2024 (Portenoy, 2024)). With the release of the recent Barcelona Declaration of Open Research Information,[12] we expect more institutions and organizations to move towards relying on more transparent and open bibliometric data and tools.

These recent movements are not only the result of the rapid development and maturation of open infrastructure and data in the bibliometrics and research information space, but also coordinated effort to reform research assessment and an increase in investment going into these infrastructures.

[7] https://www.nextgenlibpub.org/

[8] https://openknowledge.community/

[9] https://www.cwts.nl/

[10] https://www.sirisacademic.com/

[11] https://coara.eu/

[12] https://barcelona-declaration.org/

Stronger together: Shared development and adoption

One common challenge often mentioned in our conversations and engagement with institutional decision-makers is the lack of capacity (staffing, budget, and technical support) for individual institutions to maintain their own instances of open infrastructure. Many have expressed the desire for networks they are a member of, e.g. their library consortia, to provide shared resources and services to help their members adopt and maintain open infrastructure and collectively benefit from economies of scale. In that regard, as part of this section of the report, we want to highlight recent examples and cases of shared development and adoption of open infrastructure by networks of institutions that are coming together to explore how they can coordinate their adoption, maintenance, and scaling of infrastructure, improve utilization of resources and cost-effectiveness of their investments, and build representative governance.

Hyku for Consortia

Hyku for Consortia is a project with the aim to explore, develop, and pilot the Hyku open-source, multi-tenant, consortial IR to deliver ultra-low cost hosting, discovery, and access to digital materials (Pennsylvania Academic Library Consortium, Inc. (PALCI), 2019). The consortia leading this work, PALCI and Private Academic Library Network of Indiana (PALNI), heard their members and the broader community articulate the need to develop a solution with key large-scale configuration options and features that would enable shared consortial IR services. This led to the development of a collaborative repository, Hyku Commons, which currently serves over 50 repository tenants across four consortia.[13] Bringing together a broader user community has enabled the identification of gaps in Hyku, encouraged the sharing of solutions across institutions, and led to further development of additional features (PALNI & PALCI, 2023). At the end of the project in 2023, 70% of Hyku Commons users plan to continue using Hyku as a public repository service (compared to 46.7% in 2021, Hyku for Consortia Team, n.d.).

Next Generation Library Publishing project's consortial publishing pilots

In March 2022, the California Digital Library (CDL) announced the launch of its pilot of an NGLP library publishing solution (Lippincott & Mitchell, 2022). Recognizing the lack of an existing platform that can adequately support the complexity required to support the combined role of a publishing platform and IR at a consortial scale, the pilot tested and refined a robust, scalable consortial infrastructure solution for campus-based publishing. In the pilot, CDL's custom, partly OJS-based architecture was replaced by DSpace 7.0, Janeway, and NGLP technologies. In July 2023, in partnership with the Educopia Institute, Stratos, and the University of Iowa Libraries, NGLP received a grant from the Institute of Museum and Library Services to move NGLP infrastructure from pilot phase to production-ready, with consortial publishing as a primary use case. The University of Iowa Libraries is a member of the Library Initiatives of the Big Ten Academic Alliance (BTAA), which is exploring the potential of consortial approaches to campus-based publishing to achieve economies of scale and increase impact.

The Canadian shared institutional repository service

In November 2023, the Canadian Association of Research Libraries (CARL), the Ontario Council of University Libraries (OCUL), and the University of Toronto Libraries announced their collaborative intent to develop a national shared IR service (Ontario Council of University Libraries, 2023). Starting with a pilot programme, the collaboration aims to develop, test, migrate data and refine a robust and scalable IR hosting service that can be scaled out to other organizations as a nationally available service hosted by Scholars Portal. Scholars Portal is a service of OCUL; it currently provides shared technology infrastructure and shared collections for 21 university libraries in Ontario province.[14] In developing and piloting the Canadian shared IR service, Scholars Portal will provide technical hosting and support, including monitoring and security-related services, to benefitting organizations.

[13] https://hykucommons.org/

[14] https://scholarsportal.info/

ORCID's and DataCite's consortial programmes

ORCID and DataCite have separately established consortial membership programmes designed to enable a group of like-minded organizations in a nation or region to collectively participate in these open infrastructures' communities and accelerate the integration of the open infrastructure services in their contexts (Cousijn, 2019; ORCID, n.d.). By forming a consortium to access these infrastructure services, individual institutions benefit from more cost-effective and coordinated sharing of resources, as well as being part of a local community of practice, which not only allows them to access local support but also helps scale infrastructure integration and adoption efforts in their region or nation (Meadows, 2023; ORCID, Inc, 2016). For the infrastructures, the shift away from individual membership towards larger consortial membership improved their financial sustainability by closing the gap between membership dues and operating expenses (ORCID, 2015). ORCID currently has 29 Consortia members, and DataCite has 59.

Concluding remarks

These are by no means the only examples of shared adoption and coordination — we note, for example, the development of repositories and other open infrastructure services at national and regional levels (especially in Latin America, Africa, and Europe) as examples of coordination across multiple institutions and stakeholder groups. As more shared adoption efforts emerge, we find it important to continue to monitor and learn from these networks' and communities' experiences in order to better understand the challenges and impact of shared adoption.

It is also worth noting that consortial and network-level shared adoption is not the only path to addressing the issue of the lack of individual institution's capacity — indeed, as we mentioned above, successful shared adoption requires infrastructure that can handle the associated complexity and careful planning such that growing demands from consortial/network members can continue to be met. Another increasingly popular and important option is to engage service providers who can provide hosting and maintenance services for open infrastructure. These service providers play a critical role in easing and, hence, growing the adoption of open infrastructure (see the chapter on the influence of procurement and information technology governance processes on the adoption of open infrastructure in this report for more) — this is a topic that we are continuing to monitor and look forward to discussing in a future iteration of this report.

These are by no means the only examples of shared adoption and coordination. As more shared adoption efforts emerge, we find it important to continue to monitor and learn from these networks' and communities' experiences in order to better understand the challenges and impact of shared adoption.

Graceful transitions

So far, we have given examples of how open infrastructures described their key achievements and highlighted stories of increased adoption and interdependence. These are important milestones and trends to monitor and celebrate — indeed, these are often the stories and charts that we see highlighted in annual and industry reports, as well as grant proposals, to evidence the success and growth of infrastructure services.

Yet, in one-to-one conversations with OI service leaders in more informal, intimate settings, we hear a different set of stories: how services are navigating organizational transitions like spinning out from an institutional host, handling key personnel changes, mapping the next steps after the end of a major operational grant, revising their revenue models, etc. These are pivotal transitions for infrastructure services that need to be handled with care, yet the reluctance to share knowledge about these more openly can make these feel like isolated events and/or failures. In reality, with careful planning and intentional strategizing, these transitions are opportunities for a service or organization to reflect on its mission, core values, strategic direction, structure, etc. and to make changes that could kick-start its next phase of development or enable it to take on a new role in the infrastructure ecosystem.

In this section, we tell recent stories of transitions to argue that these transitions are a hallmark of health for the services/organizations that aim to be adaptive and responsive to changing community needs and evolving technological and funding landscapes (Skinner, 2022). We also hope to explore what it means and what it takes to *navigate transitions with grace*.

Key personnel departure

With key personnel departures, the reluctance to share publicly can at least be partially attributed to the fact that many open infrastructures started as an idea closely associated with a founder or founding team. As the infrastructure gains traction, the community generally ties the infrastructure service's success with the founders' abilities and reputations. Often, that founder has been the primary fundraiser, architect, or, when necessary, has supported the service with volunteer labour. When a founder or founding team member announces their move away from an initiative, the extended community (including developers, implementers, funders, members, clients, hosts, and others) may become concerned about the stability, resilience, and longevity of the service. Founding members and leadership of open infrastructure services are often reluctant to discuss such moves openly, fearing that it would cause speculation and uncertainty in the community. This is especially true when the departure timeline is short, or when the decision to depart comes with baggage like burnout or with internal friction.

This is in seeming contrast with one of the biggest advantages of working in the open, where, because the creation and development of the infrastructure service are often documented publicly and shared for reuse with an open license, anyone should be able to take the public assets and build on or continue running the service, even in the case of core maintainer or leadership departure. The reality is much more nuanced. Infrastructures and their communities also depend to greater or lesser degrees on intangible assets - the tacit knowledge embodied in experienced leaders and contributors, and the trust and relationships they have developed. A job description can articulate the tasks and responsibilities of a founder/leader, and a founder can keep thorough documentation of how the role was done in the past, but a large part of how this implicit work gets done is directly related to the person's personality, motivation, networks, circumstances, etc, making it impossible to replicate.

> In reality, with careful planning and intentional strategizing, these transitions are opportunities for a service or organization to reflect on its mission, core values, strategic direction, structure, etc. and to make changes that could kick-start its next phase of development or enable it to take on a new role in the infrastructure ecosystem.

With that in mind, a key to successfully navigating key personnel and leadership changes seems to be to decenter the founder(s) and/or leader(s) early on and to build a team that shares responsibilities (not only at the day-to-day but also at the more strategic, higher level). A recent case that exemplifies this is that of the Open Library of Humanities (OLH). OLH was launched in 2015 by Dr Caroline Edwards and Prof Martin Eve as a response to the serials crisis in academic publishing and a direct commitment to finding a way of publishing open-access journal articles in the humanities without any author-facing article processing charges (APCs) (Edwards, 2015). Today, OLH is a renowned publisher of humanities scholarship based at Birkbeck, University of London, publishing 30 diamond open-access journals and supported by more than 340 libraries worldwide (Open Library of Humanities, n.d.). Janeway, the open-source publishing solution powering OLH, was first developed in 2017 by Eve and Andy Byers.[15] Janeway is also a part of the infrastructure underlying publishing operations at institutions such as Carnegie Mellon University, California Digital Library, and TU Wien, and is a partner in the Next Generation Library Publishing project (Invest in Open Infrastructure, n.d.).

In 2022, Eve announced publicly that he would take on a new role at Crossref (Eve, 2022). It is worth noting that prior to the move, OLH's day-to-day operation was already a team effort and not dependent solely on Eve and Edwards' work. The preparation for the move created an opportunity for Edwards, now Executive Director at OLH, and the OLH team to think more deeply about its sustainability and to design a future that enables it to continue to center its core values of being academic- and scholar-led and run by a collaborative team. This included the intentional decision to embed OLH into Birkbeck. Operating out of its own unit within Birkbeck's university structure allowed OLH the freedom to manage its operations with the administrative and HR support of a large organization. At a time when OLH was witnessing other small, academic-led publishers being acquired by commercial publishers, the transition back to Birkbeck ensured that OLH would not be vulnerable to commercial acquisition. Reflecting on OLH's key organizational vulnerabilities, Edwards led the team in designing and gathering support (from the Board, Birkbeck management, OLH editors, and the community) for a new business and sustainability plan for OLH, aimed at professionalizing OLH to better serve its community of editors, academics, and libraries. The plan fostered investment into creating five new roles on the OLH team, which increased OLH's capacity to better support editors and respond to community needs. It also facilitated user research and web development that culminated in a new design for OLH's website, critical to increasing the visibility of OLH's journals and showcasing their vibrant published works (Edwards, 2024). They've recently recruited an accessibility developer and started a new, user-driven project to improve Janeway and OLH's accessibility (Byers & Driver, 2024). In 2023, OLH revived a new Library Board, which constitutes the community-based library governance of OLH's financial model and provides advice and input to the OLH team from the supporting library members (Vega, 2023). The increased capacity, strengthened governance, and enhanced UX/UI also enabled Edwards to be more ambitious with their work in flipping journals — the team confirmed that ten journals have flipped, or are in the process of flipping, from commercial publishers to OLH in the past six months.

[15] https://www.openlibhums.org/site/janeway/

Partnership conclusion

While new partnerships are often celebrated, their conclusions are seldom discussed publicly. Even where partnerships end cordially, speculation often runs rampant about the reasons for partners' going separate ways. In fact, partnerships are often formed when two or more organizations identify an opportunity where working together on a specific scope of work would advance their shared vision and/or enable them to better meet their respective communities' needs. It is, therefore, natural that that joint scope of work would one day be completed and the partnership would end, or that the community's needs would have evolved and the partnership is no longer as mutually beneficial as it once was.

Transitioning away from a partnership with an intentional process and careful preparation can create opportunities for partners to reflect on their distinct identities and work and to design a new path forward. A recent case that illustrates this well is that of Dryad. Founded in 2007, Dryad is an open data publishing platform and a community committed to the open availability and routine re-use of all research data (Dryad, n.d.). In 2018, it announced that it had entered a strategic partnership with the California Digital Library (CDL) to leverage Dryad and CDL's respective strengths to offer new products and services and build broad, sustainable, and productive approaches to data curation (The Dryad Team, 2018). This led to the launch of a modernized Dryad platform; the building of new integration with manuscript systems and more robust curation services for the benefit of journal publishers; and the development of membership pathways for academic institutions (Lippincott, 2023b).

Upon achieving the partnership's goals, Dryad and CDL announced the conclusion of the partnership in 2023. This created a new space for the Dryad team to chart their path forward and explore what it means for Dryad to operate as an independent team. In that process, the team prioritized building capacity, structures, and revenue models that would enable Dryad to grow while keeping its community's interests and needs at heart, and to advance its mission. The support and guidance from their community-elected board were critical in supporting the core team in this process. For example, the Board engaged with frameworks for value delivery (the value chain) and risk analysis, that helped it to provide informed critique and support for structural changes that were introduced. Having their codebase and content openly licensed was also critical, for it allowed Dryad to enter into and leave partnerships without losing control of these critical assets.

Having established strong service levels, platform stability, and regular community engagement programmes, the Dryad organization now has a clear view of life after the successful end of the partnership with CDL. There is a clear and full understanding of the costs associated with running the service and platform and supporting the Dryad multistakeholder community with emerging practices in research data management. To better understand the value proposition to its members, the Dryad team engaged a consultant to conduct focus groups and interviews. Building on the insights from this research, and the fresh understanding of costs, the team is exploring with its members and the broader community the levels of investment needed to allow Dryad to deliver its core services and value to meet its community's growing and changing needs (Lippincott, 2023a).

Spin-down and sunset

A spin-down or sunset can be even more challenging to share news of or talk about, not least because these sometimes happen as a surprise rather than a long-planned process. Ideally, open infrastructures have looked ahead and documented exit strategies and sunset plans, but many are so busy trying to chart a survival path that this crucial work is not undertaken until the situation becomes critical. Instead of a deliberate process that happens with an official sunset budget, the resulting shutdown experience is ad hoc, underfunded, and conducted over either a quick emergency period (e.g., Digital Preservation Network (Schonfeld, 2018), the Open Collective Foundation (Open Collective, 2024)) or unfolds as a period of long, slow decline until resources give out.[16]

A recent example of an infrastructure that is carefully planning its potential spin down and exploring options to continue is the Open Access Tracking Project (OATP)[17] and its underlying open-source software TagTeam.[18] OATP is a crowd-sourced social-tagging project that captures news and commentary on open access to research. Since its launch in 2009, OATP has tagged more than 102,000 works and amassed more than 17,000 followers on various social media platforms. It is widely regarded as the most comprehensive source of news in the open-access community. OATP is one of more than 400 "hubs" or projects hosted on Harvard's instance of TagTeam.

[16] For an analysis of the main reasons scholarly communication infrastructure providers are "Red-Queen-Racing", see Skinner (2019).

[17] https://cyber.harvard.edu/hoap/Open_Access_Tracking_Project

[18] https://cyber.harvard.edu/hoap/TagTeam

OATP and TagTeam are both created and mainly maintained by Peter Suber, Senior Advisor on Open Access in Harvard Library and Director of the Harvard Open Access Project. As Suber is planning to retire, he is seeking to turn over the responsibility of development and maintenance responsibilities of both TagTeam and OATP to a succeeding not-for-profit entity. He has carefully mapped out scenarios for how OATP can be turned over to another organization, what would that entail for the new host(s) of OATP and TagTeam, and his desired role and level of involvement after the handover.[19] The responsibility of the new host(s) would include management of OATP's tagger community (recruiting new taggers, giving feedback to taggers, keeping them motivated, etc.), moderating OATP's feeds, deciding on tag vocabulary, running OATP's social accounts, approving new TagTeam account requests, ensuring the codebase is maintained, etc.

Concluding remarks

A strength of open infrastructure is its ability to continuously evolve to adapt to changing community needs, technological and socioeconomic conditions, and policy environments. These changes can be very good reasons to sunset and retire a service (e.g. technological advancement rendering certain services obsolete), for partnerships to conclude, and for founders to move on. It is therefore as important to talk about, share knowledge around, and normalize transitions as to discuss and celebrate growth and new developments. Being able to prepare for and navigate these transitions in line with the core values of the service and its community and the vision for its future should be a hallmark of a healthy, robust open infrastructure ecosystem.

We note that the three types of transitions we've expanded on in this sub-section are not the only ones that would benefit from more careful planning and knowledge sharing. For example, we have not discussed mergers and acquisitions here — we and others have been monitoring acquisitions and consolidations in the sector. We also welcome readers' input and suggestions on other types of transitions that would benefit from normalizing and further exploration.

More open conversation is needed to help the open infrastructure community understand how best to prepare for these transitions, and for investors and contributors to understand how to support them. We draw inspiration from work done in the broader open space to change the narrative and normalize these transitions (e.g., the FOREST framework (Lippincott & Skinner, 2022), It Takes a Village,[20] the Commons Conservancy[21]) and we look forward to contributing to advancing this conversation.

Acknowledgement

The IOI team would like to thank Jennifer Gibson and Sarah Lippincott (Dryad), Dr Caroline Edwards (OLH), and Dr Peter Suber (OATP and TagTeam) for their contributions to and review of the graceful transitions section of this chapter.

[19] See Suber's succession planning document for OATP and TagTeam for more details: https://web.archive.org/web/20240411190941/https://docs.google.com/document/d/1hexl0i0L4lWN2rqGLUr364kbKlU3HPlXbml4Juhe57w/edit

[20] https://itav.lyrasis.org/

[21] https://commonsconservancy.org/

References

Badolato, A.-M. (2024, February 15). French Ministry of Higher Education and Research partners with OpenAlex to develop a fully open bibliographic tool. Ouvrir La Science. https://www.ouvrirlascience.fr/french-ministry-of-higher-education-and-research-partners-with-openalex-to-develop-a-fully-open-bibliographic-tool

Baird, L., & Meetz, J. (2022). Digital Commons to Hyku: An Institutional Repository Migration at a Small Liberal Arts University. Journal of Librarianship and Scholarly Communication, 10(1). https://doi.org/10.31274/jlsc.12916

Byers, A., & Driver, S. (2024, March 19). Our Commitment to Accessibility. Open Library of Humanities. https://www.openlibhums.org/news/694/

CNRS. (2024, January 11). The CNRS has unsubscribed from the Scopus publications database | CNRS. https://www.cnrs.fr/en/update/cnrs-has-unsubscribed-scopus-publications-database

Corrice, J., McLaren, C., Del Rosso, J., & Steinhart, G. (2021, May 31). Bepress to DSpace in PJs: Migrating two open repositories from home. OPEN REPOSITORIES 2021, June 7-10. https://or2021.openrepositories.org/2021/05/31/bepress-to-dspace-in-pjs-migrating-two-open-repositories-from-home/

Cousijn, H. (2019, August 26). The DataCite Membership model: Consortia, repositories, and more. DataCite. https://datacite.org/blog/the-datacite-membership-model-consortia-repositories-and-more/

Dryad. (n.d.). Dryad | Who we are. Retrieved 19 April 2024, from https://datadryad.org/stash/about

Edwards, C. (2015, September 27). OLH Launch. Dr Caroline Edwards. https://www.drcarolineedwards.com/2015/09/27/olh-launch/

Edwards, C. (2024, April 23). OLH unveils new website and brand design. Open Library of Humanities. https://www.openlibhums.org/news/684/

Eve, M. P. (2022, August 26). Moving On: My Infrastructural Turn. Martin Paul Eve. https://eve.gd/2022/08/26/moving-on-my-infrastructural-turn/

Goldman, J. (2022, September 1). Welcome to JeSLIB on Janeway! Journal of eScience Librarianship. https://publishing.escholarship.umassmed.edu/jeslib/news/5/

Hyku for Consortia Team. (n.d.). What Do Our Users Think? 2023 Hyku Commons Satisfaction Survey Results. Retrieved 12 April 2024, from https://hykuforconsortia.palni.org/news/what-do-our-users-think-2023-hyku-commons-satisfaction-survey-results

Invest in Open Infrastructure. (n.d.). Solution: Janeway | Infra Finder. Infra Finder. Retrieved 19 April 2024, from https://infrafinder.investinopen.org/solutions/janeway

Ioannidis, A. (2020, March 10). Dryad & Zenodo: Our Path Ahead. Zenodo | Blog. https://blog.zenodo.org/2020/03/10/dryad-and-zenodo-our-path-ahead/

Lippincott, S. (2023a, July 24). Invitation to comment: Dryad's revised membership model for academic and research institutions. Dryad News. https://blog.datadryad.org/2023/07/24/invitation-to-comment-dryads-revised-membership-model-for-academic-and-research-institutions/

Lippincott, S. (2023b, October 20). Dryad news | New at Dryad: Celebrating a milestone in the Dryad and CDL partnership. Dryad News. https://blog.datadryad.org/2023/10/20/new-at-dryad-celebrating-a-milestone-in-the-dryad-and-cdl-partnership/

Lippincott, S., & Mitchell, C. (2022, March 8). California Digital Library: Prioritizing Community and Sustainability as an NGLP Pilot Partner | Educopia Institute. Educopia Institute. https://educopia.org/nglp-cdl-pilot/

Lippincott, S., & Skinner, K. (2022). FOREST Framework for Values-Driven Scholarly Communication. Educopia Institute. https://educopia.org/wp-content/uploads/2022/05/FOREST-Framework-2.0-Spring-2022.pdf

Meadows, A. (2023, June 21). The ORCID US Consortium at Five: What's Worked, What Hasn't, and Why? The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2023/06/21/the-orcid-us-consortium-at-five-whats-worked-what-hasnt-and-why/

Nonprofit Finance Fund. (2020, May 14). NFF Selects Cohort in Mellon-Funded Initiative to Advance Financial Resiliency in the Digital Humanities | Nonprofit Finance Fund. Nonprofit Finance Fund. https://nff.org/news/nff-selects-cohort-digital-humanities

Ontario Council of University Libraries. (2023, November 22). CARL, OCUL, and University of Toronto Libraries Sign MOU Outlining Collaborative Intent to Develop a National Institutional Repository Service | Ontario Council of University Libraries. https://ocul.on.ca/carl-ocul-utl-repository-mou-announcement

Open Collective. (2024, February 28). Open Collective Official Statement—OCF Dissolution. Open Collective. https://blog.opencollective.com/open-collective-official-statement-ocf-dissolution/

Open Library of Humanities. (n.d.). Who we are. Open Library of Humanities. Retrieved 19 April 2024, from https://www.openlibhums.org/site/who-we-are/

ORCID. (n.d.). ORCID Consortia. ORCID. Retrieved 12 April 2024, from https://info.orcid.org/orcid-for-consortia/

ORCID. (2015). ORCID 2015 Annual Report [Presentation]. ORCID. https://doi.org/10.23640/07243.13000079.v1

ORCID, Inc. (2016, February 17). Launching a National ORCID Consortium—Roles and Responsibilities of the Consortium (H. Marks / L. Haak). https://www.slideshare.net/ORCIDSlides/launching-a-national-orcid-consortium-roles-and-responsibilities-of-the-consortium-h-marks-l-haak

PALNI & PALCI. (2023, October 25). Hyku for Consortia Removing Barriers to Adoption Project Wrap-up. https://docs.google.com/presentation/d/1RnJqqaSKDH6S72WbG0atrlYoE_Uhd2GMRZ6DWGvnLs4

Penn Libraries. (2017, September 19). Penn Libraries to End Partnership with bepress. Operation Beprexit. https://beprexit.wordpress.com/official-statement/

Pennsylvania Academic Library Consortium, Inc. (PALCI). (2019). Scaling Up a Collaborative Consortial Institutional Repository, March 2019 Proposal to IMLS. https://www.imls.gov/sites/default/files/project-proposals/lg-36-19-0108-19-full-proposal.pdf

Portenoy, J. (2024, March 13). OurResearch receives $7.5M grant from Arcadia to establish OpenAlex, a milestone development for Open Science. OurResearch Blog. https://blog.ourresearch.org/ourresearch-receives-7-5m-grant-from-arcadia-to-establish-openalex-a-milestone-development-for-open-science/

Schonfeld, R. C. (2018, December 13). Why Is the Digital Preservation Network Disbanding? The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2018/12/13/digital-preservation-network-disband/

Shelley, A. (n.d.). Digital Repository updates coming in 2024—University Library. Iowa State University Library. Retrieved 12 April 2024, from https://www.lib.iastate.edu/news/digital-repository-updates-coming-2024

Skinner, K. (2019, July 23). Why Are So Many Scholarly Communication Infrastructure Providers Running a Red Queen's Race? [Educopia Institute]. https://educopia.org/red-queens-race/

Skinner, K. (2022, July 15). Embracing Change | Educopia Institute. https://educopia.org/embracing-change/

The Dryad Team. (2018, May 30). Dryad partnering with CDL to accelerate data publishing. Dryad News. https://blog.datadryad.org/2018/05/30/dryad-partnering-with-cdl-to-accelerate-data-publishing/

UNESCO. (2021). UNESCO Recommendation on Open Science. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000379949

Vega, P. C. (2023, May 16). Nominations now open for the OLH Library Board. Open Library of Humanities. https://www.openlibhums.org/news/632/

Waltman, L., van Eck, N. J., Visser, M., Neijssel, M., Montgomery, L., Neylon, C., Kramer, B., Demes, K., & Priem, J. (2024, January 30). Introducing the Leiden Ranking Open Edition. https://www.leidenmadtrics.nl/articles/introducing-the-leiden-ranking-open-edition

Woodward, N. (2019). Bepress to DSpace Migration: A Case Study. https://hdl.handle.net/2249.1/156439

Regional policy developments and their implications for open infrastructure

Introduction

The 2021 adoption of the UNESCO Recommendation on Open Science marks a meaningful milestone in the global movement to make "the scientific process more transparent, inclusive and democratic" (UNESCO Open Science Advisory Committee, 2021). While there are other motivators for this shift, the time seems right to share here an overview of recent open science and adjacent policy developments across Africa, Europe, Latin America, and the United States, and to reflect on the implications for open digital research infrastructure.

Looking across these regions, we could not help but notice the diversity of the levers used to advance open science and the scale at which they operate. Europe, Africa, and Latin America have strong regional and continental initiatives and/or networks in development or in place, and we also see some trans-continental cooperation between Latin America, Europe, and Africa. Recent legal developments in the EU and their potential to both advance and also perhaps complicate open science practices are also noteworthy. In contrast, we are not aware of significant open science initiatives across North America; accordingly, our focus here is on the United States, where the U.S. federal government is one of the most important drivers of open science practice.

Another observation is that the scope and core priorities across this landscape differ. In the U.S., while equity in participation is a component of the recent government-wide mandate to promote public access to research, it is secondary to providing access to research outputs. Public access to research outputs is also a long-standing priority and driver of policy in Latin America. On the African continent, access to opportunities to *perform* research is on more equal footing, but this probably reflects a history of low investment in supporting infrastructure and access to it. Meanwhile, the EU has an ambitious open science policy that focuses on providing early access to research using digital and collaborative technology, but is also embedded within a landscape of technological policy initiatives with potentially profound implications for the entire digital ecosystem.

This is a rich space and we cannot do it justice here, and we also plan to release later this year a more comprehensive report on the policy landscape in selected regions.

Policy developments in Africa

Within the African context, open science, and by extension open infrastructure, has been pioneered and largely driven by researchers and universities. This is in part due to the fact that governmental research and development (R&D) spending on the continent has been very low. In 2007, the African Union set a target to have all African countries spend a minimum of 1% of their GDP on R&D by 2010 (Iizuka et al 2015). To date, no African country surpasses that 1% threshold, Kenya and South Africa are closest to meeting this threshold at 0.8% each (Midega et al., 2021).

While financial support for open science at the country level may be sparse, the organization of national and regional networks that are well positioned to support and advance open science policy and infrastructure is notable. In the early 2000s, national research and education networks (NRENs) started being formed in Africa with a view towards enhancing internet connectivity and shared resources for academia and research institutions. The formation of NRENs is driven by research institutions, universities, and in some instances, the government. One of the biggest barriers towards the adoption of open science in the continent was the issue of connectivity, which in the early 2000's was prohibitively expensive, and in many places non-existent. The Tertiary Education and Research Network of South Africa (TENET) and the Kenya Education Network Trust (KENET) were the first networks formed in 1998 and 1999 respectively, and since then there has been a steady formation of NRENs, some of which are being started even in 2024. These NRENS have evolved from just providing connectivity to providing services such as high-performance computing, cloud storage, and identity federation. However, even at this time, some countries do not have any NRENs to support connectivity of their universities and research institutions to high-speed networks (Mwangi et al., 2021; Foley, 2016).

Besides the establishment of NRENs, effort has also been put into forming regional research and education networks (RRENs). To date, there are three research and education networks that can be considered as RRENs in Africa, namely UbuntuNet Alliance (covering East and Southern Africa), the West and Central African Research and Education Network (WACREN), and the Arab States Research and Education Network (ASREN).

There have also been a number of initiatives on the continent that are aimed at advancing open science on the continent. Launched in 2016 by WACREN, Library Support for Embedded NREN Services and E-infrastructure (LIBSENSE)[1] aims to bring together the RENs and academic library communities to strengthen open access and open science in Africa. LIBSENSE has been holding regional policy development workshops in collaboration with regional research networks and university associations to implement the UNESCO Open Science Recommendation at the campus level. Another key dimension is that LIBSENSE has been providing open science policy templates to make it easier for institutions to enact and implement open science policies. An example of this is in Sierra Leone, where LIBSENSE helped form an NREN (SL-REN) and now is also working with stakeholders there to develop a national open science policy (LIBSENSE, 2023).

Another regional initiative driving open science in Africa is the African Open Science Platform (AOSP).[2] AOSP is an initiative formed by the National Research Foundation (NRF) of South Africa as a direct outcome of the NRF Open Access Statement that came into effect in March 2015 (National Research Foundation, 2015). It has the mission to put African scientists at the cutting edge of contemporary, data-intensive science. AOSP is developing an integrated approach involving a federated hardware, communications and software infrastructure, developing policies and enabling practices to support open science in the digital era, and a network of excellence in open science that supports scientists and other societal actors in accumulating and using modern data resources to maximize scientific, social and economic benefit.

Policy development at the regional and international levels

There are a number of instruments that have been implemented at continental and regional levels to contribute to the adoption of open science in Africa. First is Agenda 2063 (African Union Commission, 2015), which is Africa's blueprint and master plan for transforming Africa into the global powerhouse of the future. The plan, developed by the African Union, places an emphasis on science, technology and innovation (STI) as integral to Africa's transformation. In Agenda 2063, we can see the building blocks that can be used to further entrench and implement open science in the continent.Prominence is given to STI in Aspiration 1 and 2, and world-class information and communications technology is seen as crucial for Africa's transformation. One of the core areas for Agenda 2063 is the development of a pan-African e-network. This component looks at the development of terrestrial and submarine connectivity infrastructure to help enhance internet connectivity and pan-African collaboration on STI.

Second is the Science, Technology and Innovation Strategy for Africa (STISA, African Union Commission, 2014) built to operationalize sections of Agenda 2063 in relation to STI. STISA was designed as a series of policy frameworks to be renewed after 10 years to ensure that Africa is able to respond to the dynamism of the STI sector. In STISA 2024, open science is not mentioned explicitly but rather inferred through terms like open data, open innovation, and the co-creation of research and innovation.

In STISA, there are four mutually reinforcing pillars which are prerequisite conditions for its success. These pillars are: building and/or upgrading research infrastructures; enhancing professional and technical competencies; promoting entrepreneurship and innovation; and providing an enabling environment for STI development in the African continent (Hamdy, n.d.). STISA explicitly outlines the important role that NRENs have in facilitating coordinated collaboration between themselves as well as with other adjacent stakeholders across the continent to further innovation and research.

[1] https://libsense.ren.africa/en/

[2] https://aosp.org.za/about-us/

We also have the African Union Declaration on Internet Governance (Degezelle, 2022) which was ratified in 2017 in Algeria by information and communications technology (ICT) Ministers from across Africa. The Declaration advocates for an open, transparent, and inclusive strategy of internet governance based on the principles of openness, including freedom of expression, respect for private life, universal access and technical interoperability.

At the regional level, there are initiatives such as the Southern African Development Community (SADC)'s Regional Indicative Strategic Development Plan (Southern African Development Community (SADC) Secretariat, 2020) and the East African Community (EAC)'s East African Science and Technology Commission (EASTECO)[3] that are helping to drive open science.

Policy development at the national level

Over the past decade, we have seen an increasing number of national open access (OA) policies in Africa, and recently also an increase in the number of open science policies. There is also an interplay with policy development at the regional and international levels as they have influenced how national open science policies are developed. We summarize in Table 1 some of the recently implemented and upcoming open science policies in Africa.

TABLE 1.
Selected country-level open science policies in Africa

Country	Year	Policy
Ethiopia	2019	National OA policy adopted by the Ministry of Science and Higher Education of Ethiopia (MOSHE) in 2019 (Beyene et al., 2022). It mandates open access to all published articles, theses, dissertations, and data resulting from publicly-funded research and encourages open science best practices, including the use of data management plans (DMPs) by researchers and FAIR data practices. One of the implications of the policy has been the formation of the National Academic Library of Ethiopia,[4] a free, centralized repository of Ethiopian research.
South Africa	2022	South Africa has a first draft of an open science policy, discussed in a stakeholder consultation meeting in February 2022 (Merwe, 2022), which mandates open access for publicly funded research processes and outputs, including data acquired or generated by public funds. The policy proposes the establishment of a "national 4 forum" to promote best practices in open science as well as incentives for researchers to publish in open-access journals. The South African Department of Science and Innovation (DSI) in 2022 announced that it is considering creating the South Africa Open Science Cloud (SAOSC), modelled on the European Open Science Cloud (EOSC) (Cavalli, 2022).
Multiple countries	2022	Work to develop national open science roadmaps has been done in various countries across Africa to varying stages of maturity, including Côte d'Ivoire, Ethiopia, Ghana, Lesotho, Mozambique, Nigeria, Somalia, Tanzania and Uganda (Oaiya, 2022).

[3] https://easteco.org/

[4] http://ndl.ethernet.edu.et/

Conclusion

Despite the relatively late formation of open science policies in Africa compared to other regions, there is a lot of momentum for open science in the continent at the moment. There are a raft of national open science policies in development and also numerous initiatives led by NRENs, RRENs, and research organizations that are also helping drive the open science agenda. The UNESCO Recommendation of 2021 has been a big boost to efforts to advance open science on the continent. By defining open science, defining stakeholders critical to its successful implementation, as well as addressing key issues like multilingualism of science, this has de-mystified open science in Africa. The Recommendation has helped provide some parameters to factor in when making policies and this previously was an important, yet missing component.

Policy developments in Europe

Open Science is a policy priority for the European Commission. The EU's open science policy details eight areas of ambition, including the European Open Science Cloud (EOSC, more below), open data, metrics development, scholarly communication, research integrity, and more (European Commission, 2019). The policy's development and implementation are supported by the European Commission's research and innovation funding programmes[5] and synergize with the EU's support for international bodies and platforms such as Plan S, the Research Data Alliance, and the Committee on Data of the International Science Council (CODATA) (European Commission, 2019).

The EU's policy and investment into open science also impact the development of national open science and infrastructure policies and strategies in Europe, which we will elaborate on in a more comprehensive report. In this section, instead, we focus on recent EU-level legislation in the broader technology and digital infrastructure space that may have implications for open research infrastructure development and strategy in Europe and around the world. These legislative developments are motivated by the increase in power and prevalence of big tech companies and their platforms, the desire to safeguard digital sovereignty (see our Future signals editorial for more on digital sovereignty), and the growth of the open movement and digital public good conversations internationally and in Europe.

The EU Open Data Directive entered into force in July 2019. It is a key instrument mandating that EU Member States develop open access policies that must be compatible with the FAIR principles and ensure that such research data becomes available for re-use. The directive is not straightforward as to which entities are responsible for permitting the re-use of research data, but we can deduce that Research Funding Organisations (RFOs) and Research Performing Organizations (RPOs) have key roles to play (Directorate-General for Research and Innovation (European Commission) and Eechoud, 2022). This instrument also impacts the development of the EOSC, a pan-European venture with the ambition to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and reuse data, tools and services for research, innovation and educational purposes (European Commission, n.d.).

The European Data Governance Act, which entered into force in June 2022, aims to "make more data available and facilitate data sharing across sectors and EU countries in order to leverage the potential of data for the benefit of European citizens and businesses" (European Commission, 2024a). The part of the Act that potentially affects research infrastructure is Chapter III where it introduces a notification and supervision framework for data intermediation services, services aiming to "establish commercial relationships for the purpose of data sharing between an undetermined number of data subjects and data holders, on the one hand, and data users on the other hand, through technical, legal or other means, including for the exercise of data subjects' rights in relation to personal data". Scientific data repositories are not in the scope of the Data Governance Act because of their non-commercial nature, but some in the research community have noted that the case of hybrid initiatives should be clarified (League of European Research Universities, 2022, Directorate-General for Research and Innovation (European Commission) and Eechoud, 2022).

[5] First by Horizon 2020, then Horizon Europe, which is the European Commission's current research and innovation funding programme until 2027. More about Horizon Europe at https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en.

The European Data Act, which entered into force in January 2024, complements the Data Governance Act in clarifying who can create value from data and under which conditions (European Commission, 2024d). Particular aspects of the Act that can concern digital research infrastructure include provisions to enable users of connected products to access the data generated by these products and to share such data with third parties, and rules that allow customers to switch seamlessly between different cloud providers (European Commission, 2024b), which especially may substantially impact EOSC (Directorate-General for Research and Innovation (European Commission) and Eechoud, 2022). The Data Act also revisits certain aspects of the Database Directive, in particular the scope of *sui generis* database rights, which may have implications on which and under what circumstances databases and data repositories in the research and scholarship space are protected (European Commission, 2021a,b,c; Bernier et al., 2023).

The Digital Services Act package consists of the **Digital Services Act** (DSA) and the **Digital Market Act** (DMA). The package aims to "create a safer digital space in which fundamental rights of all users of digital services are protected and to establish a level-playing field to foster innovation, growth and competitiveness, both in the European Single Market and globally" (European Commission, 2024c). The DSA, which specifies additional obligations for online intermediary services and platforms, can create new costs and impacts for research infrastructure service providers. For example, if repositories fall under the DSA's definition of platforms, they will need to establish the "notice and action", internal complaint handling, and trusted flagger systems and functions (European Commission, Directorate-General for Research and Innovation, Lundqvist, B., 2022). The DMA establishes qualifying criteria for large online platforms as "gatekeepers" and their obligations. Currently, while no research infrastructure service providers nor their core platform services are designated under the DMA, gatekeepers could emerge in the digital research and scholarship industry (European Commission, Directorate-General for Research and Innovation, Lundqvist, B., 2022), affecting how users (institutions and researchers) can interact with them and the data they hold.

We also describe and note the following EU-level policies that have yet to enter into force but will potentially impact the development of open infrastructure and related policies in Europe, at both the continental and national levels.

The **EU AI Act** takes a risk-based approach to fostering trustworthy AI in Europe and guaranteeing the safety and fundamental rights of people and businesses when it comes to AI (European Commission, 2024e). Research and research infrastructures become increasingly dependent on AI tools and models (see our Future signals editorial for some of the trends we're paying attention to). Although most of the AI models used in research infrastructure are unlikely to be classified as high risk, the Act is still likely to have some effect on the research community, particularly on those who are developing AI models as part of their research or tool and infrastructure development. However, it is worth noting that a clause has been added to the draft act to exempt AI models developed purely for research, development, or prototyping (Gibney, 2024).

The EU Cyber Resilience Act (CRA) introduces mandatory cybersecurity requirements for manufacturers and retailers of products or software with a digital component to ensure that cybersecurity is maintained throughout a product's lifecycle (European Commission, 2023). When the draft text of the CRA was released in 2023, the open-source community was particularly concerned that upstream communities would be made responsible for downstream vulnerabilities (Linux Foundation, n.d.). Thankfully, the European Commission has responded to the community's feedback and adjusted the final text, which excludes free and open-source software (FOSS) products that are not monetised by their manufacturer and contributors who are not *providing* FOSS from its scope (Team NLnet Labs, 2024). However, it is likely that some open-source research infrastructures (particularly those that have the intention to monetize) will still be regulated by the CRA. Further, integrators of FOSS are obliged to share any vulnerabilities they have found in a component with the manufacturer, including any patches they might have developed (Hubert, 2023). All this would mean additional work and processes for open infrastructures to comply with the CRA. Interestingly, the CRA also specifies obligations for "open-source software stewards", who "play a main role in ensuring the viability of [FOSS which is intended for commercial activities]", in "putting in place and document in a verifiable manner a cybersecurity policy to foster the development of a secure product with digital elements as well as an effective handling of vulnerabilities by the developers of that product" (Hubert, 2023).

In conclusion, while open science priorities and policy initiatives in Europe (both at EU and national levels) impact the development and adoption of open research and scholarship infrastructure in the continent, as the infrastructure under consideration here is digital, it is worth paying attention to and considering a wider landscape of recent and emerging policies that may impact the users, service providers, and other stakeholders in research and scholarship sector, in order to achieve a more holistic understanding of where investment, resources, and support may be needed to increase the resilience and health of open infrastructure.

Policy developments in Latin America

Policy development at the regional and international levels

Within the Latin American context, there has always been a view of "science as a public good" rather than "science as a commodity" (Harris et al., 2021). This is supported by the fact that Diamond Open Access (Mounier & Rooryck, 2023) is well-established across the region as well as the relatively high investment by the government in research and scholarship. Becerril-García and Aguado-López (2019) characterizes the impact of these views on the ecosystem as follows:

The Latin American region, as a result, owns an ecosystem characterized by the fact that "publishing" is conceived as acts of "making public", of "sharing", rather than the activity of a profit-driven publishing industry (...) Latin American academic journals are led, owned and financed by academic institutions. It is uncommon to outsource editorial processes.

A series of international policies within Latin America underpins the development of open access and consequently open science within the region. The foundational policy that forms the bedrock of open science in Latin America was the **Santo Domingo Declaration Science for the 21st Century: A New Vision and Framework for Action** in 1999 (Anonymous, 1999). This declaration was the outcome of a conference organized by UNESCO and the World Federation of Scientific Workers (WFSW) held in Santo Domingo, Dominican Republic. This declaration aimed to outline a new vision and framework for scientific endeavours in the 21st century. This policy document lays the groundwork for the coordination of STI actions in the region and introduces a narrative consistent with what will later become open access and open science policies (European Commission & Directorate-General for Research and Innovation, 2023). The declaration emphasized the importance of science and technology in the region's development and called for increased collaboration, investment, and integration of scientific efforts.

The key elements of the Santo Domingo Declaration include the promotion of open science, increased funding for research and development, and the establishment of networks to facilitate knowledge exchange. The Declaration seeks to enhance the visibility of Latin American scientific research on the global stage and encourage the free flow of scientific information.

While there are other policy frameworks that were developed to support open science in Latin America after the Santa Domingo Declaration, for the purpose of this report, we focus on policies enacted in or after 2018.

The Panama Declaration on Open Science (2018)[6] was made public by members of the universities and civil society organizations gathered in Panama City (Becerril-García & Aguado-López 2019). One of the key aims of the Panama Declaration was "to move towards collaborative models of knowledge creation, management, communication, preservation and recognise that open science required going beyond open access, by repositioning society's leadership role to produce and benefit from science, technology and innovation."

The Declaration emphasizes the development of national and regional open science policies and infrastructures and ethical considerations in open science, including responsible data-sharing practices and addressing potential risks, and advocates for strengthened regional collaboration and South-South partnerships.

[6] https://forocilac.org/en/declaracion-de-panama-sobre-ciencia-abierta/

The Panama Declaration set the tone for more international collaborations to further open science. In 2021, we saw LA Referencia and RedCLARA (an RREN in Latin America) sign a memorandum of understanding with the three African RRENs with the aim of advancing open science policies, services, and infrastructure that reflect the unique needs and conditions of each continent within a framework of international cooperation (AfricaConnect3, 2021). Other examples of international collaborations are the Building the Europe Link to Latin America and the Caribbean (BELLA) I and II[7] programmes that seek to enhance internet connectivity between Europe and Latin America via submarine and terrestrial cables, and the development of LA Referencia's metadata validation service based on the OpenAIRE standard, which makes Latin American and the Caribbean (LAC) repositories and European repositories interoperable. The EU-LAC Foundation is an international intergovernmental body established in 2010 by LAC and EU states. Its mandate is to build bridges to increase mutual knowledge and promote dialogue and collaboration on issues that are high on the agenda of the strategic partnership between the two regions, such as higher education, science, technology, and innovation.

Policy development at the national level

Table 2 provides an overview of policy and planning initiatives aimed at increasing the adoption of open science in Latin America. Since 2018, we have seen a significant increase in the number of countries with open access/science policies, which provides a solid base for open infrastructure adoption in the region.

TABLE 2.
National-level open science policies in Latin America since 2018

Adapted from Appel et al., (2018) and Heredia (2022).

Country	Year	Policy
Brazil	2022	The fifth Open Governance National Action Plan of Brazil (2021-2023) commits to building a research assessment model to promote open science as an alternative to the models currently applied in Brazil.
Argentina	2021	The Argentinean Ministry of Science, Technology and Innovation (MINCyT) created the Advisory Committee on Open Science and Citizen Science, which has written the document *Diagnosis and Roadmap for an Open Science Policy in Argentina*. This builds on a 2013 law that requires that the outcomes of publicly funded research must be freely available in open access institutional repositories.
Chile	2022	In Chile, the initiative Datos Científicos (Scientific Data) has been established by the National Research and Development Agency (ANID, formerly National Commission of Scientific and Technological Research (CONICYT)), requiring that all data derived from research funded by ANID must be deposited in open repositories in adequate formats.
	2021	ANID granted 12 universities funding through the Innovation in Higher Education (InES) fund for a period of two years to strengthen their institutional capacities in open science, one of the most important dimensions being the implementation of technological infrastructure that complies with international interoperability standards.
Uruguay	2019	Uruguay does not have a national science policy. However, the National Agency for Research and Innovation (ANII), the government body that coordinates the evaluation of research activity, established its open access regulation (Agencia Nacional de Investigación e Innovación., n.d.), which mandates that grant recipients must deposit scientific publications in an institutional repository.

[7] https://www.bella-programme.eu/index.php/en/

Within the Latin American context, we have noted a very strong drive towards open science since 2010. At the national level, we are seeing an increasing number of countries having or actively developing open science policies. This is further supported by regional initiatives that aim to further open science like RedCLARA, Central America Higher Education Council (CSUCA), LA Referencia, and SciELO, among others. There is also a strong emphasis on the development of regional and shared infrastructures, for example, a shared Current Research and Information Systems (CRIS), that also portends a future of adoption and monitoring of progress towards open science and shared infrastructure within the region.

Policy developments in the United States

The primary sources of research funding in the United States are government agencies and private philanthropic foundations. There has been significant activity in both sectors, particularly in the past year or two on the federal side, to advance public access to the results of research and open science more generally. In 2023, the US Office of Science Technology and Policy (OSTP) launched the Year of Open Science, with the overall objective of advancing national open science policies. The OSTP and the National Science and Technology Council (NSTC) set the stage by drafting a definition of open science:

> *The principle and practice of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility, and equity.*
>
> *— Office of Science and Technology Policy (2023)*

Over the course of the year, federal agencies drafted or updated their public access plans in response to the Nelson memo (discussed in the next section), launched public access policies that were already in development (e.g. the National Institutes of Health (NIH) data management and sharing policy), implemented new open science infrastructures and programmes (e.g. the Department of Energy (DOE) unified access point for persistent identifier services, the National Aeronautics and Space Administration's (NASA) five-year Transform to Open Science (TOPS) mission), announced new funding opportunities (National Endowment for the Humanities (NEH) Digital Humanities Advancement Grants and the National Science Foundation (NSF) Geosciences Open Science Ecosystem), and more (Office of Science and Technology Policy, 2024).

Recent U.S. federal government public access policy developments

Much of the significant and recent policy development around open science and supporting infrastructure in the United States relates to evolving U.S. Federal government mandates that the results of government-funded research be made widely and publicly available. The 2022 directive (Office of Science and Technology Policy (2022), known informally as the "Nelson memo") issued by the OSTP is the most recent but not the first such mandate; NIH issued its first public access policy in 2005 (National Institutes of Health, 2005), NSF began requiring data management plans (including data sharing) in grant proposals in 2011 (National Science Foundation, 2011), and the OSTP issued its first multi-agency public access directive to federal agencies in 2013 (the "Holdren memo", Office of Science and Technology Policy, 2013). The Nelson memo extends the Holdren memo in several important ways:

- Eliminating embargoes on the release of research outputs,
- Extending the policy to cover all federal research funders, not only those with more than $100M USD in extramural research (as was the case with the Holdren memo),
- Requiring agencies to (eventually) extend their data sharing requirements to all research data, not only those that directly support peer-reviewed publications,
- Promoting the assignment of persistent identifiers (PIDs) to research outputs and including PIDs for authors and organizations in metadata, and
- Making an effort to address concerns related to equity both in participation in the research process[8] and in access to its results.

[8] Agencies are not tasked with addressing the issue of equitable participation in research immediately; rather the National Science and Technology Council Subcommittee on Open Science is charged with overall coordination among agencies and with considering "measures to reduce inequities in publishing of, and access to, federally funded research and data, especially among individuals from underserved backgrounds and those who are early in their careers" (OSTP 2022).

Agencies with more than $100M in annual research expenditures were given 180 days from the date of the Nelson memo (25 August 2022) to update their public access policies, and agencies with $100M or less in research expenditures were given 360 days to develop new (or extend existing) policies. These new or updated policies were to be submitted to the OSTP and Office of Management and Budget (OMB) for review, with finalized policies published by 31 December 2024, and taking effect not more than one year after publication. These policies are tracked on the U.S. government website *Public Access Plans & Guidance* (CENDI, n.d.). As of early March 2024, all agencies should have submitted drafts to OSTP and OMB, but policy availability for the agencies listed on the *Public Access Plans & Guidance* website was incomplete.

A look at several of the available draft and final policies shows that they are strikingly similar in general terms, although this comes as no great surprise as they were all crafted in response to the Nelson memo. Some of the common elements include deposit of articles (or at the very least, article metadata, with link out to an open access copy hosted elsewhere) to agency-designated repositories or catalogues, a focus on repositories as the preferred means of distributing research data and selection of data repositories that meet the criteria set forth in the National Science and Technology Council's (2022) "Desirable Characteristics of Data Repositories for Federally Funded Research" (when the choice of repository is left to the researcher), allowing some exemptions to sharing (for example in the cases of legal or ethical concerns, proprietary, controlled or classified data, or trade secrets), and allowing "reasonable" costs of compliance to be included in grant proposal budgets.[9]

A few differences across agency policies stand out:

- National Institute of Standards and Technology (NIST, 2023) allows an embargo of up to 12 months for publications under very limited circumstances (namely that a co-author has transferred copyright to the publisher).
- Some agencies are adopting a shift in terminology from "Data Management Plan" to "Data Management and Sharing Plan", to more strongly emphasize data sharing requirements.
- Multiple agencies (Agency for Healthcare Research and Quality (AHRQ, 2023), United States Department of Agriculture (2023), NASA (2023), NIST (2023), and the United States Geological Survey (U.S. Department of the Interior, 2023)) have well-developed data sharing infrastructure that will meet the needs of many of their funded researchers, while others steer researchers to external repositories.
- The NIH (National Institutes of Health, 2023) and NSF (National Science Foundation, 2023) are leading in the area of attempting to ensure equitable access to participate in the research process, even though doing so is not (yet) explicitly required. The policies of both agencies acknowledge the potential for a shift towards article processing charges (APCs) as a means of paying for open access to inequitably impact researchers with inadequate funding and/or small awards, and suggest initial strategies for at least beginning to adjust policy and practice to mitigate the possibility.
- Most agencies expect to monitor and manage compliance via existing award reporting mechanisms, although some have more mature systems in place than others. Some agencies explicitly mention the possibility of adjusting or withholding funding for non-compliance (e.g. AHRQ, DOE, NIH, NASA, U.S. Census Bureau (2024), USGS).

Many of these policies are still in draft stage, with additional policies to be released, but the overall trend is towards increasing and more immediate access to the results of federally funded research. Substantial questions remain regarding allowable and reasonable costs, as well as how potential changes in how the costs of publishing are paid might impact researchers and research funders.

[9] What is meant by "reasonable" is not clearly specified, and a current area of research for IOI. We discuss this further in Steinhart and Skinner (2024).

Implications for open infrastructure

These policy developments have wide-ranging implications for researchers, their institutions, and the research infrastructure they rely on. We share here a few thoughts on what this might mean for open infrastructures that support research and scholarship.

Increased use of some infrastructures is the most obvious consequence of these expanded mandates. We speculated in a recent report (Steinhart & Skinner 2024) that research data repositories whose sources of revenue do not scale up with an increase in deposits may eventually be faced with sustainability challenges. We might expect the same possibility for other fee-free infrastructures such as preprint services, generalist and institutional repositories that do not charge fees, and platinum open access publishers. Disciplines that are under-served by current available infrastructures will make those gaps more evident, possibly surfacing interesting opportunities to expand existing infrastructure in new directions, or develop entirely new services. Alternatively, particularly for "low-budget" disciplines, some researchers may be faced with mandates that can't be met using existing infrastructure.

The Nelson memo also directs federal agencies to ensure the results of the research they fund are accessible for machine access and for individuals using assistive technologies. In the case of funders who support their own infrastructure (for example, NIH's PubMed Central), it is clear where the responsibility lies for making good on this requirement. When researchers rely on independent infrastructure, this mandate potentially breaks down, and it is not at all clear who is responsible for ensuring compliance and where in the research process adapting materials to maximize accessibility should occur. arXiv launched an experimental project to provide an accessible HTML version of TeX/LaTeX submissions (Frankston 2023), a significant step forward in making a piece of critical scholarly infrastructure more accessible, but we can be certain that the cost of achieving this milestone was not insignificant.

Finally, institutions and funders alike will seek ways to monitor and manage compliance. Again, some funders are already better equipped than others to manage this. NIH is one such example, providing institutions with access to their Public Access Compliance Monitor (PACM) system which can be used to track a manuscript's progress through the process of publication in PubMed Central. Research Information Management and Current Research Information Systems (RIMs and CRISs, respectively) may also aid institutions in tracking outputs and their status. Aggregate sources of information include, for example, the COKI Open Access Dashboard[10] and CHORUS.[11] Whether there are open options beyond these sources and whether they will adequately meet the needs of research stakeholders remains to be seen.

This is a rich space and we cannot do it justice here, and we also plan to release later this year a more comprehensive report on the policy landscape in selected regions.

[10] https://openknowledge.community/dashboards/coki-open-access-dashboard/

[11] https://www.chorusaccess.org/

References

AfricaConnect3. (2021). Africa and Latin America agree to closer collaboration around open science. AfricaConnect3 | Welcome to the AfricaConnect3 Website. https://africaconnect3.net/africa-and-latin-america-agree-to-closer-collaboration-around-open-science/

African Union Commission. (2014). Science, Technology and Innovation Strategy for Africa 2024. https://au.int/sites/default/files/newsevents/workingdocuments/33178-wd-stisa-english_-_final.pdf

African Union Commission. (2015). Agenda 2063: The Africa We Want. African Union. https://au.int/sites/default/files/documents/36204-doc-agenda2063_popular_version_en.pdf

Agencia Nacional de Investigación e Innovación. (n.d.). Reglamento de Acceso Abierto. https://www.anii.org.uy/upcms/files/listado-documentos/documentos/anii-reglamento-de-acceso-abierto.pdf

Agency for Healthcare Research and Quality. (2023). AHRQ Public Access to Federally Funded Research. https://www.ahrq.gov/funding/policies/publicaccess/index.html

Anonymous. (1999). The Santo Domingo Declaration. Second ACP Summit, Santo Domingo, Dominican Republic,. https://www.oacps.org/wp-content/uploads/2022/05/The-Santo-Domingo-Declaration.pdf

Appel, A., Lujano, I., & Albagli, S. (2018). Open Science Practices Adopted by Latin American & Caribbean Open Access Journals. ELectronic PUBlishing, Connecting the Knowledge Commons: From Projects to Sustainable Infrastructure(Long Papers). https://doi.org/10.4000/proceedings.elpub.2018.29

Becerril-García, A., & Aguado-López, E. (2019). The End of a Centralized Open Access Project and the Beginning of a Community-Based Sustainable Infrastructure for Latin America: Redalyc.org after Fifteen Years. In L. Chan & P. Mounier (Eds.), Connecting the Knowledge Commons—From Projects to Sustainable Infrastructure: The 22nd International Conference on Electronic Publishing — Revised Selected Papers (pp. 41–55). OpenEdition Press. https://doi.org/10.4000/books.oep.9003

Bernier, A., Busse, C., & Bubela, T. (2023). Public Biological Databases and the Sui Generis Database Right. IIC - International Review of Intellectual Property and Competition Law, 54(9), 1316–1358. https://doi.org/10.1007/s40319-023-01373-0

Beyene, M., Tekle, S. M., & Alemneh, D. G. (2022). Open Access Initiatives in Ethiopia's Higher Learning Institutions. In Handbook of Research on the Global View of Open Access and Scholarly Communications (pp. 23-43). IGI Global.

Cavalli, V. (2022). South African Open Science Cloud — EGI's contribution to SA-EU Dialogue Policy Framework Report. EGI. https://www.egi.eu/article/south-african-open-science-cloud-egis-contribution-to-sa-eu-dialogue-policy-framework-report/

CENDI. (n.d.). Science.gov—Public Access Plans & Guidance. https://www.science.gov/Public-Access-Plans-Guidance.html

Degezelle, W. (2022). The Open Internet as cornerstone of digitalisation: The Global Gateway Partnership Opportunities between the European Union and Africa. Stantec. https://fpi.ec.europa.eu/system/files/2022-10/The%20Open%20Internet%20as%20cornerstone%20of%20digitalisation%20DIGITAL.pdf

Directorate-General for Research and Innovation (European Commission), & Eechoud, M. van. (2022). Study on the Open Data Directive, Data Governance and Data Act and their possible impact on research. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/71619

European Commission. (n.d.). European Open Science Cloud (EOSC): What the cloud is, how it was developed and being implemented. https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en

European Commission. (2019). Open Science—European Commission. https://research-and-innovation.ec.europa.eu/document/download/77d34a83-7ed9-4041-9fb2-3be3094f2319_en?filename=ec_rtd_factsheet-open-science_2019.pdf

European Commission. (2021a). Feedback from: OpenAIRE. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13045-Data-Act-amended-rules-on-the-legal-protection-of-databases/F2660238_en

European Commission (2021b). Feedback from: Wikimedia (FKAGEU). https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13045-Data-Act-amended-rules-on-the-legal-protection-of-databases/F2660150_en

European Commission (2021c). Feedback from: STM. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13045-Data-Act-amended-rules-on-the-legal-protection-of-databases/F2660197_en

European Commission. (2023). EU Cyber Resilience Act. https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act

European Commission. (2024a). European Data Governance Act. https://digital-strategy.ec.europa.eu/en/policies/data-governance-act

European Commission. (2024b). European Data Act enters into force, putting in place new rules for a fair and innovative data economy. https://digital-strategy.ec.europa.eu/en/news/european-data-act-enters-force-putting-place-new-rules-fair-and-innovative-data-economy.

European Commission. (2024c). The Digital Services Act package. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

European Commission. (2024d). Data Act. https://digital-strategy.ec.europa.eu/en/policies/data-act

European Commission. (2024e). AI Act. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Commission, Directorate-General for Research and Innovation, Lundqvist, B.. (2022). Study on the Digital Services Act and Digital Markets Act and their possible impact on research, Publications Office of the European Union. https://data.europa.eu/doi/10.2777/751853

European Commission & Directorate-General for Research and Innovation. (2023). Open access policies in Latin America, the Caribbean and the European Union — Progress towards a political dialogue. Publications Office of the European Union. https://doi.org/10.2777/90667

Foley, M. (2016). The Role and Status of National Research and Education Networks(NRENs) in Africa. World Bank http://documents.worldbank.org/curated/en/233231488314835003/pdf/113114-NRENSinAfrica-SABER-ICTno05.pdf

Gibney, E. (2024). What the EU's tough AI law means for research and ChatGPT. Nature, 626(8001), 938–939. https://doi.org/10.1038/d41586-024-00497-8

Hamdy, A. (n.d.). Report by ASRIC Secretariat on the implementation of STISA-2024. https://asric.africa/documents/Congress-Dec-2020/Report%20on%20the%20implementation%20of%20STISA-2024%20-%20ASRIC%20Secretariat.pdf

Harris, S., Becerril García, A., Hervé Mboa Nkoudou, T., & Dandawate, V. (2021). Global Trends in Open Access: Themes from Africa, Asia and Latin America. The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2021/05/13/global-trends-in-open-access-africa-asia-and-latin-america/

Heredia, A. (2022). A tradition of open, academy-owned, and non-profit research infrastructure in Latin America. Information Services & Use, 42(3–4), 447–452. https://doi.org/10.3233/ISU-220177

Hubert, Bert. (2023). EU CRA: What does it mean for open source? https://berthub.eu/articles/posts/eu-cra-what-does-it-mean-for-open-source/?ref=blog.nlnetlabs.nl

Iizuka, Michiko, Philippe Mawoko and Fred Gault. (2015). Innovation for Development In Southern & Eastern Africa:Challenges for Promoting ST&I Policy https://www.merit.unu.edu/wp-content/uploads/2015/02/252608902-Innovation-for-Development-in-Southern-Eastern-Africa-Challenges-for-Promoting-ST-I-Policy.pdf

League of European Research Universities. (2022). Possible impact of EU data and digital legislation on research (EC Workshop 20 June 2022 — presenting and discussing 2 studies of independent experts); LERU Intervention (focusing on recommendations and conclusions). https://www.leru.org/files/Possible-impact-of-EU-data-and-digital-legislation-on-research-EC-Workshop-20-June-2022_LERU-INTERVENTION-at-workshop-_final.pdf

LIBSENSE. (2023). Sierra Leone takes steps to ingrain open science in national research fabric. https://libsense.ren.africa/en/sierra-leone-takes-steps-to-ingrain-open-science-in-national-research-fabric/

Linux Foundation. (n.d.). EU Cyber Resilience Act: The Linux Foundation Enables Open Source Software Stewardship for the CRA. https://linuxfoundation.eu/cyber-resilience-act

Merwe, C. van der. (2022, February 17). South Africa's draft open science policy promises shake-up. Research Professional News. https://www.researchprofessionalnews.com/rr-news-africa-south-2022-2-south-africa-s-draft-open-science-policy-promises-shake-up/

Midega, J., Kyobutungi, C., Okiro, E., Okumu, F., Aniebo, I., & Erondu, N. (2021). African countries must muscle up their support and fill massive R&D gap. The Conversation. http://theconversation.com/african-countries-must-muscle-up-their-support-and-fill-massive-randd-gap-161024

Mounier, P., & Rooryck, J. (2023). Towards a federated global community of Diamond Open Access [Billet]. The Diamond Papers. https://thd.hypotheses.org/296

Mwangi, K. W., Mainye, N., Ouso, D. O., Esoh, K., Muraya, A. W., Mwangi, C. K., Naitore, C., Karega, P., Kibet-Rono, G., Musundi, S., Mutisya, J., Mwangi, E., Mgawe, C., Miruka, S., Kibet, C. K., & Collaborators, O. (2021). Open Science in Kenya: Where Are We? Frontiers in Research Metrics and Analytics, 6. https://doi.org/10.3389/frma.2021.669675

National Aeronautics and Space Administration. (2023). NASA's Public Access Plan. https://www.nasa.gov/wp-content/uploads/2021/12/nasa-ocs-public-access-plan-may-2023.pdf

National Institute of Standards and Technology. (2023). National Institute of Standards and Technology Plan for Providing Public Access to Results of Federally Funded Research. https://www.nist.gov/document/public-access-plan-public-comment-june-2023

National Institutes of Health. (2005). NOT-OD-05-022: Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-05-022.html

National Institutes of Health. (2023). NOT-OD-23-091: Request for Information on the NIH Plan to Enhance Public Access to the Results of NIH-Supported Research. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-23-091.html

National Research Foundation. 2015. Statement on Open Access to Research Publications from the National Research Foundation (NRF)-Funded Research. https://chelsa.ac.za/wp-content/uploads/2018/12/NRF_open_access_-statement_19Jan2015_V6.pdf

National Science Foundation. (2011). NSF Grant Proposal Guide Chapter II - Proposal Preparation Instructions. https://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp

National Science Foundation. (2023). NSF Public Access Plan 2.0: Ensuring Open, Immediate and Equitable Access to National Science Foundation Funded Research. https://nsf-gov-resources.nsf.gov/2023-06/NSF23104.pdf

National Science and Technology Council. (2022). Desirable Characteristics of Data Repositories for Federally Funded Research. The National Science and Technology Council. https://doi.org/10.5479/10088/113528

Oaiya, O. (2022). National open science policies in Africa. https://www.unesco.org/sites/default/files/medias/fichiers/2022/05/National_and_Institutional_OS_Policies.pdf

Office of Science and Technology Policy. (2013). Increasing Access to the Results of Federally Funded Scientific Research. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Office of Science and Technology Policy. (2022). Ensuring Free, Immediate, and Equitable Access to Federally Funded Research. https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

Office of Science and Technology Policy. (2023). FACT SHEET: Biden-Harris Administration Announces New Actions to Advance Open and Equitable Research. https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research/

Office of Science and Technology Policy. (2024, January 31). FACT SHEET: Biden-Harris Administration Marks the Anniversary of OSTP's Year of Open Science. The White House. https://www.whitehouse.gov/ostp/news-updates/2024/01/31/fact-sheet-biden-harris-administration-marks-the-anniversary-of-ostps-year-of-open-science/

Southern African Development Community (SADC) Secretariat. (2020). SADC Regional Indicative Strategic Development Plan (RISDP) 2020-2030. https://www.sadc.int/sites/default/files/2021-08/RISDP_2020-2030.pdf

Steinhart, G., & Skinner, K. (2024). The Cost and Price of Public Access to Research Data: A Synthesis. https://doi.org/10.5281/zenodo.10729547

Team NLnet Labs (2024). What I learned in Brussels: the Cyber Resilience Act. https://blog.nlnetlabs.nl/what-i-learned-in-brussels-the-cyber-resilience-act/

UNESCO Open Science Advisory Committee. (2021). UNESCO Recommendation on Open Science. UNESCO. https://www.unesco.org/en/open-science/about

U.S. Census Bureau. (2024). U.S. Census Bureau Plan for Providing Public Access to Results of Federally Funded Research—Draft for public comment. https://www2.census.gov/foia/ds_policies/PublicAccessPlan_Census_DraftforPublicComment.pdf

U.S. Department of Agriculture. (2023). Implementation Plan to Increase Public Access to USDA-Funded Research Results. https://www.nal.usda.gov/sites/default/files/page-files/USDA_Public_Access_Implementation_Plan_8_10_2023_0.pdf

U.S. Department of Energy. (2023). U.S. Department of Energy Public Access Plan. https://doi.org/10.11578/2023DOEPUBLICACCESSPLAN

U.S. Department of the Interior. (2023). Public Access to Results of Federally Funded Research at the U.S. Geological Survey: Scholarly Publications and Digital Data (ver. 2.0). https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/media/files/USGS-Updated%20PublicAccessPlan%20%28ver.2.0%2C%202023%29.pdf

The influence of procurement and information technology governance processes on the adoption of open infrastructure

Introduction and approach

In their role as builders, facilitators, and users of scholarly knowledge infrastructures, libraries and their host institutions are charged with providing cost-effective, sustainable, and mission-aligned services to their users (e.g. Goudarzi et al., 2021; ICOLC Strategies for Open Collaboration in Library Consortia Task Force, 2022; Lewis, 2017). Infrastructure solutions can be positioned along a gradient of openness and community accountability, ranging from fully open-source, community-supported and -governed applications and services to vendor-supported services that utilize open-source applications to fully proprietary commercial products. These applications and services may be sourced and developed locally (with local support and management within the institution), or procured from a vendor for a single organization or for a community (via consortia or other collaborative efforts).

Choosing among available technologies and services potentially engages an institution's procurement and/or information technology governance processes (see sidebar for definitions). In support of Invest in Open Infrastructure's mission to increase the adoption of open infrastructure, we sought to understand whether and how procurement and IT governance processes can help or hinder the adoption of open infrastructure solutions at research institutions, and to identify opportunities to ensure fair and equal consideration of open infrastructure options.

Decisions around software or service adoption can engage one or both of these frameworks, and we wanted to understand their influence on adoption decisions. Alignment across an organization as to whether "open" is a strategic priority is perhaps the critical key to ensuring it receives serious consideration, but we wanted to do a deeper dive into how these processes themselves influence adoption. Our process for investigating this question was threefold. First, we reviewed selected articles on open-source software adoption, primarily (but not exclusively) in academia. Next, we conducted 12 interviews with individuals from research institutions, consortia, and networks.[2] Interviewees were associated with libraries or library-aligned organizations, because libraries are one of several key players in supporting the research enterprise in academia. We speculated that because libraries have distinctive values and concerns and may operate at greater remove from central IT and procurement, they might be more likely to encounter friction in selecting and procuring technology. The interviewees represent a range of roles, chiefly at the director level or above, including Associate University Librarian, Assistant or Associate Vice President/Dean/Vice Chancellor, Directors and Heads of departments, Programme Managers, and IT staff. Interviewees represent the interests of institutions or consortia

IT governance and procurement defined

IT governance is an organizational process that helps align IT decisions with mission and needs, fosters communication across the organization, ensures buy-in for policy, budget, and project prioritization decisions, and integrates risk management into IT decision making (Carraway et al., 2017).[1]

Procurement refers to the policies and processes that guide the selection and acquisition of goods and services in order to maximize cost effectiveness and efficiency, and ensure alignment with institutional priorities and policies.

[1] IT governance has its origins in corporate governance, the importance of which was made clear following the corporate governance failures of the early 2000s and the passage in the United States of the Sarbanes-Oxley Act of 2002 (Bichsel & Feehan, 2014). Its uptake extends beyond the U.S., however. For example, Jisc engaged consultants to develop an IT governance framework for use in higher education (Coen & Kelly, 2007).

[2] Organizational affiliations of interviewees included: AfricaConnect, Carnegie Mellon University, Columbia University, Cornell University, Delft University of Technology (TU Delft), GALILEO, GÉANT, HELIOS Open, Massachusetts Institute of Technology, Ontario Council of University Libraries, Partnership for Academic Library Collaboration and Innovation, and University of Oregon. Please see Appendix B for a complete list.

located in the United States, Canada, the Netherlands, and Africa. Interviews were focused primarily on unit and campus-level enterprise IT decisions (rather than research IT), addressed descriptions of procurement and IT governance processes, decision-making around "build" versus "buy" solutions and what about these processes interviewees found helpful and/or challenging for the adoption of open infrastructure.[3] We also held a less structured conversation with members of the Higher Education Leadership Initiative for Open Scholarship (HELIOS) working group on Shared Open Scholarship Infrastructure (HELIOS Open, n.d.). Finally, we reviewed publicly accessible documentation on procurement and IT governance from the institutions represented in interviews. We noted and coded areas of concern for comparison across institutions.

Factors in the adoption of open-source software

The drivers of and impediments to the adoption of open infrastructure in general, and open-source software (OSS) in particular, are numerous and varied. Sanchez et al.'s 2020 systematic review of open source adoption considers three classes of factors that can influence adoption of OSS: technological, organizational, and economic.[4] Technological factors include security and reliability, data compatibility, documentation, customizability, portability, trialability, feature set, and user experience (Choi & Pruett, 2019; Petrov & Obwegeser, 2018; Sharma, 2022; Sánchez et al., 2020). The total cost of ownership is the chief economic barrier, and can encompass licensing costs, operational costs, and support costs (Sánchez et al., 2020). Organizational factors include availability of staff and staff expertise, lack of understanding or even active opposition to open source solutions in an organization's central IT unit and/or at the senior leadership level, risk aversion and lack of accountability for OSS, and brand recognition or "prestige" associated with proprietary solutions (Choi & Pruett, 2019; Linux Foundation, n.d.; OSS Watch, 2008b; Petrov & Obwegeser, 2018; Sánchez et al., 2020).

Some additional factors that argue more clearly for the adoption of OSS and open solutions include: customizability, lower costs of acquisition, interoperability and prioritizing the use of open standards and technologies, avoiding vendor lock-in, data portability, community accountability and transparency, digital sovereignty concerns, and a strategic intention to reduce reliance on proprietary software (e.g. Choi & Pruett, 2019; Günther, 2023; ICOLC Strategies for Open Collaboration in Library Consortia Task Force, 2022; Sánchez et al., 2020).

Any number of these factors might be taken into consideration in both procurement and IT governance processes. "Build versus buy" decisions bear early and critical importance, as a *buy* decision is likely to engage both IT governance and procurement, while *build* may engage only the IT governance function. Designed specifically to manage the exchange of money for goods and services, procurement processes can be especially problematic for OSS adoption, where no such transaction takes place unless a vendor-supported OSS option is under consideration (ICOLC Strategies for Open Collaboration in Library Consortia Task Force, 2022; OSS Watch, 2008b; Teal et al., 2020; Teperek & Dunning, 2020; Thompson, 2009). IT governance can also strongly influence the likelihood of OSS adoption, and depends heavily on organizational culture and strategic priorities. How well OSS is understood in the organization (what OSS is, and how it is developed, deployed and supported) often directly impacts an institution's interest in and willingness to implement OSS (ICOLC Strategies for Open Collaboration in Library Consortia Task Force, 2022; Linux Foundation, n.d.; OSS Watch, 2008b, 2008a).

[3] We include the complete interview guide in Appendix A.

[4] Our consideration of open infrastructures is intended to be broader than open source software specifically, but the findings of research into the uptake of OSS are still usefully applied here.

Results

Analysis of documentation

We present a summary of documented areas of concern for each institution or organization we interviewed in Table 1.[5] A consistent concern across all eight sets of documentation we reviewed is information security and compliance with applicable laws, including in the U.S. the Family Educational Rights and Privacy Act (FERPA), the Health Insurance Portability and Accountability Act (HIPAA), the Financial Services Modernization Act, as well as the General Data Protection Regulation (GDPR), which applies to any system or service with users located in the EU regardless of the location of the service. Issues that are commonly addressed under the terms of a service level agreement were also a top concern across the board. These may include, among other things, the responsibilities of the parties, incident management and response, and the availability of technical and end-user support services.

Technical specifications, intellectual property (retention of rights in content or code), web accessibility, support for integration with other systems and applications, and fit with overall IT strategy were the next most frequently mentioned concerns, and were flagged in five to seven of the documentation sets. Other topics that were mentioned less frequently (three to four times) include a clear exit strategy, value or cost-effectiveness, usability and user experience, and system scope (how widely used it is across an institution). Sustainability (resources required to implement a solution, total cost of ownership), governance and community accountability, and digital sovereignty concerns came up least frequently, never appearing in more than one document set among all the sets we examined.

Encouraging an organization's senior leadership to embrace open infrastructure both on its merits and as a strategic priority emerged as the single most important way to facilitate OSS adoption.

[5] No formal documentation was available for PALCI. GÉANT and AfricaConnect; instead of having their own independent policies or processes, they work within those of the institutions they support.

TABLE 1.
Summary of documented areas of concern by institution

Area of concern	Definition	Carnegie Mellon University	Columbia University	Cornell University	GALILEO	Massachusetts Institute of Technology	Ontario Council of University Libraries	Delft University of Technology	University of Oregon
Compliance: Data security	Data collection and storage, security of institutional data, privacy policy and practice, and alignment with applicable law. Examples of applicable law include the Family Educational Rights and Privacy Act (FERPA), the Health Insurance Portability and Accountability Act (HIPAA), the Financial Services Modernization Act, and General Data Protection Regulation (GDPR).	✓	✓	✓	✓	✓	✓	✓	✓
Compliance: System scope	Systems of record (systems that are the authoritative source of institutional data), and systems used widely across the institution are often subject to extra scrutiny.	✓		✓	✓		✓		
Digital sovereignty	Retaining control over a country, region, or population's data and/or technical infrastructure - for example, requiring that cloud services use infrastructure that is located in a client's country.						✓		
Exit strategy	The ability to opt to change to a different service or infrastructure. For example, the ability to export all data and metadata without limitations or cumbersome contractual obligations. Disposition of data upon termination of service.		✓		✓		✓		

Fit with IT strategy	Infrastructure characteristics are consistent with an organization's overall IT strategy. For example, an institution might have a cloud-first preference for infrastructure, or a preference for vended solutions over locally-managed ones.	✓		✓	✓		✓	✓	
Intellectual property	Retention of intellectual property rights in content or code.	✓	✓		✓	✓	✓	✓	
Service level agreement	Service level agreement attributes, for example: responsibilities of the parties, incident management and response, and availability of technical and user support services.	✓	✓	✓	✓	✓	✓	✓	✓
Sustainability	Overall effort and resources required to adopt and use the proposed solution; total cost of ownership.							✓	✓
Technical: Integration	Ease of integration with local and external systems, availability of API(s). For example: single sign-on and email.			✓	✓	✓	✓		✓
Technical: Specifications	Attributes of the system and degree to which local requirements are met.	✓	✓	✓	✓		✓	✓	✓
Governance and community accountability	Characteristics of the infrastructure and its community such as governance, open roadmap, and corporate social responsibility criteria.				✓			✓	
Usability and user experience	The user experience and usability of an application.			✓	✓			✓	✓
Value	Cost-effectiveness of the solution.	✓		✓	✓			✓	
Web accessibility	Compliance with web accessibility requirements.	✓		✓	✓		✓	✓	✓

Decision-makers may have a strong preference for open technologies, but "openness" takes many forms, and in and of itself, openness is generally not considered sufficient to justify a decision.

Analysis of interviews

Most interviewees described a technology selection process that centers identifying the problem to be solved and the desirable attributes of the solution, and selecting the one which best fits the organization's needs (via a solution's features and functionality) and circumstances (based on resources and capacity or affinity for vended versus locally developed or hosted solutions). Decision-makers may have a strong preference for open technologies, but "openness" takes many forms, and in and of itself, openness is generally not considered sufficient to justify a decision. That said, some attributes of open technologies or infrastructures — such as interoperability, open standards, and community participation and accountability — can factor into decision making.

Build versus buy

The build versus buy decision — the point at which an organization chooses to develop and/or support locally ("build") a particular technology or to contract with a vendor to provide a solution ("buy") — is a critical decision point in the selection process. Some of the organizations we engaged indicated they have the capacity to at least entertain the possibility of *build* solutions, but most mentioned that *buy* solutions are often more attractive, whether the technology under consideration is open or proprietary. Decision makers do not always have the budget flexibility to reallocate resources from staff lines to funds for service contracts, or vice versa. In addition, the skills and staff needed to support one approach or the other differ and may not be readily interchangeable. When an organization does have the freedom to choose a *build* solution, they are often frustrated by the difficulty in assessing the total cost of ownership — the resources required to install, configure, and sustain a particular technology solution over time. Attributes of *buy* solutions that make them attractive include documented up-front and ongoing costs, and having a party that can be held accountable when issues arise. Multiple interviewees noted, however, that there is a distinctive skill set that is required for managing the proposal process and implementation for a *buy* solution. This includes the ability to identify and articulate detailed technical requirements, synthesize and evaluate the proposals that are received, negotiate contracts, and manage vendor relationships.

Navigating IT governance and procurement processes

Build solutions are likely to engage IT governance or similar review, while *buy* solutions are likely to engage both procurement and IT governance reviews. In *both* IT governance and procurement, risk mitigation and compliance concerns are paramount, particularly around information security. In other areas of concern, based on interviewees' descriptions, procurement processes seem to tend towards greater formality and strictness than IT governance processes. Where IT governance is relatively flexible, there can be fewer roadblocks to adopting a particular solution, but also less rigorous scrutiny in important areas.

Interviewees often described procurement evaluation criteria that are biased toward vended and proprietary solutions and/or unsuited for assessing open solutions. These criteria clearly relate to another concern articulated in multiple interviews about a lack of understanding in IT and business units of how OSS communities function and the advantages that open source may confer. Open solutions may offer opportunities for the community to influence a product roadmap and contribute to subsequent development, but interviewees reported that there may be no way to indicate this possibility and how such activities can ensure a solution will meet (in the future, if not the present) specified requirements in an assessment rubric. Other issues included concerns over the possible acquisition of technologies and services by larger commercial entities, low spending limits that require review of even small purchases, and processes that move too slowly to take advantage of one-time funding such as year-end account closeout.

Encouraging an organization's senior leadership to embrace open infrastructure both on its merits and as a strategic priority emerged as the single most important way to facilitate OSS adoption. Such shared understanding and articulation of priorities alleviates the need for continual education around the benefits of open, and allows decision-makers to focus on the functional characteristics of technologies under consideration. Interestingly, we heard in one interview that while the individual's organization does prioritize openness for research outputs, it does not assert the same priority for open research infrastructure. This apparent disconnect might represent an advocacy opportunity, or a need to consider the institutional priorities and their alignment with the goals of open research. This is an important point, and one we regret not exploring in all of the interviews.

More prosaic ways to successfully negotiate these processes include using approved vendors if available (who can typically be engaged with less administrative overhead), offering a sole source justification whenever possible (eliminating the need for a competitive process), and providing various types of documentation to expedite review. Most organizations reported that documentation such as a Voluntary Product Accessibility Template (VPAT) or Accessibility Compliance Report (ACR) for demonstrating compliance with web accessibility standards, or a completed Higher Education Community Vendor Assessment Toolkit (HECVAT) can expedite institutional review. VPATs in particular came up as valuable to these evaluation processes, but none of the organizations said any of these documents are an absolute requirement for consideration.

The role of vended solutions

Some interviewees noted a trend in their own organizations away from *build* solutions, and an emerging preference for vendor-supported ones. Vended solutions can leverage open-source or proprietary technologies, but the distinctions go further. Some open-source communities have approved service provider programmes, often entailing financial or in-kind support for the community. Some may provide hosting services themselves, and some service providers operate completely independently from the open-source community. Noteworthy here is TU Delft's[6] experience navigating a selection process for a data repository, a process that resulted in the adoption of a commercial solution and the expression of frustration by community-based providers of open solutions who argued that Delft's tender process favoured commercial entities with the capacity to participate (Teal et al., 2020; Teperek & Dunning, 2020). Given a level playing field and a full range of choices, however, vended OSS-based options can be attractive to institutions wishing to leverage the advantages of open infrastructure even when they do not have the capacity to operate that infrastructure locally. This is consistent with what we heard in the course of our previous research (Goudarzi et al., 2021): that institutions with a history of locally developed or supported infrastructure, which has the advantage of greater customization and local integration possibilities, were starting to experience a negative impact on the sustainability of locally run and customized solutions.

Consortia and networks: Opportunities for adoption at scale

We want to highlight the unique role that alliances such as consortia and research or education networks can play in supporting the infrastructure needs of their members.[7] While one network we spoke with described the challenges of having to navigate and satisfy the diverse requirements of each and every member, we heard from others that these alliances can be a powerful means of fostering the adoption of open infrastructure at multiple institutions at once. When organizations have granted an alliance the authority to negotiate and procure services on their behalf, this can smooth the way for adoption. Even when a few participating organizations impose additional conditions or requirements on the process, the alliance usually has sufficient latitude to handle the process effectively. As a shared resource, these alliances can provide some of the skills and expertise noted earlier, such as contract negotiation and managing relationships with vendors. We heard in our interviews that vended solutions can be attractive for consortia, allowing for efficient and centralized management for a distributed community of users. Finally, moving forward as part of a larger group can alleviate concerns about being early adopters or adopting the "wrong" solution.

[6] We have avoided attributing specific comments to specific institutions throughout this report, but in this case, the details of the story are published.

[7] We offer some specific examples of successful shared adoption efforts in the section of this report, "Trends in open infrastructure performance and adoption."

Areas of opportunity

What can be done to ensure institutions are able to give serious consideration to open options? Based on what we heard from decision-makers regarding how IT governance and procurement at their institutions influence the adoption of open infrastructure, we offer the following recommendations, organized by constituent group.

For organizations seeking to adopt open infrastructure:

- Educate organizational stakeholders, particularly at the senior leadership level, on the benefits open infrastructure can confer and work to influence strategic priorities in this direction.[8]
- Encourage review and modification of procurement procedures to identify and address biases against open infrastructure or towards proprietary solutions, and to include criteria that can aid in selecting among them.
- Adapt selection criteria to allow for the distinctive characteristics of open infrastructure solutions, including possible input to product roadmaps and feature development.
- Explore and develop shared approaches to and support for open infrastructure via consortia, networks, or other kinds of alliances. Organizations with ample technical capacity might consider developing the skills to lead and support shared efforts, while those with fewer technology resources and staff may opt to participate principally as a client.

For open infrastructure communities:

- Complete and be prepared to share (or even make public) commonly requested assessment documents that apply to the infrastructure. The single most important example of this is a VPAT or ACR to demonstrate compliance with web accessibility guidelines.
- Consider completing (and making available the results of) a voluntary self-assessment using an established values framework such as the Principles of Open Scholarly Infrastructure (POSI, Bilder et al. 2020) or the FOREST Framework for Values-Driven Scholarly Communication (Lippincott & Skinner, 2022).
- Foster trustworthy service provider programmes to support a variety of vended options for open infrastructure.

For consortia, community organizations and professional associations in the space:

- Continue to identify and capitalize on opportunities to support and adopt open infrastructure at scale.
- Provide educational materials and guidance documentation to support decision making around technology adoption.[9]
- Provide professional development opportunities to skill up staff to effectively manage negotiations and operating relationships with vendors.
- Provide community fora for exploring and developing frameworks to understand the total cost of ownership for different solutions, and share resources such as contract templates.

We want to highlight the unique role that alliances such as consortia and research or education networks can play in supporting the infrastructure needs of their members.

[8] The Higher Education Leadership Initiative for Open Scholarship (HELIOS Open) community in higher education is working towards this objective.

[9] HELIOS Open's (HELIOS Open, 2023) excellent guide offers an excellent such example.

Conclusion

We have summarized here what we heard from decision makers at research institutions, and identified some possible courses of action to ensure greater consideration of open infrastructures in technology and service selection. In the course of our work, we found that:

- If not deliberately crafted to fit the characteristics of OSS, procurement and IT governance processes can be a poor fit for open infrastructure options, potentially excluding some open and community-supported options from consideration.
- Buy-in at the senior leadership level is critical.
- Vended solutions can provide viable options for institutions lacking the resources to support infrastructure in-house, and/or smooth the way for open options to be considered in the IT governance and procurement processes.
- Consortia and other alliances can both mitigate some of the complexity of navigating these processes and reduce the perception of risk among participants.

Finally, we offer some possible courses of action for adopting organizations, consortia, and open infrastructure communities.

References

Bichsel, J., & Feehan, P. (2014). Getting Your Ducks in a Row: IT Governance, Risk, and Compliance Programs in Higher Education. EDUCAUSE. https://library.educause.edu/~/media/files/library/2014/3/ers1402.pdf

Bilder, G., Lin, J., & Neylon, C. (2020). The Principles of Open Scholarly Infrastructure (v1.1, 2023). The Principles of Open Scholarly Infrastructure. https://doi.org/10.24343/C34W2H

Carraway, D., Arruda, K., Hagins, J., Martinez, B., Molina, H., Perry, J., Stingel, R., & Will, C. (2017). Higher Education IT Governance Checklist. EDUCAUSE. https://library.educause.edu/resources/2017/3/higher-education-it-governance-checklist

Choi, N., & Pruett, J. A. (2019). The context and state of open source software adoption in US academic libraries. Library Hi Tech, 37(4), 641–659. https://doi.org/10.1108/LHT-02-2019-0042

Coen, M., & Kelly, U. (2007). Information management and governance in UK higher education institutions: Bringing IT in from the cold. Perspectives: Policy and Practice in Higher Education, 11(1), 7–11. https://doi.org/10.1080/13603100601127915

Goudarzi, S., Pugh, K., Rhinesmith, V., Staines, H., & Thaney, K. (2021). Designing a Preparedness Model for the Future of Open Scholarship. Invest in Open Infrastructure. https://doi.org/10.5281/zenodo.5218968

Günther, C. (2023, November 16). Governments turn to Open Source for sovereignty. OpenSource.Net. https://opensource.net/governments-adopt-open-source-sovereignty/

HELIOS Open. (n.d.). HELIOS Open Workstreams & Outputs. Higher Education Leadership Initiative for Open Scholarship. Retrieved January 24, 2024, from https://www.heliosopen.org/workinggroups

HELIOS Open. (2023). Scholarly Communication Infrastructure Guide: Buy, Build, or Partner. https://www.heliosopen.org/s/HELIOS-Shared-Infrastructure-Decision-Making-Guide-3.pdf

ICOLC Strategies for Open Collaboration in Library Consortia Task Force. (2022). Strategies for Collaboration: Opportunities and Challenges to Build the Future We Need. https://icolc.net/system/files/Strategies%20for%20Collaboration%20-%20Opportunities%20and%20Challenges%20to%20Build%20the%20Future%20We%20Need.pdf

Lewis, D. W. (2017). The 2.5% Commitment. https://hdl.handle.net/1805/14063

Linux Foundation. (n.d.). Creating an Open Source Program. Retrieved October 19, 2023, from https://www.linuxfoundation.org/resources/open-source-guides/creating-an-open-source-program

Lippincott, S., & Skinner, K. (2022). FOREST Framework for Values-Driven Scholarly Communication. https://doi.org/10.5281/zenodo.6557301

OSS Watch. (2008a). Creating a "level playing field" for open source software options in IT selection and procurement.

OSS Watch. (2008b). Decision factors for open source software procurement. http://oss-watch.ac.uk/resources/procurement-infopack

Petrov, D., & Obwegeser, N. (2018). Barriers to Open-Source Software Adoption: Review and Synthesis. International Conference on Information Systems Development (ISD). https://aisel.aisnet.org/isd2014/proceedings2018/ISDMethodologies/3

Sánchez, V. R., Ayuso, P. N., Galindo, J. A., & Benavides, D. (2020). Open source adoption factors—a systematic literature review. IEEE Access, 8, 94594-94609. https://doi.org/10.1109/ACCESS.2020.2993248

Sharma, C. (2022). Tragedy of the Digital Commons. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4245266

Teal, T., Lowenberg, D., Smith, T., Gonzales, J. B., Nielsen, L. H., & Iannodis, A. (2020, August 26). Sustainable, Open Source Alternatives Exist. Dryad News. https://blog.datadryad.org/2020/08/26/sustainable-infrastructure-exists/

Teperek, M., & Dunning, A. (2020, August 18). Why figshare? Choosing a new technical infrastructure for 4TU.ResearchData. Open Working. https://openworking.wordpress.com/2020/08/18/why-figshare-choosing-a-new-technical-infrastructure-for-4tu-researchdata/

Thompson, M. (2009). Open Source, Open Standards: Reforming IT procurement in Government. https://markthompson1.files.wordpress.com/2012/02/open-source-and-open-standards-osborne-report1.pdf

Appendix A: Interview guide

For this interview, we are defining open infrastructure (OI) as open source software (OSS), but our working definition is broader: some combination of open source, free to use, community governed, transparent in operations, or operated by a non-profit or non-commercial entity. Feel free to keep in mind the definition that best fits your context.

1. If your organization were considering adopting OI, please describe the process(es) you would have to engage in order to do so *(additional questions as needed)*
   - **a.** What is your role in these processes? *NOTE: if short on time, might skip to question 2 and backtrack later if possible.*
   - **b.** Who else has a voice in these processes, and what types of roles do they play?
   - **c.** How do you decide whether you are managing something locally versus working with a vendor?
   - **d.** How do the processes differ if...
     - **i.** You are planning to work with a vendor
     - **i.** You are planning to host or implement the OI locally
   - **e.** Is there any supporting documentation I can see - policies, checklists, or forms, for example?
   - **f.** Are there standard documents - Voluntary Product Accessibility Template (VPAT), Accessibility Compliance Report (ACR), Software Bill of Materials (SBOM), Higher Education Community Vendor Assessment Toolkit (HECVAT), certifications (e.g. Open Source Security Foundation best practices badge), or other attributes (Internet2 service) that are
     - **i.** Required to get approval?
     - **ii.** Not required but very useful to have, to get approval?
   - **g.** Who ultimately approves (or denies) the request?
   - **h.** What happens if the request is denied?
2. What works well about these processes?
3. What is challenging?
4. If you could wave a magic wand, what procedural changes would you make that would make it easier to adopt OI?
5. Who else should we speak to, or what work should I be aware of, as we work to better understand this topic?

Appendix B: List of interviewees by organization

IOI is grateful to the following individuals for sharing their time and perspectives.

AfricaConnect: Leïla Dekkar, International Relations Project Manager - AfricaConnect3 & GÉANT

Carnegie Mellon University: Sayeed Choudhury, Associate Dean for Digital Infrastructure and Director of Open Source Programs Office

Columbia University: Rob Cartolano, Associate Vice President for Technology and Preservation, Columbia University Libraries

Cornell University: Phil Robertson, Director of Software Development and Simeon Warner, Associate University Librarian, Cornell University Library

Delft University of Technology (TU Delft): Alistair Dunning, Head, Research Services, TU Delft Library

GALILEO: Lucy Harrison, Assistant Vice Chancellor for Academic Library Services and Executive Director of GALILEO, University System of Georgia

GÉANT: Nicky Wako, Advocacy Manager

HELIOS Open's Shared Open Scholarship Infrastructure working group:

- Caitlin Carter, HELIOS Program Manager
- Alicia Salaz, Vice Provost and University Librarian, University of Oregon
- Robert Hilliker, Associate Provost for Libraries, Rowan University
- Julieta Arancio, Open Accelerator Fellow, Open Research Funders Group
- Torsten Reimer, University Librarian and Dean of the University Library, University of Chicago

Massachusetts Institute of Technology (MIT): Carl Jones, Digital Repository Services Engineer, MIT Libraries

Ontario Council of University Libraries: Kate Davis, Director of Scholars Portal

Partnership for Academic Library Collaboration and Innovation: Jill Morris, Executive Director

University of Oregon (UO): Franny Gaede, Director, Department of Open Research, UO Libraries

Future signals

In the final section of the State of Open Infrastructure report, we highlight a few emerging topics and trends that we've seen increasingly discussed in the open infrastructure and adjacent spaces over the past year. Inspired by Nesta's Future Signals (2023), our team pondered recent developments and key tensions and their impact on the investment in and adoption of open infrastructure in research and scholarship. By sharing our thoughts, we hope to hold space for further discussions with the broader community.

The impact of AI on infrastructure in research and scholarship

With the rapid advancement in artificial intelligence (AI) models, algorithms, and tools and their increasing prevalence in our day-to-day lives, it is unsurprising that AI has come up in almost every conference and community conversation in the open science, open source, and open infrastructure spaces last year. AI is redefining digital research and scholarly communication infrastructure and funders', institutional decision makers', and researchers' technological investment and adoption practices.

Redefining infrastructure for research and scholarship

Today, AI tools can readily execute or assist in many tasks that humans currently perform in the research production and communication lifecycle. They can create and enhance metadata, classify images, translate between languages and schemas, conduct data analysis, improve peer review, and more (Watkins, 2023). Similarly, AI algorithms are increasingly used by researchers and scholars in their research and academic workflows, e.g. automated transcription services to transcribe research interviews, or an image segmentation tool that is dependent on a machine learning algorithm. Inevitably, AI — the models, the training datasets, and even the hardware it depends on — is becoming part of research and scholarship infrastructure.

AI as an enabler for better open research infrastructure

Automation creates an opportunity for people, including those who are part of the human infrastructure in scholarly communication and research, to reduce effort on tedious tasks and to focus their labour on the aspects of their roles best suited to humans, exercising their creativity, social-emotional intelligence, and complex problem-solving skills.

There is a significant risk of funders and decision-makers promoting AI as a cost-reduction strategy, rather than as a way to achieve more with the same (or increased) resources and to empower human workers. While we acknowledge this risk, thinking of AI as an enhancement to rather than a replacement for human labour can expand the potential of what open research infrastructures can achieve. Take open data-sharing platforms as an example. If a curation team leverages AI to perform a series of quality checks and metadata enhancements during or immediately after the data submission process, they not only reduce human effort on tedious or repetitive tasks, but free up time to further improve data reusability: performing deeper quality checks than time might otherwise allow, providing support for authors, or developing creative approaches to promoting data reuse.

Automation creates an opportunity for people (...) to focus their labour on the aspects of their roles best suited to humans, exercising their creativity, social-emotional intelligence, and complex problem-solving skills.

Thinking of AI as an enhancement to rather than a replacement for human labour can expand the potential of what open research infrastructures can achieve.

Centering principles and values in infrastructure decision-making

The potential applications of AI in research production and communication and research stakeholders' interest in applying emerging AI tools also present novel challenges for research-performing organizations. A May 2023 UNESCO survey showed that less than 10% of schools and universities had developed institutional policies or guidance on the use of generative AI applications in education (UNESCO, 2023). While many institutions have unveiled policies since then (and EDUCAUSE has begun collecting information on its members' AI policies — see EDUCAUSE, n.d.), it remains to be seen whether policy development to encourage the responsible use of AI in research will keep pace with the proliferation and evolution of AI tools. In particular, the black-box nature of many of these tools makes it difficult, if not impossible, to validate generative AI outputs. We will be watching with interest to see how institutions, research funders, publishers, and research infrastructure providers continue to adapt to this rapidly evolving landscape.

With AI becoming part of open research infrastructure, for everyone who cares about making informed decisions about the technologies we use, including us at IOI, the important question is how can we understand the implications of our technology choices in order to make better ones? As we look across the open infrastructure and ethical AI conversations, we are excited to see the parallels that exist: the considerations around transparency, accountability, and governance, the discussions around principles and values (Decklemann, 2023), and the recognition that openness is a spectrum (Ramlochan, 2023). As AI becomes part of research infrastructure, we see opportunities to use our experience working with infrastructure service providers and adopters to surface information to help adopters make more informed decisions. Of course, the "supply chain" here makes this complicated: how does the governance model of the training data that algorithm A depends on, that is used by infrastructure B, that is depended on by infrastructure C, affect the community using infrastructure C? We have no answers at the moment, but we are encouraged by fellow travellers on this journey.

The infrastructure powering diamond open access

Diamond Open Access (Diamond OA) is, in many ways, a new name for an old phenomenon — completely free online journals have been in existence in some form since at least the late 1980s. In recent years, there has been a heightened focus and urgency (as well as the creation of the label "Diamond") to "no-pay" forms of publishing in reaction to the proliferation of the Article Processing Charge (APC) business model and the associated discussions of equity and disparities in publishing (see, for example, the 2023 reports and ongoing work from the Open Access Scholarly Publishing Association).[1]

This resurgence of interest has prompted recommendations of business models and sustainability for the publishing operations (OASPA, 2021), as well global events such as the Diamond Open Access Conference which held its second gathering in October 2023, co-organized by Redalyc, UAEMéx, AmeliCA, UNESCO, CLACSO, UÓR, ANR, cOAlition S, OPERAS, and Science Europe (Saenen et al., 2024).

Moving away from APCs

Funders have been reconsidering and evolving their approaches to open access publishing and APCs. Diamond OA is at the heart of large-scale projects such as DIAMAS in Europe, a three-year, European Commission funded effort to bring together 23 organisations across 12 countries to investigate models to advance non-profit publishing models that do not charge authors or readers as a counterpoint to APC-based OA.[2] Efforts such as cOAlition S and Plan S,[3] a consortium of research funding and performing organizations dedicated to making full and immediate open access a priority, have also converged on Diamond OA and "publish, review, curate" models of publishings as means to address current challenges in scientific publishing. The Bill and Melinda Gates Foundation recently announced that they will cease supporting APCs in 2025 (Bill & Melinda Gates Foundation, 2024). With APC inflation and proliferation being driven by funders' willingness to pay for them as part of their mission to make the outputs of their funded research available to the public, we see these movements shifting back to the original declarations of the OA movement in a much more mature ecosystem for open access publishing. The backlash against APCs may swing the funding pendulum back towards Diamond OA and the infrastructures that power it; alternatively, it may result in the further development of other ecosystems, such as preprint and peer review.

Diamond OA is described as the equitable means to making knowledge a public good (Manifesto on Science as Global Public Good: Noncommercial Open Access, 2023), calling back to the original declarations on Open Access from the early 2000s (e.g. the Budapest Open Access Initiative[4]) and recognizing that some profit-driven motives have shifted open access publishing away from the original intent of the Open Access movement. The future signal of Diamond OA is that we are revisiting the roots of the OA movement and reckoning with the reality that the current open access publishing landscape, rife with rapidly inflating and proliferating APCs, is very far from the free access ideal envisioned in the early years of the OA movement. We are keenly interested and monitoring efforts for more equitable models with investments in infrastructures as well as incentives to drive more concerted shifts towards these models.

[1] https://www.oaspa.org/news/equity-in-oa/

[2] https://diamasproject.eu/

[3] https://www.coalition-s.org/

[4] https://www.budapestopenaccessinitiative.org/

Diamond OA risks replicating this investment in new infrastructures at the expense of those that have powered these activities for decades. Investing instead in the plumbing bolsters the infrastructure for the global research community to fully participate in the Diamond OA movement.

The risks of reinventing "plumbing" infrastructure

The existing ecosystem of publishing infrastructures, human and digital, are often dependent on other infrastructures (e.g. open standards, metadata infrastructure, and collective care infrastructures). This set of behind the scenes infrastructures has been named the "plumbing" infrastructure that powers open science (Pfeiffer et al., 2024). We see an important opportunity to fortify the plumbing of the interdependent open infrastructures that already power Diamond OA in order to build a robust open toolchain that can realistically compete with proprietary publishing workflows. Commercial publishers benefit from mature (usually closed) workflow systems that enable the full spectrum of publishing activity, reliably delivering content from author to reader. Shiny new and visible tools tend to attract funding (Skinner, 2019), leaving existing open infrastructures with the challenge of growing, evolving, and investing in research and development without large infusions of funding. As a buzzy topic, Diamond OA risks replicating this investment in new infrastructures at the expense of those that have powered these activities for decades. Investing instead in the plumbing — existing interdependent infrastructures that power the full lifecycle of Diamond OA (from standards and identifiers, to repository infrastructures, access protocols, discovery, and beyond) — bolsters the infrastructure for the global research community to fully participate in the Diamond OA movement.

In order to sustain Diamond OA models and operations, it is critical to invest in the human infrastructure and labour for them in a sustainable manner by making space for them in the regular course of scholarly labour.

The importance of sustainable investment in human infrastructure and labour

Renewed investment and commitment to developing these support structures for open infrastructures is a key pathway that we see for the Diamond OA ecosystem to ensure a robust technical ecosystem into the future. Investing in Diamond OA is not only about investing in digital infrastructure, but equally (and if not more) the human infrastructure powering it. If the needs of editors and staff who operate a journal are not taken into consideration, the effects can be dire, as seen in the continued rise in journal editorial boards resigning from subscription model journals. Recent examples of this mass resignation have circled around pressures to produce higher numbers of articles at a lower quality and with lowered standards in publication services (Abels & Flynn, 2024; Oransky, 2024; Weinberg, 2023). In that regard, It is important to call out the realities that have shifted since the first online open access journals were published in the 1980s: editing and reviewing used to be a regular part of an academic's work, but increasingly, especially in the US, there has been increased pressure to publish a higher quantity of articles to achieve "success" in academia and less focus on services like editing or reviewing. This is coupled with the ongoing increase in faculty workloads at higher education institutions in response to institutional budget cuts, which provides further constraints on scholars' time and energy (Marcus, 2021). In order to sustain Diamond OA models and operations, it is critical to invest in the human infrastructure and labour for them in a sustainable manner by making space for them in the regular course of scholarly labour. An increase in investment in the technical infrastructure cannot be successful without a complementary increase in investment in the people who use the tools.

With a re-energized focus on Diamond OA, we see a path towards shoring up the infrastructures that have long powered this method of publishing and creating a viable open access publishing ecosystem that heeds the call to return to the roots of open access.

Diverse visions for digital sovereignty and the impact on open knowledge infrastructure

Digital sovereignty can be defined as the right of a nation, region, or other political entity to assert control over its digital infrastructure and data, on its own behalf and on behalf of its citizens (Pohle and Thiel, 2020). From the development of the European strategy for data to the CHIPS and Science Act in the US, recent years saw a growing, evolving conversation and rapid policy development around digital sovereignty, prompted by nations' and regions' desires to (re)gain control over and independence in their digital infrastructure and data.

Distinct motivations and visions for digital sovereignty

We provided a working definition for digital sovereignty for this section, but we would be remiss if we did not acknowledge the diversity of approaches to defining and motivations for embracing digital sovereignty. The concept has been embraced by authoritarian governments in order to exert social control over their citizens, and by more democratic regimes to protect citizens' personal data and businesses' economic interests in an information economy. In Latin America, the loss of trust in how big data corporations and developed countries utilize data is one of the root causes of the drive towards digital sovereignty (Bosoer, 2022). The enclosure of the digital commons and its integration into infrastructure developed, maintained, and licensed or sold by for-profit entities which are subject to regulations applied to their home jurisdictions and those of their users makes geography more relevant than early internet visionaries might have anticipated. Taken to its extreme, this phenomenon has the potential to fall victim to the "Galápagos Syndrome" (2024), by which infrastructures become separated and segregated into smaller, detached and non-interoperable components.

We wonder what impact the geographic "anchoring" effect of the digital sovereignty movement will have on the free flow of knowledge and research around the world, and the infrastructures that support it.

Impact on global digital research infrastructure

We wonder what impact the geographic "anchoring" effect of the digital sovereignty movement will have on the free flow of knowledge and research around the world, and the infrastructures that support it. As we think about developing and maintaining sustainable, interoperable, and global digital research infrastructure, digital sovereignty can impact how these technologies are deployed locally. For example, cloud-based technologies need to work on local servers and cloud infrastructures in order for them to be used by researchers in a certain country. Digital sovereignty regulations may also have an increasing influence on institutional procurement policies (IT-Planungsrat, 2021), affecting what technologies researchers and scholars can use. Thinking about solutions that are going to be most successful because of their abilities to aggregate from data sources worldwide, increased regulations and restrictions on how the data can be shared, motivated by digital sovereignty, may lead to additional challenges and barriers in these solutions' development.

Open-source technologies as enablers of digital sovereignty

Open-source technologies are increasingly seen as enablers of digital sovereignty as they can "cultivate trust through openness, direct involvement, and preserving entities' autonomy" (Nordhaug & Harris, 2021). In some regions and countries, there is increasing governmental support for and investment in the development of open-source technologies to advance digital sovereignty of the region/country. The Sovereign Tech Fund, supported by the German Federal Ministry for Economic Affairs and Climate Action, for example, "supports the development, improvement, and maintenance of open digital infrastructure" to ensure that it is "available, accessible and secure [...] for digitalization in the public interest."[5] We also observe an increasing number of regional and national Open Source Programme Offices, notably in Europe (Osborne et al., 2023). On the other hand, in the Majority World, where governments and the private sector simply do not have the capital or technical expertise to invest at scale for needed digital public infrastructures, we see a potential for open-source technologies to offer a viable, cost-effective alternative to the technological infrastructure provided by big tech and/or powerful, well-resourced nations.[6] All this can increase global investment in building local capacity to develop and maintain open infrastructure.

As developments surrounding digital sovereignty continue, we see a need to revisit the (also highly contextual) motivations for the pursuit of open digital research infrastructure. If enabling open and equitable access to and participation in research and scholarship is indeed the goal and motivation for the creation and deployment of this infrastructure, then we recognize that significant investments are needed in building trust in global open infrastructures. If independence and sovereignty regarding data and digital infrastructure are the goals and motivation, then significant efforts are needed to foster collaboration and cooperation in developing connections and networks using open infrastructure that can enable nations to work together while retaining their own domains of control.

[5] https://www.sovereigntechfund.de/

[6] We have already seen accelerated investment by big tech giants in Latin America, Asia, and Africa as they strive to expand their user base (Pinto, 2018; Komminoth, 2023).

References

Abels, K. & Flynn, S. (2024, March 9). Resignation of co-editors of Syntax. LINGUIST List. https://linguistlist.org/issues/35/835/

Bosoer, L. (2022). Digital Sovereignty: Voices from Latin America. https://blogs.eui.eu/latin-american-working-group/digital-sovereignty-voices-from-latin-america/

Bill & Melinda Gates Foundation. (2024). Policy Refresh 2025 Overview. https://gatesfoundationoa.zendesk.com/hc/en-us/articles/24810787662100-Policy-Refresh-2025-Overview

Decklemann, S. (2023). Wikipedia's value in the age of generative AI. https://medium.com/freely-sharing-the-sum-of-all-knowledge/wikipedias-value-in-the-age-of-generative-ai-b19fec06bbee

EDUCAUSE. (n.d.). Artificial Intelligence (AI). https://library.educause.edu/topics/infrastructure-and-research-technologies/artificial-intelligence-ai

European Commission. (n.d.). A European strategy for data. https://digital-strategy.ec.europa.eu/en/policies/strategy-data

Galápagos Syndrome. (2024, March 31). In Wikipedia. https://en.wikipedia.org/wiki/Gal%C3%A1pagos_syndrome

IT-Planungsrat. (2021). Strategy for strengthening the digital sovereignty of public administration IT: Strategic objectives, approaches and implementation measures. FITKO (Federale IT Cooperation). https://www.it-planungsrat.de/fileadmin/it-planungsrat/foederale-zusammenarbeit/Gremien/AG_Cloud/20210104_Strategy_for_strengthening_the_digital_sovereignty_of_public_administration_IT_1.0.pdf

Komminoth, L. (2023, July 21). "Can Africa Achieve 'Digital Sovereignty' in an Era of Big Tech?" African Business. https://african.business/2023/07/technology-information/can-africa-achieve-digital-sovereignty-in-an-era-of-big-tech.

Manifesto on Science as Global Public Good: Noncommercial Open Access. (2023, October 27). Cumbre Global Sobre Accesso Abierto Diamante. Retrieved April 11, 2024, from https://globaldiamantoa.org/manifiesto/#/

Marcus, J. (2021, April 30). Some universities' response to budget woes: Make faculty teach more courses. The Washington Post. https://www.washingtonpost.com/local/education/college-faculty-course-workload-budgets/2021/04/30/d5d2ee1e-a904-11eb-8c1a-56f0cb4ff3b5_story.html

Nesta. (2023). Future Signals — what we're watching for in 2023. https://www.nesta.org.uk/feature/future-signals-2023/

Nordhaug, L. M. & Harris, L. (2021). 26. Digital public goods: Enablers of digital sovereignty. In OECD, Development Co-operation Report 2021: Shaping a Just Digital Transformation, OECD Publishing, Paris, https://doi.org/10.1787/ce08832f-en

OASPA. (2021). Major OA Diamond Journals Study completed: Report emphasizes diversity and sustainable pathways for diamond Open Access. https://www.oaspa.org/news/major-oa-diamond-journals-study-completed-report-emphasizes-diversity-and-sustainable-pathways-for-diamond-open-access/

Oransky, I. (2024, February 12). Econ journal board quits en masse because Wiley 'appeared to emphasize quantity over quality'. Retraction Watch. https://retractionwatch.com/2024/02/12/econ-journal-board-quits-en-masse-because-wiley-appeared-to-emphasize-quantity-over-quality/

Osborne, C., Boehm, M., & Jiménez Santamaría, A. (2023). The European Public Sector Open Source Opportunity: Challenges and Recommendations for Europe's Open Source Future. https://www.linuxfoundation.org/hubfs/LF%20Research/European%20Public%20Sector%20Open%20Source%20Opportunity%20-%20Report.pdf?hsLang=en

Pfeiffer, N., French, A. L., Vines, T., Gibson, J., & Chandramouli, I. (2024, March 21). Open Science Plumbing: Infrastructure Enabling and Catalyzing Policy Implementation. Retrieved from osf.io/v9hy2

Pinto, R. A. (2018, July 16). "Digital Sovereignty or Digital Colonialism?" Sur - International Journal on Human Rights 27. https://sur.conectas.org/en/digital-sovereignty-or-digital-colonialism/.

Pohle, J., & Thiel, T. (2020). Digital sovereignty. Internet Policy Review, 9(4). https://doi.org/10.14763/2020.4.1532

Ramlochan, SI. (2023). Openness in Language Models: Open Source vs Open Weights vs Restricted Weights. https://promptengineering.org/llm-open-source-vs-open-weights-vs-restricted-weights/

Saenen, B., Ancion, Z., Borrell-Damián, L., Mounier, P., Oliva Uribe, D., Papp-Le Roy, N., & Rooryck, J. (2024, February 20). 2nd Diamond Access Conference Report. 2nd Diamond Open Access Conference, Toluca, México. https://doi.org/10.5281/zenodo.10684544

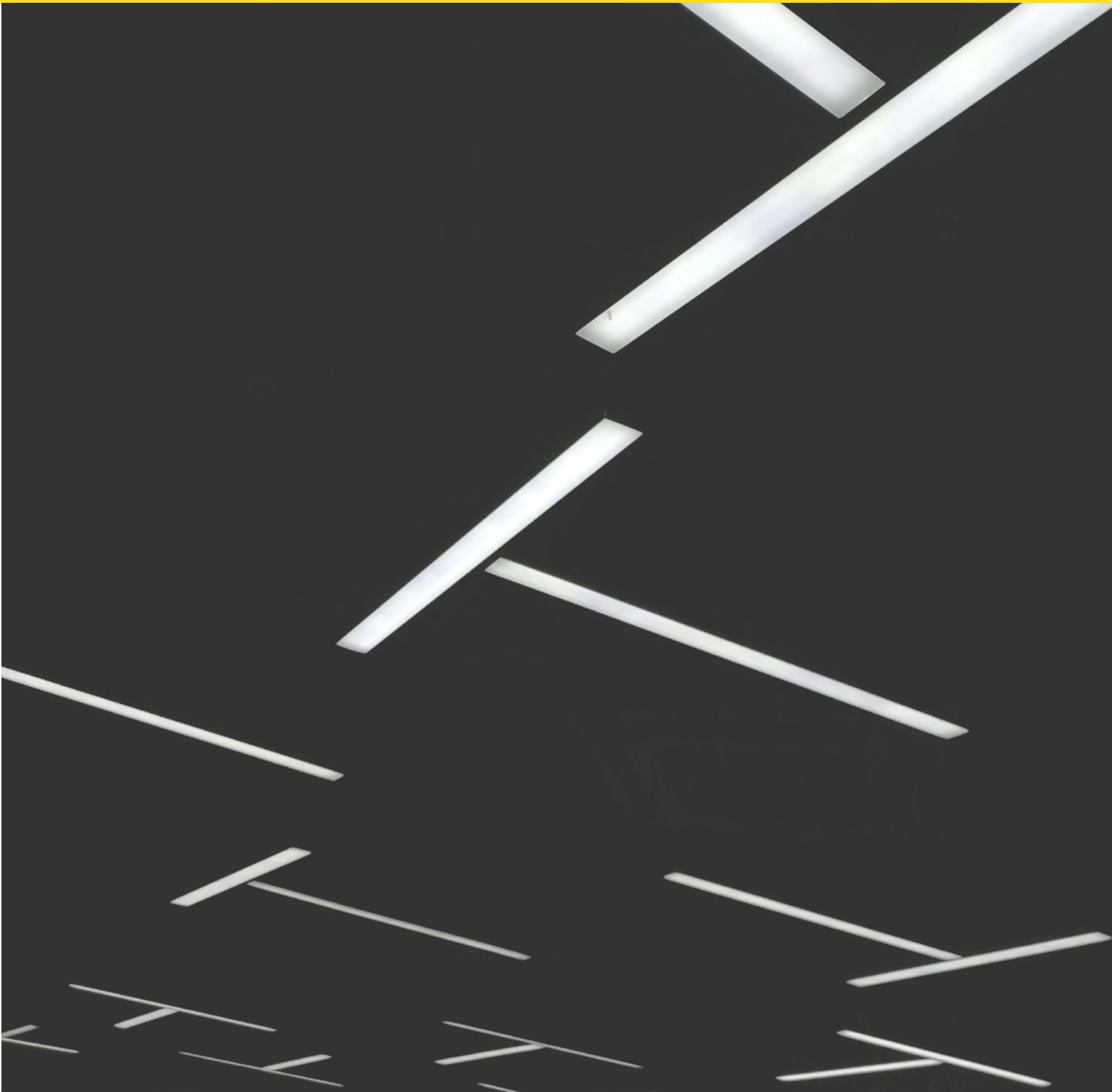
Skinner, K. (2019, July 23). Why Are So Many Scholarly Communication Infrastructure Providers Running a Red Queen's Race? https://educopia.org/red-queens-race/

Weinberg, J. (2023, May 1). More Details on the Dispute Between JPP & Wiley. https://dailynous.com/2023/05/01/the-dispute-between-jpp-wiley/

UNESCO. (2023). UNESCO survey: Less than 10% of schools and universities have formal guidance on AI. https://www.unesco.org/en/articles/unesco-survey-less-10-schools-and-universities-have-formal-guidance-ai

Watkins, E. (2023). Guest Post: AI and Scholarly Publishing - A (Slightly) Hopeful View. The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2023/05/10/guest-post-ai-and-scholarly-publishing-a-slightly-hopeful-view/

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: *data curation, formal analysis, investigation, methodology, writing — original draft, writing — review & editing*

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

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: *conceptualization, data curation, formal analysis, investigation, methodology, project administration, supervision, 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: *investigation, supervision, writing — original draft, writing — review & editing*

Kathleen Turaski, *visualization*

Funding acknowledgement

This work was generously supported by the Andrew W. Mellon Foundation and the Arcadia Fund.

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.

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.

Connect with us

linkedin.com/company/invest-in-open

indieweb.social/@investinopen

twitter.com/investinopen

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

info@investinopen.org
www.investinopen.org