

2024-02-01

Exploring National Infrastructures to Support Impact Analyses of Publicly Accessible Research: A Need for Trust, Transparency and Collaboration at Scale

Kemp, Jennifer; Watkinson, Charles; Drummond, Christina

<https://dx.doi.org/10.7302/22166>

<https://hdl.handle.net/2027.42/192166>

<http://creativecommons.org/licenses/by-sa/4.0/>

Title: Exploring National Infrastructures to Support Impact Analyses of Publicly Accessible Research: A Need for Trust, Transparency and Collaboration at Scale

Authors: Jennifer Kemp, Strategies for Open Science (Stratos) 0000-0003-4086-3196, Charles Watkinson, University of Michigan 0000-0002-9453-6695, Christina Drummond, University of North Texas 0000-0001-5794-0413

Keywords: usage data; metrics; open access; infrastructure; research funding, policy

Abstract:

Usage data on research outputs such as books and journals is well established in the scholarly community. However, as research impact is derived from a broader set of scholarly outputs, such as data, code and multimedia, more holistic usage and impact metrics could inform national innovation and research policy. Usage data reporting standards, such as Project COUNTER, provide the basis for shared statistics reporting practice; however, as mandated access to publicly funded research has increased the demand for impact metrics and analytics, stakeholders are exploring how to scaffold and strengthen shared infrastructure to better support the trusted, multi-stakeholder exchange of usage data across a variety of outputs. In April 2023, a workshop on *Exploring National Infrastructure for Public Access and Impact Reporting* supported by the United States (US) National Science Foundation (NSF), explored these issues. This paper contextualizes the resources shared and recommendations generated in the workshop.

I. Introduction

Scholarly communication infrastructures have been a topic of increasing interest in recent years. The August 2022 US Office of Science and Technology Policy (OSTP) memo *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research* (Nelson, 2022) spurred conversation on how to ensure the sustainability, interoperability, and scalability of the core systems that provide a reliable, trusted foundation for evidence-based decision-making. Throughout the US Government's 2023 Year of Open Science, scholarly communications stakeholders discussed infrastructures to support evolving requirements. While the memo itself doesn't mention infrastructure, compliance with access mandates necessitates grant reporting that requires support from publishers, content platforms and related systems. Stakeholders leverage persistent identifiers (PIDs) to enable funding agencies, researchers, and others to track and evaluate the impact of scholarly outputs.

Scholarly publishing systems and workflows are largely built around journal articles, often overlooking other outputs like books, datasets, and media, which may require their own infrastructures or modifications to existing ones. Supporting the increasing availability and use of diverse scholarship beyond articles needs further attention (Watkinson, 2023).

Infrastructures supporting specific outputs have emerged. Complex book objects (Ricci, 2023) are the focus of The Directory of Open Access Books (DOAB) and the Open Access eBook

Usage Data Trust (OAeBUDT). Data deposition, which the 2022 Nelson memo introduced as a compliance requirement, relates to efforts like DataCite, Make Data Count, and Dryad. Institutional repositories are supported by Jisc's Institutional Repository Usage Statistics aggregation service) for UK (IRUS-UK) and US (IRUS-US) institutions (Lambert, 2023).

This paper describes the US infrastructure landscape for scholarship impact analysis in 2023, with the Nelson memo and the European Open Science Cloud's (EOSC) Interoperability Framework providing context. Existing challenges to making usage and impact data more FAIR (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al., 2013) and recommendations to improve the current state are presented and reflect discussions among invited experts at the *Exploring National Infrastructure for Public Access and Impact Reporting* workshop (Drummond, 2023a).

II. The Current Landscape

Standardized content usage reporting has been a mainstay of scholarly communications for over twenty years. Project COUNTER has evolved into a community-maintained usage statistics reporting standard that governs the process of distributed data processing, i.e. 'COUNTING' across publishers, libraries, and their service providers (such as content hosting platforms). Usage and impact data must be retrieved and/or aggregated from across repositories, platforms, and services. Often this toolchain includes proprietary, selective and/or duplicative services, many of which sources don't allow for automated usage harvesting.

Use cases for scholarly usage data vary across diverse stakeholders. Publishers, librarians and their vendors make up the COUNTER membership yet technical and resource limitations combined with content distributed across multiple locations means that no stakeholder, including research funders, has a complete view of usage (Mellins-Cohen, 2023). Publishers have multiple touchpoints in this ecosystem and may rely on technology partners like hosting platforms to provide COUNTER-compliant usage to their internal editorial and marketing teams in addition to authors and librarians, who have long used usage to inform collection development (Drummond & Hawkins, 2022).

Access mandates for publicly funded scholarship coupled with the movement for responsible research metrics have led to significant interest from various stakeholders in analyzing usage across business models, disciplines, and content formats. Open Access (OA) directly involves authors and funders as primary stakeholders of usage data, changing incentive structures within this complex landscape (Ricci, 2023). For example, in the UK the Jisc IRUS service aggregates content across 31 item types, providing standardized analytics for institutional repositories (IRs), using COUNTER data in a scalable, extensible model (Lambert, 2023). The European Open Science Cloud (EOSC) uses the COUNTER standard to support interoperable data reporting for over 200 repositories (Manghi, 2023).

In this global environment, scholarship moves across borders, requiring interoperable or shared solutions; such organizations and initiatives overlap, collaborate and rely upon the same infrastructures (Stern, 2023).

Mixed Definitions and Vocabularies

Across the variety of use cases, there is no common terminology, principles or rules of engagement around the access and use of data for impact assessment. Federal policy distinguishes *public access* from *open access*; both drive the aggregation of usage data. Infrastructures that support licensing and business model identification often do not commonly or consistently distinguish between the multiple variations. OA indicators are likely to substitute for the term 'public access,' which is not commonly or reliably used in the scholarly communications supply chain. Yet common language and frameworks are key for interoperability (Manghi, 2023). The NSF workshop identified the need for a crosswalk or mapping to account for differences in terminology as well as measures and schema (Drummond, 2023a).

Metadata Interoperability and Persistent Identifier (PID) Infrastructures

Metadata is crucial to identify and link associated outputs across different platforms and to minimize broken URLs through the use of PIDs. Consider a preprint with an underlying dataset and associated software code that is published as a journal article with peer review reports. These connected outputs are not often linked together with interoperable metadata, including license information.

Metadata depends on its own set of infrastructures, such as Crossref and DataCite. For example, data citations link datasets registered with a DOI in DataCite to related content registered with Crossref. Interoperability and coordination between such organizations is necessary. The scholarly communications supply chain must work together to ensure such information is interoperable and linkable.

Usage Considerations Across Scholarship Outputs

Journal articles and books have long been part of the online scholarly usage landscape. Other outputs often don't have dedicated, standardized or reliable usage reporting. Gray literature, i.e. reports and papers produced outside of traditional publishing, is diverse, diffuse and often outside of established library and archive acquisition and reporting workflows. Yet, as more of it becomes accessible online through services like Overton and Policy Commons, the scholarly community may show increasing interest in its usage, particularly as it relates to scholarship impacts on public policy.

While the Nelson memo specifically references peer-reviewed content, linking all types of scholarly outputs to policy goals may be of growing interest for funding agencies and the larger scholarly community. Issues like climate change and the pandemic response have heightened

attention to how public access impacts the speed of innovation, raising interest in how outputs like preprints, software and protocols are accessed and used worldwide.

Data

Defining data across disciplines is a challenge (Sever, 2023) and areas like the humanities often don't generate what is traditionally considered data (Ruediger & MacDougall, 2023). Data citation practices are still developing for researchers and publishers (Lowenberg et al., 2021) and so evaluating the impact of the full breadth of outputs, publicly funded or not, remains a significant hurdle. Per Kristi Holmes, "A true understanding of the investment, reach, and impact made in publicly accessible research data is only possible with open, transparent, and responsible data metrics" (Holmes, 2023, p.3).

Books

Books and book metadata have a particularly complex supply chain rooted in print sales with book usage distributed across multiple platforms (Clarke & Ricci, 2021). While increased distribution benefits readers and authors, it complicates the process of identifying, aggregating and reporting usage for a given title with multiple URLs and digital locations. Digital book chapters also exacerbate difficulties when not given distinct Digital Object Identifiers (DOIs) and metadata to facilitate linking (Lin, 2016). This makes it challenging for research administrators to identify faculty who have authored book chapters (Bryant et al., 2021), and assess the impact of such work (Kemp & Taylor, 2020). Bibliometricians may put in considerable manual effort to get a full picture of the institutional affiliations involved in a contributed volume. Some publishers don't register DOIs for books such as professional medical titles (Conrad & Urberg, 2023). Without such core metadata the ability to link funding to outputs is challenged (Tkaczyk, 2023) and the entire landscape suffers as a result (Conrad & Urberg, 2021).

Distributed Content and Other Assessment Challenges

While digital books have long been hosted on multiple platforms, the liberal reuse licenses of publicly accessible resources mean that other output formats will also be increasingly distributed. From an impact assessment perspective, multiple content hosting providers and platforms (URLs), and different versions of works must be considered. Getting holistic usage metrics from all sources and ensuring they are comparable (if they are not COUNTER-compliant) takes time. Individuals within organizations in the usage data supply chain currently decide how to standardize and aggregate statistics for individual outputs, creating a downstream ripple effect that influences how interoperable the data is that underpins assessment of OA impacts by discipline, region, institution, output type, and format.

Citations have long been the main currency of evaluating research outputs (White, 2019) but they reflect only a subset of scholarly publication usage. Non-citation-based use can include student and researcher reading and downloads, annotations and social network sharing, and context-specific professional workforce engagement by clinicians and other 'on the job' practitioners. Sometimes usage data is explicitly acknowledged, other times it is implied, though

“It is well understood that metrics include more than citations and usage” (Lowenberg et al., 2019, p.17).

Though the focus of research assessment is often the researcher or institution, the meta-analysis fields of bibliometrics and scientometrics are growing (Organisation for Economic Co-operation and Development [OECD], 2023). Improving the evaluation of public and open access models will require timely, granular access to interoperable, high quality, ‘trusted’ impact metrics.

Yet entities that create usage and impact data may face barriers when adopting standards or engaging in multinational efforts. Challenges to providing usage reporting exist for many smaller publishing organizations, including library publishers (Mellins-Cohen, 2023) and university library repositories (Lambert, 2023).

Project COUNTER itself operates on a volunteer network managed by a single half time employee (Mellins-Cohen, 2023). It is reasonable to consider the resources needed for the interconnected infrastructures that support metrics interoperability. Workshop participants discussed operational and sustainability risks associated with thin staffing levels, noting that financing infrastructure staffing and cross-infrastructure coordination should be a priority.

National and Multinational Factors

The practicalities of access mandates, related reporting requirements, and the data exchange to support them are inherently international. However, it is unclear if global infrastructures or federated networks of domestic infrastructures are best positioned to interoperate. Multi-national infrastructures, including those that support scholarly communications, must carefully navigate layers of complex regulation. Multiple NSF workshop participants noted that the challenges of data sharing are often less technical than they are administrative, legal or policy-based (Drummond, 2023a).

Transparency, trust and participation incentives have been highlighted as requisite to advancing global usage data interoperability. Alignment among stakeholders was a theme of the NSF workshop, particularly for incentives (Drummond, 2023a). However, cultural perspectives vary on the sensitivity, security, and data management requirements. From a legal and ethical perspective, the exchange of Internet Protocol (IP) addresses tied to the ‘use’ or readers of scholarship at a specific location or by a particular author, or scholar can raise concerns over potential negative uses of such information, from misinformation campaigns to surveillance. While many countries treat IP addresses as personally identifiable information subject to privacy protections such as the European Union’s (EU) General Data Protection Regulation (GDPR); the US does not have such a comprehensive federal law. Rather, the regulation of IP addresses as personally identifiable information varies by state in a very dynamic legislative space.

Whether usage data can legally be exchanged is a separate question from whether it can ethically be exchanged. In the workshop, the CARE Principles for Indigenous Data Governance (Collective Benefit, Authority to Control, Responsibility, Ethics) (Carroll et al., 2020) were raised

as a way to consider data sensitivity. Is there a collective benefit to sharing data? If so, who has the authority to allow data sharing and use and the responsibility for ensuring such use is appropriate and legal? Is such sharing and use ethical in the eyes of the people behind the data - in this case the individuals and organizations behind the IP addresses attributed to usage data? At the workshop, organizer Christina Drummond suggested that scholarly communications is rapidly approaching a future where FAIR and CARE are necessary to ethically share usage data.

Exploring such matters across borders requires coordination and the US is widely considered to be playing catchup to Europe and other leading regions and countries on open and public access, data brokerage regulation and privacy policy. Europe is supporting infrastructure networks through funding mechanisms like the European Research Infrastructure Consortium (ERIC) approach, which allows infrastructure organizations to achieve economies of scale and better alignment as a single regional network (Stern, 2023). The approach of America's agencies and innovation networks remains an open question.

Summary

Coordinated action in the distributed environment of usage metrics involves many stakeholders who are often in collaborative competition or 'co-opetition.' This necessitates trust and transparency in addition to logistical interoperability. Accurately measuring open access publishing is itself a policy issue and highly influenced by the data sources used (Basson et al., 2022). Coordination could strengthen infrastructure networks through scale while accommodating the complexities of varied scholarship stakeholders and use cases.

III. Opportunities for Action

Invited experts at the NSF workshop identified four key areas for action to strengthen infrastructure for impact metrics for publicly accessible scholarship: 1) engaging stakeholders 2) piloting a Minimum Viable Product, 3) understanding shared values and principles and 4) addressing usage data ownership (Drummond, 2023a). This section outlines opportunities identified to better support impact and usage analytics across the research ecosystem.

1. Engage Stakeholders to Best Coordinate and Leverage Resources

Experts at the NSF workshop noted how infrastructure projects require substantial commitments of financial and staff resources. Participants emphasized that care should be taken to avoid effort duplication. The problems faced are too grand for any effort to go it alone and time cannot be wasted "recreating the wheel" (Drummond 2023a, p.18). Rather, participants strongly voiced a need for funding to support coordination and collaboration, so aligned efforts can advance together. Specific resources to engage included:

- Standards, e.g. COUNTER, Open Researcher and Contributor IDs (ORCID), Research Organization Registry (RORs)

- Policy guidance, e.g. the OSTP Nelson memo
- Open metadata, e.g. Crossref
- Related efforts and infrastructures, e.g. COUNTER, OASwitchboard, OA Book Usage Data Trust (OAeBUDT)
- Existing research metrics and dashboard providers, e.g. Digital Science

Greg Tananbaum spoke about how equity relates to usage data, noting that the Open Research Funders Group (ORFG) is looking to “build a system that enables evidence-based policymaking, that improves public engagement with and trust in science” (Tananbaum, 2023, 5:00).

Leverage Established Networks

Given the diversity of scholarly outputs across public and private stakeholders, coordinating engagement across professional networks could increase the speed of infrastructure development while fostering interoperability. As described below, different types of organizational networks and consortia within the US are well-positioned to assist in engaging their membership.

National Research and Education Networks (NRENs)

NRENs illustrate how federated networks of independent infrastructures can provide internet and identity access and authentication services for higher education worldwide. In the US, Internet2 provides such services; outside the US, peer NRENs provide support for open science. Such membership organizations offer support spanning scholarly disciplines and associated communities, including the varied output formats that are increasingly open access and distributed across platforms. An opportunity exists to explore how scholarly infrastructures could leverage the US NREN and its global peer network.

Consortia and Durable Coalitions

In the US, the culture of collaboration among libraries is well established through membership consortia. Libraries each belong, on average, to 2.5 consortia. One such consortium, LYRASIS, has over one thousand members (Lair, 2023) and serves as a US hub to support libraries connecting to global infrastructures ORCID and IRUS. This example illustrates how membership networks already support rapid, community-informed infrastructure development through scalable network-based engagement.

Funding agencies and private foundations coordinate related efforts as well. The ORFG, for example, works “to develop actionable principles and policies that promote greater dissemination and transparency and replicability and reuse of papers, data, and a whole range of other research types” (Tananbaum, 2023, 1:01). They support The Higher Education Leadership Initiative for Open Scholarship (HELIOS), an example of the “durable coalitions” the ORFG aims to facilitate (Tananbaum, 2023, 2:54).

Workshop experts agreed that engaging existing networks should be the starting point when considering allocating resources, whether for funding, policy or technology opportunities.

2. Learn from a Minimum Viable Product (MVP)

In the NSF workshop, experts documented the need to learn from an MVP related to impact and usage data exchange, such as the OAeBUDT IDS, to support multiple use cases. Specific questions they suggested an MVP could answer included:

- What users will pay for, to ensure sustainability
- Whether the focus should be on impact, use, or both
- What data is of value to exchange
- What problem would be solved through controlled data exchange

Drummond noted that OAEBUDT IDS MVP participation is possible in 2024, pending funding.

Data Intermediary Efforts

Europe's Data Governance Act empowers a network of neutral, certified data intermediaries known as International Data Spaces (IDS) to ensure the ethical exchange and controlled use of sensitive data across public and private parties (Nagel & Lycklama, 2021). Since 2020, a group of global stakeholders, including universities, libraries, book discovery platforms and publishers, have been exploring how to leverage the IDS model to address data sensitivity concerns pertaining to the international aggregation and curation of granular OA book usage data (Drummond, 2023b). Using the narrow use case of digital book-related view and download metrics, the OAeBUDT is piloting the IDS Reference Architecture Model and its associated standards to securely route and distribute sensitive usage data across public and private parties at scale. It features a community governance model, maintains data sovereignty, and accommodates the range of public access stakeholders as data contributors and recipients. While currently focused on books, stakeholders include extensibility as a core principle, paving the way to extend such infrastructure should the IDS model generate a return and make sense for other publicly accessible outputs given different data supply chains and impact vocabularies.

Federal Demonstration Project Networks

Federal agencies are coordinating on pilots to improve access to potentially sensitive federal data sets through efforts such as The National Secure Data Service Demonstration Project (NSDSD), which aims to facilitate use of US federal and non-federal data to better inform decision making. It requires collaboration across government and external stakeholders and is currently testing a pilot with the America's Datahub Consortium (Madray, 2023). Together, their pilot project aims to facilitate the secure public/private sharing and use of federal statistics that relate to the general public.

Infrastructure vs. Projects or Research Centers

While research into MVP infrastructure development is underway, scholars studying data collaboratives across the globe have found that many become one-offs as they don't scale, struggle with sustainability, or raise concerns over how the data itself is exchanged (Verhulst,

2023). Yet, successful trusted data intermediaries that exchange data for independent, cooperative, or directive use provide instructive examples (Verhulst et al., 2019).

3. Understanding Shared Values and Principles

NSF workshop groups unanimously suggested learning from existing models and emerging efforts, including the ZERO Copy Integration Framework in Canada, La Referencia, and REDALYC in Latin America, in addition to the European IDS model (Drummond, 2023a). Even with considerable national and international investment, resources will be limited yet the scope of research impact assessment is significant. Cross-stakeholder collaboration is necessary to avoid effort duplication, diverging standards, and fractured community governance.

Such collaboration is also necessary to agree on definitions and shared language, establish values and principles, inform best practices, and develop standards. Parties to bring together span research, publishing, and policy, including: funding agencies, libraries and their consortia, publishers and their service providers, federal agencies and policy makers, and scholars and their affinity groups. Example efforts include the US National Information Standards Organization (NISO), the Data Curation Network (DCN), the Research Data and Preservation Association (RDAP), the Research Data Alliance (RDA) and HELIOS; whether as models or partners, including them in planning efforts can be mutually beneficial. The experiences of efforts in the EU, with 450 million people in 27 countries using 24 languages, may be useful for US organizations navigating a patchwork quilt of state-based regulations alongside federal data policy.

European efforts that could inform and/or partner with US infrastructure development include:

- the infrastructure consortia Open Scholarly Communication in the European Research Area for the Social Sciences and Humanities (OPERAS), which is in the process of becoming a European Commission-recognized ERIC that unites the organization's members running infrastructures like IRUS (such as Jisc), DOAB and OAPEN Library (OAPEN Foundation).
- the European Open Science Cloud (EOSC) and its network of services maintained in conjunction with OpenAIRE. Over 200 repositories contribute usage information to this infrastructure, which leverages EOSC to exchange interoperable data. This usage data can be accessed by data scientists and policy makers, as well as researchers, for an aggregate of usage of their outputs across participating institutions (Manghi, 2023).

Notably, the OA Book Usage Data Trust effort is uniting OPERAS and OpenAIRE in partnership with representatives from the University of North Texas to begin adapting the data space model for scholarly impact metrics exchange with the Mellon Foundation supporting IDS governance building block development through 2025.

Given the variety of global stakeholders, NSF workshop experts encourage documenting and sharing the values and principles guiding work in this shared space.

4. Research Issues of Usage Data ‘Ownership’ and Authority over Use

Research is needed to understand which authoritative stakeholder can authorize downstream usage data sharing and use for given scholarship outputs, disciplines and political geographies. Culturally or geographically dependent policies and norms affect organization policies and workflows and there is no one-size fits all approach (Drummond, 2023a). When contractual relationships related to the scholarly publishing lifecycle span authors and universities, editors, publishers, aggregators, discovery services, digital libraries, and the many data processors and controllers behind these organizations, it’s unclear whether or how “data ownership” and “data rights” follow the data or who is positioned to authorize the application of usage data for different purposes (Drummond, 2023a, p.6). For all of these reasons, “open as possible and as controlled as necessary” (Drummond, 2023a, p.8) may be needed in the face of artificial intelligence (AI) and big data computation of public domain or harvestable data. In short, ethical, legal, and contractual research questions abound (Rabar, 2023).

Workshop participants concluded that the complexity of this issue necessitates research into applicable laws and norms and further recommended development of shared model license language.

Recommendations and Next Steps

The overarching, unanimous recommendation from workshop experts is to consider existing resources and networks first and foremost; explore, leverage and invest in existing models, platforms and providers (Drummond, 2023a). This would include library consortia as well as the use and development of standards and engaging with PID infrastructures. It could also be valuable to explore models such as the European Research Infrastructure Consortium (ERIC), which could benefit OPERAS with annual financial commitments from member states, governance support and advocacy within and beyond the EU (Stern, 2023).

The development of a data intermediary or data cooperatives to formalize infrastructure partnerships could facilitate the use of data in a responsible, trusted way (Verhulst, 2023). Leveraging the European community-governed IDS model could provide structure to address privacy and ethical concerns of downstream use of detailed usage data, for example in unbounded AI training.

Near term priority projects were identified at the end of the NSF workshop. Through Mellon Foundation support, a team is researching how existing open scholarly infrastructures could enable an IDS for scholarly communications. This could inform the piloting of an MVP data space for controlled public/private scholarship usage data exchange. Through NSF support awarded since the workshop, teams are researching impact-related vocabulary definitions and crosswalks while documenting usage data supply chains for publicly accessible journals and open data, to complement such information for books (Clarke & Ricci, 2021).

Additional identified research opportunities that are not yet underway to the authors' knowledge include: legal research on usage data "ownership," (Drummond, 2023a, p.6) intellectual property and privacy dimensions of usage and impact data, further use case and values and principles explorations, and stakeholder education on the issues involved and potential solutions. Funding coordination for both existing infrastructure efforts and emerging models was highlighted as critical to near term success and international alignment (Drummond, 2023a).

Conclusion

Momentum is growing to share and evaluate usage and impact data to support evidence-based decision making. The scholarly community shares a desire to illuminate the impact of the breadth of funded research outputs and has collaboratively defined a path forward to further explore whether shared infrastructure will benefit the varied stakeholders involved in publicly accessible scholarship publication and discovery.

Acknowledgements

The authors are grateful , to Dr. Katherine Skinner for designing and facilitating the National Science Foundation (NSF) workshop on which this paper is based and to the Coalition for Networked Information (CNI) for hosting.

Funding statement

This work was supported by the National Science Foundation (NSF) under research grant Award 2315721. The funder had no role in the workshop design, the decision to publish, or in preparation of the manuscript.

Competing interests

Jennifer Kemp is a volunteer member of the Open Access eBook Usage Data Trust (OAeBUDT) Board of Trustees and was compensated for work on this paper from the grant that supported the NSF workshop. Christina Drummond serves as the Executive Director for the Open Access eBook Usage Data Trust (OAeBUDT) through her host institution, the University of North Texas and co-organized the workshop discussed in this paper, along with Charles Watkinson, co-PI of the NSF grant and volunteer member of the Open Access eBook Usage Data Trust (OAeBUDT) Board of Trustees.

References

Basson, I., Simard, M.-A., Ouangré, Z. A., Sugimoto, C. R., & Larivière, V. (2022). The effect of data sources on the measurement of open access: A comparison of Dimensions and the Web of Science. PLOS ONE, 17(3), e0265545. <https://doi.org/10.1371/journal.pone.0265545>

Bryant, R., Watkinson, C., & Welzenbach, R. (2021, December 6). Guest Post: Scholarly book publishing workflows and implications for RIM systems. *The Scholarly Kitchen*.
<https://scholarlykitchen.sspnet.org/2021/12/06/guest-post-scholarly-book-publishing-workflows-and-implications-for-rim-systems/>

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

Clarke, M., & Ricci, L. (2021). *OA Books Supply Chain Mapping Report*. Zenodo.
<https://doi.org/10.5281/ZENODO.4681725>

Conrad, L. Y., & Urberg, M. (2023). *With or Without: Measuring Impacts of Books Metadata*. Zenodo. <https://doi.org/10.5281/ZENODO.8145260>

Conrad, L., & Urberg, M. (2021, September 30). *The Experience of Good Metadata: Linking Metadata to Research Impacts*. The Scholarly Kitchen.
<https://scholarlykitchen.sspnet.org/2021/09/30/the-experience-of-good-metadata-linking-metadata-to-research-impacts/>

Drummond, C. (2023a). *Proceedings of the Workshop Exploring National Infrastructure for Public Access Usage and Impact Reporting*. <https://doi.org/10.5281/ZENODO.8335916>

Drummond, C. (2023b, April 2). What of the OA Book Usage Data Trust approach could be extensible to other types of usage statistics? In Drummond, C. & Watkinson, C. (Chairs), *Exploring National Infrastructure for Public Access Usage and Impact Reporting*. Denver, CO, United States. https://www.mivideo.it.umich.edu/media/t/1_f8z6s9rt/296014552

Drummond, C., & Hawkins, K. (2022). *OA eBook Usage Data Analytics and Reporting Use-cases by Stakeholder*. Zenodo. <https://doi.org/10.5281/ZENODO.8325578>

Holmes, K. (2023, April 2). Make Data Count. Zenodo. <https://doi.org/10.5281/zenodo.7792180>

Kemp, J., & Taylor, M. (2020, April 29). *Crossing the Rubicon - The case for making chapters visible*. Crossref.
<https://www.crossref.org/blog/crossing-the-rubicon-the-case-for-making-chapters-visible/>

Lair, S. (2023, April 2). How do US consortia strengthen the American scholarly communication ecosystem? In Drummond, C & Watkinson, C (Chairs), *Exploring National Infrastructure for Public Access Usage and Impact Reporting*. Denver, CO, United States.
https://www.mivideo.it.umich.edu/media/t/1_mpij8vi1/296014552

Lambert, J. (2023, April 2). What of the JISC/IRUS approach could be extensible to other sources of usage statistics? In Drummond, C & Watkinson, C (Chairs), *Exploring National Infrastructure for Public Access Usage and Impact Reporting*. Denver, CO, United States. https://www.mivideo.it.umich.edu/media/t/1_hvo62zmw/296014552

Lin, J. (2016, August 25). *The article nexus: linking publications to associated research outputs*. Crossref. <https://www.crossref.org/blog/the-article-nexus-linking-publications-to-associated-research-outputs/>

Lowenberg, D., Lammey, R., Jones, M.B., Chodacki, J., & Fenner, M. (2021, April 19). *Data Citation: Let's Choose Adoption Over Perfection*. Zenodo. <https://doi.org/10.5281/zenodo.4701079>

Lowenberg, D., Chodacki, J., Fenner, M., Kemp, J., & Jones, M.B. (2019, November 1). *Open Data Metrics: Lighting the Fire* (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.3525349>

Madray, H. (2023, April 2). How could the NSDS inform multi-platform public-private usage and analytics exchange? In Drummond, C & Watkinson, C (Chairs), *Exploring National Infrastructure for Public Access Usage and Impact Reporting*. Denver, CO, United States. https://www.mivideo.it.umich.edu/media/t/1_3fpkrijvq/296014552

Manghi, P. (2023, April 2). How could the EOSC/Core Interoperability Framework approach inform multi-platform public-private usage and impact analytics exchange in the US? In Drummond, C & Watkinson, C (Chairs), *Exploring National Infrastructure for Public Access Usage and Impact Reporting*. Denver, CO, United States. https://www.mivideo.it.umich.edu/media/t/1_k8jb4tj4/296014552

Mellins-Cohen T. (2023, April 2). What, if anything, is needed beyond COUNTER to improve usage and impact data interoperability? In Drummond, C & Watkinson, C (Chairs), *Exploring National Infrastructure for Public Access Usage and Impact Reporting*. Denver, CO, United States. https://www.mivideo.it.umich.edu/media/t/1_mnb609au/296014552

Nagel, L., & Lycklama, D. (2021). Design Principles for Data Spaces - Position Paper (1.0). Zenodo. <https://doi.org/10.5281/zenodo.5105744>

Nelson, A. (2022, August 25). *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research*. <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>

OECD calculations based on Scopus Custom Data, Elsevier, Version 1.2023, March 2023. <https://www.oecd.org/sti/inno/scientometrics.htm>

Rabar, U. (2023). OAEBUDT Community Consultation Results January 2023 - July 2023. Zenodo. <https://doi.org/10.5281/zenodo.8325572>

Ricci, L. (2023, April 2). How similar are usage and impact data pipelines for OA books to pipelines for other publicly accessible scholarship? In Drummond, C & Watkinson, C (Chairs), *Exploring National Infrastructure for Public Access Usage and Impact Reporting*. Denver, CO, United States. https://www.mivideo.it.umich.edu/media/t/1_v6nxhuxp

Ruediger, D., & MacDougall, R. (2023, March 6). *Are the Humanities Ready for Data Sharing?* Ithaka S+R. <https://doi.org/10.18665/sr.318526>

Sever, R. (2030). We need a plan D. *Nature Methods* 20, 473–474. <https://doi.org/10.1038/s41592-023-01817-y>

Stern, N. (2023, April 4). Personal reflections. In Drummond, C, Stern, N, Watkinson, C (Chairs), *Workshop Report Out: National Infrastructure for Public Access Usage and Impact Reporting* [Panel presentation] CNI Spring 2023 Project Briefings. Denver, CO, United States. <https://www.cni.org/topics/ci/workshop-report-out-national-infrastructure-for-public-access-usage-and-impact-reporting>

Tananbaum, G. (2023, April 2). How are open research funders shaping the application and demand for usage and impact analytics? In Drummond, C & Watkinson, C (Chairs), *Exploring National Infrastructure for Public Access Usage and Impact Reporting*. Denver, CO, United States. https://www.mivideo.it.umich.edu/media/t/1_u5zfk5j/296014552

Tkaczyk, D. (2023, February 22). *The more the merrier, or how more registered grants means more relationships with outputs*. Crossref. <https://www.crossref.org/blog/the-more-the-merrier-or-how-more-registered-grants-means-more-relationships-with-outputs/>

Verhulst, S. (2023, April 2). Might a “data collaborative” be needed to facilitate public-private usage and impact data brokerage at scale? In Drummond, C & Watkinson, C (Chairs), *Exploring National Infrastructure for Public Access Usage and Impact Reporting*. Denver, CO, United States. https://www.mivideo.it.umich.edu/media/t/1_7hia0733

Verhulst, S.G., Young A., Winowatan M., & Zahuranec, A.J. (2019, October). *Leveraging Private Data for Public Good: A Descriptive Analysis and Typology of Existing Practices*. GovLab. <https://raw.githubusercontent.com/GovLab/data-collaboratives/master/source/static/files/existing-practices-report.pdf>

Watkinson, C. (2023, April 4). Objectives and Perspectives. In Drummond, C, Stern, N, Watkinson, C (Chairs), *Workshop Report Out: National Infrastructure for Public Access Usage and Impact Reporting* [Panel presentation] CNI Spring 2023 Project Briefings. Denver, CO, United States.

<https://www.cni.org/topics/ci/workshop-report-out-national-infrastructure-for-public-access-usage-and-impact-reporting>

White, K. (2019, December 17). *Publications Output: U.S. Trends and International Comparisons*. Science and Engineering Indicators. <https://nces.nsf.gov/pubs/nsb20206/>

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1). <https://doi.org/10.1038/sdata.2016.18>