OA usage reporting: Understanding stakeholder needs and advancing trust through shared infrastructure

Tim Lloyd[a,*], Tricia Miller[b], Christina Drummond[c] and Jennifer Kemp[d]

[a]*CEO, LibLynx, Alexandria, United States of America*
ORCID: https://orcid.org/0000-0003-0495-5691
[b]*Marketing Manager for Sales, Partnerships, & Initiatives, Annual Reviews, Sun Prarie, United States of America*
[c]*Executive Director for OA Book Usage Data Trust, University of North Texas Libraries, Denton, TX, Columbus, United States of America*
[d]*Strategies for Open Science (Stratos), New York, United States of America*
ORCID: https://orcid.org/0000-0003-4086-3196

Abstract. The complexity of usage reporting for Open Access (OA) content continues to grow, particularly with content syndication to organizations such as ScienceDirect and ResearchGate, which deliver content across multiple platforms at an unprecedented scale. What kind of usage data do diverse stakeholders (including libraries, publishers, authors, and editors) need? Can the work done to support OA book usage data analytics use cases inform OA article and data use cases? What standards and policies are required to ensure the usage data is accurate and meaningful? What infrastructure is needed to collect and disseminate this data effectively and efficiently?

This paper brings together four different perspectives to consider these questions: an OA publisher, a research infrastructure, an emerging usage data trust, and a usage analytics service provider. The authors walk through what is known, and then start to unpack the questions for which they do not yet have answers. Their goal is to inform the community's understanding of the challenges ahead and, hopefully, start to lay the groundwork for constructive policies and shared solutions.

Keywords: Open access usage reporting, subscribe-to-open, usage data trust, open access books, open infrastructure, Crossref, Annual Reviews, Liblynx, Open Science, Stratos

Section I: The increasingly complex nature of OA usage reporting

By Tim Lloyd

1. Introduction

To kick off this four-part paper, I will explore how the changing environment for scholarly publishing is driving greater complexity in open access usage reporting. A helpful model for thinking about usage reporting is to break the process down into three components (Fig. 1):

---

*Corresponding author: Tim Lloyd. E-mail: tim@liblynx.com.

Fig. 1. The components of usage reporting.

- Data capture, where we capture raw event data from the publishing and distribution platforms that are delivering OA content to users
- Processing of that raw event data, which can include custom logic for enriching the data
- Delivery of the processed output in various formats, such as aggregate reports or processed event data.

First, I am going to describe four high level drivers that impact the complexity of usage reporting across the board: scale, granularity, new use cases, and data privacy.

1.1. Scale

Usage of publicly-available content is at least an order of magnitude greater than paywalled content. Most industry applications for usage reporting were developed for paywalled content around a traditional model of month end batch processing. These reporting architectures bake in the assumption that processing can wait until all the relevant data has been assembled, and that reprocessing is a rare occurrence. In my experience, these systems struggle to transition to an environment where reporting requirements are more frequent (on-demand?) and reprocessing is more common because data isn't perfect, and some data may only become available at a later date. For example, if future re-processing is a consideration then you also need systems to make the significant volume of raw event data available, at least for a period.

1.2. Granularity

Stakeholders are increasingly interested in understanding usage at more granular levels - the chapter of a book, an article within a journal, an audio or video segment, etc. This correlates to the Item in COUNTER reporting, and it is no coincidence that COUNTER's proposed 5.1 update to their Code of Practice makes the item the default level of reporting (vs the title). Item level reporting significantly increases the volume and detail of data flowing through the system.

1.3. New use cases

The usage of open access content is also attracting interest from a broader set of stakeholders who are outside the traditional library audience for COUNTER reports. Those managing institutional research funds want to understand the impact of their funding. Those managing and negotiating publisher relationships want to understand how their organization is generating new OA content, as well as consuming it. Authors have increasing choice over where to publish, and usage reporting informs their understanding of these choices. All this is in addition to editorial staff wanting to understand usage trends as input into their editorial strategies. These emerging stakeholders drive new use cases that add additional complexity to the process.

1.4. Data privacy

COUNTER's Code of Practice includes a statement on data confidentiality [1] that is based on current ICOLC guidelines [2] (the Int'l Coalition of Library Consortia) - this statement prohibits the release of any information about identifiable users, institutions, or consortia without their permission. As usage reporting grows in scale and granularity, and data is made available to a wider range of stakeholders, we need to be thoughtful about ensuring that privacy is maintained.

2. Data capture

One major change we'll see over the next few years is an increasing number of publishers syndicating their content so that usage occurs on multiple platforms, rather than just the content owner's platform. While this model is already familiar to those in book distribution, journal publishers are experimenting with distributing content via platforms such as ResearchGate. This requires the usage from these third-party platforms to be integrated into a publisher's own usage reports to provide a comprehensive view of usage - and adds more layers of complexity.

A great example of this is my next point - diverse formats and metadata. As raw usage data is sourced from a wider range of platforms, a more diverse range of inputs is to be expected. At the format level, files can be tabular (e.g., comma delimited csv) or structured (JSON). In some cases, usage data for a single platform can be split across multiple files due to the peculiarities of the database exporting the raw events. At the metadata level, examples include the use of free-text fields vs standard identifiers, and varied conventions for time-stamping.

As a community, we have yet to coalesce around a fixed set of standards for usage data. Which is to say that plenty of standards exist, but we are not consistently selecting and using them. As a result, we are trying to build aggregated reports from diverse datasets. As is often the case, this ends up forcing us to the lowest common denominator, rather than being treated as an opportunity for best practice.

3. Data processing

The new use cases I referred to earlier will likely necessitate additional metadata to drive new reporting. For example, content topics to enable analysis against research priorities, or identifiers to enable reporting to be filtered for particular funders or authors. Anyone who has tried to disambiguate author names across multiple systems will be familiar with the huge challenges that lie there.

Similarly, we're seeing a need for new processing logic. In a paywalled world, platforms know the identity of the organization accessing content, which is how we are able to generate COUNTER reports for librarians. In an open access world, we typically do not know anything about the user, and so new processing logic is needed to affiliate usage with an organization, such as matching IP addresses to registered organizational ranges. Another example would be using third party databases to look up the funder IDs for a journal article.

4. Data delivery

These new use cases also spawn a need for new reporting formats. We need to support both machine-driven reporting needs (bulk exports; APIs) as well as formats designed for human consumption (web-pages, spreadsheets, pdfs). We will see more reporting using graphical formats that make it easy to consume at a glance, as well as the traditional tabular reporting.

And we are seeing a demand for a great frequency of reporting. COUNTER reports capture a month of usage, but increasingly usage data is flowing in real-time and enabling on-demand reporting that can cover custom date ranges, and be used to power reporting applications that are also working in real-time.

Future needs

This future has important implications for the systems and workflows needed to support open access usage reporting:

- They need to be exponentially scalable and support granular, real-time (or near real-time) reporting
- They need to be able to flexibly cope with a variety of input and output formats, and swap in custom processing logic for different use cases.
- They obviously need to be standards-based, which should help reduce the variety we have to cope with. But standards are a journey, not a destination, so we need to be practical about supporting what's possible now.
- And they need to be reliable, and auditable, so our community can understand how they are created and rely on them for decision making.

In short, this is a big change. It will not happen overnight, and it doesn't need to. But this is the future we need to prepare for.

Section II: Open access usage and the publisher perspective
By Tricia Miller

With the increase of open access usage, the data describing who and where scholarly resources are being used has changed. These changes, however, may not align with our original standards describing the value and use by institutions. The audiences, contexts, and purpose for use of scientific literature is evolving and growing as more open access content is published and therefore, the measurement and significance of open access usage data needs to be reexamined.

This puts publishers in an ambiguous position to balance paywalled and open access usage data to meet the needs of our users, ourselves, and other involved stakeholders. To start to respond to the changes in usage reporting, publishers must have access to and report upon how open access impacts our entire scholarly communications community.

For *Annual Reviews*, the challenges are amplified by our publishing model, Subscribe to Open [3], a non-APC open access business model that necessitates institutional, government, and corporate subscriptions to fund open access publishing. This means that our subscribers are acting as funders of open access and so, we must go beyond traditional usage metric reporting to describe the impact of supporting open access publishing. Not only does open access usage impact more audiences but must also be trusted to correspond with traditional usage metrics in order for subscribers to rationalize support.

How do we create a framework that correlates both usage types, can satisfy successful open access publishing, and ensure the continued financial support that is needed for any open access publishing model whether it be for access to articles, supporting author manuscripts, or fulfilling funder and government mandates?

(1) We start with the reasoning for open access usage, the access and impact to a global audience.
(2) Collaborating to create a framework for shared understanding and standards.
(3) And finally, the necessity to trust the data reported to us.

At the heart of open access is the impact and access to a global audience. Beyond the traditional audience, those accessing articles within institutional confines, open access publishing creates a complex network of users. In 2017, *Annual Reviews*, opened the *Annual Review of Public Health* [4] to help understand how open access affected usage and impact of scholarly review articles. Using the LibLynx Open Analytics Platform [5] in conjunction with COUNTER [6] compliant usage data, what we were able to find was:

- Who is using our content
- In what context
- And for what purpose

Use case

Using data from the *Annual Review of Public Health* as a use case, we could visualize the potential and impact of open access over time.

After the first year of opening access to the journal, usage increased by just over forty percent and in 2022, the usage increase was one hundred and thirty percent higher than it was when content was behind a paywall (Fig. 2).

It should be no surprise that ninety percent of article usage comes from academic institutions. What is noteworthy is the variety of usage beyond academic institutions that continues to grow. Our data shows ninety-four different types of institutions downloading full text HTML and PDF articles. The variety of institution types within academic, government, and corporate usage proves that there is a need and value of scholarly literature that non-open access publishing models leave out. This granularity of data is important to evaluating and supporting the needs of all users. We have found usage at places such as construction companies, banks, food producers, and even prisons.

Another example of the granularity of data is the range of areas of interests for users that reflects on the purpose for access and impact of articles that open access publishing can support.

For *Annual Reviews*, our data indicates three hundred and twenty-six different areas of interest by users (Fig. 3).

Finally, the granularity of open access data showed the impact of open access on global usage. Usage of Annual Reviews open access articles jumped from usage in fifty-five countries in 2016 to one hundred and eighty-seven countries in 2022 (Fig. 4).
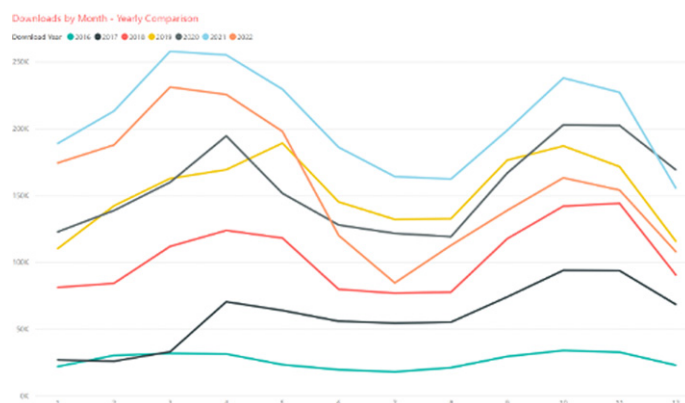
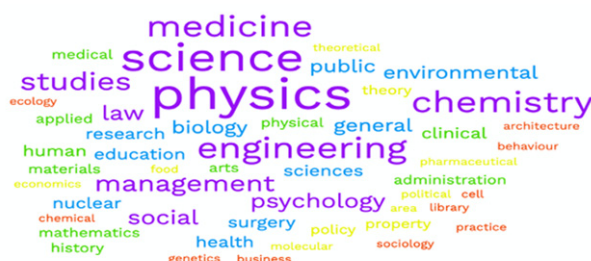Fig. 2. *Annual Review of Public Health* open access usage 2016–2022.

Fig. 3. *Annual Reviews* open access usage by user's area of interest.

While these examples are certainly not exhaustive of all the ways in which open access has an impact, from this use case, we can clearly see that open access has a significant impact on our usage. Open access data provides publishers with an opportunity to consider and create new products, services, and business models that go beyond supporting only those who can pay to participate.

Now, what we need to understand are the needs of a truly global diverse audience? How are all of our stakeholders impacted when our audience and their needs change?

Creating a collaborative framework

It leads to our need to develop a collaborative framework based on the integrity in our data, availability of data to all stakeholders, and the reproducibility and consistency of the data, all of which require transparency, but that also is subjective in interpretation without standards to accomplish both individual and collective goals.

The traditional usage framework, based around cost per download and institutional attribution is now just part of the usage interpretation that open access reporting can provide. Many stakeholders are now also aware and concerned with a mission-driven framework where the benefit to communities, society, and global knowledge sharing exists alongside institutional benefits.

The equity in access to the global audience is one significant reason why open access publishing is accelerating so quickly. What our industry needs to undertake next is how these needs can co-exist and be communicated effectively.
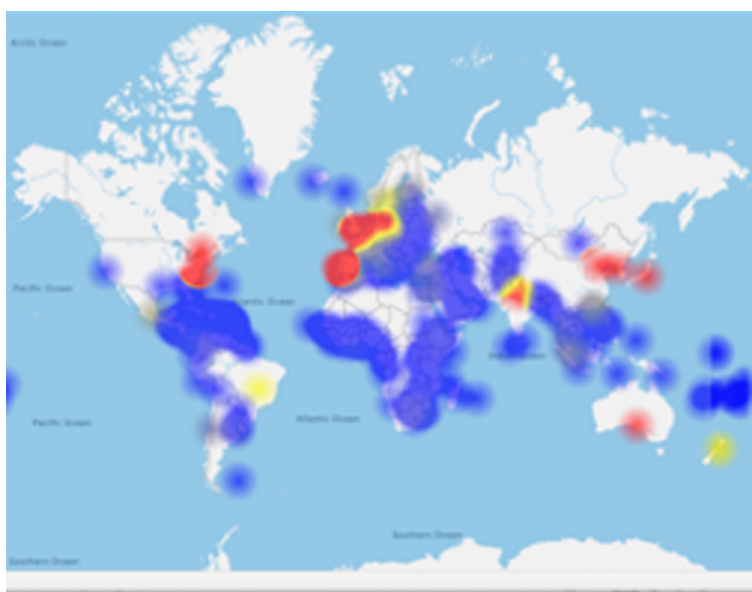
Fig. 4. 2022 Annual Reviews open access usage by country.

Building trust

The final point from the publisher perspective that matters in open access usage reporting *is trust*. Having trust that our open access data is accurate builds trust generally for a publisher and in their approach to open access publishing. We know that there are many pathways to achieving open access, but we also know that sustainability, transparency, and trust should accompany whichever model is used.

As a community, working together to achieve our collective goals, our open access data, sharing results, and interpretations helps build trust in our relationships through these collaborations to create standards and framework.

Since open access data is much more granular, we need to balance the transparency of usage data with the privacy for our users and the institutions. The Open access usage data that is available is both attributed, based on known IPs, and unattributed, meaning not associated with a particular institution, but that does not mean that information about unattributed users is unknown.

To generalize, this is different from traditional paywalled usage reporting or COUNTER reporting where the bulk of usage reports are pulled based on institutional subscriber details. Of course, there is a lot more to it, what is important is that the granularity of open access usage data, and the lack of standards mean that there may not be consistency in how data is gathered, presented, or interpreted, that can cause a lack of trust, among other issues.

So, to summarize, from the perspective of a publisher, open access data can offer evidence as to who our real audiences are, thereby allowing us to ask and understand their needs. We must also listen to institutions, libraries, funders, authors, and other publishers to help us reimagine a framework that meets the needs of all of us and is sustainable, equitable, trusted, and collaborative.

**Section III: The OA book usage data trust effort**
By Christina Drummond

Over the past few years, the global Open Access (OA) book stakeholder community has investigated how to better foster the community governed exchange of reliable, granular, on demand usage data in a way that is trusted and equitable. In my NISO+ 2023 talk, I provided project background and highlighted how we are advancing the development of organizational, policy, and technical governance mechanisms to improve interoperability and systems integrations across OA book stakeholders.

Project origins at the 2015 Scholarly Communications Institute were noted (Fig. 5). Our first sets of principal investigators were recognized as thought leaders in our field, including Charles Watkinson, Becky Welzenbach, Brian O'Leary, Cameron Neylon, Lucy Montgomery, Kevin Hawkins, and Katherine Skinner. Supported by the Mellon Foundation, these individuals led two project phases where community stakeholders came together to document shared opportunities, challenges, and use cases related to OA book usage data reporting and analytics. This foundational research prepared us to develop governance building blocks for an OA book usage data focused on data space, in line with emerging European frameworks for trusted data intermediation and exchange across both public and private organizations.

Early on, our investigators explored how data collaboratives and data trusts could support a global data exchange by providing economies of scale when it comes to usage data aggregation and benchmarking. The Data Trust model surfaced as a way to foster global cooperation among competitors through trust and reciprocity. To quote Cameron Neylon et al., the data trust model has the "potential for achieving economies of scale and for taking advantage of network effects, to address resource challenges facing individual stakeholders and allow comparison and benchmarking to benefit all stakeholders" [7]. With this value in mind, the team shifted to consider what governance and technological infrastructure could sustainably integrate standards and Permanent Identifiers (PIDs) related to OA books and usage given forthcoming EU data regulations.

In July of 2022, the Mellon Foundation supported us to strengthen the OA Book Usage Data Trust [8] community governance by engaging community stakeholders to develop ethical participation guidelines and understand the data trust ROI for participation. A diverse project advisory board [9] and Board of Trustees [10] now guide this work led by myself and my Co-PIs, Yannick Legré the Secretary General for OPERAS and Prodromos Tsiavos, Chief Legal Advisory for OpenAIRE. In our current project phase, we are creating community governing boards and policies while defining participation rules for the data trust community. This work prepares us for a technical infrastructure pilot and study of the return of investment for participation in a data trust modeled after the European Industry Data Spaces (IDS) framework.

Our current stage of work would not have been possible without the foundational research previously supported by Mellon. Focus groups resulted in *OA eBook Usage Data Analytics and Reporting Use-Cases by Stakeholder* [11], illuminating the specific data analytic and reporting queries sought by different book publishing and discovery stakeholders and the challenges faced when trying to produce them. We found that different book stakeholders wanted to use OA usage data – not just for reporting but for internal operations and strategic, data-driven decision-making. Applications of OA usage data are wide-ranging, from informing collection development or editorial strategy, to evaluating promotional campaigns and informing print editions. Publishers and libraries sought to leverage OA book usage analytics for OA program operations. We also found that organizations shared the challenge of allocating time and expertise to combining and managing COUNTER-compliant reports alongside related data and non-COUNTER compliant web-analytics provided from various providers.
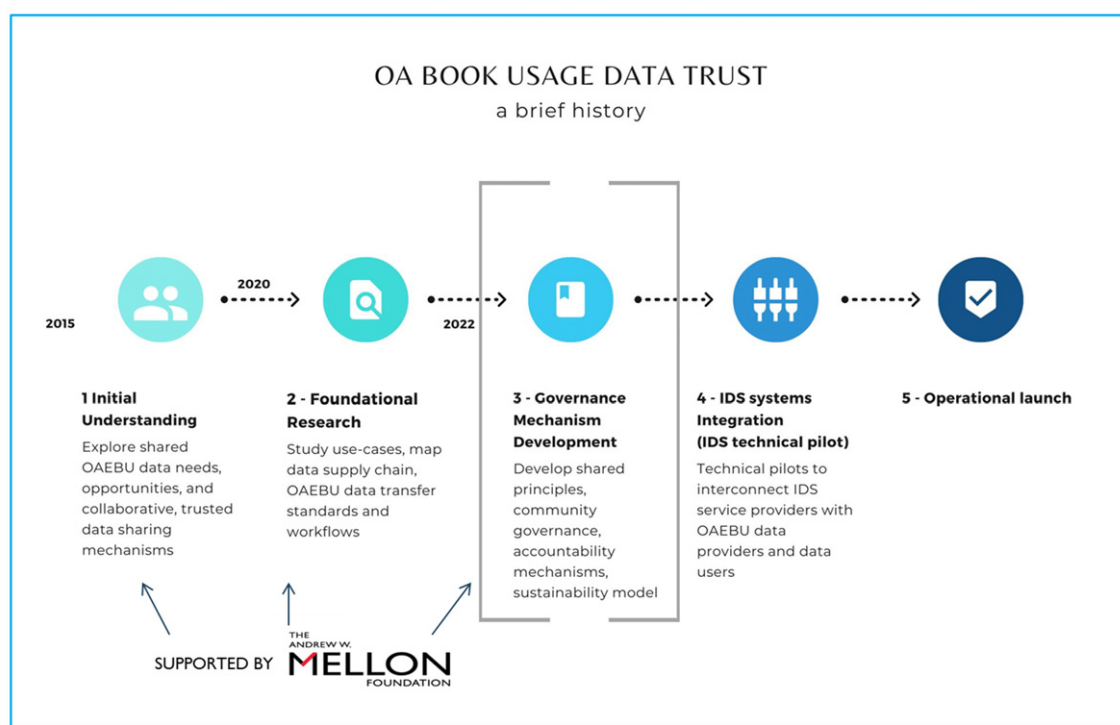
Fig. 5. OA book usage (OAEBU) data trust project timeline.

Knowing that data collaboratives and trusts could facilitate public/private data sharing, we set out to identify a technical and policy framework. In 2020, we learned both of the European efforts to regulate data intermediation and data marketplaces regulations, and of the research and development around new standards to support the new regulatory frame. Specifically, the International Data Spaces [12] or Industrial Data Spaces, (IDS for short) being developed through GAIA-X [13] and Horizon Europe [14], accessed September 24, 2023. funding surfaced as a way to facilitate the exchange and processing of data through a data governance middle-layer in cyberinfrastructure networks. At its core, the IDS model differed from data harvesting, data repositories, and data lakes in two ways: 1) IDS is oriented around maintaining data sovereignty and multi-stakeholder trust in the processes, systems, and data quality, and 2) the IDS model aimed to account for data provenance, transformation, and access logging at scale, while making it possible for data providers to join and/or leave the ecosystem with their data as they so choose [15,16]. As we aim to facilitate the exchange and aggregation of usage data across public, nonprofit, and commercial organizations, we recognized such concepts as core to our objectives.

Our effort is now focused on fostering community development and consultations on the principles and requirements for data trust participation. Issues to be addressed include community governance, technical and security requirements, data processing and usage principles, and compliance mechanisms. In April of 2023, an in-person workshop will launch the drafting of principles for broad community consultation and review. We aim to determine how to share data across the data trust community in a way that is as open and transparent as possible while as controlled as necessary to ensure granularity and quality and avoid potential harms. This means as a community we will be considering data transfer requirements, processing

Records by content type

- ● Journals
- ● Books (incl. chapters)
- ● Conferences
- ● Datasets
- ● Standards
- ● Dissertations
- ● Peer review reports =.2%
- ● Reports
- ● Posted content (includes preprints) =.8%
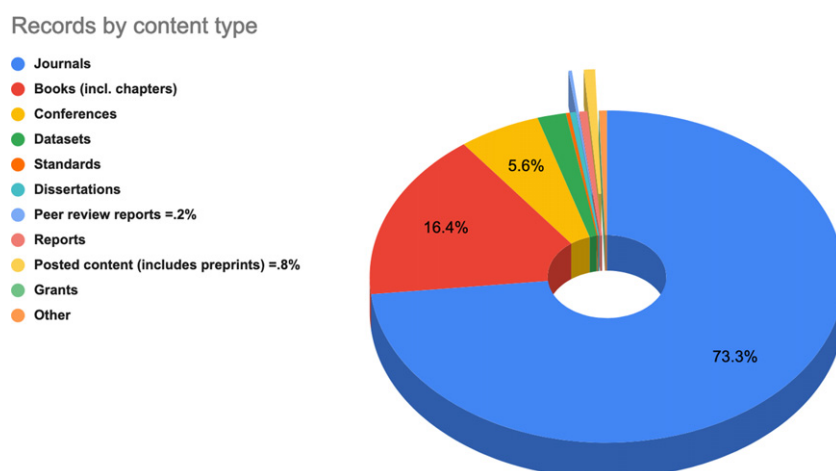- ● Grants
- ● Other

73.3%

16.4%

5.6%

Fig. 6. Percent of records registered with Crossref by content type, May 2023.

principles, and also usage policies to clarify appropriate and inappropriate uses of data accessed through the data trust network. Such principles will underlie technical requirements for participants, data policies related to ethical data use and privacy, as well as the terms for participation in the data trust. Our aim is to meet certification requirements as an IDS so that we can support our EU-based OA book stakeholders while being well positioned to meet other emergent national data brokerage regulations.

As Michael Clarke and Laura Ricci documented, usage-related data transfer metadata standards and usage data workflows travel through library management systems, book aggregators, publishing platforms, and repositories before being compiled into structured usage data reports for publishers, libraries, scholars, and funders [17]. Our OA Book Usage Data Trust is attempting to simplify today's current manual workload when multiple flows of OA usage come together.

If data processing algorithms and approaches vary from organization to organization, interoperability, reliability, and data comparability can be compromised – making aggregation and benchmarking at the system level difficult if not impossible. With this in mind, we hope that our work with OA Book Usage Data Trust can improve the FAIRness (**F**indable, **A**ccessible, **I**nteroperable, and **R**eusable) of usage data at scale.

Section IV: The role of open infrastructure
By Jennifer Kemp

Crossref doesn't generate or distribute content usage but, as open infrastructure, it provides a crucial link in a healthy usage supply chain. Like other open infrastructure providers, it facilitates a connected scholarly record and provides context for usage data.

At the time this session was presented, in February 2023, Crossref had over 142.5 million records. Just four months later, it has nearly one hundred and forty-six million records. Understanding the growth and pattern of what is registered (Fig. 6) and how this information is used outside of usage data helps explain its roles for different stakeholders in scholarly communications. For example, some of the newest content types, such as preprints and peer review reports (the slivers at the top of the chart) are among the fastest growing types of records. And even as new types emerge, the proportions of journals and books (including book chapters) have remained remarkably consistent.
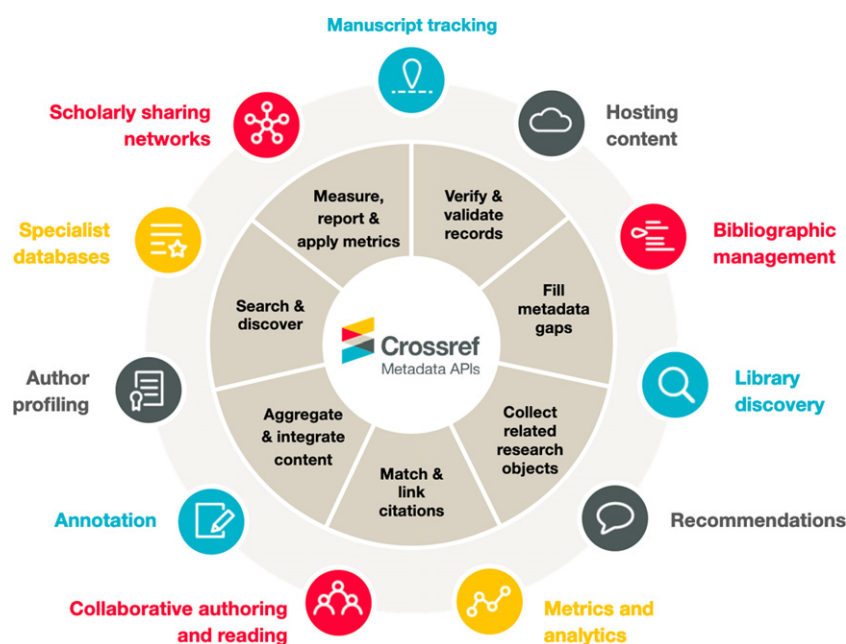
Fig. 7. *Crossref metadata use cases.*

It is also important to understand what information is and often is not in these records. What is required is generally a citation and a DOI, both useful for identifying resources to associate with usage (and other information). Members may also include a great deal of other detail with each record, which can be enormously helpful for all manner of use cases, discussed in the next section.

Books are a particular challenge to fully reflect in the scholarly record. Because they are so frequently hosted on multiple platforms, include a variety of formats such as monographs, contributed volumes and others, and have chapters that may or may not have their own records and DOIs, incomplete metadata can complicate the usage picture. Crossref has a longstanding Books Interest Group [18] that manages a best practices guide [19] for publishers that addresses issues such as versions.

Optional information is outlined for each content type [20]. In general, a few optional elements that are strongly encouraged for all record types are: abstracts, references, affiliation information (including RORs) of all contributors, funder and grant information, license information, relationships [21] and, for book publishers in particular, chapters.

Usage use cases for crossref information

Open infrastructure allows for all of this detail (and more [22]) to be collected and distributed by a neutral third party. This allows stakeholders throughout scholarly communications to connect metadata to usage. The varied use cases of Crossref records, which includes usage and related analytics (Fig. 7), highlights the importance of metadata in the scholarly communications landscape. This metadata is heavily used by machines and people, to point readers to relevant content [23]. The corpus of records is regularly used as a discovery database, which in turn can drive usage. Across all of the open APIs, Crossref gets over six-hundred million queries per month. The detail in these records, or the lack of it,
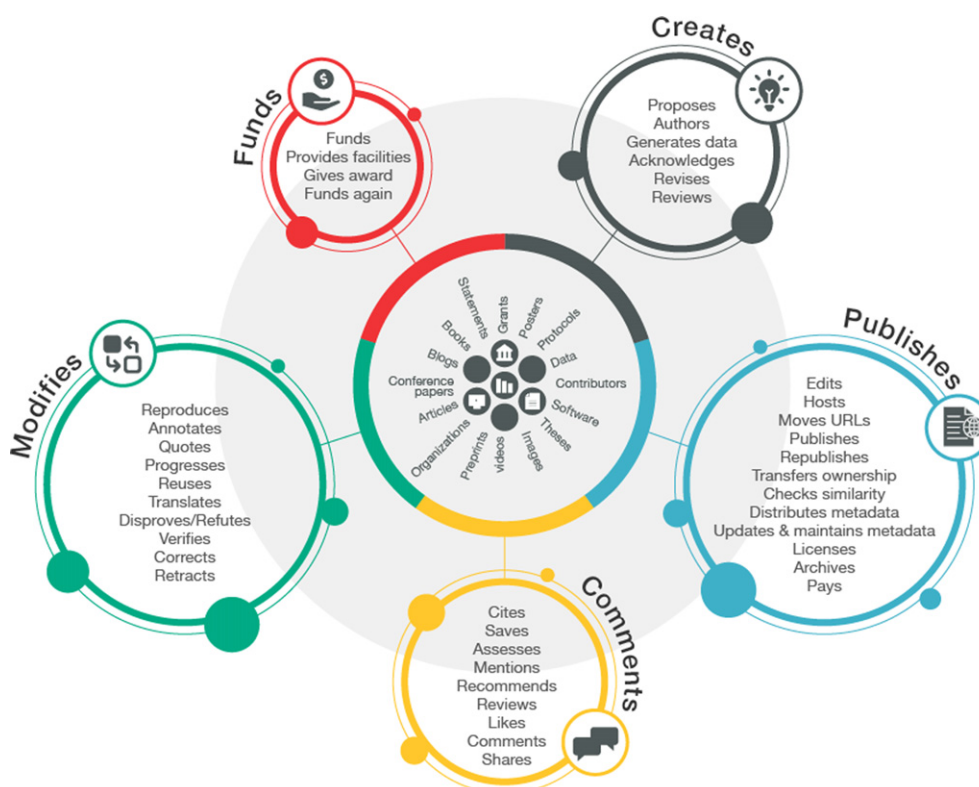
Fig. 8. Crossref research nexus vision.

is propagated through a lot of systems around the world. Verifying and enhancing existing records are other use cases that are often part of other use cases, such as recommendation services. Usage and related analytics are one use case of many. Publishers and authors may be the most directly affected by lack of complete, quality metadata.

Books are a useful example here because they are inherently so complex and because they are generally hosted on multiple platforms. As discussed earlier, the distributed landscape of scholarly content is a key challenge in gathering usage from different sources and associating it with the appropriate records. Books without chapter records, for example, may not fully reflect all contributors and their associated affiliations, making the task of funders and institutions looking to track related research outputs and their impact unnecessarily difficult.

The bigger picture

Crossref is one of many infrastructure providers committed to the Principles of Open Scholarly Infrastructure, known as POSI [24]. Transparency around infrastructure and its sustainability is key for a trusted, healthy research communications landscape. Though the focus here is on Crossref, POSI recognizes the networks, implicit and explicit, that support every stage of and every stakeholder in the research enterprise. Though POSI formalizes and documents these efforts to some degree, it is not a

necessary or isolated requirement; collaboration is. It is central to the cross-stakeholder efforts on which we're all dependent.

These efforts can culminate in the explicit, aspirational goal of the Research Nexus (Fig. 8). All research outputs should be registered and fully described in the metadata, including crucial relationship information, e.g. linking together translations, preprints to Versions of Record, post-publication commentary and so much more, reflecting the full range of touchpoints from funding to peer review to associated data, corrections and citations.

Overlaying usage information on all of these connections would be a powerful addition that benefits all the stakeholders involved.

About the authors

Tim Lloyd is the CEO of LibLynx, a company that provides cloud-based solutions to publishers, service providers and libraries to help them manage identity and access to online resources, and to better understand usage of those resources. He is also a member of the governance and outreach committees for the Coalition for Seamless Access and Project COUNTER's Open Access/Unpaywalled sub-group, and previously spent over a decade in a variety of product development and operational roles in publishing. E-mail: tim@liblynx.com; Phone: 202-888-3324 (LibLynx offices).

Tricia Miller, Marketing Manager - Sales, Partnerships, & Initiatives at nonprofit publisher *Annual Reviews* - has been working in the scholarly communications field for more than a decade to support marketing and sales efforts for society and nonprofit publishers. With a background in agricultural industry marketing, environmental horticulture, and technical and professional communication, her work emphasizes improving the user experience while supporting systems and best practices for accessing and understanding scholarly information. She is driven by supporting diversity and equity in academia and larger DEI opportunities within scholarly publishing. She recognizes that the land where she resides, in Sun Prairie, WI, sits on the unceded ancestral lands of the Ho-Chunk Nation. Tricia is a member of the SSP Continuing Education Committee, SSP Fellowship Sub-committee, C4DISC Communication and Outreach Committee, and a member of the Society for Technical Communication and Association of Teachers of Technical Writing among others supported by her organization. E-mail: tmiller@annualreviews.org.

Christina Drummond currently serves as the Data Trust Program Officer for the OA eBook Usage Data Trust project having previously served as Educopia's Director of Strategic Initiatives (2015–2017). Affiliated with Educopia since 2012, Christina has developed capacity-building resources for Educopia's affiliated communities while managing efforts relating to digital preservation, cultural heritage workforce and leadership development, and data analytics. She previously has held research faculty and administration positions at the University of North Texas and is a data policy expert, having founded the Technology and Civil Liberties Project at the ACLU of Washington and the Applied Data Ethics Lab at The Ohio State University.

She currently acts as a Co-Chair for the Professionalizing Data Stewardship Interest Group of the Research Data Alliance and serves on the Mid-Ohio Regional Planning Commission's Regional Data Advisory Committee. E-mail: christina.drummond@unt.edu.

Jennifer Kemp joined Strategies for Open Science (Stratos) in 2023 as Director, Consulting Services. She was most recently Crossref's Head of Partnerships where she worked with funders, publishers, service providers and API users to improve community participation, metadata and discoverability. Prior to Crossref, she was Senior Manager of Policy and External Relations, North America for Springer Nature, after serving in library marketing and ebook product roles. Jennifer started her career as a librarian

before joining HighWire Press where she learned the idiosyncrasies of scholarly publishing across a wide variety of subject disciplines and content formats. She is active in the research support community and serves as Board Secretary for the OA Book Usage (OAeBU) Data Trust. E-mail: jennifer@strategieos.org; https://orcid.org/0000-0003-4086-3196.

References

[1] See: https://cop5.projectcounter.org/en/5.0.2/10-compliance/02-confidentiality-of-usage-data.html, accessed September 24, 2023.

[2] https://icolc.net/content/revised-guidelines-statistical-measures-usage-web-based-information-resources-october-4, accessed September 24, 2023.

[3] See: https://annualreviews.org/S2O, accessed September 24, 2023.

[4] See: https://www.annualreviews.org/journal/publhealth, accessed September 24, 2023.

[5] See: https://www.liblynx.com/publishers/open-access-analytics/, accessed September 24, 2023.

[6] See: https://projectcounter.org, accessed September 24, 2023.

[7] C. Neylon, L. Montgomery et al., Building a trusted framework for coordinating oa monograph usage data, November 2018, doi:10.17613/36hw-gs17, accessed September 24, 2023.

[8] See: https://oabookusage.org, accessed September 24, 2023.

[9] *Project Roadmap - 2022 Forward*. OA book usage data trust. https://www.oabookusage.org/2022forward, accessed September 24, 2023.

[10] *About - OA book usage data trust*. OA book usage data trust, https://www.oabookusage.org/governance, accessed September 24, 2023.

[11] C. Drummond and K. Hawkins, OA eBook usage data analytics and reporting use-cases by Stakeholder, January 2022 Zenodo. doi:10.5281/zenodo.5572840, accessed September 24, 2023.

[12] See: https://internationaldataspaces.org, accessed September 24, 2023.

[13] See: https://gaia-x.eu, accessed September 24, 2023.

[14] See: https://commission.europa.eu/funding-tenders/find-funding/eu-funding-programmes/horizon-europe_en.

[15] L. Nagel and D. Lycklama, Design principles for data spaces, April 2021, Zenodo. doi:10.5281/zenodo.5244997, accessed September 24, 2023.

[16] B. Otto, S. Steinbuss, A. Teuscher and S. Lohmann, IDS reference architecture model (Version 3.0), April 2019 Zenodo, doi:10.5281/zenodo.5105529, accessed September 24, 2023.

[17] M. Clarke and L. Ricci, Open access eBook supply chain maps for distribution and usage reporting, doi:10.5281/zenodo.4681870, accessed September 24, 2023.

[18] See: https://www.crossref.org/working-groups/books/, accessed September 24, 2023.

[19] See: https://www.crossref.org/documentation/principles-practices/books-and-chapters/, accessed September 24, 2023.

[20] See: https://www.crossref.org/documentation/schema-library/required-recommended-elements/, accessed September 24, 2023.

[21] J. Kemp and P. Feeney, The Research Nexus vision for a more connected scholarly community, *Information Services & Use* **42**(3–4) (2022), 417–421. doi:10.3233/isu-220170, accessed September 24, 2023.

[22] See: https://www.crossref.org/documentation/schema-library/, accessed September 24, 2023.

[23] J. Kemp, In the know on workflows: The metadata user working group, *Crossref* (February 28, 2023), https://www.crossref.org/blog/in-the-know-on-workflows-the-metadata-user-working-group/, accessed September 24, 2023.

[24] G. Bilder, J. Lin and C. Neylon, The principles of open scholarly infrastructure (2020), doi:10.24343/C34W2H, accessed September 24, 2023.