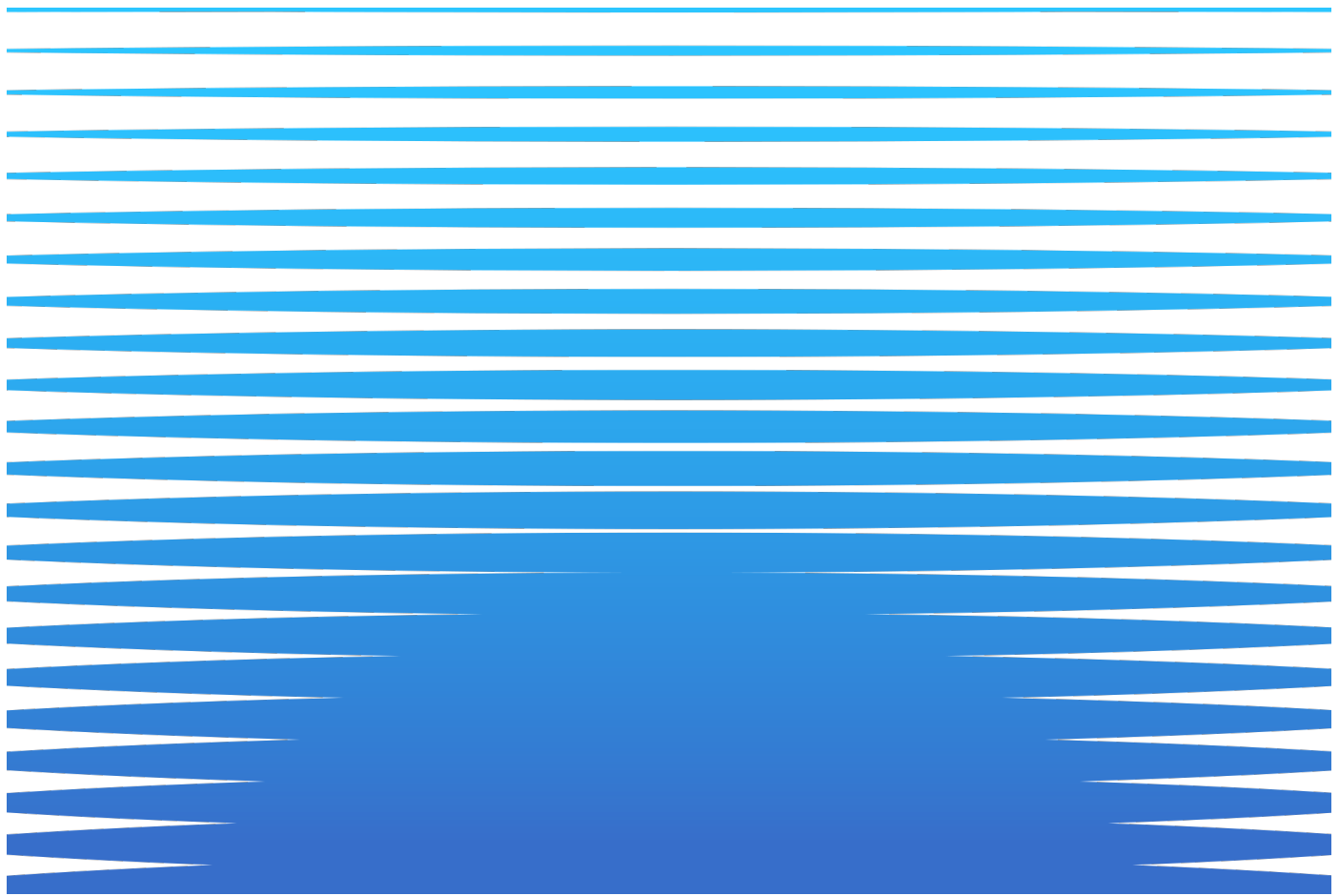


The Second Digital Transformation of Scholarly Publishing

Strategic Context and Shared Infrastructure

Tracy Bergstrom
Oya Y. Rieger
Roger C. Schonfeld





Ithaka S+R provides research and strategic guidance to help the academic and cultural communities serve the public good and navigate economic, demographic, and technological change. Ithaka S+R is part of ITHAKA, a not-for-profit with a mission to improve access to knowledge and education for people around the world. We believe education is key to the wellbeing of individuals and society, and we work to make it more effective and affordable.

Copyright 2024 ITHAKA. This work is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of the license, please see <https://creativecommons.org/licenses/by/4.0/>.

We thank STM Solutions and STM members American Chemical Society, Elsevier, IEEE, Springer Nature, Taylor & Francis, and Wiley, for providing sponsorship support that made this project possible.

Table of Contents

Introduction	3
Summary of Recommendations.....	5
Definitions and Methodology	7
Strategic Context.....	9
Transitioning towards Service Provision	9
Consolidation and Competition	11
Humans and Machines	13
Establishing a Trusted Global Public Good.....	15
The Academic Reward System.....	16
Publishing Organizations	17
Divergent Perspectives	19
Shared Infrastructure Today	21
Identifiers	21
Enterprise Publishing Systems	24
Discovery, Collaboration, and Trust.....	28
Preservation.....	30
Reflections on Organizational Models and Governance	32
New Opportunities for the Shared Infrastructure	33
Spine of the Scholarly Record.....	33
Research Integrity.....	35
Making Meaning.....	37
Supporting New Business Models	39
Alternative System	41
Concluding Remarks	42
Appendix A: List of Interviewees.....	44
Appendix B: Interview Questions	45

Introduction

Today, the scholarly publishing sector is undergoing its second digital transformation. The first digital transformation saw a massive shift from paper to digital, but otherwise publishing retained many of the structures, workflows, incentives, and outputs that characterized the print era. A variety of shared infrastructure was developed to serve the needs of this first digital transformation. In this current second digital transformation, many of the structures, workflows, incentives, and outputs that characterized the print era are being revamped in favor of new approaches that bring tremendous opportunities, and also non-trivial risks, to scholarly communication. The second digital transformation requires shared infrastructure that is fit for purpose. It is our objective with this paper to examine the needs for shared infrastructure that will support this second digital transformation.

Through these two transformations, the scholarly communication sector has grown only more complicated. At the same time, the differences between scientific (STEM) journal publishers and humanities and social sciences (HSS) publishers have grown only starker; the major commercial publishers and adjacent major STEM societies face a somewhat different set of opportunities and imperatives than the mostly smaller HSS societies as well as the university presses and commercial publishers that focus on HSS journals and monographs. The sector also includes a variety of research institutions such as universities, whose researchers are the authors of most scholarly publications, and which act strategically in scholarly communication mainly but not exclusively through their libraries.¹ Also included within the sector are consortia of these research institutions, many of which may have their roots as buying clubs but are increasingly taking on an even more strategic role in providing systems and negotiating innovative deals that stand to transform scholarly communication. And of course the sector includes funders, many of which have pursued efforts to transform scholarly publishing towards more open and in some cases less commercial directions. It additionally includes institutional and discipline-specific repositories and other alternative distribution models, operated by organizations in all of the foregoing categories, that in some cases complement the existing publishing models and in others attempt to build alternatives to them. It is no surprise that looking across these sector participants, and others, there is often a lack of strategic alignment.

A robust and nimble infrastructure is imperative to support the vital work of scholarly communication and effectively and efficiently meet the emerging service needs of different stakeholders. As we outline in the report, all kinds of publishers, repository services, and related providers that offer publishing services (hereafter collectively, “publishing organizations”) rely on various elements of infrastructure in many key parts of their work, and it forms a foundational part of their technology stack and service framework. A good deal of valuable infrastructure is currently provided on a shared basis, whether on commercial or other terms,

¹ This report focuses on academic creators and users of scholarly research, but we note that a variety of other sectors participate in this work, particularly for the scientific literature, including healthcare and corporate sectors. See for example: Roger C. Schonfeld, “Barriers to Discovery of and Access to the Scientific Literature in the Corporate Sector,” *Ithaka S+R*, 16 June 2016, <https://doi.org/10.18665/sr.283028>. We recognize the importance of engaging these user communities in implementing a number of the recommendations that follow.

although, as we recognize in the report, whether various components of the infrastructure are shared or standalone can vary across categories and market conditions.

In this project, we define shared infrastructure as broadly as possible, and we published a landscape review earlier in the project intended to show just how broad this sector has become.² Recognizing the vast array of standards, systems, and tools in many of the categories of shared infrastructure, in this paper, we examine only four critical categories of shared infrastructure in some depth, looking for successes, unmet needs, and opportunities in each. Given that many infrastructure services were designed to support the first digital transformation, it is no surprise that we find opportunities for structural and functional improvements in each of the four critical categories:

- to generate greater coordinated value from *identifier providers*,
- to ensure a competitive marketplace for publishing through *enterprise publishing systems*,
- to address opportunities at the intersection of *discovery, collaboration, and trust*, and
- to ensure long-term *preservation* of the scholarly record.

Throughout this project's interviews, we repeatedly heard that new investment in new forms of shared infrastructure is required as well as some degree of agreement on the broad purposes that this shared infrastructure should serve. What constitutes critical infrastructure, however, was a topic on which we heard many diverse opinions, indicating a lack of alignment within the field. We have identified a variety of categories where new forms of shared infrastructure should be developed, recognizing that in some of these areas there will not be unanimity across all parts of the scholarly communication community about needs or prioritization. Key reasons that new infrastructure is needed include:

- to address the growing atomization of the scholarly article,
- to ensure the trustworthiness of the scholarly record,
- to enable deeper meaning to be drawn from research outputs,
- to address new business models, and
- to provide an alternative system for scholarly communication.

Looking across the gaps that exist today, we observe that few are primarily the result of technical challenges. Rather, they are the result of stubborn strategic, governance, and business model impediments. In some categories, for example, one group of publishing organizations sees the imperative for shared infrastructure where another group sees the opportunity for competitive differentiation. In other key categories, governance of the shared infrastructure extends beyond publishing organizations, even if they are well-aligned with one another, adding a further layer of complexity. And there is the ever-present issue of the business model and investment case—

² Oya Y. Rieger and Roger C. Schonfeld, "Common Scholarly Communication Infrastructure Landscape Review," *Ithaka S+R*, 24 April 2023, <https://doi.org/10.18665/sr.318775>.

who pays, who will pay, and for what—which in turn provide incentives for innovation or inaction.

Ultimately, the gaps—and the challenges that cause or exacerbate them—are also opportunities for those prepared to invest in the future of infrastructure to support scholarly communication. It is with a strong sense of opportunity and optimism that we deliver our recommendations.

Summary of Recommendations

Throughout the paper we offer recommendations on how to build on, protect, and improve the current shared infrastructure and opportunities for creating new infrastructure services. Looking across the specific categories of infrastructure, our basic recommendations include *immediate investments* as well as *longer-term policy and research collaborations*. They are summarized here (with numbered references to the recommendations as they appear in the report).

We recommend a series of *immediate investments* by individual parties or groups of them to advance the interests of science and scientists.

- We recommend that identifier providers collaboratively create an institutional toolkit that provides concrete cost-benefit analyses and implementation strategies and tactics to guide research institutions (Recommendation 3).
- We recommend that smaller and midsize publishing organizations, looking across organizational models and governance types, find their collective voice in the enterprise publishing systems marketplace, to ensure that they are advocating effectively for their long-term strategic interests (Recommendation 4).
- We recommend that enterprise publishing systems providers invest in key functional and architectural priorities, for example to advance research integrity, open access business models, and the atomization of the research article (Recommendation 5).
- We recommend that publishing organizations participate in digital preservation services to ensure long term access and preservation, recognizing that preservation is a lifecycle process, not a final stage after publication (Recommendation 9).
- We recommend that publishing organizations and libraries invest not only in the preservation of the final published outputs of the research project but also critical and irreplaceable sources, such as observational data and primary source materials (Recommendation 10).
- We recommend that publishing organizations, universities, and infrastructure providers share trust signals with one another, through ORCID or, if appropriate, other services (Recommendation 14). We similarly recommend that publishing organizations openly share with one another threats to research integrity, for example rejections for fraud, retractions, and significant corrections (Recommendations 15 and 16).
- We recommend continued investment in automated editorial tools that can support the detection of fraud and misconduct within manuscripts and additional investment from

infrastructure providers to ensure that retractions (and perhaps even many corrections) are made more visible (Recommendations 17 and 18).

- We recommend that publishing organizations collaborate with one another directly or through third parties to distribute appropriate portions of the scholarly record in support of new types of translational and analytic services, including those utilizing AI (Recommendation 20).
- We recommend that scholarly publishing organizations dramatically increase their advocacy for investments that help the broader public understand scholarship and help them engage with trustworthy science (Recommendation 21).
- We recommend that providers of the infrastructure that supports open access business models, and their supporters, critically examine opportunities to generate additional support for their work and new models to drive its development (Recommendation 22).
- We recommend that advocates for and developers of alternative publishing models build solutions focused on how to meet the day-to-day needs of authors, researchers, and other end users more effectively, staying abreast or ideally ahead of competitors (Recommendation 24).
- We recommend that advocates for and developers of alternative publishing models assess the expense necessary to address transformative goals, including ongoing maintenance and reinvestment (Recommendation 25).

We recommend a substantial increase in *longer-term policy and research collaboration* across what have been separate interests, including publishing organizations, universities (including through their libraries and research offices), and funders. Many of these collaborative initiatives should over the medium-term result in the creation of new, improved, or revamped infrastructure. While we view these as longer-term collaborations, we recommend that they be launched immediately as they will require more time to achieve a payoff.

- We recommend that the governing boards of identifier providers form a coordinating structure to address the current fragmentation of identifiers and maximize their collective impact on strategic priorities (Recommendation 1).
- We recommend identifier organizations and publishing organizations commission a study to examine existing efforts to use identifiers as a component of the research integrity improvement framework as well as to identify further opportunities to do so (Recommendation 2).
- We recommend that publishing organizations and research libraries conduct a user-centric study of how signals of trust and authority in research are actualized in the current discovery and access ecosystem (Recommendation 6).
- We recommend that publishing organizations act to foster a competitive market among discovery and analysis providers, which are increasingly important for establishing trust in the scholarly record and collaboration among scholars (Recommendation 7).
- We recommend that universities, libraries, and open advocates convene to determine their strategic position on trust and authority in research and consider what combination of policy

advocacy and direct investment will allow them to address their interests and priorities (Recommendation 8).

- We recommend that libraries and cultural heritage organizations along with content producers review and assess the preservation gaps and investment needs for services to promote long-term access to knowledge (Recommendation 11).
- We recommend that publishing organizations, funders, senior research officers, and perhaps others design a model for how the scholarly record will be organized in an atomized environment (Recommendation 12).
- We recommend that repositories, publishing organizations, and other parties grappling with the atomization of the scholarly record ensure that preservation considerations are centered in their plans (Recommendation 13).
- We recommend that publishing organizations and university leaders together examine the best long-term models for defining the boundaries of the scholarly record (Recommendation 19).
- We recommend that universities and publishing organizations convene a strategic discussion, informed by user research, to develop new approaches to connecting content with users, both humans and machines (Recommendation 23).

In the body of the report that follows, we provide an overview of our findings and analysis that yields these recommendations.

Definitions and Methodology

Scholarly communication is the process through which research products and outputs (such as articles, audiovisual materials, data, code, and research methods) are shared, assessed, improved, disseminated, and preserved in a variety of modes including through formal and informal publications. In the digital environment, shared infrastructure has emerged as the key enabler for delivering the services that authors and readers need. It is composed of standards, platforms, technologies, policies, and the communities that enable and support them.³ Services like reference linking, repositories, identifiers, single sign-on, and digital preservation have supported the digital transformation of scholarly publishing, achieving real efficiencies for all stakeholder communities. The ultimate goal of publishing organizations and the shared infrastructure is to support researchers, first as authors and creators with ideas and results to share and to receive credit for and second to discover, access, share, and use relevant and trustworthy materials, in all cases as effortlessly as possible.

Developing, maintaining, and sustaining fit-for-purpose community infrastructure is a challenge, particularly when the technology, policy, and business environments are in flux and user behavior and needs are evolving. It is necessary to sustain and, in some cases, improve

³ For a survey of definitions that explain infrastructure and describe its attributes, see: Saman Goudarzi and Richard Dunks, “Defining Open Scholarly Infrastructure: A Review of Relevant Literature” (Version 2), *Zenodo*, 21 June 2023, <https://doi.org/10.5281/zenodo.8064102>.

existing shared infrastructure as some elements of it become more important while others decline in value. Given the proliferation of providers, the sustainability of some shared infrastructure elements will be dependent on the competitive differentiation among them. Infrastructures are known to be embedded in ways that only become visible upon breakdown or glitches. It is natural for outside observers to take an infrastructure component for granted if it is working well and has become well-established and embedded, such as metadata standards that facilitate discovery and access. This makes it more difficult to focus on the seamless aspects of shared infrastructure than it is to describe its gaps and challenges.

Scholarly communication and its supporting infrastructure pose a complex and in some ways contentious set of topics even in individual countries and geographic regions. At the same, many view the scholarly communication system as actually, or at least ideally, global. In this project, we made the scoping decision to focus on a group of countries that have a good deal of commonality in their research and scholarly communication practices. Thus, we principally focused on Anglophone and EU countries, with more limited engagement in Africa, Asia, and Latin America.

The aforementioned high-level overview of the shared scholarly communication infrastructure published in April 2023 provided scoping for the various forms of shared infrastructure examined in this report. It is based on the interviews we conducted with 49 infrastructure service providers in spring and early summer 2023. The interviewees included publishers, librarians, advocates, analysts, funders, and policy makers to further examine the strategic context behind the development of shared infrastructure and to assess its effectiveness. Given the vast array of standards, systems, and tools in many of the categories of shared infrastructure, the aim of this review was to provide illustrations of representative elements by category. Appendix A includes a list of the individuals we interviewed.

The interviews were framed around several key questions:

- What factors are driving change in scholarly communication and its infrastructure?
- What elements of infrastructure provision are working well?
- Where are there gaps or new opportunities for shared infrastructure?

See Appendix B for the full list of questions explored during the interviews.

We thank STM Solutions and STM members American Chemical Society, Elsevier, IEEE, Springer Nature, Taylor & Francis, and Wiley, for providing sponsorship support that made this project possible. We benefited from project-specific advice or feedback on earlier drafts from a number of individuals, including IJsbrand Jan Aalbersberg, Sophia Krzys Acord, Laird Barrett, Jean-Claude Burgelman, Tara Cataldo, Philip Cohen, Perry Collins, George Cooper, Martin Paul Eve, Erin Gallagher, Steven Heffner, Chelsea Johnson, Rose L’Huillier, Lauren Kane, Hylke Koers, Kimberly Lutz, Clifford Lynch, Mark McBride, Courtney Pyche, Pongracz Sennyey, Chris Shillum, Jasper Simons, Katherine Skinner, Eefke Smit, Laura Spears, Suzanne Stapleton, Todd

Toler, Paul Tuten, Craig Van Dyck, Ralph Youngen, and Charles Watkinson. We thank Joanna Dressel for her substantial contributions in scheduling and coordinating our interviews.

The research and analysis are solely the work of the project team members, and we accept all responsibility for the report.

Strategic Context

In this section, we provide a high-level outline of some of the key strategic contexts faced by publishing organizations and infrastructure providers. We acknowledge here, but do not further examine below, some of the broader factors at play, such as the macroeconomic uncertainty that was growing as we conducted this project. Our focus is on the sense of great opportunity—and substantial uncertainty, grave challenges, and diverging perspectives—amidst this second digital transformation.⁴

Transitioning towards Service Provision

Scholarly publishing is in the midst of a substantial transition, from a model that was more centered on selection functions and editorial work, and towards a model in which funders and research producers seek services to distribute their research. In the most complete version of such a transition, one can imagine a publishing organization that serves essentially as a “white labeled” service provider for a research producer, but such a complete outcome may not be an end state we will see very often. Today, publishing organizations fall into a variety of places along this spectrum, and the transition is occurring at a different pace in different parts of the sector. There is a rich complexity of factors driving this transition, including scholarly incentives and the rise of open access, as well as broader changes in the environment for digital content distribution and protection.

The strategic context includes the transformation of business models for content broadly in the digital environment. These include ad-supported free access (for example through YouTube), as well as traditional subscriptions and hybrid models (for example through Substack). Copyright enforcement can remain a tactic for content producers and distributors, particularly for piracy at scale, particularly as a delaying move, as we saw with music labels and Napster. In the long run, though, the winning strategy has proved to be business model innovation that responds to particular customer categories and market sectors, as we saw first with Apple’s music business and more recently with Spotify. Of course, even this kind of success brings enormous disruption to the overall ecosystem, as musicians, journalists, and many others have experienced. The question for incumbent consumer businesses is not how to stop this transformation altogether but rather how to adapt. To be sure, there are important differences between consumer content businesses and scholarly publishing, which is not consumer facing and has functions that go well beyond content distribution. Still, scholarly publishers have been coming around to this

⁴ As one important source for strategic context, we recommend the STM Trends reports, including the recent STM Trends 2027 output, “Level Up!” available at <https://www.stm-assoc.org/trends2027/>.

mode of thinking about how to adapt their businesses, particularly as they consider how to respond to open access policy dynamics.

Open access is now a factor for a growing share of scholarly publications.⁵ In the United Kingdom and European Union, government and funder policy initiatives have driven change largely through gold-based models which have matured into transformative agreements that now seem likely to be challenged through rising interest in diamond and repository-based green open access models. In the United States, the policy landscape that will result from the Nelson Memo will also drive an increase in immediate open access, certainly for federally funded materials, although there is some contention about which models will be utilized for compliance.⁶ Other geographies include a range of models as well. Given the resource disparity between the research institutions, including across regions and countries and within them, one of the critical issues is how to ensure that all research can be communicated efficiently in ways that ensure trustworthy discovery and access.

Business models that provide fairly direct rewards for growing the volume of articles have changed the marketplace in a variety of ways. Perhaps most importantly, the interest in sourcing articles has translated into a growing strategic focus on improving the author experience.⁷ This is a critical element that several major publishers see as a vital strategic priority in terms of competitive differentiation. Additionally, transactional processes such as managing APCs, transformative agreements, and compliance with funder mandates require new platforms, skill sets, and organizational structures. These new business models motivate publishing organizations to publish articles they otherwise might not have solicited or accepted, resulting in the inclusion of authors and works of scholarship in the scholarly record while at the same time introducing a threat vector or risk for research integrity, as we discuss in a section below. Finally, the transition to these new business models has provided an opportunity for new entrants into the marketplace, bringing new forms of competition including, at least in some cases, on price. While new entrants enter the market, transformative agreements, in particular,

⁵ There are several models for open access. Green is a version of the publication archived online not through the publisher. Gold denotes immediate open access by the publisher usually based on a publishing charge. Diamond refers to a publication model in which journals and platforms do not charge fees to either authors or readers. Transformative agreements are negotiated between institutions (e.g., libraries, national and regional consortia) and publishers to bundle together subscription and open access publishing fees with a goal to transition to greater levels of open access. For more expansive definitions of these and other terms, and some discussion of them, see: Lisa Janicke Hinchliffe, "Seeking Sustainability: Publishing Models for an Open Access Age," *The Scholarly Kitchen*, 7 April 2020, <https://scholarlykitchen.sspnet.org/2020/04/07/seeking-sustainability-publishing-models-for-an-open-access-age/>.

⁶ Lisa Janicke Hinchliffe, "The Double-Cost of Green-via-Gold," *The Scholarly Kitchen*, 25 April 2023, <https://scholarlykitchen.sspnet.org/2023/04/25/green-via-gold/>. See the full memo as published by the Office of Science and Technology Policy: Alondra Nelson, "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research," *OSTP*, 25 August 2022, <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-access-Memo.pdf>.

⁷ Colleen Scollans, James Butcher, and Michael Clarke, "Author Experience (AX): An Essential Framework for Publishers," *C&E Perspectives*, Clark & Esposito, 22 September 2022, <https://www.ce-strategy.com/2022/09/author-experience-ax-an-essential-framework-for-publishers/>.

are seen by many interviewees to continue to drive consolidation among incumbents, as described in greater detail in the following section.

This strategic context has also raised questions about the nature of the publishing services that are needed and alternative models for how to organize them. There is a long-standing effort to build repositories, at institutional, disciplinary, funder, and in some cases national or supra-national levels, to promote open access, reduce costs, and improve speed relative to Gold models. In some fields and geographies, these have been transformative, while in others they have thus far been additive. Most of these services offer more streamlined review compared with traditional editorial and peer review processes, but both traditional and repository-based models have a wide range of standards and processes. Low-cost publishing models and alternatives, including some diamond models and repository models, may yet prove to be a source of disruptive competition for the sector. Some interviewees contemplate a future in which major publishers are not meaningfully distinguished from one another by brand and other factors but rather come to serve as contracted suppliers for open access publishing services, believing that this second digital transformation cannot be complete without this essential shift.

Notwithstanding these broad transitions, there are key exceptions as well. So long as publishing is used by universities as a key component of their academic personnel evaluations for recruitment, tenure, and promotion, there will be an essential need for the sector to retain many of its underlying characteristics that are ultimately about assessment. This presents a foundational drag on whatever transformations advocates might wish to see. Separately, while the decline of print as a format for publications in many STEM fields is a foundational element of many of these transitions, print remains important for monograph publishing in the humanities and social sciences. The challenge of monographs going “out of print” in response to low levels of demand has eroded in response to new printing technologies such as print-on-demand. There nevertheless persist substantial questions about the extent to which monograph publishing will transition to open access, let alone some of these second order digital business transformations that we discussed in this section.

Consolidation and Competition

The strategic directions of scholarly publishing, and marketplace dynamics of publishing organizations, are essential to understanding the landscape for shared infrastructure. For traditional publishing activities, this includes two major forms of consolidation that have swept the marketplace, even as there are several important new entrants that have brought renewed competitive vigor. Additionally, particularly for some of the largest publishers, strategy is as much around the development of a platform, analytics, or services business, as it is around scholarly publishing. Some interviewees see this marketplace dynamic as the problem that is to be fixed—and envision alternatives to it.

During the first digital transformation of scholarly publishing, there were a number of notable acquisitions and mergers, including Elsevier’s acquisitions of Academic Press and Cell Press, Wiley’s acquisition of Blackwell, and the deal that brought together Springer and Nature. This

activity created the small set of largest houses that currently populate the landscape. Today, further acquisitions are most likely either to bring scale to open activities (witness Wiley’s purchase of Hindawi) or among other publishers (witness DeGruyter’s acquisition of Brill).⁸

But another form of consolidation is very much still in play. Many independent societies have housed their publishing with some of the largest publishers. During the first digital transformation, this was particularly important because it allowed them to gain access to services such as journal hosting and other infrastructure as well as to a global sales team. During this second digital transformation, societies are seeking to access the umbrella of global transformative agreements that large publishing houses have negotiated. Experts believe it is difficult for a society to leave their publisher partnerships, meaning that this form of consolidation may be no less permanent than an actual acquisition.⁹

Consolidation of both types leads some to forecast a much more concentrated marketplace than we have today, with less than a dozen independent publishers and far less of a long tail remaining. Should this consolidation scenario play out, each remaining independent publisher could be expected to have an essentially separate technology stack that it develops itself. This could in turn lead to questions about the implications for certain types of shared infrastructure, particularly the enterprise software categories we discuss below, as well as the possibility of disjointed user experiences.

At the same time, there are new categories of competition. The transition to gold open access models created an opportunity for a new group of pure OA publishers, several of which experienced skyrocketing growth rates through a combination of mega journals and more traditional offerings, almost all driven through article processing charges (APCs).¹⁰ While these new entrants use a variety of shared infrastructure, they are probably more likely to build aspects of their own infrastructure than similarly sized traditional publishers, given their distinctive OA-only needs.

As several of the largest publishing organizations have sought to diversify their lines of business over the past decade or more, a new category of competition has been developing in platform, analytics, and services businesses. This includes several discrete strategies, including Wiley’s “solutions” for publishers, Elsevier’s “research management” for universities, and Sage’s

⁸ Roger C. Schonfeld, “Will Humanities and Social Sciences Publishing Consolidate?” *The Scholarly Kitchen*, 22 February 2023, <https://scholarlykitchen.sspnet.org/2023/02/22/hss-publishing-consolidate/>.

⁹ Michael Clarke, “Navigating the Big Deal: A Guide for Societies,” *The Scholarly Kitchen*, 4 October 2018, <https://scholarlykitchen.sspnet.org/2018/10/04/navigating-the-big-deal/>.

¹⁰ See Figure 10 in Philip A. Sharp et al., “Access to Science and Scholarship: Key Questions about the Future of Research Publishing,” November 2023, <https://assets.pubpub.org/d535ifal/Access%20to%20science%20and%20scholarship%20-%20MIT%20report%20v1.4-41701631814319.pdf>.

“technology” for libraries.¹¹ As a result, these publishers now also compete with pure-play infrastructure providers (such as Clarivate and Digital Science among others).

As the examples above suggest, the largest publishing houses are not by any means identical in their forays into new business lines. Each has a strategy that is distinguishable from its peers, both because of the different nature of their publishing lists and even more importantly because of their different ventures into platform, analytics, and services businesses. The result is that their interests, while sometimes aligned, in other cases are not. This has tremendous implications for where they can align as both users as well as creators and developers of shared infrastructure.

University presses are an important type of publishing organization, especially for monographs. They have different kinds of organizational dynamics, as their university-based identity has tended to reduce otherwise natural opportunities to seek scale through consolidation. While there have been efforts made at cross-university press models,¹² these have been exceptions that have not always proved sustainable. Several university presses have developed solutions businesses to serve one another in areas such as fulfillment, warehousing, and digital distribution. Others have been integrated into their universities’ libraries with the goal of aligning on values and strategy.

Humans and Machines

The longstanding purpose of scholarly communication—to support the dissemination of knowledge through human readership and authorship—is itself in transition as we move to a future where more and more creation and consumption of the scholarly record involves machine-to-machine communication and artificial intelligence. Over time, infrastructure will be a key enabler of this transition.

Many interviewees observe that the scholarly record may be atomizing into component elements such as data, code, and research methods. A publication may be understood to incorporate the article identified as the version of record, related to a preprint in a disciplinary server, and multiple datasets in various repositories, as well as code, methods, protocols, and so forth. Increasingly, medical and other fields are assessing the viability of “article extenders,” meant to recap the contents of research publications for non-specialist audiences via videos, infographics, or other formats. Ideally, all these individual components will be connected together to represent the complete set of outputs of a given research project. Large publishers in particular believe that remaking the scholarly record for computational analysis and reproducibility is an important priority.

¹¹ See Wiley Partner Solutions at <https://www.wiley.com/en-fr/business/partner-solutions>; Research & Academic Management at <https://www.elsevier.com/solutions/research-and-academic-management>; and Technology from Sage at <https://www.technologyfromsage.com/>.

¹² See for example: “USU Press Joins University Press of Colorado,” University Press of Colorado, 24 November 2014, <https://upcolorado.com/about-us/news-features/item/30-usu-press-joins-university-press-of-colorado>.

It should be noted that the atomization of the scholarly record raises questions about assessing the quality, integrity, and impact of research at various levels beyond the version of record alone. The traditional model involves the editorial and peer review of a single manuscript submitted with the intention of it becoming the version of record and impact measurements around the resulting publication. By contrast, in an atomized ecosystem, it could be that each constituent element would need to be reviewed and measured separately. Today, editorial and peer review is an essentially human activity that cannot scale indefinitely. Moreover, it is unclear how humans might review components of the scholarly record that are ultimately created by machines for machines—or if other forms of review could stand in for us. It is comparatively easy to pose questions, but few models have emerged that provide very much clarity about what we might expect.

Another approach, which some see as complementary and others as an alternative, is allowing users to conduct analysis of the literature using text and data mining techniques. Rather than atomizing the literature into a series of linked objects, a number of services and tools support non-consumptive analysis of the text by providing secure access that respects rights and aggregates at scale. While aggregators provide one set of services for such work by scholars, there is growing interest from the commercial sector, for example from pharmaceutical companies, to access such materials at scale for their own analytical purposes.

Over the past year, strategic considerations (and in some cases moral panic) around generative AI have gripped many sectors, and scholarly publishing is no exception. For publishing organizations, the strategic considerations appear to be clustering into several categories.

First, there is product and monetization. A number of major publishing organizations are investing in discovery and analysis tools and techniques that utilize large language models and other techniques. Some of these will be monetized as products while others will add value to existing products. Organizations best prepared to productize these kinds of features seem to have two advantages: first, access to large amounts of content and/or metadata; and second access to the resources, including talent, necessary to develop them. Relatedly, publishers are interested in licensing access to their content to third parties as training data or otherwise, which is related to the text mining access mentioned above. With this in mind, rights and restrictions is becoming an important topic, including how open access connects with this type of monetization. There are also business and technical questions about how to provide content to third parties for machine analysis, perhaps especially for the long tail of smaller publishing organizations that are unable to do so on their own. Both on the product and the third-party monetization side, there is a risk of reinforcing the drive to scale, which is powering consolidation across the sector, as discussed above.

Another impacted area is for authorship, fraud, review, and integrity. AI-powered applications have been in use for a number of years to support the review process, including both peer review as well as copyediting. Some interviewees anticipate that there is greater opportunity here, even suggesting that AI applications could one day replace peer review. In some fields, interviewees see great potential for scientists, in particular, to leverage large language models to prepare scholarly manuscripts, raising both the promise of greater efficiencies in accelerating the

transition away from human authorship of a version of record as the key output of a given research project—as well as a bevy of potential ethical issues.¹³ A number of interviewees shared with us concerns about a substantial growth in fraudulent submissions generated with AI techniques, which one termed “the third wave of fraudulent content submission,” for which they felt unprepared.

Establishing a Trusted Global Public Good

Scholarship is, at least in an idealized form, a trusted global public good. It is said to be a global public good insofar as increasing access to knowledge globally does not diminish the value of knowledge to anyone else, particularly in light of globally shared interests in pursuing societal goals and human understanding. It is expected to be high quality and trustworthy, generating broad public support for the expertise that it reflects. All this to say, it has become clear in recent years how far away we are from the ideal.

Sadly, some of the challenge comes from within academia. It has become clear that the incentives to commit academic fraud and misconduct far outstrip the consequences, and the result is a rash of manipulated images, fabricated data, paper mills, and other threats to the integrity of the scholarly record. Today, responsibility for addressing these threat vectors and investigating suspicious activity is often all too diffuse, with universities, publishing organizations, and funders, among others, all having a role but rarely collaborating effectively. Maleficent actors using generative AI will introduce a new set of challenges. Quality publishing organizations are investing in expertise, tools, and processes designed to block the vectors through which misconduct can otherwise enter the scholarly record, but as recent examples illustrate, new business models operating at scale globally can nevertheless pose real challenges. Today, large publishers are existentially concerned about research integrity because the scale at which they operate introduces a number of threat vectors, but all publishing organizations have a shared interest in ensuring trust in and authority of the scholarly record, in their own work, and in scientists and science itself.¹⁴ The services for preprints and other forms of pre-publication distribution also seem ill prepared to safeguard the literature in the face of fraud, misconduct, and misinformation.

Although serious, fraud and misconduct are probably not the primary drivers of growing public mistrust in science and scholarship. In the polarized political environment in the United States and other countries, higher education and scientific institutions are among the organizations that are no longer uniformly trusted. This has manifested itself in conspiracy theories, fake

¹³ Brady D. Lund, Ting Wang, Nishith Reddy Mannuru, Bing Nie, Somipam Shimray, and Ziang Wang, “ChatGPT and a New Academic Reality: Artificial Intelligence-written Research Papers and the Ethics of the Large Language Models in Scholarly Publishing,” *Journal of the Association for Information Science and Technology* 74, no. 5 (10 March 2023): 570–581, <https://doi.org/10.1002/asi.24750>.

¹⁴ Jeffrey Brainard, “Fake Scientific Papers are Alarming Common,” *Science*, 9 May 2023, <https://www.science.org/content/article/fake-scientific-papers-are-alarming-common>; Jay Flynn, “Guest Post — Addressing Paper Mills and a Way Forward for Journal Security” *The Scholarly Kitchen*, 4 April 2023, <https://scholarlykitchen.sspnet.org/2023/04/04/guest-post-addressing-paper-mills-and-a-way-forward-for-journal-security/>.

news, and the politicization of science, among other unfortunate outcomes. In the long run, taxpayer support for science through public funding agencies underpins a great deal of research activity and publishing. It is in the common interests of universities and publishing organizations to bolster public trust in these institutions and this expertise.

The contemporary expectation that all scholarship will be shared freely, as a kind of global public good, needs to be situated historically within the era of globalization beginning in the 1990s. The era of globalization has been winding down for several years, particularly given splits that have taken place with China and Russia. These obvious divergences have placed limits on certain forms of scientific collaborations and industrial partnerships, either outright because of new regulatory frameworks or indirectly through severe chilling effects. The extent to which expectations for fully integrated global science and scholarly communication will hold up in the face of an increasingly fractious international environment remains uncertain.

Even without the emerging national security and other regulatory limits to globalization, there are several different “geographies” for scholarly communication, for example the US, Latin America, Europe, and India, as well as China. Within each “geography,” there is a discrete package of funding models and levels and the associated policy that incentivizes and shapes scholarship and scholarly communications. Beyond the policy environment specific to scholarly collaboration and scholarly communication, there are other types of regulatory divergences, for example on data, privacy, and artificial intelligence, which can serve as impediments to global solutions for publishing services and/or shared infrastructure. Also, local systems of knowing and knowledge-making are another related consideration, for example with respect to Indigenous knowledge, where the entire research and communication lifecycle needs to adhere to the principles of knowledge sovereignty. These can sometimes diverge from other practices and overall policy directions. Divergent collaboration and communication geographies and practices place real limits not only on the global reach of publishing organizations but also the global distribution of certain forms of knowledge. Open access has reduced the barriers to distributing scholarship across “geographies,” but it has had less of an impact in integrating scholarship from across geographies into a single global and trustworthy scholarly record.

The Academic Reward System

The academic research system is organized to a very great degree around competition—for research positions, for grant funding, among universities, and so forth. Peer reviewed publications continue to serve as a vital coin of the realm, and “publish or perish” continues to be a dominant concept as researchers, especially from academic institutions, continue to feel the pressure to maintain the strong research and publication record required for review, promotion, and tenure. This competitive system, reliant in certain disciplines to a great degree on citation metrics, establishes journal hierarchies (publishers in the case of books) that serve as measurements and signals of excellence. The system reinforces the value of traditional publications.

Various stakeholders have been calling for revisions to the traditional tenure and promotion process and indicators to build greater equity and flexibility in the rewarding system. This would

require a direct assessment based on the scholarly content of an article rather than publication metrics at the journal- or article-level. While initiatives such as the San Francisco Declaration on Research Assessment (DORA) are driving the change, at an academic institutional level, the assessment process and institutional prestige continue to be tightly coupled with traditional evaluation metrics.¹⁵ Progress has been slow in aligning the tenure and promotion indicators with the values and missions of research institutions and the emerging open science principles.¹⁶

As publication formats and platforms have proliferated, there have been efforts to incorporate other signals that might indicate the impact of research, for example media mentions and social media engagement, with the incorporation of these newer signals sometimes termed “altmetrics.” However, these have served to complement rather than replace traditional impact metrics and the applications for research evaluation remain underdeveloped. Although some research-funding agencies are relying on alternative metrics such as social media mentions to assess the full reach of the research that they support, the use of altmetrics at higher education institutions is immature when compared to traditional bibliometrics.¹⁷

Publishing Organizations

The publishing landscape remains rich with divergent organizational types despite the compact nature of the field. The field is dominated by several large publishers, many of whom have built their own underlying technological infrastructure to manage internal processes. Midsize and university presses can't utilize the functionality of these systems, and the market isn't robust enough to warrant extensive commercial investment and innovation for the needs of smaller publishing organizations. Because serial publications command more market share than single titles, functionality for journals drives innovation more so than monographic publishing; much of the recent investment in monographic platforms, for instance, has come from foundations, underscoring the resourcing and sustainability issues of monographic publishing. STEM publications also command a larger market share, so their needs currently drive change more so than those of humanities or social science publishing.

¹⁵ Another aspect of this call for change is providing an increased granularity in the recognition of author contributions and rewarding a broader range of contributions beyond authors to enhance the reliability of research and recognize the collective efforts of research participants. Initiatives such as the Contributor Roles Taxonomy (CRediT) aims to foster a culture of openness within the scholarly community, providing a framework for authors to document individual contributions to the work such as data curation, investigation, project administration, software development, reviewing, etc.

¹⁶ The Declaration on Research Assessment (DORA, <https://sfdora.org>) recognizes the need to improve the ways in which researchers and the outputs of scholarly research are evaluated. It criticizes the practice of correlating the journal impact factor to the merits of a specific scientist's contributions as it may create biases and inaccuracies when appraising scientific research. Initiatives such as DORA's Tools to Advance Research Assessment (TARA) are important to facilitate the development of new policies and practices for academic career assessment and to make visible the criteria and standards universities use to make hiring, promotion, and tenure decisions.

¹⁷ Wenceslao Arroyo-Machado and Daniel Torres-Salinas, “Evaluative Altmetrics: Is There Evidence for its Application to Research Evaluation?” *Frontiers in Research Metrics and Analytics* 8 (25 July 2023), <https://doi.org/10.3389/frma.2023.1188131>.

Large publishers also have the ability to mix commercial, in-house, and open infrastructure to create a best-in-breed approach as the landscape continuously shifts. Several publisher interviewees told us that the adoption of open tools is often advantageous to their workflows to fill specific needs or gaps in processes. Some are also evaluating open infrastructure components for the benefit of the field at large to increase transparency and functionality between systems. But it is most often that business needs, rather than altruism, drive adoption of open tools at this level.

Smaller, university, or mission-driven publishing organizations have various constraints that often force a competitive rather than collaborative environment. Branding, especially for university presses, is critical to their success; this, however, impedes mergers that might help stabilize multiple small presses. These publishing houses are also often thinly staffed, which makes innovation difficult. In our interviews, several small publishers spoke of their challenges in managing their back-office functions, including title management and payments. Outsourcing some functions such as customer service may be an option, but this inevitably adds costs at a time when revenue is likely diminishing.

The rise of open access publishing has disrupted the landscape with new considerations. A number of individuals representing multifaceted perspectives spoke about the specific challenge of APCs, both from a payment and receivable perspective. While larger and more agile publishing houses are able to respond to new challenges like collecting APC payments by establishing new business models and building responsive systems, small and university presses often struggle.¹⁸ Better infrastructure is needed to ensure open access is viable and sustainable. Several interviewees also commented that the current dominance of open access and the exponential need for its support have pushed other essential discussions to the background, including on the collective need for stronger linked data and the industry's continued dependence on the PDF as a format.

During the last several decades, several library publishing initiatives have emerged to support the publication of digital academic journals, monographs, and conference proceedings. Such initiatives sometimes involve partnership with a university press that is based in the same institution. Over the years several open source publishing applications also have been developed to support library-based initiatives that aim to transition to community-owned and led open access publishing programs and promote ethical practices.¹⁹ As this type of publishing advances, one of the goals is to implement sustainable and inclusive models and support the development of new publishing platforms, practices, and shared expertise.

¹⁸ Nancy Maron, "TOME Stakeholder Value Assessment: Final Report," with input from Peter Potter and the TOME Advisory Board, *Association of American Universities, Association of Research Libraries, and Association of University Presses*, August 2023, <https://doi.org/10.29242/report.tome2023>.

¹⁹ Tina Baich et al., "An Ethical Framework for Library Publishing: Version 2.0," *Library Publishing Coalition*, May 2023, <https://doi.org/10.5703/1288284317619>.

Divergent Perspectives

Within the work of “scholarly communication,” there are an array of publishing models and perspectives at play. But the perspectives have evolved substantially over time, and today we do not find that there remains a major fissure between promoters of open access versus subscription models. Ultimately, we did not find so much a clear set of coherently differentiated schools of thought about shared infrastructure as we did a series of spectra on which perspectives varied significantly. This is vitally important for understanding whether alliances can be built and sustained and where separate efforts may need to continue to be advanced in parallel.

Everyone we interviewed expressed a strong belief in the importance of creating and sustaining a shared infrastructure for scholarly communication, but the commercial and strategic rationale for doing so varies, as does the actual sense of what elements of a shared infrastructure are most important.

Some interviewees were highly critical of capitalism and/or commercial organizations, at least with respect to scholarly communication and its infrastructure. Several are concerned about the profitability of commercial providers and worry that commercial interests diverge from those of the academy, of researchers, or of science. The extreme monetization of scholarly research, including within varying open access models, creates an unsustainable model that precludes many from participating.

As part of these conversations, some interviewees advocated for the creation of new modes of publishing that disrupt the traditional publishing ecosystem. They see these new modes—for example, a certain kind of repository ecosystem—as a key form of shared infrastructure. Others believe outright disruption to existing businesses is infeasible. They are looking instead for specific interventions that might dilute the long-term control of the scholarly record by commercial firms while working with these firms today.

We also heard about the importance of funding innovation within the academic and research sectors, even as interviewees acknowledged that these sectors on their own may not be able to provide the resources necessary to launch and sustain new infrastructure. Some of the people we interviewed are advocating for new workflows for research and communication—and platforms for facilitating rapid sharing of research and data to facilitate communication before publication—that would dilute the primacy of formal publishing and publication branding. Developing and supporting such transformative tools require reliable funding sources in both prototyping and deployment stages.

A number of our interviewees work for commercial, in some cases publicly traded, firms involved in publishing or infrastructure provision. They tended to focus on the importance of creating systems and processes at scale, marketplace dynamics, and the strategic directions of science and the academy. Interviewees from traditional publishers spoke about their own commercial and strategic rationale for creating and maintaining a common infrastructure. Many commercial providers are comfortable thinking as community members interested in

governance, representation, and business model issues for the shared infrastructure, at least for the portions of it that benefit them.

Most interviewees from disruptive and academy-owned publishers also described their own commercial and strategic rationale for creating and maintaining shared infrastructure. While some noted preferring certain forms of governance or business models for this infrastructure, they focused mostly on that strategic and/or commercial rationale given their own goals in the marketplace.

In considering the shared infrastructure, some interviewees, particularly those from the academy or other not-for-profit organizations, express a particular focus on the values that underpin this infrastructure (beyond business requirements). We heard from them about the importance of values and principles such as transparency, open systems, equity, and a form of academy or non-commercial community control.

In our interviews, a real tension emerged between those who believe that publishing should primarily be a truly open system that maximizes inclusive participation and those who believe that publishing should primarily be responsible for securing the boundaries of the scholarly record to ensure that it is validated and trustworthy. Ultimately, most publishing organizations seek to achieve a combination of both of these visions, but it does not seem to be possible to prioritize both equally. “Quality” as a principle in scholarly publishing has profound implications for shared infrastructure. In some cases, interviewees believe strongly in quality as a core value of certain forms of publishing. Others express a strong view that quality cannot be determined objectively or that when invoked it serves too often as a mechanism of inappropriate exclusion, with some feeling that as a result nearly any contributions should be welcomed into the scholarly record. Assessment of the quality of publications, however, remains a core activity of the tenure and review process and as such remains critical for infrastructure to support.

We heard several different perspectives on privacy and anonymity in terms of data practices. Some feel very strongly that absolute anonymity is the only position that adequately protects users. Others feel that appropriate safeguards can be put into place to protect user privacy while adding new services that they believe users expect from their broader digital experiences.

Several interviewees take a decidedly globalist perspective, believing that science cannot respect borders. Other interviewees expressed concerns that, whatever their own preferences, global science and scholarly communication may be fracturing.

Several interviewees used the term “open science” but hold different definitions for the term. For some, it means making research products openly available and reusable for everyone and making the research process transparent. For others, the emphasis is on increasing international and cross-institutional scientific collaboration to address societal problems and communicating scientific outcomes to the public beyond the traditional scientific community. For others, it is about extending free and open principles from open access and open source into other parts of the scientific process. Each of these aims has different implications for infrastructure.

Finally, some interviewees believe that identifying the common layer of shared interests is the best starting point for creating shared infrastructure, even if that results in somewhat of a least common denominator outcome. Other interviewees are interested in building “coalitions of the willing,” in some cases to challenge incumbents and drive change or in other cases to carry forward what is seen to be a pressing community priority.

While there is some clustering together of groups of these views, we were not able to identify two or three or four coherent overall schools of thought on shared infrastructure.

Shared Infrastructure Today

The foregoing strategic context informs our understanding of the needs for shared infrastructure. Next, we turn to an examination of the current state of shared infrastructure. At a high level, shared infrastructure for scholarly communications is extensive and impressive. It addresses many foundational needs of publishing organizations and leverages collective action, while its providers serve as vehicles for common (or in some cases outside) investment.

Our analysis finds that there is not a single infrastructure with common purposes, key stakeholders, funding models, and related characteristics. Instead, there are a number of categories of infrastructure each of which has distinguishable characteristics from one another. Indeed, not all stakeholders agree on which categories should be included in a comprehensive understanding of the shared infrastructure. Offering a single comprehensive overview is not, in our estimate, actually possible.

Instead, our approach is to provide an in-depth analysis of several key categories of shared infrastructure, recognizing that this is not an exhaustive treatment of every possible category: **Identifiers; Enterprise Publishing Systems; Discovery, Collaboration, and Trust; and Preservation.** For each, we review their development and current dynamics, examining key themes, strategic issues, and gaps, without assessing individual infrastructure services or providers, which was outside the scope of this project. Our objective is to inform an understanding of what is working well and where there are limitations, offering recommendations for future investment in each category.

Identifiers

As scholarly communication increasingly becomes more complex, the need for common identifiers and standards continues to grow in order to facilitate discovery, access, linking, rights management, and assessment of scholarly content. Starting with object identifiers, and more recently identifiers for researchers and research organizations, the research community has benefited tremendously from the availability of what are generally termed persistent identifiers (PIDs).

PIDs are particularly well developed for the publishing and science sectors, where they show promise of enhancing automation in ways that can streamline workflows, improve user experiences, enhance trust, and drive evaluative processes. At the same time, like other

categories of shared infrastructure, they are less well developed beyond publishing and science. For example, a variety of types of cultural outputs have no native PID infrastructure and can be a challenging fit with PIDs that are not purpose built. Certain forms of Indigenous knowledge may also require different solutions.²⁰

PIDs are provided by an array of organizations. For objects, the basic standard has become the digital object identifier (DOI), which is stewarded by the DOI Foundation. Several community-governed not-for-profits such as CrossRef and DataCite, as well as several national consortia, issue DOIs and provide a variety of enabling and related workflow and metadata services. Somewhat more recently, proprietary efforts to provide interoperable researcher identity gave way to the ORCID identifier, which is stewarded and issued by the ORCID organization. Today, there is widespread agreement about the need for an identifier for research organizations; no single model has yet achieved consensus, perhaps because of divergent use cases such as commercial integrations and research attribution.

Membership organizations are prominent in identifier work, though other models exist, including services run by for-profit companies and wider collaborations between member organizations and library consortia. These and other identifier organizations are joined by somewhat similar kinds of organizations that develop and steward both formal standards and community-wide recommended practices. We group them together in examining them organizationally and strategically because they have some key dynamics in common.

First, several of these organizations were founded with a strategic purpose—or a strategic constraint. For example, a framework for linking across publishing platforms drives interoperability with the effect of reducing what otherwise would be the strong network effects advantages of the largest publishing houses. Or to take another example, if a provider focuses on the identifier alone rather than the broader service framework that could be developed, it can avoid interfering with current or future business directions of key commercial providers. The implication is that several identifier organizations were founded not simply as neutral stewards of a shared infrastructure but rather as reflections of the political economy of the marketplace: to achieve a strategic objective or with a constraint to enable the compromise necessary to their creation.

Further, whoever the original founders of these membership organizations, they were developed with the expectation, or at least the ability, to serve the widest possible array of members and communities. To take just one example: despite the challenges posed by certain key geographies, to say nothing of geopolitical tensions, there remains a dream of a single identifier encompassing every researcher across the globe. Several identifier organizations have been

²⁰ For instance, Africa PID Alliance is in its beta phase for replicating DOI for the purposes of Indigenous and cultural heritage knowledge. From its website: “The Africa PID Alliance’s mission is to secure the future of African innovation, Indigenous Knowledge, and Cultural Heritage. We are also striving to support scientists and inventors to disseminate and commercialize their research innovations. Digitizing of research outputs is a real challenge in Africa. Through the Africa PID Alliance innovative projects, we provide reliable open research infrastructure services which provides access to knowledge and metadata about digital objects closer to the wider communities, including indigenous knowledge and patent metadata, starting from Africa,” <https://africapidalliance.org/>.

characterized by growth not only in the number but also the type of members, enabling the identifiers they issue to become widely accepted standards. This growth leads to some foundational questions about organizational purpose, governance, and business models, including about membership fees, voting privileges, and the extent to which they attempt to maintain their membership or make it easy for members to depart.

As the reach of some of these organizations has grown well beyond the original founders, they have become accepted as the basic infrastructure for all sorts of organizations that are not always in strategic alignment with one another. This requires the push and pull of compromise and, if no party is fully satisfied, perhaps that is by design. Still, some large publishers express concern about “losing control” of governance for organizations that they had originally established, when their developing strategic needs are, as they see it, insufficiently prioritized. Several commercial publishing interviewees wondered whether it remained in their organization’s interest to maintain membership in community identifier organizations, pointing, for example, to their need for these identifiers to provide services that advance research integrity, such as a trusted digital identity for researchers.²¹ Other interviewees wondered if these organizations are too commercially dominated.

Since each identifier organization operates and sets direction independent of one another, there is also no shared, common direction for organizations to move forward collectively. There are also some concerns about the governance bandwidth and overhead burdens imposed by multiplying membership organizations. These considerations help to explain why several recent identifier and infrastructure initiatives have been set up inside of, or acquired by, one or more existing organizations. Even so, several interviewees questioned whether more was needed, specifically whether a model of shared governance or services might act as an umbrella over multiple projects with similar purviews to mitigate overhead and provide common objectives. A small but notable example of how a shared mindset could help is the need to link a research program with its participants, contributions made, and outcomes. There are broader goals as well.

Shared infrastructure, and the interconnectedness of research content through a network of PIDs, has tremendous benefits for researchers but also introduces complexity. For instance, if an error exists in metadata, it is difficult to rectify once the error has proliferated across systems, and it is unclear whose responsibility this may be.

Recommendations

Identifier work is highly strategic, yet the fragmentation into multiple organizations and projects can reinforce outdated objectives and models. Governance models focus on the strategic needs of the identifier organization in ways that balance those of its members. One effect is omitting key cultural products as well as certain forms of Indigenous knowledge from the

²¹ Paula Demain and Tom Demeranville, “Summarizing ORCID Record Data to Help Maintain Integrity in Scholarly Publishing,” August 2023, https://info.orcid.org/summarizing-orcid-record-data-to-help-maintain-integrity-in-scholarly-publishing/?utm_source=mailpoet&utm_medium=email&utm_campaign=blog-newsletter.

identifier landscape. Another is the opportunity to provide more organized analytic products. And, as we note in another recommendation, there are important opportunities for advancing research integrity as well. Working in a more integrated fashion could have many positive outcomes for a number of parties.

- Recommendation 1: We recommend that the governing boards of identifier providers form a coordinating structure that can draw in representatives from across the sector, for example through the advocacy/professional organizations of libraries, universities, and publishers. This structure should create a vision for how identifiers can contribute collectively to current and emergent sector-wide strategic priorities and address gaps that the more siloed structure leaves unfilled. Individual providers should align themselves to this broader vision.

An important question for the broader community involving almost all stakeholders is about the role of identifiers in advancing research integrity, for example a trusted digital identity for researchers. Identifiers can facilitate reliable connections between researchers, their scholarly contributions and collaborations, and their organizational affiliations. Persistent digital identity also plays an important role in ensuring the reproducibility and transparency of research data by identifying researchers and their affiliations. Additionally, publishing organizations can rely on identifiers to validate manuscript submissions to ensure that the content meets some of the research integrity standards.

- Recommendation 2: We recommend identifier organizations and publishing organizations commission a study to examine existing efforts to use identifiers as a component of the research integrity improvement framework as well as to identify further opportunities to do so. Such a study should not only focus on existing identifiers and identifier organizations separately but also (building on our previous recommendation) examine opportunities for a collective approach.

The challenge of adoption for research institutions is often not trivial. Understanding the landscape of identifiers and their benefits can be complicated. Individual benefit to researchers is sometimes obscure as well. Also complicated is apprehending the requirements for research institutions as they implement and integrate different PID schemes in their own infrastructures.

- Recommendation 3: We recommend that identifier providers collaboratively create an institutional toolkit that provides concrete cost-benefit analyses and implementation strategies and tactics to guide research institutions. Once again, thinking holistically across identifier types seems foundational as one means to drive adoption.

Enterprise Publishing Systems

Publishing systems offer integrated workflows for authors, editors, and designers to facilitate content management and distribution. Some systems specifically focus on facilitating the submission and processing of manuscripts and managing the review process while others provide extended functionality to support delivery, distribution, analytics, and e-commerce. In recent years, some publishing systems have been expanded to allow for newer features such as preprint deposit and dataset submission and review, the management of open access

submissions, and other functions. Platforms that couple hosting, publishing, and delivery are deployed in many scholarly societies, university presses, and other publishing organizations. Nevertheless, there is no turnkey solution that can handle all the requirements of every publishing organization as the technical infrastructures, priorities, and resources vary.

Although there are a number of enterprise publishing systems in the marketplace, large publishers increasingly develop and maintain customized in-house systems to support their large operation. Such internal investment enables them to position services such as manuscript submission and editorial management for competitive differentiation. On the other hand, the majority of mid-size or small publishing operations rely on off-the-shelf publishing platforms in order to control in-house maintenance and development costs.

Since the early 1990s, with funding from foundations or governmental agencies, several open-source publishing and hosting platforms have been developed with the goal of reforming scholarly publishing.²² This category of publishing systems aims to promote and enable community-based, collaborative, and academically-driven publishing and supports new workflows for content creation and review. The sustainability of these systems is based on adoption rates and successful fundraising efforts. The library publishing initiatives aim to combine the values alignment of open source software with open source publishing systems and institutional repositories.²³ The associated publishing systems are designed to address the needs of library publishers and values-based assessment frameworks.

On the commercial side, most of the innovation has been introduced by start-up initiatives that attract investors who provide the capital for development. If these start-ups are successful, they mature into service providers with the potential to achieve substantial market share and revenue. As the market share of an enterprise software provider develops, their publishing organization clients often find that they become “monolithic,” in the sense that the dependencies of a vast array of clients impede strategic innovation either by the enterprise provider or, as a result, among the clients. Typically, strategic innovation is enabled by a new enterprise platform—either from an existing provider willing to disrupt its own product line or a new entrant. But before this happens, the general trend is for large publishers or other commercial organizations with a strategic interest in the business to acquire mature enterprise providers, raising unavoidable questions from clients about their independence and neutrality.

Some small or not-for-profit publishing organizations feel that the overall industry suffers from a lack of innovation because of its small size. While major players can develop and maintain

²² Jean Francois Lutz, Jeroen Sondervan, Xenia van Edig, Alexandra Freeman, Bianca Kramer, and Claus Hansen Rosenkrantz, “Knowledge Exchange Analysis Report on Alternative Publishing Platforms,” *Alternative Publishing Platforms*, 21 September 2023, <https://knowledge-exchange.pubpub.org/pub/d9h2tp1x/release/1>; Kristen Ratan, Katherine Skinner, Catherine Mitchell, Brandon Locke, and David Pcolar, “Library Publishing Infrastructure: Assembling New Solutions,”

Educopia Institute, March 2021, <https://educopia.org/nglp-lib-pub-infrastructure>.

²³ The Next Generation Library Publishing project aims to transform the scholarly communication landscape by empowering academic libraries and their collaborators and providing alternatives to commercial publishing: <https://www.nglp2022.org/>.

systems for advanced features, small or not-for-profit publishing organizations feel the pressure to utilize cost-efficient out-of-the-box solutions, leaving a potential gap in innovation.

Regardless of the publishing system innovation coming from the commercial or not-for-profit sector, existing legacy systems, editorial workflows, and hosting platforms often can slow down adoption and transformation at the local level.

Many publishing organizations are agnostic about the innovation of any one particular provider. Indeed, we heard directly from some interviewees that their goal is decidedly not to sustain individual infrastructure providers but rather to ensure that they have access to needed infrastructure, a pointed distinction. The key factor for them is to ensure that market forces are adequate to enable a disruptive innovator to compete. Unfortunately, today there is some concern that such market forces are not working all that effectively, and some publishing organizations feel at least somewhat trapped by existing providers. Additionally, as they adopt new systems, smaller independent and society publishers tend to lack economies of scale to afford and sustain new system deployment.

Broadly speaking, there are several key functional weaknesses among the different enterprise providers. These include, in varying degrees:

- The need for tools as well as workflows that identify and block fraudulent submissions and those containing research misconduct;
- The need to incorporate business logic seamlessly to support an open access environment, including transformative agreements, APCs, multi-payer models, and more;
- The need to integrate preprint submission more deeply with manuscript review for publication; and
- The need to integrate manuscript submission and review more concretely with deposit and in some cases review of additional research artifacts such as datasets.

Beyond these functional needs, there are also profound architectural questions. For example, many manuscript submission and management systems were designed around the needs of a single journal and its editors. The result is that sharing reviewers and articles across a portfolio of journals—even within a single publisher—can require an array of workarounds. While some newer systems are being designed around the article rather than the journal, this model may also bring architectural limitations. To take another example, content hosting and delivery is mostly organized today around publishers, when researchers have little desire to traverse multiple publisher websites in search of their content. Services like SeamlessAccess and GetFTR have sought to reduce some of the greatest burdens on users, but it could be that a small number of high-traffic platforms will emerge to use the syndication model. In this case, we may look forward to a reduction in the need for publisher-specific content delivery sites, even while being concerned about the data policies of syndication platforms and competitive risks of this emergent market structure.

For smaller publishers, supporting back-office functions is a big challenge as there are different platforms and title management systems for different audiences. For instance, for some

university publishers, insufficient integration between their payment systems and a university's financial framework makes it complicated to manage the revenue generation process. For smaller or mission-driven publishers, there are opportunities to outsource administrative support, but each new organizational layer is likely to add new costs. Some have to operate multiple publishing platforms as no single system meets all their needs. For instance, book publishers that need to fulfill certain funder OA mandates need to use a system that will support not only chapter-level but also volume-level creation of a PDF version. Publishers in the textbook sector may need a supplemental system that allows rentals and low-level DRM use. In terms of publishing workflow support systems, there is no perfect solution yet, but small publishing organizations are looking for predictable hosting fees, good customer service, and a system that can be managed internally using existing resources and expertise.

The strategic context of consolidation is a particularly important factor for enterprise systems. On the one hand, a strong marketplace for shared enterprise systems makes it possible for smaller publishing organizations to thrive, and new publishers to enter the marketplace, without needing to create their own publishing platforms. In this sense, enterprise systems are a vital element of ensuring a competitive marketplace. But a downward spiral is also a real risk: a more consolidated marketplace reduces demand for shared enterprise systems and in turn the ability for their providers to make the investments necessary to enable the remaining independent publishers to resist consolidation.

Recommendations

Smaller and midsize publishing organizations are unable to create enterprise software on a standalone basis and are reliant on shared offerings. The largest publishers have either purchased, or created on a standalone basis, key components of common enterprise software in recognition of its strategic importance. Smaller and midsize publishing organizations that wish to maintain their independence may find the infrastructure marketplace more challenging for them over time.

- Recommendation 4: We recommend that smaller and midsize publishing organizations, looking across organizational models and governance types, should find their collective voice in this marketplace, to ensure that they are advocating effectively for their long-term strategic interests.

Enterprise publishing systems have, in varying degrees, an array of architectural and functional weaknesses and opportunities, including better supporting: research integrity, open access business models, the atomization of the research article, integrating multiple article versions such as preprints, sharing articles and reviewers across a publisher portfolio, and platform-driven syndication models. Enterprise systems enable, or thwart, the development of publishing organizations' ability to adapt to, and advance, the new strategic directions of the second digital transformation.

- Recommendation 5: We recommend that enterprise systems providers address key functional and architectural priorities. We recognize that in many cases they are doing so already, but in other cases perhaps not with sufficient investment, in which case we

recommend that their customers exercise their purchasing power to ensure that systems address their priorities.

Discovery, Collaboration, and Trust

In this section we explore a set of categories of the shared infrastructure where we see particular opportunity for innovation and disruption alike, including the discovery of scholarly materials, scholarly collaboration, efforts to establish trust and broadcast the trustworthiness of scholarship, and methods for analyzing the impact of scholarship. The strategic importance of this infrastructure is that researchers—scholars, students, and indeed machines as well—need to be able to discover the content they need, gauge its trustworthiness before utilizing it, and then collaborate in incorporating it into future scholarship. And publishing organizations have a strong incentive to maximize the discovery—and, they therefore hope, measures of usage and impact—of their publications.

The discovery category is essential precisely because of the fragmentation of the delivery/access environments, both among publishing organizations as we have discussed and also for other content types often housed in separate services. Several providers market “library discovery,” generally understood as a tool that provides discovery for all the library’s collections through its website. There are also multiple “research discovery” services, including tools that originated as citation indices, as well as free services that as a result display uncertain long-term sustainability, including Google Scholar. Research discovery is often integrated into or offered alongside various impact metrics, which are also in some cases integrated into library discovery. Some services offer tools that are intended to provide discovery of “everything” a user might reasonably seek, while others are bounded based on analyses of quality, openness, or other characteristics.

Content discovery and assessment are critical functions in the researcher journey and there is tremendous value in providing these services. Many organizations also express serious concern when a competitor, or a party with divergent interests, controls these functions. Some of the recent efforts to ensure the integrity of the scientific record have exposed the market tensions that ordinarily lay further under the surface (as we discuss in further detail in a section below on Research Integrity).

A newer model has grown through what are sometimes called scholarly collaboration networks, which host content and facilitate the sharing of scholarly resources across researcher grids. The current offerings in this category have built what can be understood as researcher-centric platforms, with a core focus on the researcher experience that generates traffic and engagement that can ultimately be monetized not only through advertising but also by providing services to publishing organizations (or competing with them), as well as to other parties interested in research and researchers.

In recent years, it has become clear that the fractured nature of publisher-specific websites has been an impediment to a strong user experience. Instead, there is a growing effort to place the content where the users are, to drive a more seamless discovery to access experience.

Syndication models and some of the standards that enable them are a major driver. While it is still early days in how this overall shift will play out, it is possible to imagine substantial opportunities to bring together discovery, trust markers, access, and potentially even collaboration through the same platform experience.

The existing infrastructure services described above are provided through strikingly different service and business models, an illustration of the unsettled nature of this market. Some are provided as or closely connected to enterprise software bundles sold to academic libraries. Others rely strongly on brand, particularly for impact measurement, and are sold as databases to academic libraries. And scholarly collaboration network platforms are monetized through publisher charges for what is essentially content marketing or through advertising models.

As mentioned, we see a particular opportunity in this space for innovation and, perhaps, disruption. New forms of AI will lead to new ways of understanding the purpose and nature of search, discovery, and academic collaboration. And, it will be essential to address integrity, trust, and authority in ways that respond not only to changes within science and scholarly communication but also external threats and risks as well. Existing infrastructure providers in these categories have access to content, which is an absolutely foundational requirement but not one that has, in recent years, impeded new competitors from entering the marketplace.

Recommendations

Trust and authority in research is becoming a more complex topic as information types proliferate and distribution and aggregation of content from different sources becomes more and more complex. Most user studies remain at the level of a single publisher, platform, or library website, rather than addressing the complicated dynamics that users experience moving across these environments. All parties have a shared interest in understanding how signals of trust and authority are functioning so that system design and information literacy approaches can be adapted accordingly.

- Recommendation 6: We recommend that publishing organizations and research libraries collaborate to conduct a user-centric study of how signals of trust and authority in research are actualized in the current discovery and access ecosystem, along with a projection of how these may develop in several different future scenarios such as greater distribution or greater centralization.

In the years to come, we expect to see substantial product innovation and disruption in this Discovery, Collaboration, and Trust infrastructure category, which is vital to everyone with a strategic interest in scholarly communication. Incumbents may wish to control or even resist some of these innovative or disruptive forces. Start-ups will continue to develop particularly as AI enables new opportunities for search, discovery, and collaboration. Publishing organizations can retain, or at least generate, substantial leverage in how this category develops given the need for metadata and content to power it. It will be important to ensure a focus on the needs of research and researchers, even if commercial interests do not directly align with these needs.

- Recommendation 7: We recommend that publishing organizations work individually, and to the extent possible collectively, to foster a competitive market for this category of infrastructure services.²⁴
- Recommendation 8: We recommend that universities, libraries, and open advocates convene to determine their strategic position on this infrastructure category and consider what combination of policy advocacy and direct investment will allow them to address their interests and priorities.

Preservation

As one of our readers pointed out, preservation is in many ways the poster child for shared infrastructure investment. Preservation benefits everyone societally, to some degree, and it needs to be done in a trustworthy way, yet every stakeholder wants to minimize their own investments.

Preservation relies on trusted, sustainable environments to provide critical functionality for the long-term stewardship of research outputs. Although enduring access has a fundamental primacy in scholarly communication, the concept of preservation as necessary and shared infrastructure is often overlooked or taken for granted. The interviewees rarely mentioned the role of infrastructure in supporting long-term access to digital content. Yet concerns persist about the stability of digital preservation across the landscape and preparedness for accommodating different types and formats of publications. More commonly adopted, shared infrastructure could both mitigate some concerns and be more agile for future development as scholarly communication continues to evolve. Some disciplines are also clearly better prepared and organized in this regard.

Libraries and archives took the lead in ensuring the preservation of print cultural heritage through selection, acquisition, and long-term storage—in some cases with redundancy by making multiple copies of materials available at different institutions. The digital sphere requires different models as maintaining a centralized infrastructure is not feasible due to the characteristics of digital content, which is both pervasive and ephemeral at the same time. The challenge is not only preserving the bits of digital objects but also being able to transition over time their affordances, software environment, and the context required for interpretation and consumption.

There are several successful initiatives and services to support the preservation of scholarly journals and monographs, including at the article and chapter level. That said, there continue to be gaps in the service infrastructure for addressing the requirements of other content types, including those that are emerging such as interactive multimedia and embedded visualizations, as well as categories of societally important content such as journalism and modern cultural production accessed via streaming services. We also note concerns deriving from the potential further atomization of the scholarly record and the reality that research materials connected to

²⁴ One model for this work might be NISO's Open Discovery Initiative, adapted to incorporate the broader aperture and strategic dynamics identified in this section.

one another may not be preserved together. Digital preservation remains a critical challenge for many institutions, but especially for cultural heritage organizations from low- and middle-income countries. To this end, international, national, and regional preservation services, advocacy organizations, and technology initiatives continue to play a critical role as needed infrastructure.

Recommendations

As the scholarly record becomes more heterogeneous, variable, dynamic, and distributed, keeping track of and archiving various versions and components (article, preprint, underlying data, etc.) will become even more complex. From a future proofing perspective, preservation needs to be seen not as a final stage in the life cycle of scholarly materials but a functional requirement that is considered at multiple stages.²⁵ This is an important collaborative effort that engages many stakeholders. Preservation should be recognized as a lifecycle process that spans from the creation to publication of information.

- Recommendation 9: We recommend that, regardless of size or disciplinary coverage, every publisher should participate in digital preservation services that aim to protect and preserve digital content for long-term access and use and that libraries act to ensure that content of importance to their communities is preserved and invest to support such work.
- Recommendation 10: We recommend that publishing organizations and libraries invest not only into preservation of the final published outputs of the research project but also critical and irreplaceable sources, such as observational data and primary source materials.

There are important categories of materials that are insufficiently preserved in the digital environment, often because the nature of how they are produced, and how they can be collected and preserved, has changed substantially. Journalism is one societally vital example. Similarly, various types of cultural production are accessed through streaming services. The ambiguous roles and unclear responsibilities of different stakeholders are compounded by copyright and technology challenges. Although they continue to support various preservation collaborations, libraries are not in a position to take leadership in preserving scholarly content that they no longer collect but rather license.

- Recommendation 11: We recommend libraries and cultural heritage organizations along with content producers initiate a major review and assessment of preservation gaps and develop an investment case so that appropriate services can be developed to promote long-term access to this knowledge.

²⁵ With funding from the Mellon Foundation and led by New York University Libraries, the Enhancing Services to Preserve New Forms of Scholarship initiative examined a variety of enhanced ebooks and developed recommendations for publishers to create digital publications that are more likely to be preservable. See: Jonathan Greenberg, Karen Hanson, and Deb Verhoff, "Guidelines for Preserving New Forms of Scholarship," *NYU Libraries*, September 2021, <https://doi.org/10.33682/221c-b2xj>.

Reflections on Organizational Models and Governance

There is an enormously rich terrain of shared infrastructure, as we documented in our landscape review and analyzed in the sections above. The providers of this infrastructure, as well as those who fund and govern it, can take real pride in this achievement. We have studiously avoided addressing individual providers and their particular strengths and weaknesses as such an assessment is outside our scope. Instead, in this section, we offer some overall reflections on the current state of the shared infrastructure from an organizational perspective.

Different categories of shared infrastructure seem to accrete providers with discrete blends of organization types and governance models. In some cases, we believe there is a basic logic to the current models, for example the membership organizations that provide identifiers and standards, as well as the commercial start-up ecosystem for enterprise systems which periodically disrupts the status quo. In other cases, such as in the areas of discovery, trust, and impact, the infrastructure is unstable or still being formed, and the appropriate governance and business models have not yet necessarily been accepted across the community.

At the individual infrastructure provider level, we note that few have achieved all of the key characteristics that stakeholders would like them to entail:

- *Trustworthiness* among clients and other stakeholders,
- *Financial returns* that yield breakeven or sufficient profit margin, depending on the model, to ensure sustainability;
- *Access to capital*, i.e. working capital for maintenance and growth capital for additional development;
- *Sustainable and expedient innovation* in response to emergent needs; and
- *Agility and flexibility* as the market changes.

To be sure, not all providers may need to possess all five of these characteristics in equal measures. For example, a start-up enterprise systems provider may optimize differently than a community governed PID provider.

As the scholarly publishing sector is rapidly transitioning to open with the increasing prominence of open access and open source, some interviewees question the business models and operating principles of the commercial infrastructure providers, particularly the large multinational organizations with significant market share. To these interviewees, the profit motivation does not align with core values of academia and they contend that the scholarly communication ecosystem should not only be open but also governed by not-for-profit organizations that are controlled by the academy. Organizations that support scholarly communication, by this principle, should not earn profit for their owners, and money generated and funds raised through business activities should go right back into running the entity.

However, as documented in several previous studies, there are obstacles to transitioning the infrastructure to a not-for-profit model that is agile, responsive, and sustainable.²⁶

In some geographies, most—not all—of the infrastructure we have examined is paid for, in whole or in part, by publishing organizations. On some level, it makes good sense for publishing organizations to pay for their own infrastructure, either individually or collectively. On the other hand, like any other cost, this is a kind of barrier to entry to the marketplace. We note this because in some countries, the government pays for a substantial portion of this infrastructure, enabling domestic publishing organizations to focus on adding value on top of it.

New Opportunities for the Shared Infrastructure

In the above section, we examined some of the challenges and opportunities for four key categories of the existing shared infrastructure. In this section, we build upon the strategic context discussed above to examine opportunities to create new categories of shared infrastructure. In some categories, shared infrastructure will be highly contested as there will be meaningful strategic consequences from the choices that are made.

Spine of the Scholarly Record

Many interviewees expect to see traditional articles atomizing into their component parts in ways that will enhance machine-to-machine communication. We are already seeing the growing importance of dataset deposit, as well as indications that research protocols, code, and other elements may themselves become first-class research artifacts. The re-use that such deposit can enable is an important shift in how science can be practiced both by humans and machines. The vision is that ultimately these various research artifacts can be re-used and recombined to advance not only replicability but also new research and interdisciplinary collaborations.

Up to now, we are mostly seeing different services used for the deposit of these different types of artifacts. For example, datasets end up in various repositories rather than “in” the journal publishing the article itself. There is a need for a set of standards, and possibly a technology layer, to help link all of these elements together. It is probably not just a matter of a link or DOI from an article to a dataset as it is. Instead, we may see the need for the elements of the scholarly record to be persistently connected with one another and actively managed through versions, corrections, retractions, and the like. Ultimately, we may see the need for a new way of understanding and visualizing the scholarly record.

²⁶ Some recent related reports include: Claudio Aspesi and Amy Brand, “In Pursuit of Open Science, Open Access Is Not Enough,” *Science* 368, no. 6491 (May 2020): 574-577, <https://www.science.org/doi/10.1126/science.aba3763>; David W. Lewis et al, “Funding Community Controlled Open Infrastructure for Scholarly Communication: The 2.5% Commitment Initiative,” *College & Research Libraries News* 79, no. 3 (2018), <https://crln.acrl.org/index.php/crlnews/article/view/16902>; “Funding Open Infrastructure,” Invest in Open Infrastructure, 5 July 2022, <https://investinopen.org/research/funding-open-inf>.

Publishing organizations tend to envision that the version of record will link out to each of the constituent elements. Even if the version of record grows shorter as its component parts end up elsewhere, in this vision it still maintains its value as the element of the scholarly record that undergoes peer review, that counts for academic career advancement, and that ultimately serves as the spine of the scholarly record. It is understandable that publishing organizations would envision this model, given that it maintains the centrality of their brands and businesses.

Others might see the world differently. For a funder, for example, the key unit of measure is the funded research project. The funded project typically will have several different articles resulting from it, as well as datasets, various coded programs, a number of methods, and so forth. Each of these has a many-to-many relationship with one another. And of course the researcher may instead see themselves as the organizing unit of measure, no matter how collaborative their research may be.

If the foreseen atomization of the research article comes to pass, there could be opportunities to rethink and possibly even disrupt the nature of the scholarly record and its leading custodians. There is a risk that the shared infrastructure supporting this atomization will have an embedded business logic that could favor one type of party or another.

This is ultimately a tremendous innovation opportunity and vehicle for disruption that no publishing organization has the ability to address on its own.

Recommendations

In an atomized ecosystem, it is not clear how all research outputs (for example, the datasets, methods, and code along with the article itself) will fit together and be organized and whether (and if so how) the concept of version of record will continue to apply. It seems that a new kind of standard and associated infrastructure is needed to ensure that the scholarly record can be effectively organized and maintained even if it is atomized. This is what we call the “spine” of the scholarly record.

- Recommendation 12: We recommend that publishing organizations, funders, senior research officers, and perhaps other stakeholders work collaboratively to design a model for the spine of the scholarly record in an atomized environment. This model should be implemented in proprietary and shared infrastructure systems across the research community. We anticipate substantial differences between parties in this model, forcing a tradeoff between each group’s strong interests and the compromises necessary to yield consensus, perhaps enabling all the different organizing models for the scholarly record to be accommodated.

Components of the atomized scholarly record currently reside or are preserved in a variety of locations. This presents challenges to ensuring stability of content over time, to long-term preservation, and to providing secure linkages between components.

- Recommendation 13: We recommend that repositories, publishers, and other parties grappling with the atomization of the scholarly record ensure that preservation

considerations are centered in their plans. Considerations about preservation workflows should be incorporated into new systems and models regarding the spine of the scholarly record, and preservation organizations should be drawn into strategic engagements about the spine of the scholarly record to ensure alignment.

Research Integrity

The publishing community has woken up to concerns about research integrity. Shared infrastructure in support of research integrity to provide confidence and trust in the methods and the findings of the research is being developed both by enterprise software providers, particularly those that work on manuscript submission and editorial review, as well as on a shared basis through community organizations. Existing infrastructure and that currently in development will address some but not all threat vectors to research integrity.²⁷

The issue that we identified as the biggest gap today is the need to identify and qualify legitimate researchers, to help editors triage submissions into more and less trusted categories. This has proved to be an extraordinarily complex challenge in an environment characterized by greater openness, globalization, and scale. For some interviewees, a key element of this gap is perceived to rest in the need for a secure digital identity for legitimate scholars, or to associate a trusted digital identity with one's validated scholarly record. We see opportunities for researcher identifiers and ORCID in particular to be used as the hub for much greater information about digital identity, in part by allowing publishing organizations and other parties to submit markers of identity into identifier records. As examples, publishers that have processed APC transactions using credit cards have substantial signs of verified identity, and they also have validated records of who has published what, as do universities that have securely linked an email address.

Another vital issue is the review of individual articles and components prior to publication as well as post publication. Prior to publication, there is a growing array of tools designed to support editorial review processes, for example to check for plagiarism or image manipulation, and these tools need to grow more sophisticated as quickly as possible. And post-publication, the growth in retractions suggests the need for increasingly strong signals throughout workflows and discovery tools to ensure that signals of mistrust accompany these retracted materials, along with the need for strong metrics and analysis for accountability. CrossRef has an important role to play in this work, as indicated by its recent acquisition of the Retraction Watch database.²⁸ Rejections or retractions on the basis of fraud or misconduct (perhaps even many corrections)

²⁷ There are a range of global challenges related to research integrity, including practices. For instance, see: Karen MacGregor, "Global Research Integrity Statement Calls for Fairness and Equity," *University World News*, 29 March 2023, <https://www.universityworldnews.com/post.php?story=20230329135629222>.

²⁸ Dalmeet Singh Chawla, "Crossref Acquires Retraction Watch Database," *Chemical & Engineering News*, 19 September 2023, <https://cen.acs.org/research-integrity/Crossref-acquires-Retracton-Watch-Database/101/web/2023/09>.

should feed back into future signals about the trustworthiness of the author upon article submission.

The boundaries of the scholarly record represent another aspect of research integrity that requires new forms of infrastructure. Of course the record has never had absolute boundaries. But in a subscription landscape, libraries played an important role in establishing the metes and bounds of the scholarly record (and what would be preserved over time) based on their selection decision-making. In a gold or diamond open access environment, libraries may have a reduced role and so other forms of boundary-setting may be required. Journal rankings increasingly have come to serve to set the boundaries of the scholarly record, although whether that is the best approach for this vital function, or whether it has the right governance and business model to allow it to serve this role without fear or favor, is not yet settled.

Finally, we note that the atomization of the scholarly record discussed above will only drive additional challenges in research integrity. It is imperative that infrastructure providers consider this carefully in their development work.

Recommendations

The work of qualifying researchers through digital identity and identity frameworks is already underway through ORCID, but there is far more that should be done with urgency to ensure these develop as strongly as possible.

- Recommendation 14: We recommend that publishing organizations, infrastructure providers, and universities partner with ORCID (or if appropriate with other parties) to share the signals of trusted identity across the community of researchers.
- Recommendation 15: We recommend that ORCID (or if appropriate other services) enable publishing organizations to incorporate retractions (and perhaps even many corrections) prominently into an author's record so that these can be made available as future signals of lowered trustworthiness.
- Recommendation 16: We recommend that publishing organizations share with one another (perhaps through ORCID or another infrastructure) information about the author identities associated with manuscripts that were rejected during the editorial process on the basis of suspected fraud or misconduct, as a future signal of lower trustworthiness.

Article review processes are developing dynamically today and will benefit from further investments.

- Recommendation 17: We recommend continued investment in automated editorial tools that can support the detection of fraud and misconduct within manuscripts.
- Recommendation 18: We recommend additional investment from infrastructure providers to ensure that retractions (and perhaps even many corrections) are made visible throughout the scholarly record and that metrics about them are made available transparently.

The boundary of the scholarly record—the line between which materials count as scholarship and which are of insufficient trustworthiness to count—are always unsettled. We note that the

mechanisms for determining this boundary have shifted without very much systematic attention, and in many ways the Web of Science services have become the default mechanism for establishing these boundaries.

- Recommendation 19: We recommend the commissioning of a project, run by a task force drawn from publishing organization executives and university leadership through their senior research officers and library leaders, to examine the best long-term models for defining the boundaries of the scholarly record.

Making Meaning

Research publishing shares scholarship within a community of peers. Scholars are trained to read the primary literature with a critical eye and to recognize that any single new study is part of a broader tapestry. Basic literature reviews, systematic reviews, and meta-analyses are among the formal techniques that are used to draw meaning from the literature. There are vast opportunities to make meaning more readily from scholarly literature, for scholars themselves, for other primary consumers of research and scholarship, and for a broader public.

Not only scholars themselves but also other researchers and practitioners, for example in corporate research environments or clinical settings, must make meaning out of the scholarly record. For some of these researchers and practitioners, existing services and features are important; for example, the clinical decision support services that are widely used in medicine. For other use cases, fields, and segments, the nature of the infrastructure needed will differ. For scholarly uses in particular, tools will build on specialized repositories, especially for some parts of biomedicine. In some cases we will see scholarly collaboration networks as the home for such infrastructure, in other cases other kinds of research analysis environments. For applied research and development and practitioners, other kinds of tools and products are required. Well tagged and structured content, for instance, allows for machine entity extraction and inference of content that promote new types of discoverability. In all cases, the ability to draw connections across disparate research objects and outputs will be foundational, and the questions we have already posed about the spine of the scholarly record may prove relevant also.

Another important part of this category will reach a broader public.²⁹ In an information ecosystem that is driven by troubling degrees of political polarization, trusted interpretations and translations of science and scholarship to inform current awareness and policymaking is absolutely vital. While some scholarly associations see scientific communication as an important element of their work, the research publishing sector as a whole tends to see this work as outside its ambit. And yet the environment of misinformation and declining trust in science means that scientific communication is not just important for its own sake but as a key element in the political process that ultimately leads to public funding of science itself. As one of us recently argued, “the very structure of the biomedical literature is not fit for purpose for patients and their advocates. What is needed is distillation and synthesis into formats that will be useful and

²⁹ “Communicating Science Effectively: A Research Agenda,” *National Academies of Sciences, Engineering, and Medicine*, 2017, <https://doi.org/10.17226/23674>.

situated within workflows that will be convenient, incorporating the medical literature and addressing misinformation, through whatever combination of expert human creation, curation, and artificial intelligence can be most effective.”³⁰

One critical enabling element across this category is that it must bring together the publications and perhaps other research objects and outputs in a single field or set of fields, across the publishing organizations active in that field, in order to be successful. While good results may be possible with something short of 100 percent coverage in a field, there are few if any fields where a single publishing organization has sufficient coverage on its own.

Additionally, both for services for fellow researchers and for the general public, there are a number of positive roles that emerging technologies such as artificial intelligence can offer, and also a number of risks. In introducing new technologies to the work of making meaning, it is essential to bear in mind key elements of trust and authority. For example, clear indications of provenance of knowledge claims will remain foundational. We also foresee at minimum a period where human curatorial and analytical contributions will remain significant components of the work of making meaning.

We note several limitations in existing and developing models. Building services on top of versions of record and other vetted research objects and outputs may have substantial advantages, yet in the marketplace today it is not simple to access these versions comprehensively, and compromises abound. We also note the array of start-ups offering platforms to help scholars express their findings using “plain language,” to provide AI-driven syntheses for interested readers, to offer audio interviews or summaries, among other models. These services face several common challenges. One is the risk of giving too much attention to the supply side of scholarly content and publishing organizations than the demand side of individuals interested in accessing scholarly discovery and expertise. A second is the question of whether they are providing value to publishing organizations as a form of content marketing or if not how they interact with the economic value chain.

Recommendations

Innovations in making meaning typically require access to a larger proportion of the scholarly record than is available through any single publishing organization or distribution channel. The desire to monetize new forms of discovery and machine readership and the current challenges in doing so readily are major impediments.

- Recommendation 20: We recommend that a streamlined aggregate technical and licensing mechanism be developed to enable publishing organizations to distribute appropriate portions of the scholarly record, both to one another and to third parties, for new kinds of translational and analytic services.

³⁰ See Roger C. Schonfeld, "The Problem at the Heart of Public Access," *The Scholarly Kitchen*, 5 December 2023, <https://scholarlykitchen.sspnet.org/2023/12/05/problem-heart-public-access/>.

In our current environment of rising misinformation and declining trust in science, there is a growing need for new thinking about how the broad general public makes meaning out of science—and what can be done to enable and improve this process. To do so, it is important to consider different use cases, for example the needs of patients and their advocates to interpret the best relevant biomedical knowledge or the needs of a broad general public for scientific literacy around climate change or vaccine efficacy. We see little evidence that existing models are up to the task and instead see the need for entirely new approaches that embrace modern digital and social workflows, the nature of the contemporary political environment, and our education systems.

- **Recommendation 21:** We recommend that all parties, including governments, with mandates for or interests in supporting the needs of the broad public in these areas, develop strategies and devote resources to building services and products that address these needs. We recommend that scholarly publishers should see themselves as key stakeholders in this work and therefore devote lobbying resources to developing the policy and funding environments that are necessary enablers.

Supporting New Business Models

The transition to open access has introduced new business models for scholarly publishing. As publishing organizations have developed and implemented these new business models, they have discovered that certain key elements of their shared infrastructure are incapable of supporting them. Some important elements of the shared infrastructure were, perhaps unwittingly, designed primarily to support the subscription model. As a result, publishing organizations have been faced with a dilemma about how to provide the infrastructure necessary to support open access business models. Many existing providers have adapted, but in some cases publishing organizations have been forced to utilize manual processes as a stopgap. We recognize the challenges of building shared infrastructure when business models are evolving rapidly in a competitive environment, leaving so much unsettled. That said, there remain substantial opportunities for the development of shared infrastructure that would support and enable open access business models.

One major category of infrastructure would support the transfer of money necessary to pay for publishing services. This includes systems necessary for individual universities to bundle together their resources to pay for consortial transformative agreements. It also includes the systems necessary to tie together individual institutional customers of transformative agreements with the workflows for manuscript management, where authors unaffiliated with such institutions are otherwise charged an APC. And it also includes the related systems necessary to support the multi-payer model that bundles divergent resource budgets together within a single institution to pay for transformative agreements. Major publishing organizations (and in some cases universities) have invested in building their own systems to enable such services, some of which are mere stopgaps and others of which serve as a point of competitive differentiation. Some shared solutions have also emerged.

There is also infrastructure enabling libraries to analyze and manage their collections budgets to help them optimize acquisitions, holdings, impact, and spending, all of which supports the transition to open access (as well as other institutional priorities). There are some existing products in this category but libraries do not have anything approaching comprehensive collection analysis tools, either for content or for spending.

Several kinds of infrastructures have developed in support of content syndication models. This has included tracking content usage on a distributed basis, important in order to ensure that the impact generated is assigned appropriately. (Interestingly, there is also a completely separate effort in the academic ebook and university press community for tracking usage in a distributed open access environment). Shared information about content “entitlements,” as GetFTR has developed, as well as common approaches to authentication and authorization are other elements that can ease the shift towards syndication. The competitive dynamics in the market for syndication posed some meaningful challenges in developing shared infrastructure, many of which more recently appear to have been overcome.

Recommendations

We have identified several categories of infrastructure that can be more fully developed to support new business models in support of open access. In the case of the transfer of money to support publishing services, we hypothesize that the biggest impediment is the dynamic nature of open access business models; to the extent that this environment matures the infrastructure will catch up with it. In the case of library collections analysis, we hypothesize that libraries have under-valued business intelligence tools and services, which as a consequence is an under-developed sector. In the case of usage reporting, different mechanisms are being developed for journals and for books, which will likely cause operational challenges.

- Recommendation 22: We recommend that current and potential providers of specifically open access infrastructure critically examine the opportunities for greater investment, given the needs for services and tools to be developed. It is possible that there are opportunities for collaborative efforts to develop strong standalone offerings or prototypes designed to stimulate commercial investment. Foundations and funding agencies interested in promoting the transition to open access should see this area of shared infrastructure as one of several where direct investments could have a useful enabling role.

A number of categories of infrastructure have been developed in recent years to support syndication models. These include efforts to streamline authentication and authorization, to share entitlement information, and to expedite usage reporting. We suspect that syndication models are not the only innovation in content dissemination and access.

- Recommendation 23: We recommend that universities, through their libraries and research offices, and publishing organizations convene a strategic discussion, informed by user research, about how content and users, including machines, will best engage with one another in the future. The outcomes of this strategic discussion should result in new approaches to connecting content with users, and ultimately new infrastructure to support those approaches.

Alternative System

Some interviewees and other observers favor the creation of an alternative scholarly communication system that will overcome some of the deficiencies of the current system. Many of those with such perspectives are particularly troubled with the academic incentives that have led to a prestige economy enabling so much scholarly publishing to become dominated by commercial and marketplace incentives.

A good deal of foundational elements already exist for an alternative model.³¹ Some repositories for green versions of open access have proved highly successful, particularly in major physical and biomedical scientific fields. Indeed, major publishing organizations have invested in preprints, bringing them into submission workflows.³² Several publishing organizations have developed models that enable post-publication peer review, and there has been a long dream of adding “overlay journals” atop repository items.

The latest proposal from Plan S pushes towards an approach characterized by diamond open access models, repositories, and other alternatives.³³ These elements, which have a strong foundation in some fields and geographies, could be expanded elsewhere. Even before adoption let alone implementation, the proposal has motivated a number of commercial as well as academy-affiliated innovators to begin building services and infrastructure that would be suited for such an environment. Intriguingly, there are a number of infrastructure elements beyond repository and diamond publishing models themselves that are emerging as possible points of leverage, for example within the manuscript submission and editorial management systems.

Alternative models are often designed with a transformative purpose, but the infrastructure elements that support them will still face some of the same logics as traditional publishing organizations. For example, there is little reason to believe that current challenges such as research integrity, with its attendant and growing costs, will not present itself for alternative models. And, while it may be possible to reduce some costs by reducing or eliminating the profit motive, services will still need to be sustained, which may be more challenging if scale and a stable business model are not present. There is an increasing fragmentation in a marketplace that offers limited funding opportunities for not-for-profit scholarly communication initiatives, especially after the initial start-up phase. Indeed, forces of consolidation are equally present, both to save faltering initiatives and to drive the scale that is often essential to sustainability, as

³¹ See for example Kristen Ratan, Katherine Skinner, Catherine Mitchell, Brandon Locke, and David Pcolar, “Library Publishing Infrastructure: Assembling New Solutions,” *Educopia Institute*, March 2021, <https://educopia.org/nglp-lib-pub-infrastructure>.

³² Roger C. Schonfeld and Oya Y. Rieger, “Publishers Invest in Preprints,” *The Scholarly Kitchen*, 27 May 2020, <https://scholarlykitchen.sspnet.org/2020/05/27/publishers-invest-in-preprints/>.

³³ “Towards Responsible Publishing: A Proposal from cOAlition S,” *Plan S*, 31 October 2023, https://www.coalition-s.org/wp-content/uploads/2023/10/Towards_Responsible_Publishing_web.pdf.

the merger of DuraSpace into Lyris served to illustrate.³⁴ And of course commercial participants in this sector will follow the incentives that policy environments establish for them.

Recommendation

Alternative models are often mission driven, yet they nevertheless operate in a competitive marketplace. To attract authors and users, they must provide a set of services that not only align on mission or transformative vision but that also meet the day-to-day needs of end users. In today's market, author and reader experience are vital, while looking ahead computational use cases (including AI) will be increasingly important. Some of the other new developments we discuss in this report, such as the atomization of the scholarly record, will also likely require attention.

- Recommendation 24: We recommend that advocates for and developers of alternative publishing models build solutions focused on how to meet the day-to-day needs of authors, researchers, and other end users more effectively, staying abreast or ideally ahead of competitors.

Sustainability not only involves covering operating costs but also securing ongoing investment to enable continued maintenance, innovation, and uptake. Several interviewees stated that launching a new not-for-profit initiative should be seen as a start-up venture, requiring a talented leadership team with a strong understanding of the scholarly landscape and entrepreneurial skills. Not-for-profit organizations, similar to their commercial counterparts, need to continuously assess the market to understand the competitive environment and users' actual needs and behavior to market their products accordingly. Building and maintaining a social infrastructure to ensure the promotion, adoption, and adaptation of new systems and tools can be expensive and time-consuming, especially in a competitive marketplace.

- Recommendation 25: We recommend that advocates for and developers of alternative publishing models assess potential benefits not only on the basis of assumed opportunities for resource reduction but also recognizing the expense necessary to address transformative goals in ways that are grounded in and informed by the logics of and challenges facing publishing. Successful models will account for the expenses not only of ongoing operations but also maintaining and reinvesting in infrastructure over time.

Concluding Remarks

We are experiencing the Second Digital Transformation, an incredibly demanding and innovative time for scholarly communication. The competitiveness of the shared infrastructure market influences the competitiveness of the research publishing sector—and vice versa. There are an array of growing and unmet needs for shared infrastructure, requiring investment and strategic governance. In this report, we have tried to address not only the rich value of shared

³⁴ Roger C. Schonfeld, "More Scholarly Communications Consolidation as Institutional Repository Provider DuraSpace Merges into Lyris," *The Scholarly Kitchen*, 25 January 2019, <https://scholarlykitchen.sspnet.org/2019/01/25/lyris-duraspace-merger/>.

infrastructure but also the complicated set of perspectives, sometimes diverging, that characterize demand and advocacy for it.

We have distinguished between types of infrastructure and infrastructure needs with divergent characteristics. For example, access to consistent and reliable identifiers and metadata matters to everyone involved in order to facilitate access, discovery, reuse, analytics, preservation, and many new functions. Here the challenge is to benefit from the fractured control of metadata and identifiers. On the other hand, enterprise publishing systems are vital to every publishing organization itself individually. Here the challenge is how to maintain a competitive and versatile marketplace given the forces driving several different forms of consolidation. And in some of the areas of new opportunity, we note varying priorities from determining how to make markets for making meaning to how to build an alternative system for scholarly communication.

We wish to note the opportunity to think in terms of greater interoperability for existing and new types of infrastructure. Different elements of the infrastructure can be somewhat siloed as a result of the specific needs of each. There are opportunities for greater attention to interoperability across competing infrastructure providers and infrastructure categories. In addition, thinking in global terms, we note that infrastructure will not always be developed or sustained on a truly global basis, particularly given increasingly stormy geopolitical dynamics and values mismatches. Opportunities may exist in some cases for increasingly sophisticated forms of interoperability across not only infrastructure categories but also geographies.

In the end, the Second Digital Transformation offers profound strategic opportunities and challenges for publishing organizations and infrastructure providers alike. Recognizing the complexity of this marketplace and all its many participants, how do we pursue not just individual differentiated interests but account for the larger needs of science, scholarship, and society?

Appendix A: List of Interviewees

- Clare Appavoo, Executive Director, Canadian Research Knowledge Network
- Allison Belan, Director for Strategic Innovation and Services, Duke University Press
- Laird Barrett, Head of Product, Springer Nature
- Amy Brand, Director, MIT Press
- Rachel Bruce, Head of Open Research, UK Research and Innovation
- Johannes Buchmann, Chief Operating Officer, De Gruyter
- Adrian Burton, Director, Data, Policy, and Services, Australian Research Data Commons
- Matt Buys, Executive Director, DataCite
- Ana María Cetto, Research Professor, Universidad Nacional Autónoma de México
- Angela Cochran, Vice President, Publishing, American Society of Clinical Oncology
- Raym Crow, Senior Consultant, SPARC
- Chris Freeland, Director of Open Libraries, Internet Archive
- Nicko Goncharoff, Managing Director, Osmanthus Consulting Ltd
- Joshua M. Greenberg, Program Director, Digital Information Technology, Sloan Foundation
- Sören Hofmayer, Chief Strategy Officer, ResearchGate
- Daniel Hook, Chief Executive Officer, Digital Science
- Hannah Hope, Open Research Lead, Wellcome Trust
- Wolfram Horstmann, Director of the Göttingen State and University Library and University Librarian, Georg August Universität Göttingen
- Lauren Kane, President and Chief Executive Officer, BioOne
- Thane Kerner, Chief Executive Officer, Silverchair
- Richard Kidd, Head of Chemistry Data, Royal Society of Chemistry
- Hylke Koers, Chief Information Officer, STM Solutions
- Rose L'Huillier, Senior VP for Researcher Products, Elsevier
- Karel Luyben, Rector Magnificus Emeritus, Delft University of Technology
- Babis Marmanis, Executive Vice President & CTO, Copyright Clearance Center
- Salvatore Mele, Senior Advisor, CERN
- Kate McCready, Visiting Program Officer, Big Ten Academic Alliance
- Alison Mudditt, Chief Executive Officer, Public Library of Science
- Ritsuko Nakajima, Director of the Department for Information Infrastructure, Japan Science and Technology Agency
- Josh Nicholson, Co-founder and Chief Executive Officer, scite
- William Nixon, Deputy Executive Director, Research Libraries UK
- Darby Orcutt, Assistant Head, Collections and Research Strategy, North Carolina State University Libraries
- Joy Owango, Founding Executive Director, Africa PID Alliance

- Ed Pentz, Executive Director, Crossref
- Kristen Ratan, Founder, Strategies for Open Science (Stratos)
- Howard Ratner, Executive Director, CHORUS
- John Sherer, Spangler Family Director, The University of North Carolina Press
- Chris Shillum, Executive Director, ORCID
- Jasper Simons, Chief Publishing Officer, American Psychological Association
- MacKenzie Smith, University Librarian and Vice Provost of Digital Scholarship, University of California, Davis
- Carly Strasser, Senior Program Manager for Open Science, Chan Zuckerberg Initiative
- Kaitlin Thaney, Executive Director, Invest in Open Infrastructure (IOI)
- Todd Toler, Vice President, Product Strategy & Partnerships, John Wiley & Sons
- Herbert Van de Sompel, Research Fellow, Data Archiving and Networked Services
- John M. Unsworth, Dean of Libraries and University Librarian, University of Virginia
- Dan Whaley, Chief Executive Officer, Hypothesis
- John Wilkin, CEO, Lyris
- Alicia Wise, Executive Director, CLOCKSS Archive
- Ralph Youngen, Senior Director of Technology Strategy and Partnerships, American Chemical Society Publications

Appendix B: Interview Questions

1. What do you see as the most important drivers of change in scholarly communication?
 - What are the key trends in [specialty area of interviewee]?
 - For example, trends like
 - Growth of new form of content such as datasets, code, methods, and implications for DOIs
 - Protecting the integrity of the scientific record
 - Widespread shift to open
 - Increasing importance of computational forms of content consumption and content production
 - Divergent demands from users for such features as providing seamless discovery and access; broadening participation through movements like “citizen science; producing translational outputs that meet the needs of the general public
 - Successes registered and roadblocks and impediments

2. What kinds of shared infrastructure do you rely upon in your role? What kinds of shared infrastructure do you see as most important in the scholarly communications sector?

- Prompt with categories of shared infrastructure drawn from the landscape review (ie Assessment, Metrics, and Reporting; Authentication and Authorization; Discovery, Syndication, Hosting, Delivery, and Aggregation; Licensing and Rights Management; Manuscript Submission and Editorial Management; Metadata; Peer Review, Annotations, and Commenting; Persistent Identifiers; Preservation; Publishing Platforms and Repositories; Research Information Management; Researcher Identity and Marketing; Research Data Curation and Management; Usage Data)
3. What is working especially well in this shared infrastructure?
 4. What are some of the challenges and pain points that you see with this infrastructure provision?
 - Are you satisfied with technical abilities; product strategy, governance, business model, etc.
 5. What kinds of infrastructure doesn't yet exist but would be valuable to your work or goals?
 - You mentioned XXX strategic direction. Do you see infrastructure needs associated with that?
 6. Do you have any other comments or suggestions to inform our study?