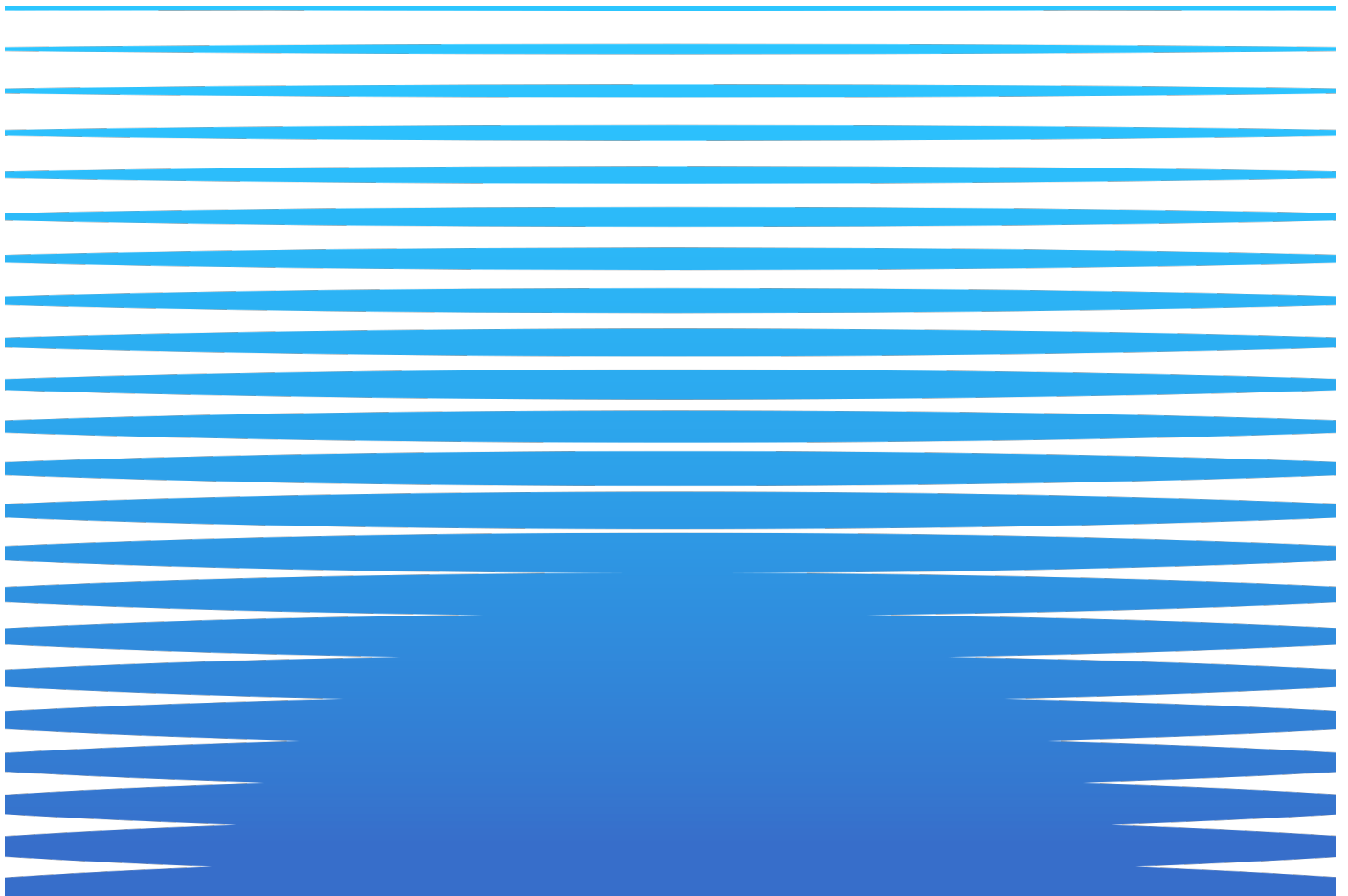# Common Scholarly Communication Infrastructure Landscape Review

Oya Y. Rieger
Roger C. Schonfeld

ITHAKA S+R

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.

ITHAKA is interested in disseminating this brief as widely as possible. Please contact us with any questions about using the report: research@ithaka.org.

Table of Contents

Scholarly communication is a complicated sector, with numerous participants and multiple mechanisms for communicating and reviewing materials created in an increasing variety of formats by researchers across the globe.[1] In turn, the researcher who seeks to use the products of this system wishes to discover, access, and use relevant and trustworthy materials as effortlessly as possible. The work of driving efficiency into this complex sector while bringing its multiple strands together seamlessly for the reader (or, increasingly, for a computational user) rests on a foundation of infrastructure, much of it shared across multiple publishers. In this landscape review, we seek to provide a high-level overview of the shared infrastructure that supports scholarly communication. The purpose of this landscape review is to provide scoping for the array of shared infrastructure that we intend to examine in a larger project about the strategic context that has driven and will continue to drive the development of this infrastructure. That project will include a needs analysis on what parts of the shared scholarly communication infrastructure are working well and where they can be improved, culminating in recommendations for where additional or revised collective action and community investment is indicated.[2]

Introduction and Scope

Scholarly communication is the process through which research products and outputs (such as articles, audiovisual materials, data, code, and research methods) are created, assessed, improved, shared, disseminated, and preserved in a variety of modes including through formal and informal publications, conferences, and other academic networking methods. Shared infrastructure is a 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.

The global landscape for this work includes a complex mixture featuring several different kinds of academic systems, a number of well-established commercial publishing houses, an array of not-for-profit publishers such as university presses and scholarly societies, several disruptive innovators leveraging new business models, a steady pace of consolidation along with a rich start-up environment, and much more.

A robust and nimble infrastructure is imperative to support the ongoing digital transformation of scholarly communication, enabling new and improved services and achieving real efficiencies for all stakeholder communities. Developing, maintaining, and sustaining fit-for-purpose community infrastructure is a challenge particularly when the technology and policy environments are in flux. Some elements of the shared infrastructure are rising in importance, while others may be declining in value, as a result of change in scholarly communication and the broader ecosystems in which it occurs. In this report, we do not attempt to review or assess the

---

various pieces of the shared infrastructure or their future trajectories, but simply to describe their current role.[3]

Given the vast array of standards, systems, and tools in many of the categories of shared infrastructure, we make no claim to comprehensiveness. Our aim is to provide illustrations of representative elements in each category of the shared infrastructure, and this effort should not be confused with an inventory.

In scoping this piece, perhaps our most notable exclusion is the myriad commercial, open source, and community-based scholarly workflow tools for conducting and managing research individually and with collaborators. Among the workflow tools that fall into this category are Executable Research Article, Jupyter Notebook, Overleaf, EndNote, SkyPortal, and CedarWorkbench.[4] We have included a detailed section on Research Data Curation and Management services, some of which could be said to fall into this otherwise excluded category.

We also want to acknowledge the many advocacy, funding, standard-setting, and community groups that are extremely active in the shared infrastructure landscape but are not themselves currently infrastructure providers. This group includes the Committee on Data of the International Science Council (CODATA), Confederation of Open Access Repositories (COAR), Force11, Coalition for Advancing Research Assessment (COARA), Global Sustainability Coalition for Open Science Services (SCOSS), Invest in Open Infrastructure (IOI), LA Referencia, Library Publishing Coalition, LIBSENSE, National Information Standards Organization (NISO), Open Access Scholarly Publishing Association (OASPA), Research Data Alliance, SPARC, World Intellectual Property Organization (WIPO), among others.[5]

We wrestled at some length with how best to organize the shared infrastructure into categories. We ultimately elected to prefer categories that describe current purpose and structure rather than those that might skew towards directions we can foresee them taking in the future. We acknowledge different ways of organizing these categories may make more or less sense to different readers.

Notwithstanding these many caveats, we hope that our work of bringing together this landscape will be of use to others in the community. It is our effort to bring some shape to a complicated

---

[3] We benefited in our work from several directories and other efforts to provide an overview of parts of this landscape. See (among others): Katherine Skinner, "Mapping the Scholarly Communication Landscape – 2019 Census," Educopia, https://educopia.org/2019-Census/; SComCat: Scholarly Communication Technology Catalogue, 2020 https://www.scomcat.net/functions.

[4] Executable Research Article, https://stenci.la/blog/2020-08-24-executable-research-article-launch; Jupyter Notebook, https://jupyter.org; Overleaf, https://www.overleaf.com; EndNote, https://endnote.com; SkyPortal, https://bids.berkeley.edu/research/skyportal; CedarWorkbench, https://metadatacenter.org; Omeka, https://omeka.org.

[5] Coalition for Advancing Research Assessment, https://coara.eu; CODATA, https://codata.org; Confederation of Open Access Repositories, https://www.coar-repositories.org); Force11, https://force11.org; Global Sustainability Coalition for Open Science Services, https://scoss.org; Invest in Open Infrastructure, https://investinopen.org; LA Referencia, https://www.lareferencia.info; Library Publishing Coalition, https://librarypublishing.org; LIBSENSE, https://libsense.ren.africa; National Information Standards Organization, https://www.niso.org; Open Access Scholarly Publishing Association, https://oaspa.org; Research Data Alliance, https://www.rd-alliance.org; SPARC, https://sparcopen.org; World Intellectual Property Organization, https://www.wipo.int.

landscape as part of a broader analysis of the shared infrastructure that we will be publishing later in 2023. Particularly given this broader project currently underway, we welcome suggestions and observations about the shared infrastructure we have profiled in this piece and how we have organized it.

Components of the Shared Infrastructure

Assessing Impact and Value

Description

A wealth of data from scholarly communication are used to measure and assess the impact and value of research, researchers, universities, publishers, and products. Of the underlying sources that are used to calculate these metrics, two stand out: citations for measuring research and usage for measuring value.

Research metrics are quantitative tools used to help assess the quality and impact of research at article, researcher, journal, and institutional levels.[6] Citation-based metrics have long been the most common indicator of research productivity and impact. In recent years, citation metrics have become more widely available through Crossref, which has reduced the barriers to entry through a basic citation-driven research metric.[7] In addition, 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."[8] However, these are meant to complement rather than replace traditional impact metrics. The work to develop and steward research metrics can be automated to some degree but often has real costs, including highly skilled labor, which is supported directly or indirectly through products that are sold to libraries and publishers.

While research metrics tend to be drawn from article-based measures and are used to analyze the impact of researchers and journals, the category of usage data is particularly important for measuring and establishing the economic value of journals, journal bundles, and aggregations,

---

[6] The founding provider of these metrics, the Journal Impact Factor, warns against their misuse. See: "Time to Remodel the Journal Impact Factor," *Nature* 535, no. 466 (July 2016) https://doi.org/10.1038/535466a. 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.

[7] Amy Brand, "Guest Post — Crossref at a Crossroads: All Roads Lead to Crossref," *The Scholarly Kitchen*, 22 October 2019, https://scholarlykitchen.sspnet.org/2019/10/22/crossref-at-a-crossroads-all-roads-lead-to-crossref/.

[8] Although there is strong evidence of the desire to improve the way research is assessed based on evolving research and communication practices, it is complicated to transform the system with multi-stakeholders. Inspired by the Leiden Manifesto (http://www.leidenmanifesto.org) and originally implemented by Altmetrics.org (Digital Science), Altmetrics is being proposed as an alternative to traditional citation impact metrics to include social media, online reader behavior, network interactions with content, and social media.

among other content subscription packages. Usage data have become essential to the digital subscription model. But whereas Clarivate, a publicly traded corporation, stewards what is generally seen as the most important research metric, the most important usage data metric is stewarded by COUNTER, a not-for-profit organization.

Examples

Journal-Level Metrics

These metrics indicate the level of significance and impact an academic journal has within its field of research, usually via an algorithm that takes into account the number of articles published per year and the number of citations to articles published in that journal.

- Several citation-based impact measures are productized in a variety of ways, including through citation databases that are sold to academic institutions and research performance analytic services, as well as through bespoke consultancy engagements at the funder, university, and national level. They can also feature prominently in league tables for universities and departments.
  - Journal Impact Factor (JIF) is a measure of the average number of citations made to articles within an academic journal over the course of a year.[9] The most widely established and frequently discussed measure of impact, it is included as part of Clarivate's Journal Citation Reports (JCR) product (https://clarivate.com/products/scientific-and-academic-research/research-analytics-evaluation-and-management-solutions/journal-citation-reports), which provides a number of impact measurements for journals in the sciences and social sciences. A recent direction for this product is expanding the coverage of impact factors and associated rankings to cover the humanities.[10] JIF is distinguished from some other metrics because it has been offered selectively only to journals that meet certain standards.
  - CiteScore reflects the annual average number of citations to recent articles published in a journal. Provided as part of Elsevier's Scopus product, it is an alternative to the Journal Impact Factor. It is calculated as the ratio of citations to documents published over a four-year period over the number of documents in the same four-year period.
- SCImago Journal and Country Rank includes the journals and country scientific indicators developed from the information contained in the Scopus database (Elsevier) to assess and analyze scientific domains.
- There are also a number of other journal-level metrics, some of which are designed to promote particular goals for scholarly communication. For example, TOP Factor (https://www.topfactor.org/) is a metric to assess how a journal is implementing open science practices. It is a tool developed by the Center of Open Science to rate journals' policies adherence to the Transparency and Openness Promotion (TOP) Guidelines

---

[9] See: "Journal Citation Reports," Clarivate, https://clarivate.com/products/scientific-and-academic-research/research-analytics-evaluation-and-management-solutions/journal-citation-reports.

[10] See: Nandita Quaderi, "Mapping the path to future changes in the Journal Citation Reports," Academia and Publishing Blog, Clarivate, March 7, 2023, https://clarivate.com/blog/mapping-the-path-to-future-changes-in-the-journal-citation-reports.

(https://www.cos.io/initiatives/top-guidelines) and promote the implementation of open science practices to support transparency and reproducibility.

Author and Article Metrics

Author and article metrics measure an author's impact on their field or discipline using the number of academic publications authored and the number of times these publications are cited by other researchers. These metrics typically combine article-level metrics such as citation counts with standardized authorship metrics, which can be generated algorithmically, for example by Google Scholar or through a community standard like ORCID.

The following are examples of researcher profile services with naming systems that are not considered to be persistent identifiers (PIDs) but are often mentioned within the context of PIDs.

- Google Scholar Citations allows authors to set up a profile page that lists their publications and citation metrics and track citations to their publications over time. Google Scholar Metrics (https://scholar.google.com/intl/en/scholar/metrics.html) allows authors to gauge the visibility and influence of recent articles in scholarly publications. Used in Google's My Citations feature, i10-Index counts the number of publications with at least 10 citations.
- ResearchGate Profiles (https://explore.researchgate.net/display/support/Profile) provide a snapshot of an individual's research, affiliations, and experience, and show who is reading, citing, and mentioning the works. It is maintained by ResearchGate, a commercial social networking service provider.
- h-Index measures the cumulative impact of a researcher's publications through both quantity (number of publications) and quality (number of citations).[11] The index has been applied to the productivity and impact of a scholarly journal as well as a group of scientists. A common criticism is that it is not an accurate measure for early-career researchers. It is made available automatically through services like Google Scholar and can also be calculated manually.
- Altmetrics application examples include:
  - Altmetric Bookmarklet (https://www.altmetric.com/solutions/free-tools/bookmarklet) by Digital Science is a free browser tool to track how much attention recent papers have received online.
  - Impactstory (https://profiles.impactstory.org) is an open-source website that helps researchers explore and share the online impact of their research to build a new scholarly reward system (funded by the National Science Foundation and the Alfred P. Sloan Foundation and incorporated as a 501(c)(3) nonprofit corporation).
  - PlumX Metrics (https://plumanalytics.com) from Elsevier is a suite of metrics to provide insights into the ways readers interact with individual pieces of research output in the online environment.
- Contributor Role Taxonomy (ANSI/NISO, CRediT https://www.niso.org/publications/z39104-2022-credit) enables the range and nature of

---

[11] The h-index was suggested in 2005 by Jorge E. Hirsch, a physicist at UC San Diego, as a tool for determining theoretical physicists' relative quality and is sometimes called the Hirsch index or Hirsch number.

contributions to scholarly published output to be captured in a transparent, consistent, and structured format to improve accessibility and visibility.

Related Resources

Das, Anup Kumar. *Guide to Research Evaluation Metrics*. Paris: United Nations Educational, Scientific and Cultural Organization, 2015. https://unesdoc.unesco.org/ark:/48223/pf0000232210.

Jones, Phill. "How Do We Make Research Assessment More Responsible? – A Multi-stakeholder Discussion." *The Scholarly Kitchen*. 9 February 2022. https://scholarlykitchen.sspnet.org/2022/02/09/how-do-we-make-research-assessment-more-responsible-a-multi-stakeholder-discussion.

Jones, Phill and Fiona Murphy. "Openness Profile: Modeling Research Evaluation for Open Scholarship." *Zenodo*. 31 March 2021. https://zenodo.org/record/4581490#.Y_foRnbMJPY.

"Metrics Toolkit." Metrics Toolkit. 2021. https://www.metrics-toolkit.org.

Mudditt, Alison. "Reforming Research Assessment: A Tough Nut to Crack." *The Scholarly Kitchen*. 18 February 2020. https://scholarlykitchen.sspnet.org/2020/02/18/reforming-research-assessment-a-tough-nut-to-crack.

Price, Robyn. "Are Research Organisations Ready for Open Science Indicators?" *The Bibliomagician*. 16 February 2023. https://thebibliomagician.wordpress.com/2023/02/16/preparing-for-open-science-indicators.

Authentication and Authorization

Description

Authentication and authorization infrastructure involves protocols, technologies, and standards to enable members of different institutions to access scholarly information that requires verification of identity and/or affiliation for access. Early models for authorizing on-campus access to site-licensed services to academia relied largely on IP addresses, which served the function effectively (including effectively anonymizing individual users) despite not being intended for this purpose. Subscription-based business models face the challenge of ensuring that authorized users (such as students, faculty, researchers, and staff) can access content regardless of their location, not limited to a campus or other specific workplace. The proliferation of handheld devices has made this task even more complicated. Another key challenge is protecting the privacy of users and allowing them and their institutions to decide what personal information, if any, is released to the content provider. Some of these services are

used not only in conjunction with subscription-based business models, but also to generate metrics that are shared with authors, institutions, and funders for open access materials as well.

Examples

- IP Registry (https://theipregistry.org) is a single repository of the validated IP addresses for over 70,000 content licensing organizations worldwide, accessible by both publishers and libraries. IP Registry is developed by PSI Ltd, an independent UK company.

- EZproxy (https://www.oclc.org/en/ezproxy.html) is an OCLC product that provides the authorized users with remote access to subscription resources by authenticating their identity and delivering e-content.

- SeamlessAccess (https://www.seamlessaccess.org) enables users to sign in using their preferred log-in credentials. Formerly RA21, it is based on Federated Identity Management (FIM), an identity arrangement made between multiple online domains/applications to allow users to access several domains/applications without going through multiple logins. The federated authentication system describes how information is exchanged about the rights to facilitate a seamless user experience. It also relies on Security Assertion Markup Language (SAML), an open standard designed for secure single sign-on managed by the Organization for the Advancement of Structured Information Standards (OASIS). Governance of SeamlessAccess service is through the Coalition for Seamless Access, a collaboration between GÉANT, Internet2, the National Information Standards Organization (NISO), and the International Association of Scientific, Technical and Medical Publishers (STM).

- Federated Credential Management (FedCM) is a web API for privacy-preserving identity federation. Originally developed by Google, the technology has gathered the support from other major browser vendors to replace the existing tracking functionalities. It is not currently in wide usage in the scholarly communication sector.

- European Open Science Cloud (EOSC) Authentication and Authorization Infrastructure (AAI https://op.europa.eu/en/publication-detail/-/publication/d1bc3702-61e5-11eb-aeb5-01aa75ed71a1) is an example of a proposed architecture to streamline researchers' access to services, both provided by their own infrastructure and shared with other communities. It was developed by the EOSC Architecture Working Group.12

- Knowledge Bases (KBART https://www.niso.org/standards-committees/kbart) provide vital information for authentication and authorization systems.

- Campus Activated Subscriber Access (CASA, https://journals.ala.org/index.php/ltr/article/view/7852), which builds on Google Scholar's Subscriber Links program) are used to support library link resolvers in order to enable users' access to licensed content they are entitled to use through subscriptions held by their institutions.

- Order management and fulfillment is an important infrastructure category particularly for subscription publishers. System and services such as AdvantageCS, SiteManager

---

12 Based on AARC blueprint as underlying standard: https://aarc-project.eu/architecture/.

(Silverchair), and Klopotek's O2C Apps (Klopotek) support order, subscription, fulfillment, distribution, and account management.[13]

Related Resources

Carpenter, Todd A., Hylke Koers, and Heather Flanagan. "Security, Safety, Seamless Access." *The Scholarly Kitchen.* 7 June 2021. https://scholarlykitchen.sspnet.org/2021/06/07/security-safety-seamlessaccess.

Schonfeld, Roger C. "Dismantling the Stumbling Blocks that Impede Researcher Access to E-Resources." *The Scholarly Kitchen.* 13 November 2015. https://scholarlykitchen.sspnet.org/2015/11/13/dismantling-the-stumbling-blocks-that-impede-researcher-access-to-e-resources.

Tay, Aaron. "Improving Access to and Delivery of Academic Content from Libraries." *Library Technology Reports.* American Library Association, 2022. https://www.alastore.ala.org/LTR58n6.

Wierenga, Klaas, Leif Johansson, Christon Kanellopoulos, David Groep, Davide Vaghetti, and Nicolas Liampotis. "EOSCN Authentication and Authorization Infrastructure (AAI)." Edited by the EOSC Executive Board. Publications Office of the European Union, January 2021. https://op.europa.eu/en/publication-detail/-/publication/d1bc3702-61e5-11eb-aeb5-01aa75ed71a1.

Discovery, Syndication, and Aggregation

Description

In a user-centric analysis, the journey quite often starts not at a publisher platform but rather at a discovery service or aggregator, with syndication providers attempting to similarly serve as the starting point for a user journey. Through these services, researchers identify scholarly materials of interest, which in some cases they access through the site through which they discovered them and in others are routed to another site for access.

Although the services that we discuss in this section often serve as starting points for discovery, other services, such as academic and professional networking sites, also play an increasingly important role in helping scholars to identify and access scholarly content. In addition, email alerts and feed-based social media can also be quite important.

Discovery tends to rely on metadata standards, which are discussed under Persistent Identifiers and also below in this section. Semantic technologies such as natural language processing, data

---

[13] AdvantageCS, https://www.advantagecs.com; SiteManager, https://www.silverchair.com/the-silverchair-platform/tools-technology; and Klopotek O2C Apps, https://www.klopotek.com/o2c.

mining, artificial intelligence (AI), category tagging, and semantic search are increasingly used to improve metadata and lead to better discovery and access.

Examples

Discovery

These services enable a user to conduct a search, register for alerts, or use other approaches to provide a result to a user either in response to a specific query or predictively. Also see: Publishing Platforms and Repositories, to which researchers are often routed from a discovery service.

- Library Discovery Systems enable library users to search and access other discovery services (in addition to online public access catalogs, or OPACs), the full text of sources, subject guides, or subject-oriented abstract and indexing products. Examples include EBSCO Discovery Service, Primo and Summon (Clarivate), and WorldCat Discovery (OCLC).[14]
- Citation Indices provide discovery and detailed analytics of scholarly outputs, focused originally on research articles but more recently including other materials such as patents. Examples include Web of Science (Clarivate Analytics), Scopus (Elsevier), Dimensions (Digital Science), and Lens (from social enterprise Cambia).[15]
- Abstracting and Indexing (A&I) Databases focus on individual fields of study, providing an often human-curated index of materials, including field-specific subject indexing. Such A&I Databases include the MLA Bibliography, the Bibliography of Asian Studies, and Chemical Abstracts, among many others.
- Commercial Consumer Search Engines such as Google, Google Scholar, Google Books, and Bing allow users to conduct web searches in a systematic way for information in different formats.
- Academic and Professional Networking Websites enable scholars to communicate, collaborate, share their work, find collaborators, and interact with peers. Prominent examples are run by vendors include ResearchGate, Academia, and LinkedIn.[16]
- scite (https://scite.ai) is a platform by a US-based startup for discovering and evaluating scientific articles via Smart Citations. Smart Citations allow users to see how a publication has been cited by providing the context of the citation and a classification describing whether it provides supporting or contrasting evidence for the cited claim.

Syndication

Syndication enables users who are entitled to access subscription-based materials to do so on sites other than those provided by the publisher. In turn, usage data is shared back to the

---

[14] EBSCO Discovery Service, https://www.ebsco.com/products/ebsco-discovery-service; Primo, https://exlibrisgroup.com/products/primo-discovery-service; Summon https://exlibrisgroup.com/products/summon-library-discovery/; WorldCat Discovery, https://www.oclc.org/en/worldcat-discovery.html.

[15] Web of Science, https://clarivate.libguides.com/home; Scopus, https://www.elsevier.com/solutions/scopus); Dimensions, https://www.digital-science.com, Lens, https://www.lens.org.

[16] ResearchGate, https://www.researchgate.net; Academia, https://www.academia.edu; LinkedIn, https://www.linkedin.com.

publisher to control leakage into COUNTER-trackable data that would be shared with libraries. Several providers such as ResearchGate and Elsevier (through ScienceDirect) are offering syndication for (other) publishers through their sites. Several emerging infrastructure elements can be used to enable syndication, including Distributed Usage Logging (discussed in Assessing Impact and Value) and GetFTR, which is provided by STM Solutions and enables a syndicator service to link readily to the best copy of the item.[17]

Aggregation

Aggregation is distinctive from syndication. In aggregator models, the publisher is paid a licensing fee by the aggregator which in turn uses a separate license to provide the content to institutions. Aggregations are particularly useful in assembling materials necessary for specific use cases such as undergraduate education and can in some cases provide substantial additional distribution beyond publisher channels. Materials can enter and leave some of the services in this category as licensing arrangements evolve over time, while in other cases the provider negotiates permanent access rights from the copyright holder.

- Examples of service providers in this category include ProQuest (Clarivate), EBSCO (EBSCO Industries, Inc), Gale (Cengage Group), and JSTOR (ITHAKA).[18]

Metadata Standards for Enabling Discovery, Exchange, and Transfer

Metadata continues to play a crucial role in enabling discovery and access as well as facilitating information encoding, exchange, and transfer through interoperability.

- Discovery and Access
  - Machine Readable Cataloging (MARC/MARCXML, http://www.loc.gov/standards/marcxml) is a standard to create catalog records for integrated library systems to describe both digital and print resources. Academic libraries have been shifting from Integrated Library Systems (ILS) to Library Services Platforms (LSP) designed to manage all collection formats.
  - Dublin Core (https://www.dublincore.org) Metadata is a schema with 15 different properties for use in resource description. It is supported by the Dublin Core Metadata Initiative (DCMI), which is an organization supporting innovation and best practices in metadata design.
  - Metadata Object Description Schema (MODS, https://www.loc.gov/standards/mods) is a descriptive schema developed by the Library of Congress to bridge the complexity of the MARC format and the simplicity of Dublin Core.
- Encoding and Transmission
  - Document Type Definition (DTD is a set of markup declarations that define a document type for an SGML-family markup language (GML, SGML, XML, HTML). Newer XML namespace-aware schema languages (such as W3C XML Schema and ISO RELAX NG) have largely superseded DTDs.

---

[17] Get Full Text Research, https://www.getfulltextresearch.com.

[18] ProQuest, https://www.proquest.com; EBSCO, https://www.ebsco.com; Gale, https://www.gale.com; JSTOR, https://www.jstor.org. Ithaka S+R is part of the ITHAKA not-for-profit organization, which also includes JSTOR, Portico, Artstor, Reveal Digital, and Constellate.

▫ Metadata Encoding & Transmission Standard (METS, http://www.loc.gov/standards/mets) is commonly used in digital libraries to code descriptive, administrative, and structural metadata using XML schema.
  ▫ ONIX for Books Product Information Format (https://www.editeur.org/83/Overview) is an XML-based standard for book and other book-related products' metadata in order to provide a consistent method to share product information for publishers, retailers, and supply chain partners. It is maintained through EDItEUR and a network of national user groups across a number of countries.
  ▫ Journal Article Tag Suite (ANSI/NISO JATS, https://www.niso.org/standards-committees/jats) is an XML format to enable the exchange of journal content. It provides a set of XML elements and attributes for describing the textual and graphical content of journal articles as well as some non-article material such as letters, editorials, and book and product reviews.

Related Resources

Asmi, Nowsheeba and Madhusudhan Margam. "Academic Social Networking Sites: What They Have to Offer for Researchers?" *Journal of Knowledge & Communication Management* 5 (2015): 1-11. https://www.researchgate.net/publication/280069078_Academic_Social_Networking_Sites_What_They_Have_to_Offer_for_Researchers.

Breeding, Marshall. "2022 Library Systems Report: An Industry Disrupted." *American Libraries*. American Library Association, 2 May 2022. https://americanlibrariesmagazine.org/2022/05/02/2022-library-systems-report.

Bide, Mark. "Identifier and Metadata Standards in the Publishing Industry." *International Publishers Association*. 28 October 2021. https://www.internationalpublishers.org/state-of-publishing-reports/identifier-and-metadata-standards-in-the-publishing-industry.

Conrad, Lettie Y. and Michelle Urberg. "The Experience of Good Metadata: Linking Metadata to Research Impacts." *The Scholarly Kitchen*. 30 September 2021. https://scholarlykitchen.sspnet.org/2021/09/30/the-experience-of-good-metadata-linking-metadata-to-research-impacts/.

Kemp, Jennifer, Lettie Conrad, and Michelle Urberg. "Measuring Metadata Impacts: Books Discoverability in Google Scholar." *Crossref*. 25 January 2023. https://www.crossref.org/blog/measuring-metadata-impacts-books-discoverability-in-google-scholar/.

Open Discovery Initiative Standing Committee. "Open Discovery Initiative: Promoting Transparency in Discovery." *National Information Standards Organization*, 22 June 2020. https://www.niso.org/publications/rp-19-2020-odi.

Schonfeld, Roger C. "What is Content Syndication?" *Ithaka S+R*. 1 March 2019. https://sr.ithaka.org/blog/what-is-content-syndication.

Smith-Yoshimura, Karen. "Transitioning to the Next Generation of Metadata." *OCLC Research*. 2020. https://doi.org/10.25333/rqgd-b343.

Licensing, Reading Ecosystems, and Rights Management

Description

Licensing and rights management tools and systems come in several forms. Some protect online content from illegal or otherwise unwanted downloads and sharing to protect the intellectual rights of online content producers. Digital rights management (DRM) is the management of legal access to restricted or copyrighted digital content through various tools or technological protection measures. Others provide frameworks that enable and even encourage widespread sharing and reuse, some of which are foundational to the licensing and copyright expectations of the open science initiatives of key funders and research institutions. Some are closely linked with digital reading platforms, which have other features in addition to rights management. Several of the services listed in this section play an important role in helping smaller and niche audiences, for example small and medium enterprises and unaffiliated scholars, to access subscription materials.

Examples

- Adobe Digital Editions (https://www.adobe.com/solutions/ebook/digital-editions.html) software allows publishers to manage the copying, printing, and sharing of ebooks with the option of implementing digital rights management (DRM). It is developed by Adobe Systems and is incorporated into various front-end services.

- Amazon's Kindle ecosystem is absolutely vital to many book publishers. It applies Digital Rights Management controls to ebooks to prevent unauthorized sharing and copying and requires using Amazon's hardware (Kindle device) or software app. Kindle also provides many value-added services such as dictionaries, highlighting, annotation, and more.

- Palace Project (https://thepalaceproject.org/about), managed by Lyrasis and a strategic partner of DPLA) is a free library-centered platform and e-reader app for digital content and services to allow libraries to purchase, organize, and deliver ebooks and other digital content to their patrons quickly and easily while protecting patron privacy.

- Creative Commons (CC, https://creativecommons.org) provides a standardized way to grant copyright permissions to creative work produced by individuals, organizations, and large companies. They allow users to make decisions about how their creative work will be copied, distributed, edited, remixed, and built upon, all within the boundaries of copyright law. It is run by CC, which is an international nonprofit organization.

- Copyright Clearance Center (https://www.copyright.com), owned by a US-based company, helps organizations integrate, access, and share information through licensing, content, software, and professional services. With its subsidiary RightsDirect (https://www.rightsdirect.com), it provides collective copyright licensing services for corporate and academic users of copyrighted materials.

- Research Solutions (https://www.researchsolutions.com), a subsidiary of Reprints Desk (R&D), provides workflow solutions for R&D-driven organizations by providing on-demand access to scholarly journal articles and other scientific, technical, and medical content. Reprints Desk focuses on electronic document delivery to institutional users (such as students and faculty).
- DeepDyve (https://www.deepdyve.com) provides access to scientific and scholarly articles, allowing users to buy individual papers or get a subscription that offers reading access from publishers in their network (including publishers like Wiley, Springer Nature, JAMA, and Wolters Kluwer). It is owned by DeepDyve, a US-based technology company.

Related Resources

"Copyright for Librarians." Berkman Klein Center for Internet & Society at Harvard University. https://cyber.harvard.edu/copyrightforlibrarians/Main_Page.

Harington, Robert. "Copyright, Creative Commons, and Confusion." *The Scholarly Kitchen*. 20 April 2020. https://scholarlykitchen.sspnet.org/2020/04/20/copyright-creative-commons-and-confusion/.

"Plan S Rights Retention Strategy." *cOAlition S*. https://www.coalition-s.org/rights-retention-strategy/.

"Rights & Licensing Hub." Rights & Licensing Hub. https://www.rightsandlicensing.co.uk/.

Manuscript Submission, Editorial Management, and Research Integrity

Description

Editorial management systems provide workflow tools for submitting and processing manuscripts, including assigning and managing peer reviewers. They also provide content management repository and distribution platforms to streamline tasks and support integrated workflow for writers, editors, and designers. In recent years, these systems have been expanded to allow for newer features such as preprint deposit and dataset submission and review, among others. Peer review is designed to assess the validity, quality, and originality of scholarly works for publication. It aims to maintain the integrity of science by filtering out poor quality, manipulated, or fraudulent content. There are enormous challenges to research integrity today and among the efforts to police the scholarly record include those intended to address threat vectors that have emerged in the manuscript submission and peer review processes.

Examples

- Aries Editorial Manager (https://www.ariessys.com/solutions/editorial-manager) is a cloud-based manuscript submission and peer review system for scholarly journals, reference

works, books, and other publications. It was developed independently and subsequently acquired by Elsevier.

- ScholarOne (https://clarivate.com/products/scientific-and-academic-research/research-publishing-solutions/scholarone) offers a generally similar set of series to Aries Editorial Manager. It is operated by Clarivate.

- Manuscripts.io (https://www.manuscripts.io/about) is an open-source collaborative editing environment, tailored for scholarly and research papers. It has been developed by Atypon Systems (Wiley).

- Authorea (https://www.authorea.com) is an online collaborative writing tool that allows researchers to write, cite, collaborate, host data and publish. It was previously owned by Atypon Systems (Wiley).

- Publishers have been relying on AI-powered applications (and machine-learning) in publishing for a number of years (e.g., copy editing and proofreading) but have recently developed more sophisticated practices, especially to bring efficiencies to the peer-review process and improve the quality of published research. Some examples include:

  - Frontiers Artificial Intelligence Review Assistant (AIRA, https://publishingpartnerships.frontiersin.org/our-platform) is used as a peer review technology to support in-house teams, editors, and reviewers in assessing and making decisions about the quality of manuscripts.

  - Writefull's Manuscript Categorization API (https://blog.writefull.com/acs-integrates-writefulls-manuscript-categorization-api-for-post-acceptance-classification) is integrated into the publishing workflow of the American Chemical Society (ACS) to automate the classification of manuscripts after acceptance based on language quality.

  - DeepAI (https://deepai.org/publication/what-makes-a-scientific-paper-be-accepted-for-publication) helps automate the peer-review process by using machine learning algorithms to identify potential flaws and weaknesses in research papers.

- Manuscript Exchange Common Approach (MECA, https://www.niso.org/standards-committees/meca) is a NISO project to develop a common means to easily transfer manuscripts between and among manuscript systems, such as those in use at publishers and preprint servers.

- Research integrity

  - Committee on Publication Ethics (COPE, https://publicationethics.org) is a UK based organization that educates and supports editors, publishers, universities, research institutes, and all those involved in publication ethics.

  - STM Integrity Hub (https://www.stm-assoc.org/stm-integrity-hub) by STM offers best practices and tools to detect research-integrity-offending manuscripts.

  - iThenticate (https://www.ithenticate.com) is a plagiarism detection service from Turnitin, LLC. It allows licensed publishers to check submitted manuscript documents against its database and the content of other websites with the aim of identifying plagiarism.

- A variety of services have been developed to support identifying peer reviewers or to support their work. These include:

  - Peer Review Taxonomy (https://osf.io/68rnz) development is led by the STM and the International Association of Scientific, Technical and Medical Publishers to help make

the peer review process for articles and journals more transparent and enables the community to better assess and compare peer review practices between different journals.

- ▫ Prophy (https://www.prophy.science) creates profiles for scientists (currently covering physical sciences and engineering, life sciences and medicine, economics, and social sciences) and ranks them according to their semantic similarity to a refereed manuscript or proposal. Prophy provides Referee Finder services for the European Research Council. For each grant proposal, Prophy generates a ranked list of relevant experts based on semantic and bibliographic information.
- ▫ Reviewer Credits (https://www.reviewercredits.com) is a peer review recognition platform that certifies and rewards the activity of peer reviewers by assigning them virtual credits that can be spent in a Reward Center. It is run by a startup company from Germany.

- ▪ Alternative infrastructures have been developed to support reviews of preprints or for the discussion of scholarship following publication. These services include:

  - ▫ Peer Community In (PCI, https://peercommunityin.org) is a non-profit scientific organization that aims to create thematic communities of researchers reviewing and recommending articles posted on preprint servers and other open-access repositories for free.
  - ▫ Review Commons (https://www.reviewcommons.org) is a platform for journal-independent peer-review in the life sciences. It also facilitates author-directed submission of Refereed Preprints to affiliate journals in order to streamline publication.
  - ▫ PubPeer (https://pubpeer.com) was developed by the PubPeer Foundation to facilitate discussion and review of scientific research after publication (post-publication peer review).

Related Resources

Awati, Mriganka. "Guest Post — A Case for Universal and Simplified Journal Systems." *The Scholarly Kitchen*. 20 August 20 2019. https://scholarlykitchen.sspnet.org/2019/08/20/guest-post-a-case-for-universal-and-simplified-journal-systems.

Carpenter, Todd A. "What Constitutes Peer Review of Data? A Survey of Peer Review Guidelines." *The Scholarly Kitchen*. 11 April 2017. https://scholarlykitchen.sspnet.org/2017/04/11/what-constitutes-peer-review-research-data/.

Coalition for Diversity and Inclusion in Scholarly Communications. "Guidelines on Inclusive Language and Images in Scholarly Communication." October 2022. https://c4disc.pubpub.org/guidelines-on-inclusive-language-and-images-in-scholarly-communication.

Horbach, Serge, and Willem Halffman. "The Changing Forms and Expectations of Peer Review." *Research Integrity Peer Review* 3, no. 8 (14 November 2018). https://doi.org/10.1186/s41073-018-0051-5.

Kaltenbrunner, Wolfgang, Stephen Pinfield, Ludo Waltman, Helen Buckley Woods, and Johanna Brumberg. "Innovating Peer Review, Reconfiguring Scholarly Communication: An Analytical Overview of Ongoing Peer Review Innovation Activities." *SocArXiv Papers*. 6 June 2022. https://doi.org/10.31235/osf.io/8hdxu.

Macdonald, Stuart. "The Gaming of Citation and Authorship in Academic Journals: A Warning from Medicine." *Social Science Information*. 7 February 2023. https://doi.org/10.1177/05390184221142218.

Mayo-Wilson, Evan, Sean Grant, Lauren Supplee, et al. "Evaluating Implementation of the Transparency and Openness Promotion (TOP) Guidelines: The TRUST Process for Rating Journal Policies, Procedures, and Practices." *Research Integrity Peer Review* 6, no. 9 (2021). https://doi.org/10.1186/s41073-021-00112-8.

Michaelmar, Ann. "Ask The Chefs: How Can We Improve the Article Review and Submission Process?" *The Scholarly Kitchen*. 26 March 2015. https://scholarlykitchen.sspnet.org/2015/03/26/ask-the-chefs-how-can-we-improve-the-article-review-and-submission-process.

Publons. "2018 Global State of Peer Review." Clarivate, 2018. https://clarivate.com/lp/global-state-of-peer-review-report.

Zhou, Hong and Sylvia Izzo Hunger. "Guest Post - Enabling Trustable, Transparent, and Efficient Submission and Review in an Era of Digital Transformation." *The Scholarly Kitchen*. 31 January 2023. https://scholarlykitchen.sspnet.org/2023/01/31/guest-post-enabling-trustable-transparent-and-efficient-submission-and-review-in-an-era-of-digital-transformation/.

Non-Consumptive Use

With growing demand for text and data mining (TDM) from scholars, a number of cross-publisher services have been developed with various features. Many of these services have been operated by aggregators (see Discovery, Syndication, and Aggregation), but there has also been at least one effort to create another model for TDM. The infrastructure created for TDM may eventually be repurposed for other goals, including providing training data for large-language models.

▪ Aggregator-provided services include TDM Studio (ProQuest), Digital Scholar Lab (Gale), Nexis Data Lab (LexisNexis), HathiTrust Research Center (Indiana University and University of Illinois with HathiTrust), Constellate (JSTOR).[19] TDM services and tools support non-consumptive text analysis by providing secure virtual environments and access

---

[19] Aggregator-provided services include TDM Studio, https://about.proquest.com/en/products-services/TDM-Studio; Digital Scholar Lab, https://www.gale.com/primary-sources/digital-scholar-lab; Nexis Data Lab, https://www.lexisnexis.com/en-us/professional/academic/nexis-data-lab.page; HathiTrust Research Center, https://www.hathitrust.org/htrc; Constellate, https://constellate.org/.

to right-cleared and public domain materials. Some also provide analysis and visualization tools.

- Crossref TDM API (https://www.crossref.org/documentation/retrieve-metadata/rest-api/text-and-data-mining) allows researchers to harvest full-text documents from participating members, regardless of whether the content is open access or subscription.

Persistent Identifiers

Description

A persistent identifier (PID) is a unique and long-lasting reference to digital objects, contributors, and organizations to facilitate discovery, access, linking, rights management, and assessment of scholarly content. Publishers, funders, and other organizations rely on PIDs to create trusted digital connections. Proliferation of digitally available research outputs are leading to the development of new machine-readable and interoperable PID types, such as the ones to support research instruments. Also, PIDs play an important role in enabling AI systems to access, integrate, and analyze data from multiple sources.

Examples

Object identifiers encompass a broad range of resources including books, articles, white papers, chapters, datasets, tables, figures, and videos. A single resource may have multiple identifiers associated with its different components (e.g., entire book, each chapter, individual figures, etc.).

- Handle (http://www.handle.net) is a decentralized identifier resolution system operated by the Corporation for National Research Initiatives (CNRI). These identifiers can be used to create URLs to access the resource without concern that its location may change.
- Digital Object Identifier (DOI, https://www.doi.org) identifies an information object persistently and allows it to be uniquely identified and accessed reliably. The DOI Foundation is a not-for-profit organization that governs the DOI system. Crossref is an official DOI registration agency of the International DOI Foundation. In addition, DataCite (https://datacite.org) is a global non-profit organization that provides DOIs for research data and other research outputs.
- There are many other types of object identifiers, including URIs, URNs, URLs, PURL, ARK.[20]

Contributor identifiers encompass establishing a profile for researchers, authors, and scientists to disambiguate them from others.

---

[20] Digital Preservation Coalition, "Persistent Identifiers," *Digital Preservation Handbook*, https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers.

- Open Researcher and Contributor ID (ORCID, https://orcid.org) provides a registry of unique researcher identifiers (including professional information such as affiliations, grants, publications, peer review) as well as a transparent method of linking research activities and outputs to these identifiers. It provides a persistent digital identifier that researchers can have and control to distinguish them from others and ensure recognition for all contributions. It is managed by ORCID, a global not-for-profit organization sustained by fees from member organizations.

- International Standard Name Identifier (ISNI, https://isni.org) is a global standard number for disambiguating and identifying contributors to creative works and those active in their distribution, including researchers, inventors, writers, artists, visual creators, performers, producers, publishers, aggregators, etc., and serves to disambiguate contributor names to improve search and discovery. ISNI identifiers and ORCID iDs are interoperable. ISNI is a global standard governed by ISO while ORCID iDs is an open registry where researchers can edit their own identifier page.

- ResearcherID (https://publons.com/wos-op) is a unique identifier to enable researchers to manage their publication lists, track citations, identify potential collaborators, and avoid author misidentification. It was developed by Thomas Reuters (used in Web of Science) and was integrated with Publons (Clarivate Analytics-owned platform) where researchers can track their publications, peer reviewing activity, and journal editing work.

Organization, funder, and grant identifiers encompass research institutions, funders, corporations, government agencies, etc., with the goal of enabling long-term linking between research/funding organizations to researchers and research outputs.

- Open Funder Registry IDs (https://www.crossref.org/services/funder-registry) is an open registry of grant-giving organization names and identifiers. Managed by Crossref, it is a freely downloadable RDF file (CCO-licensed and freely downloadable RDF file).

- Research Organization Registry (ROR, https://ror.org) is a global community-led registry of open persistent identifiers for research organizations. It is based on the Global Research Identifier Database (GRID), which collects and disambiguates institutional information and assigns a PID and metadata to each research institution. ROR is operated as a collaborative initiative by California Digital Library, Crossref, and DataCite.

- CCC Ringgold Identity Database (https://www.copyright.com/solutions-ringgold) provides persistent identifiers and associated metadata for organizations and institutions. Originally it was built to enable publishers to manage customer lists, given the complex nature of institutional names and organizational hierarchies. It is overseen by the Copyright Clearance Center, Inc.

Linking and Interoperability

- Standards such as DataCite Event Data (https://datacite.org/eventdata.html) and Crossref Event Data (https://www.crossref.org/services/event-data) enable links between

publications, research data, citations, software, reuse, documentation, and facilitate linking research outputs to funder information and other services.

▪ Scholix (http://www.scholix.org) aims to establish a high-level interoperability framework and guidelines for exchanging information about links between scholarly literature and research data.

Research Activity and Conference Identifiers

▪ Research Activity Identifier (RAiD, https://www.raid.org.au/) provides identifiers for research projects based on the global handle system. It also collects and records descriptive metadata about the project activities including funders and grants, organizations, articles, tools, etc. It is a not-for-profit service delivered by the Australian Research Data Commons (ARDC).

▪ Conference IDs (https://www.crossref.org/categories/conference-ids) aims to establish a PID system for registry of scholarly conferences. Longer term, it also aims to identify fraudulent and/or low-quality conferences. DataCite and Crossref will be implementing this metadata set to allow conferences to be registered as DOIs.

Related Resources

"Comparing ARKs, DOIs and other identifier systems." *ARK Alliance*. 3 November 2022. https://arks.org/about/comparing-arks-and-other-identifiers.

Cousijn, Helena, Ricarda Braukmann, Martin Fenner, Christine Ferguson, René van Horik, Rachael Lammey, Alice Meadows, and Simon Lambert. "Connected Research: The Potential of the PID Graph." *Patterns* 2, no. 1 (2021) https://doi.org/10.1016/j.patter.2020.100180.

De Castro, Pablo, Ulrich Herb, Laura Rothfritz, and Joachim Schöpfel. "Persistent Identifiers for Research Instruments and Facilities: An Emerging PID Domain in Need of Coordination." *Zenodo*. 1 February 2023. https://doi.org/10.5281/zenodo.7330372.

Jones, Phill and Alice Meadows. "Why Publishers Should Care About Persistent Identifiers." *The Scholarly Kitchen*. 21 June 2021. https://scholarlykitchen.sspnet.org/2021/06/21/why-publishers-should-care-about-persistent-identifiers.

Macgregor, George, Barbara S. Lancho-Barrantes, and Diane Rasmussen Pennington. "Measuring the Concept of PID Literacy: User Perceptions and Understanding of Persistent Identifiers in Support of Open Scholarly Infrastructure." *arXiv*. 21 February 21, 2023. https://arxiv.org/abs/2211.07367.

"Risks and Trust in Pursuit of a Well-Functioning Persistent Identifier infrastructure for Research." *Knowledge Exchange*. 1 September 2021 – 2 February 2023. https://www.knowledge-exchange.info/event/pids-risk-and-trust.

Schonfeld, Roger C. "Who Is Competing to Own Researcher Identity?" *The Scholarly Kitchen*. 6 January 2020.   https://scholarlykitchen.sspnet.org/2020/01/06/competing-researcher-identity.

Preservation

Description

There are a range of risks involved in managing digital content, including technical malfunctions, media obsolescence, and organizational failures—just to name a few. In light of such threats, digital preservation involves the maintenance of digital objects to ensure their authenticity, accuracy, and usability over time. It also requires taking into consideration information security, privacy, and compliance policies. The digital preservation and curation process involves a series of technical, intellectual, and managerial activities to enable discovery, access, and use of content by designated user communities over time. Administrative and preservation metadata includes technical information to support long-term management and preservation of digital collections (e.g., file formats, software dependencies, rights management, funder information, etc.). There are a number of organizations, systems, services, and standards that support the digital preservation lifecycle as well as individuals involved in digital preservation. In addition to the tools and standards discussed below, which focus on the preservation of works of scholarship, there is another set of systems that are focused on addressing the preservation of digital and digitized special collections of libraries and archives.

Examples

- Programmatic preservation is undertaken by initiatives and services to curate and preserve specific content types or collections, and in these cases of works of scholarship, is based on the establishment of trusted repositories.
    - CLOCKSS (https://clockss.org) is a global archive that preserves content on behalf of all libraries and scholars worldwide based on an implementation of the LOCKSS infrastructure.
    - HathiTrust (https://www.hathitrust.org) preserves and provides both open and controlled access to digitized books, journals, and government documents, among other material, and is governed by a partnership of libraries and operated through the University of Michigan.
    - Internet Archive (https://archive.org) preserves and provides both open and controlled access to an array of materials, including the open web as well as books, serials, and audiovisual materials. It is a not-for-profit organization that conducts extensive public fundraising.
    - Portico (https://www.portico.org) is an ITHAKA service that preserves scholarly journals, books, and digitized primary source materials for participating libraries.
- Metadata Standards and Registries:
    - PREMIS Data Dictionary for Preservation Metadata (https://www.loc.gov/standards/premis) is an international standard to support the

preservation of digital objects and ensure their long-term usability. It is hosted by the Library of Congress.

- ▫ Metadata Encoding and Transmission Standard (METS https://www.loc.gov/standards/mets) is an XML encoding standard which enables digital materials to be packaged with archival information. It is maintained by the Library of Congress.
- ▫ PRONOM (https://www.nationalarchives.gov.uk/PRONOM) is a file format registry in support of preservation planning activities for digital records. It was developed by the National Archives (UK) to support the accession and long-term preservation of electronic records held.
- ▫ Perma.cc (https://perma.cc/about#developer-overview) ensures that material cited by authors will always be accessible to readers, preserving the foundation of scholarship and reference online. It is developed and maintained by the Harvard Law School Library in conjunction with university law libraries across the country and other organizations in the "forever" business.
- ▫ Audit and certification processes enable organizations to evaluate their digital preservation infrastructures against an assessment framework. Examples include ISO 16363 (https://www.iso.org/standard/57950.html), Data Seal of Approval (DSA, https://www.coretrustseal.org/), and Digital Preservation Peer Assessment (https://www.nedcc.org/assets/media/documents/nedcc-DPA-Peer-5.16.pdf).

Related Resources

Kussman, Carol, Matt Schultz, Lauren Work, Nathan Tallman, and Paige Walker. "National Digital Stewardship Alliance (NDSA)." 20 February 2023. https://osf.io/4d567.

Cramer, T., Chip German, Neil Jefferies, and Alicia Wise. "A Perpetual Motion Machine: The Preserved Digital Scholarly Record." *Learned Publishing* 36, no. 2 (April 2023) https://doi.org/10.1002/leap.1494.

Handbook: Digital Preservation Handbook. Edited by Neil Beagrie. Digital Preservation Coalition, 2015. https://www.dpconline.org/handbook

NASIG Digital Preservation Committee. "NASIG Model Digital Preservation Policy." NASIG, 2022. https://nasig.org/NASIG-model-digital-preservation-policy.

--. "NASIGuide: Talking Points and Questions to Ask Publishers about Digital Preservation." NASIG, January 2020. https://nasig.org/Talking-Points-and-Questions-about-Digital-Preservation.

Levels of Digital Preservation Working Group. "Levels of Digital Preservation." National Digital Stewardship Alliance (NDSA), 2019. https://ndsa.org/publications/levels-of-digital-preservation/.

Pendergrass, Keith L., Walker Sampson, Tim Walsh, and Laura Alagna. "Toward Environmentally Sustainable Digital Preservation." *The American Archivist* 82, no. 1 (March 2019): 165–206. https://doi.org/10.17723/0360-9081-82.1.165.

Rieger, Oya Y., Roger C. Schonfeld, and Liam Sweeney. "The Effectiveness and Durability of Digital Preservation and Curation Systems." *Ithaka S+R*. 19 July 2022. https://doi.org/10.18665/sr.316990.

Vallières, Nathalie. "Open Access Journals Must Be Preserved Forever." *Public Knowledge Project*. 4 November 2021. https://pkp.sfu.ca/2021/11/04/open-access-journals-must-be-preserved-forever/.

Publishing Platforms and Repositories

Description

Content hosting and delivery platforms are utilized by publishers to provide access to publications for researchers and other users. Larger publishers such as Elsevier and Springer Nature will typically develop and maintain themselves on a homegrown basis. Beyond these large publishers, other houses will generally rely on a shared infrastructure provider.

Since the early 1990s, several alternative publishing and hosting platforms have emerged that are led by various research communities, funders, and libraries. Some aim to reform scholarly publishing, some want to speed up the process, and some encourage new forms of peer review. For instance, repositories facilitate content management services to support discovery, access, rights management, and archiving. Institutional repositories focus on facilitating the discovery and showcasing of institutional digital assets and encouraging open access to scholarly research. Disciplinary and preprint repositories (such as arXiv, bioRiv, and SSRN) enable access in ways that are typically organized by subject or field.

Examples

Hosting, Publishing, and Delivery Platforms

- Atypon and Silverchair operate the infrastructure for many dozens of scholarly societies, university presses, and other publishers.
- Several open source alternatives have been developed, including Open Journal Systems (OJS, https://pkp.sfu.ca/software/ojs) and Open Monograph Press (https://pkp.sfu.ca/software/omp), run by the Public Knowledge Project (Core Facility of Simon Fraser University), and JaneWay (https://janeway.systems/).
- There have been several initiatives by university presses to develop the infrastructure necessary to explore different kinds of publishing models, including Fulcrum (https://www.fulcrum.org), which was developed by the University of Michigan Library and Press working with partners from Indiana, Minnesota, Northwestern, and Penn State, as well as Manifold (https://manifoldapp.org), a collaboration between the CUNY Graduate Center, the University of Minnesota Press, and Cast Iron Coding.

Repositories

In this section, we focus on systems that are used for scholarly communication such as article preprints, rather than other kinds of materials such as library special collections.

- Repository systems
  - Digital Commons (Elsevier, https://bepress.com/products/digital-commons)
  - Invenio (https://www.tind.io) is an open-source digital repository framework that provides the tools for management of digital assets in an institutional repository and research data management systems. It is maintained by TIND, which is a CERN spin-off.
  - Other open source and community-based examples include Samvera (https://samvera.org), Islandora (https://www.islandora.ca), and Mukurtu (https://mukurtu.org).
  - LYRASIS brings together several open-source technologies (e.g., ArchivesSpace, DSpace, Fedora, etc.) to provide and sustain a shared infrastructure.[21]

- Directories and databases
  - OpenDOAR (https://v2.sherpa.ac.uk/opendoar/) is a global directory of open access repositories that allows users to search and browse through registered repositories based on a range of features, such as location, software or type of material held.
  - Registry of Open Access Repositories (http://roar.eprints.org/) is hosted at the University of Southampton and provides information about the growth and status of open access repositories throughout the world.
  - CHORUS (https://www.chorusaccess.org/about/faq) facilitates access to and information about articles and data reporting on funded research by collecting funder and institution metadata. It is an initiative of CHOR, Inc., a membership-based nonprofit organization.

Related Resources

Arnold, Denis, Bernhard Fisseni, Felix Helfer, Stefan Buddenbohm, and Peter Kiraly. "Repository Solutions - Technology Watch Report 1." *Zenodo*. March 2020. https://doi.org/10.5281/zenodo.3873027.

Lin, Dawei, Jonathan Crabtree, Ingrid Dillo, et al. "The TRUST Principles for Digital Repositories." *Scientific Data* 7, no. 144 (2020). https://doi.org/10.1038/s41597-020-0486-7.

Macgregor, George. "Digital Repositories and Discoverability: Definitions and Yypology." In *Discoverability in Digital Repositories*, edited by Liz Woolcott and Ali Shiri. London: Routledge, 2023. https://doi.org/10.4324/9781003216438-3.

Maxwell, John W., Erik Hanson, Leena Desai, Carmen Tiampo, Kim O'Donnell, Avvai Ketheeswaran, Melody Sun, Emma Walter, and Ellen Michelle. "Mind the Gap: A

---

[21] Lyrasis Community Programs, https://www.lyrasis.org/programs/Pages/default.aspx.

Landscape Analysis of Open Source Publishing Tools and Platforms." Simon Frasier University, July 2019. https://mindthegap.pubpub.org/.

Puebla, Iratxe, Jessica Polka, and Oya Y. Rieger. "Preprints: Their Evolving Role in Science Communication." *Against the Grain*, 2022. https://doi.org/10.3998/mpub.12412508.

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

Schöpfel, Joachim and Otmane Azeroual. "Current Research Information Systems and Institutional Repositories: From Data Ingestion to Convergence and Merger." In *Future Directions in Digital Information*, edited by David Baker and Lucy Ellis. Chandos Publishing, 2021. https://www.sciencedirect.com/science/article/pii/B9780128221440000021.

Sondervan, Jeroen, Jean Francois Lutz, and Bianca Kramer. "Alternative Publishing Platforms." *Knowledge Exchange*. 2022. https://knowledge-exchange.pubpub.org/pub/73tb00rf/release/3.

The National Science and Technology Council. "Desirable Characteristics of Data Repositories for Federally Funded Research." *Smithsonian Libraries and Archives*. White House Office of Science and Technology Policy, April 2022. https://doi.org/10.5479/10088/113528.

US Repository Network. "Desirable Characteristics of Digital Publication Repositories, Version 1.0." 31 March 2023. https://sparcopen.org/wp-content/uploads/2022/10/Desirable-Characteristics-of-Digital-Publication-Repositories-APPROVED-20230331.pdf.

Research Data Curation and Management

Description

Research data curation entails the documentation, organization, and maintenance of data sets to facilitate their discovery, sharing, and preservation. Funders are increasingly emphasizing the importance of curation and quality assurance when choosing a repository.

Examples

Repositories

There are a number of community-based and open source repositories that offer services and tools to host, publish, and preserve research data, including figshare, Dryad, Dataverse, ICPSR, https://www.icpsr.umich.edu), Open Science Framework (Center for Open Science, https://www.cos.io), Mendeley Data, and Zenodo. Initiatives such as the Research Data Alliance

(RDA), Data Curation Network (DCN, https://datacurationnetwork.org) are examples of community-based partnerships that facilitate the developing and adopting infrastructure that promotes data-sharing and data-driven research.

Principles and Standards:

- FAIR Principles (https://www.go-fair.org/fair-principles/guidelines) provide guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets. The principles emphasize machine-actionability (automated processes to find, access, interoperate, and reuse data) to deal with the increase in volume, complexity, and creation speed of data. Persistent identifiers (PIDs) play a central role in the FAIR ecosystem by enabling ways to refer to entities in a permanent way.

- CARE Principles for Indigenous Data Governance (CARE, https://www.gida-global.org/care) complement the FAIR principles by encouraging open and other data movements to consider both people and purpose in their advocacy and pursuits.

- The Scientific Knowledge Graph - Interoperability Framework (SKG-IF, https://rd-alliance.org/group/scientific-knowledge-graphs-interoperability-framework-skg-if-wg/case-statement/scientific) targets the definition of a framework to enable a seamless exchange of information among diverse initiatives regarding Scientific Knowledge Graphs, intended as knowledge bases of scholarly knowledge content (e.g. repositories, databases, catalogues, knowledge graphs, LOD collections). SKG represents research data as structured, interlinked, and semantically rich.

Directories and Tools

- Repository Finder (https://repositoryfinder.datacite.org) aims to help researchers locate an appropriate repository to deposit their research data. The pilot project is led by the American Geophysical Union (AGU) in partnership with DataCite and the earth, space and environmental sciences community.

- DataSeer (https://dataseer.ai) enables researchers to comply with data sharing policies and allows them to monitor compliance with data policies. Funders, journals, and institutions can use DataSeer to find all of the data associated with a corpus of articles, or use it to promote compliance with their data sharing policies.

- DMP Tool (https://dmptool.org) is an open-source online application that helps researchers create data management plans (DMPs) to comply with funder requirements. It also has direct links to funder websites. DMP Tool is a service of the California Digital Library, a division of the University of California.

- Research Electronic Data Capture (REDCap, https://projectredcap.org/about) is a web application for building and managing online surveys and databases. It is run by Vanderbilt University and overseen by the international REDCap Consortium.

Other Content Types

Beyond research data, we have begun to see efforts to identify other research artifacts that might benefit from being made computable as first-class research objects, for example research

protocols and software code. Some of the services that provide platforms for such objects include:

- Code Ocean (https://codeocean.com/) is a cloud-based computational reproducibility platform that enables researchers and developers to share, discover, and run code published in academic journals and conferences. It provides open access to the published software code and data.
- Protocols.io (Protocols.io) is an open-access platform for detailing, sharing, and discussing molecular and computational protocols that can be useful before, during, and after publication of research results.

Related Resources

Arguillas, Florio, Thu-Mai Christian, Mandy Gooch, Tom Honeyman, and Limor Peer. "10 Things for Curating Reproducible and FAIR Research (Version 1.1)." *Zenodo*. Research Data Alliance, 2022. https://doi.org/10.15497/RDA00074.

Aryani, Amir. "Data Description Registry Interoperability Model." Research Data Alliance (RDA), 2018. https://www.rd-alliance.org/group/data-description-registry-interoperability-ddri-wg/outcomes/data-description-registry-interoperability

Chiarelli, Andrea. "The Role of Publishers in the Evolving Research Data Landscape." *ALPSP* (Video). February 2023. https://www.youtube.com/watch?app=desktop&v=7zyvj9FRyW0&t=1s.

Jonesmar, Phill. "Is it Finally the Year of Research Data? – The STM Association Thinks So." *The Scholarly Kitchen*. 5 March 2020. https://scholarlykitchen.sspnet.org/2020/03/05/is-it-finally-the-year-of-research-data-the-stm-association-thinks-so/.

Manghi, Paolo, Andrea Mannocci, Francesco Osborne, Dimitris Sacharidis, Angelo Salatino, and Thanasis Vergoulis. "New Trends in Scientific Knowledge Graphs and Research Impact Assessment." *Quantitative Science Studies* 2, no. 4 (2022): 1296–1300. https://doi.org/10.1162/qss_e_00160.

"NOW Roadmap Grant for Digital Infrastructure Social Sciences and Humanities." Open Data Infrastructure for Social Science and Economic Innovations (ODISSEI), February 2023. https://odissei-data.nl/en/2023/02/nwo-roadmap-grant-for-digital-infrastructure-social-sciences-and-humanities.

Ruediger, Dylan. "Guest Post — The Outlook for Data Sharing in Light of the Nelson Memo." *The Scholarly Kitchen*. 6 September 2022. https://scholarlykitchen.sspnet.org/2022/09/06/guest-post-the-outlook-for-data-sharing-in-light-of-the-nelson-memo/.

The National Science and Technology Council. "Desirable Characteristics of Data Repositories for Federally Funded Research." *Smithsonian Libraries and Archives*. White House Office of Science and Technology Policy, 2022. https://doi.org/10.5479/10088/113528.

Supporting New Business Models and Policy Environments

Description
New business models, such as APC-based open access, transformative agreements, subscribe-to-open, and more, require or may benefit from a variety of shared infrastructure. Additionally, policy guidance and mandates, including on open access initiatives, are similarly producing various kinds of shared infrastructure. The shared infrastructure in this category is often leveraged by libraries to more effectively manage the open access activities and evolving business models.

Examples
- Open Access Management Solution for Institutions (OABLE, https://oable.org/) was developed by Knowledge Unlatched (a Wiley brand) to provide a suite of open access tools to help institutions manage open access agreements with publishers.
- SHERPA ROMEO (https://v2.sherpa.ac.uk/romeo) is an online resource to aggregate and analyze publisher open access policies and provide summaries of publisher copyright and open access archiving policies on a journal-by-journal basis. It is operated and funded by Jisc.
- Plan S Journal Checker Tool (https://www.coalition-s.org/resources/journal-checker-tool-jct) is an online tool to enable authors funded by research organizations supporting Plan S to find compliant routes for publishing their articles. Plan S Journal Comparison Service (https://www.coalition-s.org/journal-comparison-service) enables libraries, library consortia, and funders to assess if the fees they pay are commensurate with the publication services delivered. Both tools are supported by cOAlition S (https://www.coalition-s.org/about).
- OurResearch (https://ourresearch.org) is a non-profit organization that provides open source tools to uncover, connect, and analyze research products. Examples include unsub to help libraries cancel subscriptions, unpaywall to provide a database of open access scholarly articles with an API and browser extension, and Impactstory to promote investment in nontraditional research outputs (such as datasets, code, and blogs) by showcasing impact beyond traditional citation.[22]

---

[22] "unsub," Our Research, https://ourresearch.org/projects#unsub; "unpaywall," Our Research, https://ourresearch.org/projects#unpaywall; Impactstory, https://profiles.impactstory.org.

Usage Data

Description

Usage data provides insights about users' interactions with information resources such as publisher and aggregator platforms. Libraries analyze usage statistics to monitor and assess how their resources are being used to ensure that they are providing the right resources and getting good value out of their budget. Some have speculated that such institutional usage reporting may decline in value in an open access environment. Beyond reporting to subscribers, publishers and information vendors leverage usage data for a variety of reasons, for instance for market analysis to understand usage, readers, and assess distribution channels. In an increasingly open access environment, there is speculation that usage data will be used to demonstrate the impact of journals to authors and for publishing services agreements to universities.

Examples

- The Counting Online Usage of Networked Electronic Resources (COUNTER, https://www.projectcounter.org) standard enables vendors and publishers to supply their library customers with consistent and comparable usage data. Using COUNTER reports, libraries can get statistics about the number of downloads, searches, sessions, and turn-aways. The COUNTER standard is maintained by an international non-profit membership organization of libraries, publishers, and vendors, which provides an audit function to ensure the reliability of reporting.

- Standardized Usage Statistics Harvesting Initiative (SUSHI ANSI/NISO, https://niso.org/standards-committees/sushi) is a protocol to automate the harvesting of COUNTER reports from SUSHI-compliant providers. It is based on an automated request and response model for the harvesting of electronic resource usage data that can replace the user-mediated collection of usage data reports.

- Google Analytics (https://analytics.google.com/analytics) is a web analytics tool offered by Google that collects, analyzes, and reports website traffic data. It is used by some resource providers to monitor user activities with research databases by looking at bounce rates, visits, selections per page view, average time on page, and visit depth, etc.

- One challenge that has emerged for incorporating usage data into value assessments is the case of materials that are distributed through multiple platforms.

  - For open access materials, there are efforts to aggregate these usage data to provide a more complete dataset. Such efforts include Open Scholarly Communication in the European Research Area for Social Sciences and Humanities Metrics (OPERAS, https://operas-eu.org/services/metrics-service/) and the OA Book Usage Data Trust.[23]

  - Additionally, in support of the developing syndication model (see Discovery, Syndication, and Aggregation), publishers have created the Distributed Usage Logging

---

[23] OA Book Usage Data Trust, https://www.oabookusage.org.

standard (originally developed through Crossref and now maintained through STM Solutions).[24]

- Hum (https://www.hum.works) is a very different kind of service, a Customer Data Platform for publishers and media organizations to collect, connect, and act on their first-party data. Owned by Hum LLC, it generates insights on how publishing audiences interact with content (whether journal, book, blog, video, or content marketing) and creates user profiles of readers, authors, and librarians.

Related Resources

Barnes, Lucy. "What We Talk About When We Talk About Book Usage Data Metrics." *Open Book Publishers Blog*. 2019. http://doi.org/10.11647/OBP.0173.0101.

Hinchliffe, Lisa Janicke and Roger C. Schonfeld. "Diverting Leakage to the Library Subscription Channel." *The Scholarly Kitchen*. 16 July 2019. https://scholarlykitchen.sspnet.org/2019/07/16/diverting-leakage-to-subscription.

Michael, Ann. "Ask The Chefs: What Is The Most Important Data For A Publisher To Capture?" *The Scholarly Kitchen*. 5 April 2017. https://scholarlykitchen.sspnet.org/2017/04/05/ask-the-chefs-data-for-publisher.

Wang, Jian and Hannah McKelvey. "Usage Statistics." *SPARC*. 2021. https://sparcopen.org/our-work/negotiation-resources/data-analysis/usage-statistics.

---

[24] Distributed Usage Logging (DUL) Public-Key Registry, https://www.stm-assoc.org/stm-solutions/dul; Distributed Usage Logging Collaboration, https://www.crossref.org/community/project-dul.