

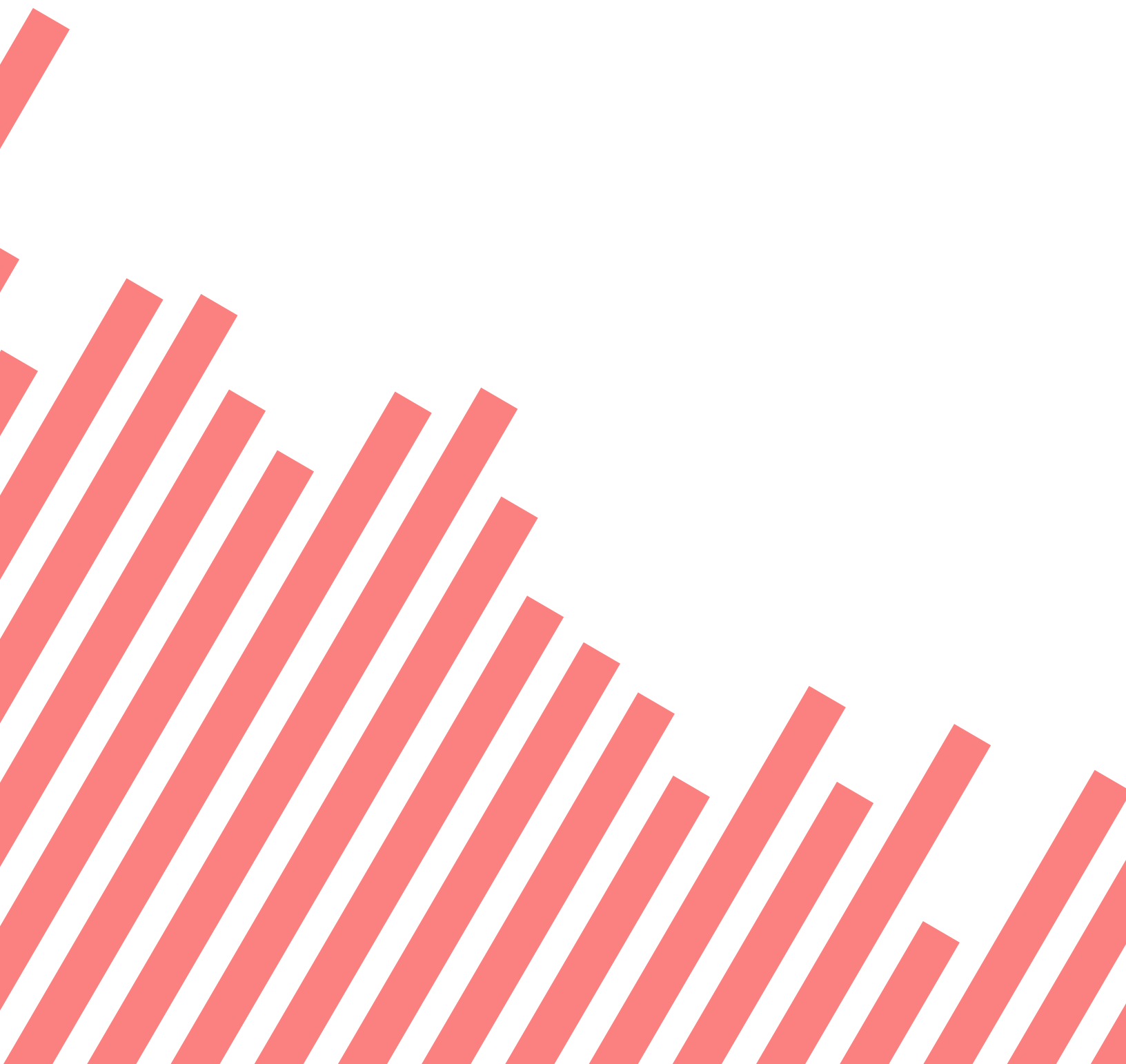
Research Library Issues

RLI 302:

Sustaining Open Content
and Infrastructure

2021

ASSOCIATION
OF RESEARCH
LIBRARIES



In This Issue

Introduction	3
Open Persistent Identifiers: The Building Blocks of Sustainable Scholarly Infrastructure	5
Introduction	
Persistent Identifiers: Unlocking Discovery	
Library Publishing and Institutional Repositories	
Data Services	
Research Data Management	
Conclusion: The Path to Unlocking Discovery	
Planning for Digital Accessibility in Research Libraries	19
1. Introduction	
2. The Basics: Plan Ahead and Avoid After-the-Fact Remediations	
3. Planning for Pivoting to Fully Virtual Operations	
4. Formats for Accessible Content	
5. Captioning	
6. Creating a Stronger Infrastructure for Digital Accessibility	
7. Looking towards the Future	
8. Summary	
Sustainability Funding for Scholarly Infrastructure Needs Infrastructure of Its Own	33
The Importance of Data	
Frameworks of Principles and Values for Sustainable Scholarship	
Conclusion	

Introduction

Judy Ruffenberg, Association of Research Libraries

This year has been dominated by discussions of infrastructure—social, technical, and physical. Concerns over the sustainability of scholarly infrastructure are complex and urgent for research library leaders, with issues of openness, inclusivity, and, of course, financing in the forefront of many conversations. Research libraries are engaged in institutional policies to strengthen scholarly infrastructure, in public policy considerations including but not limited to funding, and with the research and learning community as it navigates a diverse and often fragile landscape of infrastructures to do its work.

This issue of *Research Library Issues* takes a look at three important requirements for scholarly infrastructure from the perspective of openness, inclusivity, and sustainability: (1) Maria Gould and Maria Praetzellis examine how consolidation in the publishing industry encompassing platforms and services, as well as content, has inspired advances and commitments to **open infrastructure, specifically persistent identifiers (PIDs)**; (2) Jonathan Lazar presents a study of higher education’s surge-readiness for online services, as necessitated by COVID-19, with respect to **digital accessibility for people with disabilities**; and (3) I look at developments in **sharing data and information about the cost and financing of scholarly infrastructures** in order for individual institutions to plan and commit to sustaining those infrastructures.

© Judy Ruffenberg



This article is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>.

To cite this article: Ruttenberg, Judy. “Introduction.” *Research Library Issues*, no. 302 (2021): 3–4. <https://doi.org/10.29242/rli.302.1>

Open Persistent Identifiers: The Building Blocks of Sustainable Scholarly Infrastructure

Maria Gould, California Digital Library

Maria Praetzellis, California Digital Library

Introduction

In May 2021, Microsoft circulated an announcement that it would be shutting down its Microsoft Academic Services (MAS) by the end of the year. The news of this decision reverberated through the open-scholarship community, raising questions and concerns among the many stakeholders who relied on the free service for tracking research activities in various contexts.¹

At a time when research discovery is more necessary than ever, it is also becoming more complicated. The work of tracking and identifying publications and other research outputs is taking place in a context of increased technological complexity, competing motivations and priorities, and constrained resources. As exemplified by the Microsoft case, one of the fundamental challenges and risks in the scholarly infrastructure landscape is the unpredictable availability of the platforms and services we rely upon to perform this work. When these platforms and services go away, what do we have left?

Such challenges and risks might be overcome or at least mitigated if and when scholarly infrastructure is built with open components that can persist beyond their packaging. “The Principles of Open Scholarly Infrastructure” (POSI), which were initially outlined in 2015 and are seeing a revival in 2021, provide a set of guidelines for open infrastructure for research and scholarly communications.² Within this framework, open infrastructure is a strategy for sustainability. Using the POSI principles as a backdrop, we examine one essential ingredient of open infrastructure: persistent identifiers, or PIDs. We explore ways in which the use of openly available PIDs, and investments in

the services that support them, can enable the discovery of research outputs while promoting the sustainability of data and information.

Research libraries have an opportunity to adopt a “PID-centric” approach to tracking, sharing, and publishing research. PIDs have the potential to address pain points, increase efficiencies, and save time. Promoting the implementation of open PIDs and the metadata associated with them serves a broader goal of improving information connectivity.

While this article does not aim to offer an exhaustive discussion of the many complexities of funding, maintaining, and connecting the multiplicity of scholarly systems, nor does it promise a comprehensive survey of all persistent identifiers, we want to share our first-hand perspective on the dynamics of building and planning for open and sustainable scholarly infrastructure and we want to outline ideas and strategies to advocate specifically for prioritizing open PIDs and open metadata to ensure research sustainability.

Persistent Identifiers: Unlocking Discovery

Overview: Core Persistent Identifiers for Scholarly Communication

Persistent identifiers in the scholarly communication context serve as stable, long-lasting unique references to core components of the research enterprise. These components include but are not limited to publications and other research outputs, researchers and contributors, institutions and facilities, instruments and materials, funders, and grants and awards.

PIDs help to provide long-term unambiguous identification of and access to research (and information about research). This is useful in today’s dynamic and diffuse research landscape: for example, a publication’s URL is likely to change over time, multiple researchers have the same name, and researchers’ affiliated institutions or funding organizations might be written in multiple ways across different

outlets. PIDs enable disambiguation and discovery by providing machine-readable data that can be used to track individual components of research and establish connections between these components at a given point and over time. They can associate researchers to publications, capture networks of research collaborators, link a set of related publications to each other, or identify the downstream products of a grant-funded project, among other uses.

PIDs can therefore help answer questions that are crucial for effective research discovery and management, such as:

- How can I find all of the research published at my institution?
- How can I identify the publications that resulted from a specific research project?
- How can I locate the data set associated with a publication?
- How can I track the downstream outcomes and impacts of a research project?
- How can I record collaborations with other research institutions?
- How can I ensure compliance with funder requirements for data sharing?

As scholarship proliferates across digital platforms and discovery systems, PIDs have become the essential building blocks of the scholarly communications infrastructure for finding, accessing, and tracking research outputs. In this context, the PIDs most commonly used include:

- PIDs for people
- PIDs for outputs
- PIDs for organizations
- PIDs for funders and grants

Within these categories, there may be more than one type of identifier. For instance, the Open Researcher and Contributor ID (ORCID) is a well-known global identifier for researchers. The ORCID registry is open and managed by a community-governed nonprofit organization.

However, there are other identifiers besides ORCID that can be used to identify researchers. ResearcherIDs and Scopus Author IDs are two examples; unlike ORCID, they are used in commercial databases (Web of Science and Scopus, respectively) and are not openly available. In the case of institutional identifiers, Research Organization Registry (ROR) IDs are freely and publicly available in the ROR registry, which provides an open data set and includes additional tools for working with institutional data, such as an open application-programming interface (API). Other identifiers for institutions also exist but they are not openly available, such as those in the Ringgold database and in Web of Science and Scopus.

It is important to understand the differences between open and non-open PIDs because they speak to real risks and inconsistencies in our current landscape. Therefore, our focus is on those PIDs that have broad adoption globally and that allow use and reuse of the metadata they contain.

The Importance of PIDs and Connected Metadata

A PID itself should not be seen as the end goal. Instead, the power of PIDs is not so much what they identify as the connections they enable. These connections, and the insights they offer, can only be fully realized through open metadata and open infrastructure.

As co-authors of a September 2020 report, *Implementing Effective Data Practices: Stakeholder Recommendations for Collaborative Research Support*,³ we presented a set of recommendations for implementing and advocating for PIDs in research infrastructure as a way to “unlock discovery.” A premise of the report is that PIDs are an essential element in building a more open research ecosystem. To fully realize this vision, systems and services that use or provide PIDs must follow open practices, particularly in terms of the open licensing of metadata:

Organizations that sustain identifier registries are essential pieces of scholarly infrastructure, and beyond adoption and use of PIDs, these

organizations need the support of the research community. The research community is also best served by **open licensing of metadata that enables interoperability across systems**. Libraries, IT professionals, and research offices that develop or purchase research support systems can help accelerate the adoption of PIDs by requiring that these systems be designed to integrate with identifier registries, and by advocating for open metadata and open code.⁴

When relying on PIDs to track and connect research, we need to be aware of the opportunities and limitations of the PIDs and the underlying research infrastructure that we use to do this. A proprietary identifier in a closed system is only useful to that system and its user base. This is a sustainability concern. PIDs developed as open infrastructure and for use **in** open infrastructure afford the greatest potential to implement efficient, cost-effective, long-lasting scholarly communication practices. An investment in open PID infrastructure is a strategy for making research—and insights about research—more accessible to all, and serves as a sort of insurance policy against the unpredictable events that can arise in a commercialized scholarly communication landscape. Frameworks like POSI can help distinguish which organizations or tools follow open principles and surface information about governance and sustainability. These considerations are significant factors if we are to prioritize investments in open infrastructure.

Below, we explore three critical areas for investment in sustainable, open PID infrastructure: (1) library publishing and institutional repositories, (2) data services, and (3) research data management. We focus on use cases featuring digital object identifiers (DOIs), for two reasons. First, we expect that many institutional stakeholders will be familiar at least in principle with DOIs, as they are commonly visible in publications, reference lists, and databases. Second, DOIs exemplify how PIDs can be enriched with metadata and how PIDs can work in concert with each other to make research more discoverable.

Library Publishing and Institutional Repositories

Institutions worldwide are sites of and incubators for transformations in research dissemination. Library publishing and institutional repository services contribute to the increase in digital scholarship artifacts that need to be managed and made available for discovery, access, and use.

These areas of growth also present some challenges and pose questions. Fundamentally, how can library publishers and repositories ensure the discoverability of their content? How can usage be tracked to understand how this content is being used and cited? How can the metadata in publishing and repository platforms generate insights and reports on research activities? How can content be networked to identify connections between researchers and the outputs they generate?

Incorporating open and PID-based infrastructure in these initiatives is one way to address these challenges and answer these questions. Institutions have several options in this regard. A starting point would be to make sure library publishers and institutional repositories are registering DOIs for their content and taking advantage of opportunities to enrich the DOI metadata with information to aid in tracking and discovery.

For library publishers, becoming a Crossref member or choosing a platform or service provider that integrates with Crossref means that DOIs can be registered for publications. For institutional repositories, becoming a member of DataCite or using a platform or service provider that integrates with DataCite means that DOIs can be registered for repository content.

However, registering DOIs is about more than just getting a DOI. While a DOI alone is useful insofar as it can provide a permanent reference to an object regardless of whether the object's location changes over time, a DOI becomes much more useful when it includes rich metadata.⁵

This metadata includes:

- Information about who published the work (names, unique IDs), their roles (creators, contributors), and where they are affiliated
- Information about the work itself (abstract, work type)
- Information about related works (data, older versions, series, dissertations, preprints)
- Information about who funded and/or sponsored the underlying research
- Information about copyright and licensing
- Information about referenced works

Enriching DOI metadata provided to Crossref and DataCite optimizes the work for greater discoverability and therefore reusability. Systems that harvest from Crossref and DataCite can index the additional metadata in the DOIs. Works can then be searched to find specific authors, or works associated with a particular institution or funder. Reference lists and related works can be analyzed to provide a fuller picture of the work in context.

Data Services

Researchers and research stakeholders today must navigate an array of policies and requirements around sharing data and following best practices for data publication—the FAIR principles.⁶ While the landscape of data management and data sharing policies has been widely discussed, less frequently addressed is the role that persistent identifiers can play in navigating these requirements and adhering to best practices.

To illustrate how this can work, let's take the example of a single DOI for a data set. At a bare minimum, this DOI could function as a unique, long-lasting reference to the data set in case the location where it is stored changes over time. The existence of the identifier alone can make a significant difference in promoting the stability of this scholarly resource over the long term.

The DOI string itself is just the beginning, however. The discoverability and usability of the data set will be limited unless the DOI metadata contains additional information about the resource. This metadata is best enriched with other persistent identifiers that can optimize the DOI for discovery and usability. For example, including an ORCID ID in metadata for the data creator—as opposed to just including the creator’s name as a text string—allows for the creator to be unambiguously identified with the work, and for scholarly reporting systems to better locate all of the research associated with this particular ORCID ID. In a similar vein, including a ROR ID—as opposed to a text string—for the creator’s institutional affiliation allows for this institution to be linked to the work, and for systems to better track all of the research associated with the institution.

Enriching data-set metadata with identifiers also enables best practices with data-citation and data-usage tracking. When a data-set DOI is used in citations, services can capture this usage information.⁷ Rich metadata included with the DOI provides more context about the research. When other PIDs are included as part of this metadata, that optimizes the metadata for machine-readability and for more efficient and comprehensive aggregation and reporting.

Institutions can pursue the following concrete steps to maximize the potential of identifiers in data publishing.

- If your institution hosts one or more data repositories, make sure the repository assigns DOIs to the data, and guide researchers and data managers to supply rich metadata when they register the DOIs.
- If you do not host a repository, make sure researchers are guided to submit their data to repositories that do follow best practices when it comes to DOIs.
- Researchers should be encouraged to obtain ORCID IDs so they can provide these identifiers with their data publications.
- Researchers preparing and publishing manuscripts should include data citations in their manuscripts.

Research Data Management

Data-management plans (DMPs) contain a wealth of information about research projects, including, amongst other things, project plans for access, preservation, and storage. Historically, DMPs have been two-page narrative documents that outline proposed data practices during a research project and detail where investigators will deposit research outputs upon project completion. Over the past few years, there has been a concerted push towards creating machine-actionable DMPs (maDMPs).⁸ These next-generation DMPs are designed to move past the static narrative format and facilitate the creation of a living document that can guide research by integrating data-management activities with related systems and workflows in the research life cycle. Demonstrating their support for this work, the US National Science Foundation (NSF) recently recommended that researchers utilize PIDs for their data outputs and generate DMPs that allow for automated information exchange (maDMPs).⁹

PIDifying the DMP

Utilizing identifiers within DMPs allows information within a DMP to be shared across stakeholders, linking metadata, repositories, and institutions, and allowing for notifications and verification, with reporting taking place in real time. A vital goal of this system is to reduce the burden on researchers by generating automated updates to a plan and facilitating seamless integration with systems and groups that support research. Networked DMPs are a vehicle for reporting on the intentions and outcomes of a research project that enable information exchange across relevant parties and systems. They contain an inventory of crucial information about a project and its outputs (not just data). With a change history, stakeholders can query for updated details on the project over its lifetime.

The recent development of a new PID for DMPs, the DMP-ID, was a fundamentally important step toward creating Networked DMPs.¹⁰ With the development of this new PID, which is built on DOI infrastructure, we can expose connections between the rich metadata within a DMP and related works such as project outputs, individuals, affiliations, and publications.¹¹

Beyond the DMP-ID

Simply receiving a DMP-ID or creating a machine-readable DMP does not realize the true potential of the Networked DMP. Connections between DMPs and their eventual outputs are made possible through the linking of open identifiers, which form an interconnected web of research components in the form of a graph.¹² In the same way that an ORCID record will be empty if researchers do not provide their ORCID IDs when publishing works (and if publishing systems do not collect this information), a DMP-ID needs to be utilized and recorded to build on the networking capabilities of PIDs. Capturing these assertions on the DMP-ID enables the tracking of data-management activities as they occur during a grant project. Again, to facilitate these connections, we need both rich metadata records that include related identifiers and to build systems that enable seamless ways for researchers to include identifiers such as the DMP-ID.¹³

Use Case: FAIR Island Project

The FAIR Island Project, in which our organization is a lead collaborator, is an attempt to showcase how best to maximize the information-rich potential of the Networked DMP. FAIR Island addresses the current challenge of discovering and accessing research connected to field stations. Administrators generally do not have precise methods for tracking the research outcomes resulting from work conducted at their facilities. The FAIR Island Project utilizes a working field station as a controlled environment to test the implementation of optimal FAIR data policies and workflows built around the Networked DMP that address discovery and access to

research outputs.¹⁴ The project builds interoperability between pieces of critical research infrastructure—DMPs, research practice, DOIs, and publications—to facilitate the advancement and adoption of open science. Through the Networked DMP, the project will promote the quantification of productivity of field stations, which has proven difficult despite qualitative assessments of the immense value of these centers of research.

How Libraries Can Support the Networked DMP

While the work to develop a Networked DMP is ongoing, libraries can now promote the adoption of this new best practice in data management by encouraging researchers at their universities to get DMP IDs. Currently, DMP-IDs can be generated via the DMPTool or Zenodo, or through DataCite member services.¹⁵ As we work to build a more connected ecosystem, there will be increasing ways that researchers will be able to utilize their DMP-ID and cite this ID when publishing outputs related to their project.

Conclusion: The Path to Unlocking Discovery

This article has outlined specific opportunities that research libraries can pursue to incorporate or advocate incorporating persistent identifiers into workflows, infrastructures, and policies. Extending the premise that PIDs can “unlock discovery,” we have discussed ways to choose and leverage identifiers to best achieve this goal. We emphasize that PIDs alone are not the solution, but rather that strategies and policies regarding PIDs should focus on what metadata is used with PIDs, and how PIDs are connected to each other. A core principle in this vein is one of openness—the openness of the identifiers and metadata themselves, as well as of the infrastructure in which they are embedded. Scholarly infrastructure should be open from the inside out, and PIDified from beginning to end. With open metadata and open infrastructure, we can build connections and support the long-term stability and usability of scholarship to promote open knowledge practices, save time and resources, and develop more meaningful insights about research.

In this framework, we claim that investing in PIDs and open PID infrastructure should be recognized and adopted as a core sustainability strategy that can insulate research stakeholders from an unpredictable landscape in which scholarly communication services come and go. Furthermore, openly available PIDs containing rich metadata give tool providers and builders a connected ecosystem to work from and offer the library community the flexibility and assurance that the information contained within these systems is not dependent on a single provider or platform.

Endnotes

1. Shortly after Microsoft's announcement, OurResearch—the organization behind such tools as Unpaywall and Unsub—announced that they were building a similar service: “We’re Building a Replacement for Microsoft Academic Graph,” *OurResearch Blog*, May 8, 2021, <https://blog.ourresearch.org/were-building-a-replacement-for-microsoft-academic-graph/>. For more on the shifting landscape of scholarly discovery services in the wake of the Microsoft announcement, see: Dalmeet Singh Chawla, “Microsoft Academic Graph Is Being Discontinued. What’s Next?,” *Nature Index*, June 15, 2021, <https://www.natureindex.com/news-blog/microsoft-academic-graph-discontinued-whats-next>.
2. Geoffrey Bilder, Jennifer Lin, and Cameron Neylon, “The Principles of Open Scholarly Infrastructure,” 2020, <https://doi.org/10.24343/C34W2H>.
3. John Chodacki et al., *Implementing Effective Data Practices: Stakeholder Recommendations for Collaborative Research Support* (Washington, DC: Association of Research Libraries, September 2020), <https://doi.org/10.29242/report.effectivedatapactices2020>. An additional related work also explores concrete steps libraries can take to accelerate the sharing of research data: Tobin L. Smith et al.,

Guide to Accelerate Public Access to Research Data (Washington, DC: Association of American Universities and Association of Public and Land-grant Universities, 2021), <https://doi.org/10.31219/osf.io/tjybn>.

4. Chodacki et al., 5, emphasis added.
5. For a good example of the power of rich metadata see: Kristian Garza, “Are You There, Metadata? It’s Me, the Bibliometrician,” *DataCite Blog*, January 25, 2021, <https://doi.org/10.5438/j4xv-y945>.
6. Mark D. Wilkinson et al., “The FAIR Guiding Principles for Scientific Data Management and Stewardship,” *Scientific Data* 3 (March 2016): 160018, <https://doi.org/10.1038/sdata.2016.18>.
7. Daniella Lowenberg et al., *Open Data Metrics: Lighting the Fire* (version 1), computer software, Zenodo, 2019, <http://doi.org/10.5281/zenodo.3525349>.
8. For an excellent overview of maDMPs see: Tomasz Miksa et al., “Ten Principles for Machine-Actionable Data Management Plans,” *PLoS Computational Biology* 15, no. 3 (2019): e1006750, <https://doi.org/10.1371/journal.pcbi.1006750>.
9. Joanne S. Tornow et al., “Dear Colleague Letter: Effective Practices for Data,” NSF 19-069, National Science Foundation, May 20, 2019, <https://www.nsf.gov/pubs/2019/nsf19069/nsf19069.jsp>.
10. Kristian Garza and Matt Buys, “A Brave New PID: DMP-IDs,” *DataCite Blog*, April 7, 2021, <https://doi.org/10.5438/j22a-5d79>.
11. For more, see: “Machine Actionable Data Management Plan Connections,” Jupyter nbviewer, accessed November 24, 2021, <https://nbviewer.org/github/datacite/pidgraph-notebooks-python/blob/master/dmp/user-story-single-dmp-connections.ipynb>.
12. Helena Cousijn et al., “Connected Research: The Potential of the PID Graph,” *Patterns* 2, no. 1 (January 2021), <https://doi.org/10.1016/j.patter.2020.100180>.

13. For more information on linking DMP-IDs to other resources, see: “Link DMP IDs to Other Resources,” DataCite, accessed November 24, 2021, <https://support.datacite.org/docs/link-dmp-ids-to-other-resources>.
14. Maria Praetzellis, “Introducing the FAIR Island Project,” February 10, 2020, University of California Curation Center (UC3), <https://uc3.cdlib.org/2020/02/10/introducing-fair-island/>.
15. For more information on creating a DMP-ID, see: “DataCite DMP IDs,” DataCite, accessed November 24, 2021, <https://support.datacite.org/docs/datacite-dmp-ids>.

© Maria Gould and Maria Praetzellis



This article is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>.

To cite this article: Gould, Maria, and Maria Praetzellis. “Open Persistent Identifiers: The Building Blocks of Sustainable Scholarly Infrastructure.” *Research Library Issues*, no. 302 (2021): 5–18. <https://doi.org/10.29242/rli.302.2>.

Planning for Digital Accessibility in Research Libraries

Jonathan Lazar, College of Information Studies (iSchool) and Trace Research & Development Center, Human-Computer Interaction Lab (HCIL), University of Maryland

1. Introduction

In a recent NBC News story, Carey Jaros, CEO of GOJO Industries (better known as the maker of Purell), stated that while they were caught off guard by the increased customer demand during the COVID-19 pandemic, “we are [now] committed to being surge-ready.”¹ How many research libraries are surge-ready when it comes to planning for increased demand for digital accessibility? This is not just a philosophical or a theoretical question. Planning for digital accessibility in a research library takes time. It’s not something that can occur on the fly. Many organizations, including libraries, were caught off guard by the switch to strictly virtual services during the pandemic. In-house accessibility operations couldn’t be scaled up, and external professional accessibility services experienced a mismatch between greatly increased demand and an insufficient supply during the pandemic. Unfortunately, this made live captioners in high demand and they were rarely available, just as hard to find and acquire as hand sanitizers like Purell. Even though the COVID-19 pandemic may (hopefully) be waning due to vaccines, booster shots, and new treatments such as pills, we can expect future dramatic changes due to climate change, political instability, disease, or other events. It is therefore expected that there may again be unplanned surges in demand for accessible and remote library services. This paper will describe some of the challenges and lessons learned from the COVID-19 pandemic related to digital accessibility. In addition to strengthening the infrastructure and surge capability related to digital accessibility, it is equally important for research libraries to look ahead to future developments in the area of digital accessibility.

2. The Basics: Plan Ahead and Avoid After-the-Fact Remediations

Digital accessibility means providing an equal user experience for people with disabilities, and it never happens by accident. It is important to plan ahead for digital accessibility and include accessibility requirements in planning for building or acquiring any new digital technologies or content. Planning for digital accessibility is like planning for accessibility in building a new house. When you build the house from scratch and you have included accessibility in the design specs, the accessibility costs are minimal. However, if you have built a house in an inaccessible manner, or if you purchased an existing inaccessible house, the costs of making that house accessible can become enormous. You may need to re-grade, to add ramps, to make the doorframes wider, and you may need to move the plumbing because there isn't enough space in the bathroom for the turnaround radius. Similarly, when you build a technology from scratch with accessibility as a key design goal, the costs are minimal.² Yet if you build a technology in an inaccessible way, and then want to retrofit it after-the-fact, the amount of code required, and the costs, increase exponentially.³ To be clear, accessibility by itself is not expensive. Choosing to add accessibility later on, as a retrofit, is when the costs of accessibility increase. But that increased cost is due to poor design decisions, not inherent to accessibility itself.

It is important to plan for accessibility in a website design or redesign, for digitizing existing paper materials, for acquiring a license for digital library materials, or for materials specifically for a university course. At this point, many research libraries have processes in place, staff who are familiar with accessibility, and policies to encourage or force digital accessibility. That's the good news. In many research libraries, the infrastructure of human expertise for digital accessibility already exists. The bad news is that the massive shift to virtual operations, and the increase in demand for professional accessibility services due to the pandemic, hit research libraries just as hard as other organizations. No one had planned for a shift to strictly virtual operations. In many ways,

the infrastructure related to policy, process, and capacity, was insufficient.

While there are many legal requirements specifically addressing digital accessibility for people with disabilities, it is important to note that digital accessibility benefits the broader population, not only people with disabilities. First of all, by following technical standards that ensure accessibility, the content also becomes more portable across platforms, browsers, and operating systems. Making digital content or interfaces accessible basically means making them flexible, and everyone benefits from flexible interfaces, as they allow content to be correctly rendered across a broader range of devices and platforms.⁴ Second, many people who might not consider themselves people with disabilities, are likely to benefit from using accessibility features. A large survey done by Microsoft estimates that 57% of computer users are likely or very likely to benefit from the use of accessibility features.⁵ Third, there is evidence in the research that making web content accessible makes that content easier for everyone to use.⁶ So ensuring that digital content, websites, and software are accessible benefits the entire population of users. Due to the pervasive use of captioning on videos by a large percentage of the population (including people in places that are quiet or noisy, people learning English, people who want to search video, etc.), professionally captioned video is already perceived to have wide benefits for the broader population beyond people who are deaf or hard of hearing. But those benefits don't only occur in the area of captioning, and while there are differences in the definitions of accessibility and universal design, accessibility does have universal benefits.⁷

3. Planning for Pivoting to Fully Virtual Operations

As of late 2021, the COVID-19 pandemic seems to be entering another resurgence wave due to the Omicron variant. It is unknown whether there will be new variants or new increases in infection rates, or how often lockdowns will need to take place again in the future. For students with disabilities, the impacts of remote learning, often

inaccessible remote learning, were massive. Data from surveys and social media studies reported high levels of concern and stress among students with disabilities related to whether online learning would be accessible,⁸ and a broad population of university students (not only those with disabilities) experienced increased stress and anxiety during the COVID-19 pandemic.⁹ However, even without a future pandemic, there is likely to be increased demand for accessing library resources without physically coming to the library, and instead delivering that content in digital format, even for people without disabilities. How can you plan for virtual access to library collections? How can you plan for surges in accessibility demand?

In a recent study of university directors of digital accessibility during the COVID-19 pandemic, there had been no “surge” plans or policies in place for how to address increased demands for accessibility.¹⁰ Even when accessibility services are generally done within the university (“in-house”), contacts should be made, and perhaps retainer contracts should be put in place with outside vendors, so that when increased needs occur for services such as captioning video, remediating documents, or American Sign Language (ASL) interpretation for online events, the surge capacity is already in place. Even by the end of 2020, none of the directors of digital accessibility interviewed in the study had created any new policies for managing the shift to virtual operations, or for triaging and prioritizing the various accessibility needs in a situation where all of the needs could not be met.¹¹

The results of that study demonstrate good approaches and bad approaches that were used during the pandemic. Some universities had success with offering training courses on digital accessibility to faculty and staff during semester breaks when the enrollment and interest was much higher.¹² In addition, small (but competitive) grants were made to faculty who proposed ways to rework their classes for a strictly virtual format, and most of those proposals included accessibility components. In the area of procurement, however, there were many failures during the pandemic. Most research libraries are familiar by now with how

procurement processes can be used as a lever for ensuring digital accessibility. When digital materials are procured, or when a license is being negotiated for access to an outside digital library, it is important to only procure accessible materials, include accessibility requirements in the procurement contract, and use that purchasing power to pressure vendors on accessibility. However, during the COVID-19 pandemic, multiple university officials responsible for digital accessibility reported that their procurement controls were being bypassed using “emergency” or “fast-track” procurement authority, and much inaccessible software and digital content was procured.¹³ One accessibility director noted, “we are going to be reaping the rotten effects of those contracts signed without accessibility, for years to come.”¹⁴

4. Formats for Accessible Content

Research libraries tend to be known for the quality and extensiveness of their collections. While public libraries may lend out devices (such as e-readers), typically, the focus of a research library is on the content, the resources, the collections that they can provide, whether it is for research purposes or for classroom use. It is therefore important to have an understanding of the three core formats for ensuring the accessibility of digital content: HTML, EPUB, and PDF, and their corresponding accessibility guidelines: WCAG, Accessible EPUB3, and PDF U/A.

HTML format is best known as the markup language used for web pages. The corresponding guidelines for accessible web pages are the Web Content Accessibility Guidelines (WCAG). WCAG is the international standard for creating accessible web-based content. WCAG started as the Trace Center Unified Web Accessibility Guidelines in the mid-1990s, and WCAG version 1.0 was issued by the World Wide Web Consortium (W3C) Web Accessibility Initiative in 1999. WCAG 2.0 was issued in 2008 and WCAG 2.1, the current standard, was issued in 2018.¹⁵ In later sections, I will talk about the next steps for the WCAG.

EPUB3 is now the predominant format for e-books. Originally developed as a project of the International Digital Publishing Forum in 2010, it is now a standard run out of the Web Accessibility Initiative. EPUB3 allows for multiple resources in a single file, using a specified reading order or another reading order. The current version is EPUB 3.2, which was approved and published by W3C as a Final Community Group Specification (slightly different from a standard) in 2019,¹⁶ however, earlier versions of EPUB were adopted as international technical standards by the International Organization for Standardization (ISO). EPUB3 was designed to be easy to make accessible, and over the past few years, there has been a major shift from publishers putting textbooks in PDF format, to instead publishing their books in EPUB3 format.

Of the three formats for digital content described here, PDF format is considered to be the hardest format to make accessible. While HTML and EPUB3 were built with accessibility in mind from the start, the guidelines for creating accessible PDF content (known as PDF U/A, PDF Universal Access, or sometimes as the Matterhorn Protocol) were created long after the creation of the PDF format.¹⁷ While most web content development and management tools have some features built in to encourage accessibility, there are very limited tools available for making PDF files accessible, and they are often hard to use. The limitations in the existing tools have even caused some universities to try and limit or eliminate the PDF format from their campus (affectionately named the “Great PDF Purge” by North Carolina State University).¹⁸ This seems to be a valid concern, as some campus leaders were concerned that during the COVID-19 pandemic, their level of PDF accessibility on campus actually decreased, as paper forms were quickly scanned as graphical PDF files without considering accessibility.¹⁹ Until better tools and solutions are in place to assist with creating and remediating PDF files for accessibility, this will continue to be the hardest format to make accessible.

5. Captioning

All videos, livestreaming, and other multimedia on web pages or provided in other ways (such as teleconferencing via Zoom, Skype, or Microsoft Teams) must be captioned. Captioning, as a technical concept, is not hard, and captioning of video and television shows has been done for nearly 50 years. Recent legal cases against universities have brought attention to captioning. What changed during the pandemic is the greatly increased demand for captioning at universities and libraries, which had a greatly increased demand for captioning videos, course materials, and livestreaming events, but often faced two challenges, related to budget and capacity. Due to an overall increase in expenses, many academic units at many different universities reportedly pushed back on human captioning, and instead wanted to use automated (AI-based) captioning, because it's less expensive (and in many cases, free).²⁰ However, the quality of automated captioning is much lower than professional, human captioning. And even when there was the desire and the budget to professionally caption a video, there was often an inability to get a video captioned within a reasonable timeframe. Hiring someone to do live (real-time captioning) seemed to be the most challenging, with universities reporting that their usual vendors were unable to provide live human captioning. Finding ASL interpreters available for real-time work was equally challenging. One university described how when they started using automated captioning, they found that it was insufficient for their needs and had to switch to human captioning midway through courses.²¹ Whenever possible, human captioning is always preferable to automated captioning, especially in content where technical or complex terminology is present, where automated captioning is even less effective.

6. Creating a Stronger Infrastructure for Digital Accessibility

As we approach two years since the pandemic started, universities and research libraries have often not created new policies and processes to address the situations that may arise. For instance, if there are

three library events that need to be live captioned, and only one human captioner is available during that time, who gets the human captioner? Is it based on attendance? Or is it based on whether the event is a “public event” vs. a “private event” (behind a password wall for employees)? Or is it based on whether someone with a disability specifically requests an accommodation? If video captioning now takes two weeks for turnaround time instead of three days, have policies changed to note that? Can the standard promises of how quickly library materials will be scanned and emailed still remain? When new print collections are received, are they immediately scanned in an accessible manner, so that they are available when needed, not requiring a wait when a patron request is made? Does the library need additional scanners or new software tools to assist with, for example, PDF remediation? Have new collaborations with outside vendors been formed? For captioning of videos, is there a plan in place to add outside capacity when in-house, in-university services are already being utilized at full capacity? Are vendors keeping their promises about fixing accessibility barriers as promised in existing procurement agreements and settlement agreements?

Overall, the question is, “have libraries created an infrastructure for digital accessibility, to deal with the current challenges and future surges in demand?” There need to be policies created, resources allocated, and plans implemented, for describing how digital accessibility will be addressed in the future. As it stands right now, some states will require that everyone on a university campus be vaccinated (with exceptions for religious or health-related reasons), and other states will not require vaccination (or are preempted from requiring vaccination by state law).²² Regardless of what the political and public health situation is for a research library related to vaccination, it is likely that some people will not yet feel comfortable coming to campus and instead will request that all materials be delivered virtually. Research libraries need to plan for those surges in requests, and the associated accessibility needs.

7. Looking towards the Future

It is important to stay aware of new developments in the area of digital accessibility. Right now, there are major changes underway with the WCAG. While the current version of WCAG is version 2.1, the Web Accessibility Initiative (WAI) is already working on the development of WCAG versions 2.2 and 3.0. WCAG 2.2 is a minor extension of WCAG 2.1, using the same structure and format, with a focus on additional success criteria to meet the needs of “users with cognitive or learning disabilities, users with low vision, and users with disabilities on mobile devices.”²³ To do so, WCAG 2.2 includes nine new success criteria, related to Accessible Authentication (level A), Dragging Movements (level AA), Consistent Help (level A), Page Break Navigation (level A), Focus Appearance (Minimum) (level AA), Focus Appearance (Enhanced) (level AAA), Visible Controls (level AA), Target Size (Minimum) (level AA), and Redundant Entry (level AA). In addition, Focus Visible, a success criteria already in WCAG 2.1, has moved from level AA to level A.²⁴

Beyond WCAG 2.2, there’s another effort underway: WCAG 3. Note that as people have been using WCAG to understand accessibility beyond just web content, WAI is currently planning to rename WCAG to mean “W3C Accessibility Guidelines” instead of the “Web Content Accessibility Guidelines.” While the WCAG 3.0 document is currently in draft format, and is still subject to change, the current draft of WCAG 3.0 describes “additional tests and different scoring mechanisms.”²⁵ Future guidelines may also enhance accessibility for people with cognitive, language, and learning disabilities.

It’s also important to note that W3C recently published a working draft of EPUB 3.3.²⁶ And in PDF accessibility, the long-neglected format is finally garnering attention in the area of accessibility. Two parallel efforts may provide future benefits in making it much easier to remediate PDF documents for accessibility. The Trace Research and Development Center at the University of Maryland (the nation’s oldest research center on technology and disability) is collaborating with

Adobe to improve Adobe's tools to make it much easier for content creators to remediate PDF documents for accessibility. And researchers at the Allen Institute for AI are working on AI-based approaches for remediation.²⁷ Both of these efforts work on the current problems related to remediation, but future tools may offer functionality where content creators can ensure accessibility when the PDF file is created, rather than an after-the-fact remediation (which, as already discussed, is not the most efficient way to do it).

Aside from technical developments, it's also important to be aware of legal and policy developments.²⁸ Many in the digital accessibility realm have been watching the two ongoing legal cases related to web accessibility and public accommodations, *Robles v. Domino's Pizza* (in the Ninth Circuit) and *Gil v. Winn-Dixie* (in the Eleventh Circuit). Both of these cases, involving blind individuals challenging inaccessible websites and apps, are primarily about web accessibility for public accommodations under Title III of the Americans with Disabilities Act, and both deal with issues of the nexus between a physical location and a website. Both of these cases have had major court decisions during 2021, yet do not directly impact the existing accessibility requirements for research libraries, which, as recipients of federal funding, are also covered under Section 504 of the Rehabilitation Act, as well as potentially other statutes, regulations, and agency interpretations relating to education. It is important to note that after a four-year absence (since Maria Town last served in this role in 2016), the White House again has a lead policy person on disability policy, Kimberly Knackstedt. While the Biden-Harris administration has already signaled stronger support for disability-related issues, it is not known at this time whether that will result in any administrative actions related to digital accessibility.

8. Summary

The first two years of the COVID-19 pandemic were challenging for all organizations, including research libraries. Yet the surges in demand for accessibility services are likely to occur again in the next

few years. Research libraries need to more fully integrate accessibility requirements into technology development and procurement, and plan ahead for how to address the increasing requirements for accessibility and the surges in demand.

This work was funded in part by the National Institute on Disability,

Acknowledgements

This work was funded in part by the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR) at the Administration for Community Living, US Department of Health and Human Services under grant #90REGE0008. The opinions herein are those of the author and not necessarily those of the funding agency.

Endnotes

1. “Inside a Purell Factory and How the Company Adapted amid the Pandemic,” *NBC Nightly News*, May 19, 2021, 2:25 video, <https://www.nbcnews.com/nightly-news/video/inside-a-purell-factory-and-how-the-company-adapted-amid-the-pandemic-112514629612>.
2. Brian Wentz, Paul T. Jaeger, and Jonathan Lazar, “Retrofitting Accessibility: The Legal Inequality of After-the-Fact Online Access for Persons with Disabilities in the United States,” *First Monday* 16, no. 11 (2011), <https://doi.org/10.5210/fm.v16i11.3666>.
3. Roland Ossmann, Klaus Miesenberger, and Dominique Archambault, “A Computer Game Designed for All,” in *Computers Helping People with Special Needs: ICCHP 2008*, eds. Klaus Miesenberger et al., Lecture Notes in Computer Science, vol. 5105 (Berlin, Heidelberg: Springer, 2008), 585–92, https://doi.org/10.1007/978-3-540-70540-6_83; Maurice Dawson et al., “Integrating Software Assurance into the Software Development Life Cycle (SDLC),” *Journal of Information Systems Technology and Planning* 3, no. 6 (2010): 49–53, https://www.researchgate.net/publication/255965523_Integrating_Software_Assurance_into_the_Software_Development_Life_Cycle_SDLC.

4. Sven Schmutz, Andreas Sonderegger, and Juergen Sauer, “Effects of Accessible Website Design on Nondisabled Users: Age and Device as Moderating Factors,” *Ergonomics* 61, no. 5 (2018): 697–709, <https://doi.org/10.1080/00140139.2017.1405080>.
5. Forrester Research, *Accessible Technology in Computing—Examining Awareness, Use, and Future Potential* (Redmond, WA: Microsoft Corporation, 2004), <http://download.microsoft.com/download/0/1/f/01f506eb-2d1e-42a6-bc7b-1f33d25fd40f/researchreport-phase2.doc>.
6. Sven Schmutz, Andreas Sonderegger, and Juergen Sauer, “Implementing Recommendations from Web Accessibility Guidelines: Would They Also Provide Benefits to Nondisabled Users,” *Human Factors* 58, no. 4 (2016): 611–29, <https://doi.org/10.1177/0018720816640962>.
7. Hans Persson et al., “Universal Design, Inclusive Design, Accessible Design, Design for All: Different Concepts—One Goal? On the Concept of Accessibility—Historical, Methodological and Philosophical Aspects,” *Universal Access in the Information Society* 14, no. 4 (2015): 505–26, <https://doi.org/10.1007/s10209-014-0358-z>.
8. Cole Gleason et al., “Disability and the COVID-19 Pandemic: Using Twitter to Understand Accessibility during Rapid Societal Transition,” in *ASSETS ’20: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, eds. Tiago Guerreiro, Hugo Nicolau, and Karyn Moffatt (New York: ACM, 2020), 1–14, <https://doi.org/10.1145/3373625.3417023>; Han Zhang et al., “How Does COVID-19 Impact Students with Disabilities/Health Concerns?,” preprint, submitted May 11, 2020, last revised May 6, 2021, <https://arxiv.org/abs/2005.05438>.
9. Changwon Son et al., “Effects of COVID-19 on College Students’ Mental Health in the United States: Interview Survey Study,” *Journal of Medical Internet Research* 22, no. 9 (2020): e21279, <https://doi.org/10.2196/21279>.

10. Jonathan Lazar, “Managing Digital Accessibility at Universities during the COVID-19 Pandemic,” *Universal Access in the Information Society* (2021): 1–17, <https://doi.org/10.1007/s10209-021-00792-5>.
11. Lazar, “Managing Digital Accessibility.”
12. Lazar, “Managing Digital Accessibility.”
13. Lazar, “Managing Digital Accessibility.”
14. Lazar, “Managing Digital Accessibility.”
15. Andrew Kirkpatrick et al. eds., *Web Content Accessibility Guidelines (WCAG) 2.1: W3C Recommendation*, World Wide Web Consortium, June 5, 2018, <https://www.w3.org/TR/WCAG21/>.
16. Matt Garrish and Dave Cramer, eds., *EPUB 3.2: Final Community Group Specification*, World Wide Web Consortium, May 8, 2019, <https://www.w3.org/publishing/epub3/epub-spec.html>.
17. Aravind Jembu Rajkumar et al., “PDF Accessibility of Research Papers: What Tools Are Needed for Assessment and Remediation?,” in *Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS)* (Manoa, HI: University of Hawaii, 2020), 4185–94, <http://hdl.handle.net/10125/64254>.
18. “Participate in the Great PDF Purge,” NC State University Office of Information Technology, August 15, 2019, <https://oit.ncsu.edu/2019/08/15/participate-in-the-great-pdf-purge/>.
19. Lazar, “Managing Digital Accessibility.”
20. Lazar, “Managing Digital Accessibility.”
21. Lazar, “Managing Digital Accessibility.”
22. Nick Anderson et al., “Colleges Want Students to Get a Coronavirus Vaccine. But They’re Split on Requiring the Shots,” *Washington Post*, June 23, 2021, <https://www.washingtonpost.com/education/2021/06/23/colleges-covid-vaccine-mandates-divide/>.

23. Chuck Adams et al., eds., *Web Content Accessibility Guidelines (WCAG) 2.2: W3C Working Draft*, World Wide Web Consortium, May 21, 2021, <https://www.w3.org/TR/WCAG22/>.
24. Adams et al., WCAG 2.2.
25. Jeanne Spellman et al., eds., *W3C Accessibility Guidelines (WCAG) 3.0: W3C Working Draft*, World Wide Web Consortium, June 8, 2021, <https://www.w3.org/TR/wcag-3.0/>.
26. Matt Garrish, Ivan Herman, and Dave Cramer, eds., *EPUB 3.3: W3C Working Draft*, World Wide Web Consortium, November 25, 2021, <https://www.w3.org/TR/epub-33/>.
27. Lucy Lu Wang et al., “Improving the Accessibility of Scientific Documents: Current State, User Needs, and a System Solution to Enhance Scientific PDF Accessibility for Blind and Low Vision Users,” preprint, submitted April 30, 2021, <https://arxiv.org/abs/2105.00076>.
28. US Congress, “H.R. 1100—Online Accessibility Bill,” introduced in House February 18, 2021, <https://www.congress.gov/bill/117th-congress/house-bill/1100>.

© Jonathan Lazar



This article is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>.

To cite this article: Lazar, Jonathan. “Planning for Digital Accessibility in Research Libraries.” *Research Library Issues*, no. 302 (2021): 19–32. <https://doi.org/10.29242/rli.302.3>.

Sustainability Funding for Scholarly Infrastructure Needs Infrastructure of Its Own

Judy Ruffenberg, Association of Research Libraries

For decades, research libraries have understood that the current system of scholarly communications—both content and infrastructure—would become unsustainable. Consolidation among large commercial publishing services has driven up prices faster than inflation or library allocations, and open source or community-based alternatives are often fragile by comparison: under-resourced, reliant on volunteer labor, and lacking in stable business models. Digital content and digital infrastructure are intertwined in the world of platforms and databases; but infrastructure is what makes scholarship possible to do, to disseminate, to engage, and to preserve.

A growing number of Association of Research Libraries (ARL) member libraries have recently launched “sustainable scholarship” initiatives and/or made commitments to values-based investing—supporting scholarly publishers and services that are aligned with library values of openness and equitable access. Goals of such efforts are to ensure ongoing access to and preservation of content, durable, affordable infrastructure, and to contain costs. Under a values-based investing framework, research libraries intend to divert a portion of their collections funds from the excesses of commercial publishing to sustain open or community-based scholarly infrastructures. They are doing so in alignment with the scholarly community and with local research priorities.

Research libraries describe sustainable scholarship initiatives as more “open, affordable, and transparent”¹ than the current system, which is dominated by an ever smaller number of large commercial publishing services including platforms, workflow tools, analytics, and computational environments. In libraries, supporting open and community-based infrastructures involves a complicated mix of both

collections budgets and staff expertise—as decision-makers, metadata specialists, acquisitions managers, advisors, developers, or through participation in the infrastructure services’ governance. In order to make informed decisions that help address the sustainability of such resources, libraries need better data about potential investments in order to assess them against agreed-upon criteria. They also need better data about contributions from their own organizations and those of their peers. Initiatives like Invest in Open Infrastructure (IOI), SCOSS (the Global Sustainability Coalition for Open Science Services), and the Open Access Community Investment Program (OACIP) are welcome and emergent, with intention and promise to address this critical data gap.

The Importance of Data

David W. Lewis’s 2017 article “The 2.5% Commitment”²—suggesting that academic libraries commit 2.5% of their total budgets to supporting the common infrastructure needed to power an open scholarly commons—marked a watershed moment in discussions of library collective action. The paper was widely discussed online and at professional conferences, and cited in the Association of College and Research Libraries (ACRL) “2018 Top Trends in Academic Libraries.”³ In “Top Trends,” ACRL’s Research Planning and Review Committee said that to meet the 2.5% objective (or any funding target), collections managers would need to “establish clear policies that outline parameters for the support and funding of specific open access initiatives and programs.” Lewis acknowledged the need for such policies. Among the first steps necessary to achieving widespread commitment, he suggested, were the establishment of criteria to identify worthy open scholarly infrastructure, and the creation of “a digital platform...to provide academic libraries a means of accounting for their contributions.” In other words, investing in open infrastructure would require an infrastructure of its own. Several large-scale national and international projects have emerged to understand and address this need, and in the process help research libraries

articulate selection criteria for investment. The success of these projects is crucial to the sustainable scholarship endeavor.

In 2019, a national forum funded by the US Institute of Museum and Library Services (IMLS), “OA in the Open,” focused on the information decision-makers needed to acquire—and under what conditions they would prioritize—open content or services. Forum participants from small and very large libraries alike found the landscape confusing with:

too many projects (some of which are very similar to each other) and too many different models and no clear way to determine worthy projects or initiatives. Librarians at large universities, even those with dedicated scholarly communication librarians, were as likely to cite being overwhelmed and under-informed as were those from smaller institutions.⁴

In addition to selection criteria for content and infrastructure, participants expressed desire for an information clearinghouse (like a digital platform) in order to determine where peer institutions were contributing.

Also in 2019, the Andrew W. Mellon Foundation–funded report *Mapping the Scholarly Communication Landscape* addressed the complexity and opaqueness of the open scholarly infrastructure environment by creating an instrument for collecting data about scholarly communication infrastructure providers (SCIP) for a proposed regular census. In 2020, the project produced case studies of several participating infrastructures, as well as a composite data set of more than 100 scholarly communication resource providers.⁵ Data from this project can be used to understand the landscape of open and community-based scholarly infrastructure providers with respect to their “forms, functions, structures, and models,” as well as their financial stability. In the meantime, SCOSS, the Global Coalition for Sustainability in Open Science Services, discussed later in this paper, has been functioning as a highly effective international crowdfunding broker, doing outreach and fundraising with libraries and consortia for

two or three key infrastructures at a time. Finally, individual libraries, consortia, and professional associations are articulating their own values frameworks to guide decision-making.

Frameworks of Principles and Values for Sustainable Scholarship

Many ARL member libraries and consortia, in partnership with both faculty and administrators, have pledged to align their spending with statements of values and principles that have implications for open and/or community-based infrastructures. Members of the ARL community have created high-level decision-making frameworks that can form the basis of local policy and criteria for infrastructure investments against which they can assess investment opportunities. These frameworks⁶ are based on alignment with institutional mission to provide wide and equitable access to scholarship, and generally promote the following criteria for scholarly communication services: (1) fair and sustainable pricing, (2) community input or governance in the infrastructure, and (3) transparency of financial operations and costs. The Confederation of Open Access Repositories (COAR) and SPARC's "Good Practice Principles for Scholarly Communication Services,"⁷ endorsed by ARL, also include easy migration (no lock-in), and open standards.

Data within Organizations

In order for libraries to advance sustainable scholarship initiatives through their own values frameworks and in partnership with scholars, the library community needs mechanisms for information-sharing within and across organizations. If the data were easy to collect, individual institutions, consortia, or membership organizations could recognize and measure financial contributions to open scholarship, which in turn would help make such contributions normative, as proponents of the 2.5% commitment had envisioned. But contributions

across an entire library budget—people, collections, technology, and services—are diffuse and difficult to capture. While the academic library community has grappled with “The 2.5% Commitment” over the past five years and wondered whether it was the right target, the community also struggled with how to measure contributions to open services.

In 2020, the Canadian Association of Research Libraries (CARL) undertook a study of its members’ contributions to open initiatives, including services, staff, and infrastructure.⁸ The study’s findings are instructive for library leaders and for the community:

Staff: By far, the largest category of investment is in local staff, with an average of 74% of the libraries’ open investments going toward salaries. On average, respondent libraries have about 7 FTEs working in open activities, scattered across a number of areas: digitized content, scholarly communications, open repositories, and research data management (including staff contributing to the national Portage project).

Content: The second largest category of spending on open were funds directed to publishers through several means: consortial licences via the Canadian Research Knowledge Network (CRKN) or, in Ontario, the regional association Ontario Council of University Libraries (OCUL) via Scholars Portal, institutional membership with open access publishers, and payment of article processing charges (APCs). This amounted to an average of 14% of total open spending, or approximately \$3.2 million CAD in total, 80% of which was directed toward licences with open access publishers or platforms.

Infrastructure: The rest of the open investments, approximately 12%, were spent on a wide variety of other types of open services, platforms and infrastructures.⁹

Just as the academic library community desires a digital platform or clearinghouse to help understand how much support open and community-based infrastructures are receiving and from whom, individual academic libraries need tools to help measure and report their own contributions across their organizations. Questions about open access collections and staff time are now part of CARL's member statistics program for the first time, which will be an important international contribution. In the absence of widely adopted tools for such disclosure in the US, some libraries, such as The University of Arizona Libraries, maintain public web pages listing their contributions to open content and open infrastructure and services.¹⁰ These interim reporting mechanisms could provide the basis for more standardized data collection that libraries can use to assess their own contributions relative to peers.

Data across Organizations

LYRASILS is making progress on the clearinghouse concept through a pilot program called the Open Access Community Investment Program (OACIP). Among OACIP's objectives are to "centralize the administration and funding of open access initiatives or programs at multiple scales and make transparent to the community at large who is participating in each investment community."¹¹ According to Sharla Lair, LYRASILS senior strategist of open access and scholarly communication initiatives, "OACIP is a funding infrastructure that supports a multi-stakeholder funding community. What does this mean? Academic libraries (of all sizes), public libraries, museums, archives, funding agencies, provost offices, & departments can collectively fund these programs."¹²

Invest in Open Infrastructure (IOI) has developed a rigorous research agenda based on investigating critical gaps, funding structures, and vulnerabilities and risks related to the development and maintenance of open infrastructure. In its recently released strategic plan, IOI commits to "increase our collective understanding of the funding and infrastructure landscape...[and] provide strategic support &

investment guidance for those looking to adopt, build, and sustain open infrastructure.”¹³ IOI has also published key findings from the Future of Open Scholarship project to better understand key decision points, costs, and funding models to maintain, sustain, and scale open infrastructure projects.¹⁴

Partnership with the Scholarly Community

The critical motivation for research libraries to contribute sustainability funding to scholarly infrastructures is that the infrastructures are in use by the research community. Participants in the IMLS-funded “OA in the Open” forum in 2019 indicated that open infrastructure—such as publishing platforms or repositories—was easier to support than open content because infrastructure was viewed as a local priority, used by local constituents. Open content on the other hand (free to all) was harder to justify except as an explicit collective contribution to the commons,¹⁵ or what Raym Crow characterizes as altruism.¹⁶ Library support for locally used infrastructure can include hosting of institution-led society journals on open platforms (such as the University of Pittsburgh Library System E-Journal Publishing Program¹⁷), the development of overlay journals based on preprint services (such as Queen’s University Library’s *Advances in Combinatorics*¹⁸), or the embrace of locally led preprint services themselves (such as the recent acquisition of SocArXiv by the University of Maryland Libraries¹⁹).

Another way that research libraries contribute to a more sustainable ecosystem of scholarly infrastructure is through teaching and supporting open source or non-proprietary tools, software, and platforms. By training graduate students and faculty in open statistical software packages and the Open Science Framework, for example, or publishing platforms like Omeka and Scalar, libraries contribute to a virtuous circle of open scholarly practices by influencing scholars to adopt such tools. COVID-related shifts to a digital-first environment

further underscored the importance of open infrastructure and services. In the early months of the COVID pandemic, demand for training surged, especially among graduate students who didn't have access to their labs or other means of doing their research. Duke University Libraries noticed a marked increase in viewing of their recorded training workshops. Libraries at Penn State, the University of Florida, and The Ohio State University all reported and responded to the increased demand for such training.²⁰

Adriene Lim, dean of libraries at the University of Maryland, expressed her decision to support SocArXiv in terms of scholarly partnership and public engagement:

SocArXiv fits into the UMD Libraries' strategies related to enhancing open access and supporting academy-owned infrastructure for scholarly communication....[and] we're proud to be the institutional home and sustain this valuable resource for the entire research community.²¹

The University of Maryland Libraries noted that:

The Libraries also manages the [Digital Repository at the University of Maryland](#) (DRUM), which hosts material from UMD researchers, including theses and dissertations as well as research articles. In the future, SocArXiv hopes to integrate submission of Maryland researchers' content with DRUM, extending the reach of UMD's research output, as well as leveraging other benefits offered by SocArXiv.²²

SCOSS: A Partnership between Infrastructure Services, National Funding Agencies, and Libraries

The Global Coalition for Sustainability in Open Science Services (SCOSS) formed in 2017 to:

help essential, noncommercial services for open science in need of immediate financial support. In other words, SCOSS formed to

address the imbalance of community overreliance on particular resources, with community underinvestment in their sustainability. SCOSS functions as a global crowdfunding call for vetted services. Open access and open science infrastructure providers apply for consideration and are evaluated by an expert advisory panel appointed by the SCOSS Board. If selected, SCOSS works with the provider to establish a three-year funding target meant to transition the service to stable funding. Then SCOSS leverages the participation of key organizations—primarily national consortia—to promote the funding call in their respective regions. SCOSS also strongly encourages the community governance of these structures.²³

SCOSS thus functions as a partnership among scholarly infrastructure providers, national funding agencies, and libraries by identifying worthy services that researchers use, and marshalling national funding and individual or consortial research library funding to transition these services to sustainable financial footing.²⁴ In a 2021 survey of SCOSS-contributing institutions, the top reason for contributing funds to a service was that people in their institutions were using it.²⁵ The second reason cited for support was that national consortia had both promoted the service and handled the administrative work of invoicing on behalf of their member libraries, another burdensome aspect of supporting the nascent open ecosystem. Perhaps due to the absence of a central licensing consortium in the United States, the US has lagged behind the rest of the world in contributing to SCOSS funding, despite demonstrable use of the services.²⁶ This is an area requiring further advocacy and visibility to remedy.

Conclusion

SCOSS, OACIP, the SCIP census, and IOI are all promising initiatives to address the library community's need for data, criteria, and transparency that would enable the operationalization of maintenance

funding for community-based infrastructure necessary for sustainable scholarship. Research libraries can work with these new projects to supply and help standardize data about which scholarly infrastructures are used by their local communities and how their organizations are contributing to the infrastructures' sustainability.

Endnotes

1. See, for example, "Sustainable Scholarship," The University of North Carolina at Chapel Hill, <https://sustainablescholarship.unc.edu/>; "Sustainable Scholarship," River Campus Libraries, University of Rochester, <https://www.library.rochester.edu/services/collection-strategies/sustainable-scholarship>; and "Sustainable Open Scholarship Working Group," Office of the Executive Vice President and Provost, The University of Texas at Austin, <https://provost.utexas.edu/initiatives/sustainable-open-scholarship-working-group/>, all accessed November 30, 2021.
2. David W. Lewis, "The 2.5% Commitment," IUPUI ScholarWorks, September 11, 2017, <http://doi.org/10.7912/C2JD29>.
3. ACRL Research Planning and Review Committee, "2018 Top Trends in Academic Libraries: A Review of the Trends and Issues Affecting Academic Libraries in Higher Education," *College & Research Libraries News* 79, no. 6 (June 2018): 286, <https://doi.org/10.5860/crln.79.6.286>.
4. Rebecca Kennison et al., *OA in the Open: Community Needs and Perspectives*, LIS Scholarship Archive, September 11, 2019, <https://doi.org/10.31229/osf.io/g972d>.
5. Katherine Skinner, *Mapping the Scholarly Communication Landscape 2019 Census* (Atlanta, Georgia: Educopia Institute, 2019), <https://educopia.org/2019-census/>.

6. MIT Ad Hoc Task Force on Open Access to MIT's Research, MIT Faculty Committee on the Library System, and MIT Libraries, "MIT Framework for Publisher Contracts," Massachusetts Institute of Technology, last updated May 19, 2020, <https://libraries.mit.edu/scholarly/publishing/framework/>; "CRKN Licensing Principles," Canadian Research Knowledge Network, accessed November 29, 2021, <https://www.crkn-rcdr.ca/en/crkn-licensing-principles>; "Good Practice Principles for Scholarly Communication Services," Confederation of Open Access Repositories (COAR) and SPARC, January 29, 2019, <https://www.coar-repositories.org/news-updates/good-practice-principles-for-scholarly-communication-services-2/>.
7. "Good Practice Principles for Scholarly Communication Services."
8. Kathleen Shearer, *Investments in Open: Canadian Research Libraries' Expenditures on Services, Staff, and Infrastructures in Support of Open Scholarship*, Canadian Association of Research Libraries, July 2020, https://www.carl-abrc.ca/wp-content/uploads/2020/07/CARL_open_investments_report_July2020.pdf.
9. Shearer, *Investments in Open*, 2.
10. "UA Libraries Investments in Open," The University of Arizona Libraries, accessed November 29, 2021, <https://new.library.arizona.edu/research/open-access/initiatives>.
11. OACIP landing page, LYRASIS, accessed November 29, 2021, <https://www.lyrasis.org/content/Pages/oacip.aspx>.
12. Sharla Lair (@liblalair), "OACIP is a funding infrastructure that supports a multi-stakeholder funding community..." Twitter, July 29, 2021, 1:47 p.m., <https://twitter.com/liblalair/status/1420803275918565378>.

13. “IOI’s Strategic Plan for 2021–2024,” Invest in Open Infrastructure, accessed November 29, 2021, <https://investinopen.org/about/strategic-plan-2021-2024/>.
14. “Furthering the Future of Open Scholarship: Final Report & Key Findings,” Invest in Open Infrastructure, accessed November 29, 2021, <https://investinopen.org/blog/furthering-the-future-of-open-scholarship-final-report-key-findings/>.
15. Kennison et al., *OA in the Open*.
16. Raym Crow, Richard Gallagher, and Kamran Naim, “Subscribe to Open: A Practical Approach for Converting Subscription Journals to Open Access,” *Learned Publishing* 33, no. 2 (April 2020): 181–5, <https://doi.org/10.1002/leap.1262>.
17. “ULS E-Journal Publishing Program,” University of Pittsburgh Library System, accessed November 29, 2021, <https://www.library.pitt.edu/e-journals>.
18. “Innovative Math Journal Publishes First Articles,” Queen’s University Library, October 31, 2019, <https://library.queensu.ca/about-us/news-events/innovative-math-journal-publishes-first-articles>.
19. “University of Maryland Libraries Becomes the Institutional Home of SocArXiv,” University of Maryland Libraries, May 05, 2021, <https://www.lib.umd.edu/news/2021/05/socarxiv>.
20. Jessica Aiwuyor, “How Research Libraries Are Supporting Research during Remote Operations,” *ARL Views* blog, July 13, 2020, <https://www.arl.org/blog/how-research-libraries-are-supporting-research-during-remote-operations/>.
21. “University of Maryland Libraries Becomes the Institutional Home of SocArXiv.”

22. “University of Maryland Libraries Becomes the Institutional Home of SocArXiv.”
23. Judy Ruttenberg, “How SCOSS Is Strengthening Global Open Access and Open Science Infrastructure,” *ARL Views* blog, October 21, 2020, <https://www.arl.org/blog/how-scoss-is-strengthening-global-open-access-and-open-science-infrastructure/>.
24. Ruttenberg, “How SCOSS Is Strengthening Global Open Access and Open Science Infrastructure.”
25. Global Sustainability Coalition for Open Science Services (SCOSS), “SCOSS Strategy Survey,” [data set], Zenodo, September 29, 2021, <https://doi.org/10.5281/zenodo.5537416>; “Hot off the Press! SCOSS Publishes Its Strategy for 2022–2024,” SCOSS, November 17, 2021, <https://scoss.org/hot-off-the-press-scoss-publishes-its-strategy-for-2022-2024/>.
26. Judy Ruttenberg, “SCOSS 2021 Update: Library Consortia Create Path for Funding Open Infrastructure,” *ARL Views* blog, June 17, 2021, <https://www.arl.org/blog/scoss-2021-update-library-consortia-create-path-for-funding-open-infrastructure/>.

© Judy Ruttenberg



This article is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, [visit https://creativecommons.org/licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/).

To cite this article: Ruttenberg, Judy. “Sustainability Funding for Scholarly Infrastructure Needs Infrastructure of Its Own.” *Research Library Issues*, no. 302 (2021): 33–45. <https://doi.org/10.29242/rli.302.4>.

Research Library Issues

Research Library Issues (RLI) focuses on current and emerging topics that are strategically important to research libraries. The articles explore issues, share information, pose critical questions, and provide examples. Suggestions for potential themes, articles, and authors are welcome. Please submit suggestions to commteam@arl.org.

ISSN 1947-4911 <https://doi.org/10.29242/rli>

Editor-in-chief: Mary Lee Kennedy

Guest editor: Judy Ruttenberg

Copy editor: Kaylyn Groves

Layout editor: Samantha Musser

© 2021 Association of Research Libraries

ARL policy is to grant blanket permission to reprint as long as full attribution is made. Exceptions to this policy may be noted for certain articles. This is in addition to the rights provided under sections 107 and 108 of the Copyright Act. For more information, contact ARL Publications, pubs@arl.org.

Current and back issues are available on the ARL Digital Publications website, publications.arl.org/rli. The website is also where you may sign up for alerts to new releases of *Research Library Issues*.

Association of Research Libraries

21 Dupont Circle, NW
Suite 800
Washington, DC 20036
T 202.296.2296
F 202.872.0884

ARL.org
pubs@arl.org