

Institutional repositories and the item and research data metrics landscape

The success of COUNTER in supporting adoption of a standard to measure e-resource usage over the past 15 years is apparent within the scholarly communications community. The prevalence of global OA policies and mandates, and the role of institutional repositories within this context, prompts demand for more granular metrics. It also raises the profile of data sharing of item-level usage and research data metrics. The need for reliable and authoritative measures is paramount. This burgeoning interest is complemented by a number of initiatives to explore the measurement and tracking of usage of a broad range of objects outside traditional publisher platforms. Drawing on examples such as OpenAIRE, IRUS-UK, Crossref's Distributed Usage Logging and Event Data service and COAR Next Generation Repositories, this article provides a brief introduction and overview of developments in this area.

This paper forms the basis of a session that was originally presented at the UKSG Annual Conference in Telford in April 2019.

Keywords

Usage; metrics; research data; open access; OA; COUNTER



PAUL NEEDHAM

Research and
Innovation Manager
Cranfield University



JO LAMBERT

Jisc Service Manager
Jisc

Context

Open access (OA) to scholarly publications promotes visibility and discoverability of research, with OA policies and mandates providing impetus to support a change in academic practice and encourage a transition to OA. With this comes a requirement for data and tools that facilitate openness and reuse. Institutional repositories (IR) are a critical part of the OA infrastructure, facilitating knowledge sharing and enabling academics to disseminate their research widely.

'measuring the reach of research is of fundamental importance'

Against this context, measuring the reach of research is of fundamental importance. Data and metrics can provide evidence that an institution is meeting funder requirements, it can be used to advocate for and promote use of the IR and it can help to monitor and plan for use of institutional infrastructure. For funding bodies, collecting information about the outcomes of the research they fund can inform research strategies and future funding decisions, as well as demonstrating return on investment (ROI). For researchers, data and

2 metrics are an important indicator of the impact and reach of their research. The increasing use of metrics to support decision-making and dissemination in these areas requires effective tools and appropriate data to calculate ROI and to demonstrate value and impact.

There is no single, perfect measure to assess value and impact; institutions may use a range of metrics including citations, page views and altmetrics. However, download statistics are one of several measures used to demonstrate value and are the focus of this article. Usage metrics are a key aspect in terms of understanding how publicly available research is being used. Tracking, monitoring and benchmarking usage of scholarly resources supports understanding of an institution's research, helps to identify emerging trends against a broader context and informs policy and process.

'There is no single, perfect measure to assess value and impact'

Over the past 15 years the COUNTER¹ standard has been integral to facilitating the recording and reporting of online usage statistics in a consistent, credible and comparable way. COUNTER statistics support librarians to compare vendor statistics, generally at the title or package level, to make better informed purchasing decisions. However, in an OA environment tracking use of research and the data underpinning that research prompts an increasing interest in more granular metrics, including item-level and research data metrics. OA content is likely to be available in multiple places; from the author accepted manuscript in an institutional repository, or the publisher version of record on a supplier platform, to harvested publications available through content aggregators such as CORE² (the world's largest collection of OA research papers) and scholarly networks. Facilitating access to research is the ultimate goal, but how do you track usage and monitor the success of your OA content or services when usage is occurring across multiple platforms? How do you gain a complete picture of usage when availability of content is fragmented?

'an increasing interest in more granular metrics, including item-level and research data metrics'

This article offers a brief overview of tools, initiatives, standards and protocols that seek to address these questions. Intended as an introduction to the topic, we outline recent work in this area. Although there is as yet no single perfect approach, a combination of effective partnerships and use of standards and technical developments can offer solutions towards developing a more coherent picture of usage wherever that usage occurs.

Standards and protocols

Common standards and protocols are integral to work in this space; they form the bedrock supporting standardized and transparent approaches to data exchange. The COUNTER standard and associated data transfer protocols, which facilitate automated machine-to-machine communications – SUSHI (Standardized Usage Harvesting Initiative),³ tracker⁴ and Distributed Usage Logging (DUL)⁵ – are central to supporting development of services, and these are described below.

Standards: COUNTER

COUNTER provides an infrastructure to support publishers, libraries and third parties who wish to create or access statistics or build services to support access to those statistics. Since the first Code of Practice (CoP) was published in 2003, COUNTER has been instrumental in bringing together publishers, vendors and librarians to develop and maintain a standard for counting usage of networked e-resources. Collaboration is key to the development and maintenance of an effective standard supporting consistent and comparable measurement and intended for global adoption and use.

The COUNTER standard, now in its fifth iteration, has evolved over time in response to a changing environment and evolving requirements. The COUNTER CoP release 5 (R5) standardizes usage metrics for e-resources, including journals, books, databases, platforms,

3 multimedia and articles. R5 focuses on improving the clarity, consistency and comparability of usage reporting by reducing the overall number of reports from Release 4 to include a smaller number of flexible reports. R5 defines Master Reports and Standard Views; Master Reports are large reports containing many metrics and attributes which can be filtered to show Standard Views and user-selected views that meet a wide range of use cases. While tabular versions of reports can be manually downloaded by consumers, easy, automated machine-to-machine retrieval of reports is increasingly important, e.g. for populating statistics in ERM (electronic resources management systems). In keeping with modern thinking and practices for machine-to-machine communications, R5 supports and closely integrates SUSHI, a RESTful API (representational state transfer application program interface) which makes reports available in JSON (JavaScript Object Notation) format.

'R5 focuses on improving the clarity, consistency and comparability of usage reporting'

Building on the success of COUNTER, a more recent development, the CoP for Research Data Usage Metrics release 1,⁶ supports consistent reporting of research data by standardizing the generation and distribution of usage metrics. COUNTER collaborated with members of the Make Data Count⁷ team (California Digital Library, DataCite, and DataONE) in drafting the CoP for Research Data Usage Metrics release 1. The CoP is aligned with COUNTER CoP R5 as clearly there are many areas of commonality. However, research data involves certain unique aspects. Text and data mining (TDM), for example, is a more common route to accessing research data than is the case with traditional scholarly research. These aspects are handled specifically through this CoP, enabling data repositories and providers to report usage metrics according to a common standard and supporting best practice.

Protocols: SUSHI, tracker and Distributed Usage Logging (DUL)

The SUSHI protocol was originally developed, in a collaboration between NISO (National Information Standards Organization) and COUNTER, as a SOAP (simple object access protocol)/XML service. It was designed to facilitate the automated harvesting and consolidation of usage statistics from different vendors. This is undoubtedly essential to handling a large amount of usage data, making the automated retrieval of COUNTER reports into local systems quicker and easier for consumers of reports. The SOAP/XML version of SUSHI was utilized in COUNTER releases up to release 4. As mentioned above, with COUNTER R5, in keeping with current practices for machine interfaces, SUSHI has been reimaged as a RESTful API that returns JSON, which is much simpler, quicker and easier to implement and work with than SOAP/XML.

Again, developed in collaboration with COUNTER, the tracker protocol defines a simple transmission mechanism. When a user downloads a file from a platform with the tracker in operation, an OpenURL-like log entry is generated and sent, as a query string appended to a URL, to the party responsible for creating and consolidating the usage statistics. This protocol was initially developed as part of the Publisher and Institutional Repository Usage Statistics (PIRUS) project but was subsequently incorporated into the COUNTER CoP for Articles. PIRUS demonstrated the facility to harvest and consolidate usage data from multiple sources, whether hosted by repositories or publishers, to offer a comprehensive picture of usage. Unfortunately, both PIRUS and the CoP for Articles received little or no interest from publishers at the time. However, the tracker protocol was adopted by IRUS-UK⁸ and is now used to gather usage data from over 140 UK institutional repositories (IRs). The protocol has also been adopted by OpenAIRE for their work in consolidating repository statistics across Europe and beyond.

More recently, a complementary protocol, the DUL Protocol, has emerged in a collaborative initiative between Crossref, publishers and service providers. Recognizing that scholarly research is increasingly available from repositories, aggregator platforms, researcher-oriented networking sites and reading environments and tools, DUL addresses this by providing 'a private peer-to-peer channel for the secure exchange and processing of COUNTER-compliant private usage records from hosting platforms to publishers'.

- 4 The DUL protocol serves a similar function to the tracker protocol in terms of moving raw usage data to somewhere it can be processed into standardized statistics, but it can also be used to transmit snippets of pre-processed COUNTER-compliant statistics. DUL is more rigorous than the tracker protocol about authentication, verification and provenance of data; and the implementation of the DUL protocol is more technically demanding than the tracker.

DUL allows publishers to capture traditional usage activity related to their content where usage might be occurring on sites other than their own. This facilitates the reporting of overall usage regardless of where that usage occurs. DUL is now being used by a number of publishers, including Elsevier and others.

'facilitating the gathering and reporting of usage data'

These standards and protocols form the basis of a variety of tools and initiatives that are outlined below. These initiatives provide the building blocks for development of services, facilitating the gathering and reporting of usage data whilst serving to provide a more complete picture of usage of scholarly resources.

Tools, initiatives and partnerships

IRUS-UK

Although most products designed for use within IRs provide some form of usage statistics, making comparisons across organizations or products is often difficult or impossible as different products process raw usage data in different ways. Part of Jisc's⁹ OA offer, IRUS-UK and its accompanying programme of services addresses this problem by enabling IRs to share and compare usage data based on the COUNTER standard. IRUS provides IRs with access to authoritative, standards-based statistics that are created on the same basis and comparable with scholarly publishers, supporting participating organizations to gain a better understanding of the breakdown and usage of their institution's research.

Used by virtually all UK IRs, the service supports national comparison and benchmarking, offers a unique source of data for organizations such as funders and policymakers, and serves as an intermediary between UK repositories and other agencies, for example OpenAIRE.

IRUS services work by adding a small piece of code to repository software which employs the tracker protocol described previously. This supports collection of raw usage data from repositories which are then processed and consolidated into COUNTER-conformant statistics by following the rules of the COUNTER CoP.

The standards-based approach to repository usage that IRUS-UK pioneered is easily replicable and has been broadly adopted. IRUS-UK is part of a family of services that currently include instances for CORE, OpenAIRE, the University of Amsterdam and OAPEN in addition to pilot instances that have been developed in Australia, New Zealand and the USA. The value of a standards-based approach is in being able to look across a range of repositories and services that use the standard, measure usage wherever it occurs and assess the impact of various tools. For instance, by combining data from CORE and IRUS-UK, repository managers can evaluate what proportion of their usage is via their native repository and what proportion is via a content aggregator such as CORE.

IRUS for research data

An extension to IRUS-UK, IRUSdataUK, was a pilot project to provide COUNTER-compliant download metrics for research data held in research repositories. It was intended specifically for repositories that host research data, acknowledging that there are implications specific to data repositories; many scholarly items typically consist of metadata and an associated item, whereas data sets are typically comprised of multiple files. IRUSdataUK, used the same tracker protocol as IRUS-UK and data processing is similar but with reporting at the individual file level rather than at the item level as is the case with IRUS-UK.

The IRUSdataUK pilot led to work involving various bodies including Jisc, Making Data Count and COUNTER, and this subsequently led to development of an experimental COUNTER CoP for Research Data. The pilot initially emerged as part of a wider scheme of

5 work that Jisc conducted in 2016, Research at Risk, that dealt with a range of issues around research data management (RDM) and data sharing. At the time of writing, IRUSdataUK is a Beta service, but will be rolled out by Jisc as part of the IRUS family of services in 2020, enabling participating IRs to gain an accurate picture of use of both research items and accompanying data sets.

OpenAIRE Usage Statistics Service

Building on developments described above, IRUS-UK also provides a central source of data which is subsequently utilized by OpenAIRE. OpenAIRE is an initiative supported by the European Commission (EC) with a general remit to implement EC Open Access and Open Data policies and mandates. The OpenAIRE Usage Statistics Service gathers and consolidates usage statistics from a distributed network of data providers (including IRUS-UK, services in Europe, and in South America via La Referencia) through the use of open standards and protocols such as the COUNTER Code of Practice. Its value is in contributing towards impact evaluation of OA usage activity. The benefit of working with IRUS-UK is as a NOAD (National Open Access Desk) for UK IRs rather than needing to interact with individual IRs. OpenAIRE also uses the tracker protocol, although a slightly different implementation from IRUS. Usage data (and more) are available to participants via the OpenAIRE Content Provider Dashboard.¹⁰

'scholarly content is discussed outside the formal literature and beyond the academic community'

Crossref Event Data service

A further Crossref initiative, seeking to provide greater understanding of how scholarly research is used, developed from a project which was to become the Crossref Event Data service.¹¹ Recognizing that scholarly content is discussed outside the formal literature and beyond the academic community, the service collates information and tracks activity surrounding research from potentially any source where an event is associated with a DOI (digital object identifier). The service currently takes data from multiple sources: DataCite, Twitter, Tumblr, Facebook, etc. and provides an open, common infrastructure to track activities around DOIs. It offers raw data available via an API for anyone wishing to build tools and see a fuller picture of activity around an article. It potentially offers value for a range of stakeholders from funders tracking usage of the research they fund that is occurring outside traditional scholarly platforms, to publishers using data to inform business planning.

Each of the tools and initiatives noted above collate and aggregate data from multiple sources, exploiting use of common standards and protocols in order to achieve this. Finally, the Confederation of Open Access Repositories' (COAR's) Next Generation Repository Working Group is helping to drive developments that underpin many of these types of initiatives.

'collate and aggregate data from multiple sources, exploiting use of common standards and protocols'

COAR's Next Generation Working Group

COAR's vision is 'to position repositories as the foundation for a distributed, globally networked infrastructure for scholarly communication'. Aspirations to achieve cross-repository interoperability reinforce the benefits of openness, inclusivity and collective approaches. In 2017 the COAR Next Generation Working Group published a report¹² that defined priority functionalities that repositories should support. These included aspects such as exposing identifiers, declaring licences at the resource level, resource transfer, batch discovery, identification and authentication of users, exposing standardized usage metrics and preservation. Broader, widespread adoption of these principles and functionalities is recommended and would facilitate improved metadata and development of new services on the basis of that. The report highlighted the potential for repositories 'to promote the transformation of the scholarly communication ecosystem, making it more research-centric, innovative, while also managed by the scholarly community', provided that repositories function according to common technologies, standards and protocols.

Conclusion

With a scholarly communications environment in transition, the requirements of researchers, funders and libraries continue to evolve. In order to respond to those changing needs, a collaborative and unified approach is key. As COAR's Next Generation Repositories report indicates, a network of repositories offers a comprehensive view of research globally, and exposing standardized usage metrics supports understanding of the ways in which research is used.

'the requirements of researchers, funders and libraries continue to evolve'

There is a variety of tools and initiatives to support common approaches to measurement and data sharing, and, as we have shown, a significant level of collaboration under way. The success of the COUNTER CoP over the past 15 years in supporting delivery of consistent, comparable statistics is apparent from wide-scale adoption and use of the standard. A collaboratively developed and agreed standard, informed by practical and focused requirements, has resulted in an engaged community using a standard that meets collective needs. Continued engagement with and adoption of the COUNTER CoP for Research Data will help to drive greater standardization and support understanding of how publicly available research is being used. To support comparison and use of data-level metrics, data repositories need to engage and help to refine the CoP Research going forward.

'The success of the COUNTER CoP ... is apparent from wide-scale adoption and use of the standard'

With the adoption of standards and common approaches, there are opportunities to exploit new sources of data and information. The IRUS family of services, which are easily replicable, and OpenAIRE, which is ingesting data from central services, provide effective models and demonstrate the benefits of shared approaches in addition to potential for international measurement and benchmarking. Additionally, Crossref's Event Data is helping to provide a broader picture of the conversations happening around scholarly research.

'there are opportunities to exploit new sources of data and information'

It is through an engaged user community developing these types of practical examples through the application of standards and technologies that the vision of achieving a more coherent and joined-up picture of usage can progress.

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'full list of industry A&As' link: <http://www.uksg.org/publications#aa>

Competing interests

Jo Lambert works for Jisc, a not-for-profit company which undertook some of the work described in this article.

References

1. COUNTER: <https://www.projectcounter.org> (accessed 29 July 2019).
2. CORE: <https://core.ac.uk> (accessed 29 July 2019).
3. "SUSHI (Standardized Usage Harvesting Initiative) Protocol," NISO: <https://www.niso.org/standards-committees/sushi> (accessed 29 July 2019).
4. "Tracker protocol," Jisc: <https://irus.jisc.ac.uk/documents/TrackerProtocol-V3-2017-03-22.pdf> (accessed 29 July 2019).
5. "Distributed Usage Logging," Crossref: <https://www.crossref.org/community/project-dul/> (accessed 29 July 2019).
6. "COUNTER Code of Practice for Research Data Usage Metrics release 1," COUNTER: <https://www.projectcounter.org/counter-code-practice-research-data-usage-metrics-release-1/> (accessed 29 July 2019).
7. "Making Data Count," Lagotto: <http://mdc.lagotto.io/> (accessed 29 July 2019).
8. "IRUS-UK," Jisc: <https://irus.jisc.ac.uk> (accessed 29 July 2019).

9. Jisc:
<https://www.jisc.ac.uk> (accessed 29 July 2019).
10. "Content Provider Dashboard," *OpenAIRE*:
<https://provide.openaire.eu/landing> (accessed 29 July 2019).
11. "Event Data service," *Crossref*:
<https://www.crossref.org/services/event-data/> (accessed 29 July 2019).
12. *Next Generation Repositories: Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group*, November 28, 2017:
<https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf> (accessed 29 July 2019).

Article copyright: © 2019 Paul Needham and Jo Lambert. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.



Corresponding author:

Jo Lambert

Jisc Service Manager,

Jisc

6th Floor, Churchgate House, 56 Oxford Street, Manchester M1 6EU, UK

E-mail: jo.lambert@jisc.ac.uk

ORCID ID: <https://orcid.org/0000-0002-5808-6172>

Co-author:

Paul Needham

ORCID ID: <https://orcid.org/0000-0001-9771-3469>

To cite this article:

Needham P and Lambert J, "Institutional repositories and the item and research data metrics landscape," *Insights*, 2019, 32: 26, 1–7; DOI: <https://doi.org/10.1629/uksg.478>

Submitted on 15 July 2019

Accepted on 25 July 2019

Published on 25 September 2019

Published by UKSG in association with Ubiquity Press.