

Understanding OA Ebook Usage: Toward a Common Framework

Rationale

Journals and books are not equal in their migration into the digital environment, and nowhere is this difference clearer than in the area of usage tracking and reporting. For journals, a range of bibliometrics and altmetrics exist, powered by almost ubiquitous stable identifiers (DOIs), sophisticated abstracting and indexing services, and agreed standards of reporting. For books, no index can claim any degree of comprehensiveness, and standards for usage data such as COUNTER are inconsistently applied.

A particular challenge for book publishers is that most ebooks are not sold directly to customers from publisher platforms but rather pass through a supply chain of intermediaries and aggregators. These include library providers (e.g., EBSCO, ProQuest, JSTOR, Project MUSE) and director-to-consumer retailers (e.g., Rakuten Kobo, Amazon Kindle, Google). To a greater or lesser degree these organizations view information about use as proprietary and share information inconsistently and in formats that cannot easily be compared: For example, some library aggregators report chapter downloads while others report whole book downloads.

At the same time, information about use and engagement is the currency of OA publishing. The promise of OA is that a publication will be more downloaded, more used, and more cited than a comparable restricted-access title. OA publishers need to show such impact to continue to receive support, funders look for this information to demonstrate return on their investments, and authors are eager to show evidence of their reach. The existing challenges of tracking ebook usage are magnified by a proliferation of third-party hosting platforms (e.g., OAPEN, Unglue.it, Internet Archive), which sometimes acquire OA books through formal relationships with publishers but also make use of their Creative Commons licenses to harvest titles.

Looking to the future, the challenge of understanding OA ebook usage may only get worse. New forms of scholarship that explore formats beyond the book are starting to proliferate and make tracking even more complicated because there is no single trackable container. Meanwhile there are an increasing number of platforms beginning to deliver OA books, including FigShare, ResearchGate, Academia.edu and a multitude of institutional and disciplinary repositories.

Support for OA book publishing will not grow unless a compelling case is made for investment. Little of the promise for increasing the reach of long-form digital scholarship through OA can be

demonstrated without the availability of aggregated ebook usage data, agreement on what is important to analyze, and tools for reporting to authors and funders. What is needed is a cross-industry recognition of the opportunities of collaboration, alignment around the challenges of implementation, and a framework for moving forward.

After discussion with a number of stakeholders who also recognize the picture described above, the authors of this proposal intend to convene a structured community conversation around usage tracking for OA ebooks. This conversation would focus on the challenges of identifying and aggregating relevant information from different platforms, analyzing what has been gathered in ways that respect user privacy, and communicating relevant information about usage to stakeholders (including authors, their publishers, their parent institutions, and their funders). As well as looking at strictly quantitative information (such as COUNTER-compliant usage data) we will explore more qualitative “story-telling” indicators, such as the altmetrics which are being collected by providers such as Altmetric.com, Plum Analytics, and CrossRef. Altmetrics are indicators of online engagement derived from tracking identifiable mentions of a particular publication in a variety of digital venues including social media platforms, mainstream and niche news publications, syllabi, and policy documents.

The proposed conversation would bring together three groups of stakeholders who have been actively working on these issues largely independent from one another. While the affiliations are relatively loose, the clustering of individuals and organizations into three groups reflects two important divisions that have emerged in work on OA ebook usage up to now: Firstly, between activity in Europe supported by European Commission funds and activity in North America supported by NEH and Andrew W. Mellon Foundation funds; and secondly, between non-profit and for-profit organizations.

North American non-profits: This group is loosely centered around Project Meerkat. This project was initiated as a 2015 Triangle Scholarly Communication Institute project that investigated the issues around “collecting and analyzing usage data for online scholarly publications” and proposed to develop a “publishing analytics data alliance” for all scholarly monographs (<https://educopia.org/research/meerkat>). Over the last three years, the participants in the Meerkat group have continued to refine their ideas but the scope of the project’s ambition (all ebook usage data, not just OA usage data) has made it difficult to create an actionable plan to move forward.

European non-profits: This group is centered around the Horizon 2020-funded HIRMEOS initiative (<http://www.hirmeos.eu/>). HIRMEOS stands for “High Integration of Research Monographs in the European Open Science Infrastructure.” The HIRMEOS

project is developing shared infrastructure to support the effective integration of monographs into the European open science infrastructure, including the development of tools and services for the collection, aggregation, and visualization of usage data from multiple platforms. The initial project (January 1, 2017 to June 30, 2019) is working with five scholarly publishing platforms in the EU. These are OpenEdition Books (France), The OAPEN Library (Netherlands), EKT ePublishing Services (Greece), Göttingen University Press (Germany), and Ubiquity Press (UK).

Multinational aggregators and platform providers: Members of this disparate group have conducted a number of internal analyses as well as publishing some public reports that contribute to our understanding of the problem as well as hinting at solutions. In 2017 JSTOR made available its ebook data to Knowledge Unlatched Research to explore OA book usage of titles from the University of Michigan Press, University of California Press, UCL Press, and Cornell University Press (http://www.kuresearch.org/PDF/jstor_report.pdf). Also in 2017 Springer Nature presented a report on the “OA effect” that it observed for books on its platform, including titles from leading humanities publisher Palgrave (<https://www.springernature.com/gp/open-research/journals-books/books/the-oa-effect>). Both reports noted that their analyses were preliminary and a larger aggregation of data from other platforms was needed to make more substantive claims.

Schedule of Major Activities

The proposal is structured as a three-part engagement that would (a) identify the challenges in understanding the usage of OA ebooks and recommend a strategy for resolving them (research component); (b) refine these recommendations and create buy-in through a consultation exercise, conducted via online survey and an invited workshop (consultation component); and (c) publish a white paper including a framework for further action (dissemination component).

The research component would be led by Curtin University associate professor Lucy Montgomery of KU Research. The KU Research team includes Lucy Montgomery, professor Cameron Neylon, and associate professor Nic Suzor. Together, the members of this team have been doing important work to investigate issues of OA visibility, discovery and use of open access scholarly books; as well as the strategies and principles needed to support effective governance and mobilization of community data resources. Recent work includes an investigation of the uses of OA books made available via the JSTOR Platform (Montgomery et al 2017: <http://dx.doi.org/10.17613/M6CV52>); a study of the digital visibility of OA books in a

European Context, carried out as part of the European Commission funded OPERAS-D project (Neylon et al 2018: <http://doi.org/10.17613/M6156F>); as well as work on platform governance and legitimacy (Suzor et al 2018: <https://eprints.qut.edu.au/112749/>).

Suzor (Queensland University of Technology) is examining questions of data governance and management, to identify a route toward “data trusts.” The concept of a data trust builds on the observation that data relating to the uses of scholarly works is held by many different stakeholders within the scholarly communication system. Individual libraries, open access repositories, journals, indexing services, and aggregators each have access to some data about how scholarly resources are being used. Each of these stakeholders also possesses different internal technical capacities. Comprehensive and reliable approaches to measuring the uses of OA books requires shared protocols and understanding for data management, exchange, and access.

We will investigate how the needs of various stakeholders can be best reconciled in building a “data trust” or shared data commons, with clear rules for participation and access to data. We will work with participants to identify specific requirements and a road-map for the development of consensus-based protocols that address these needs. Ultimately, the successful development of a sustainable data trust will require a set of access and interchange protocols that protect sensitive information while ensuring that new participants have a strong incentive to provide access to the data they hold in exchange for access to aggregated benchmarking and longitudinal, geographic, and sector-specific insights.

The three purposes of this consultancy by KU Research would be to produce two documents (a landscape survey and an “action plan provocation” – also sometimes referred to as a “straw man”) that together:

- articulate the main issues around collecting, aggregating, analyzing and sharing OA book usage data;
- describe initiatives currently engaging with the issues of OA ebook usage measurement and analysis, especially those outside North America. This would include HIRMEOS but also IRUS (Institutional Repository Usage Statistics) a standards-based statistics aggregation service for repositories created by JISC in the UK (<https://www.jisc.ac.uk/irus>);
- provide a framework presenting issues, options and requirements for governance and technical details moving forward.

Kevin Hawkins, Charles Watkinson, Brian O’Leary, and Katherine Skinner (in her role as facilitator for the potential summit) would provide feedback on drafts of the documents. Hawkins and Watkinson would invite participation in and plan the summit, refine the documents into a white paper, and participate in writing the final report on the project for the Foundation.

Circulation of the documents for feedback would include the posting of an electronic draft to major email discussion lists such as ALA’s ScholComm list, the AUPresses directors list (AAUP-D), Liblicense, the Association of European University Presses list, and Read 2.0. The draft would be posted online by University of Michigan as a Google Doc allowing inline suggestions alongside a Qualtrics survey that will capture meta-discussion through questions that will ascertain whether the landscape survey has missed important initiatives, whether the recommendations seem reasonable and feasible, and whether there are other considerations around the measurement of OA ebook impact and engagement that need to be considered. Although we are not applying for funds for dissemination travel, project team members are frequent participants in industry conferences and plan to publicize the documents and encourage feedback at meetings they will be submitting proposals too, such as at the COASP (Conference on Open Access Scholarly Publishing) meeting in September, the FORCE2018 conference to be held in October, and the Charleston Library Conference in November 2018.

The summit would be held in New York City in early December 2018. It would last 1.5 days starting at 9:30 am on day 1 and ending with dinner, then on day 2 ending at 12 pm with a “to go” lunch. We have selected the participants based on the following criteria: (a) that the individuals represent a particular stakeholder group; (b) that the individuals have a track record of experience with open access books; (c) that, if they are not the leads of their organizations, the individuals are people who are close enough to the issues that they can talk about the details (both strategic and technical) but are respected enough in their organizations to get commitments for further action. We would invite five participants from each of the three stakeholder groups, with only one representative from each organization. The list of participants is shown below, ranked in the order in which we will approach them. There are additional names in each category in anticipation of some invitees being unable to attend.

The three purposes of the summit would be:

- To agree on the issues around aggregating and analyzing OA book usage data;
- to hear from project teams working on related projects, as identified in the environmental scan;
- to discuss and refine the governance and technical framework.

Following the summit, Hawkins and Watkinson would take the lead on reshaping the action plan provocation into a white paper, to be published by the BISG, that would refine the plan for a governance and technical framework for further action by the community.

Brian O’Leary, Executive Director of the Book Industry Study Group (BISG), would convene the summit and act as logistical host. BISG is an appropriate convener since (a) it is an intermediary that is trusted by commercial as well as non-profit organizations, (b) it is heavily embedded in relevant standards conversations with a focus on the overall book supply/value chain, and (c) it represents a diverse range of different types of supply chain partners including a variety of types of partner. As well as attracting a diverse range of participants, we expect that BISG would play an important role in promoting the findings of the project among its member organizations and structuring a continued conversation. As part of the BISG subgrant Katherine Skinner, executive director of the Educopia Institute, would provide professional facilitation for the event.

The respective roles of the different team members are described more in the “Staffing” section below.

Staffing

Charles Watkinson and Kevin Hawkins will divide the responsibilities of coordinating the project, with Watkinson focused on administration (including managing the budget and travel logistics) and Hawkins focused on writing and revising documents and project-managing the different stages (the more time consuming part of the project’s leadership). Both have long-standing interests and demonstrated experience of issues surrounding ebooks and usage measurement.

Charles Watkinson is Associate University Librarian for Publishing at the University of Michigan and Director of the University of Michigan Press. He has been involved in open access book publishing for a number of years and has increased the OA publishing output of UMP (in 2017 over 15% of new books published by UMP were OA). He has also been involved in a number of initiatives focused on usage and engagement measurement. He is chair of the impact assessment working group for the AAU/ARL/AUPresses “Toward an Open Monograph Ecosystem” project, has served as a member of the Altmetric Advisory Board, and was PI on the Foundation-funded project “Mapping the Free Ebook Supply Chain” which focused on how OA ebooks are discovered and used.

Because he is employed by another institution, the University of North Texas, Kevin Hawkins's time commitment to the project would be accounted for through a subgrant to UNT. This administrative arrangement underplays Hawkins' shared leadership of the project with Watkinson. Hawkins has been a pioneer in the publication of OA ebooks, coordinating the design and production of one of the earliest OA ebook imprints, Digital Culture Books, during his employment at University of Michigan Library. Over 50 titles have been published under this imprint since 2006, free-to-read online but financially sustained by sales of print and downloadable ebook editions. Hawkins is now assistant dean for scholarly communication at UNT Libraries, where he founded the library publishing operation, supervises consultation and outreach on copyright, and contributes towards consultation and outreach on research data management. He also leads planning of UNT's annual OA Symposium where impact measurement is a recurrent topic.

The research component of the project will be led by Lucy Montgomery, Director of Research at KU Research. KU Research is an independent research and analysis group focusing on strategy and analytics that support the ecosystem of scholarly monographs. It was selected by Watkinson and Hawkins as the most appropriate consultant for the project because of its focus on OA ebook business modeling and impact measurement, its global outlook and breadth of international relationships, and the proven quality of its information scientists. Montgomery also serves, for example, as the Director of the Centre for Culture and Technology at Curtin University, Australia and a member of the Curtin Institute for Computation. She was a key member of a small team responsible for developing and successfully piloting Knowledge Unlatched: a globally coordinated, collaborative model for enabling OA for specialist scholarly publications at scale. Her current projects include the EU-funded Horizon 2020 OPERAS-D project, where Montgomery led the team responsible for mapping the digital visibility of OA books in a European context, and she has published a number of important studies on OA impact measurement.

Core to this project is the bringing together of a diverse group of expert stakeholders at a summit in New York City. Brian O'Leary is uniquely qualified to convene such a group in his capacity as executive director of the Book Industry Study Group, a U.S.-based trade association that works to create a more informed, effective and efficient book industry supply chain. He oversees the work BISG does to disseminate information, create and implement standards, conduct research and grow membership from companies working throughout the supply chain. He has deep industry experience, having worked for over 30 years in a variety of trade publishing and consultancy capacities. BISG was invited to participate in this project because of its nodal role in advancing the development of ebooks and deep relationships/credibility with a range of relevant stakeholders.

To ensure that the meeting is productive requires an expert facilitator who also has a deep understanding of the issues. Katherine Skinner is the Executive Director of the Educopia Institute, a not-for-profit organization that builds networks and collaborative communities to help cultural, scientific, and scholarly institutions achieve greater impact. Skinner was invited both because she provides meeting facilitation for a range of cross-sector initiatives, drawing up on such methodologies as “Boundary-Spanning Leadership” and “Collective Impact,” but also because of her experience developing long-lasting collaborative frameworks for advancing important scholarly communications themes. Examples of her leadership include the MetaArchive Cooperative, a community-governed digital preservation network with more than 60 member institutions in four countries; the Library Publishing Coalition, a membership organization supporting library publishing and scholarly communications activities across more than 75 academic libraries; and the BitCurator Consortium, a community-led membership association focused on digital forensics practices in libraries, archives, and museums.

Timeline

June – August 2018

- Set-up of funds, consultancy and sub-contract agreements at Michigan (Watkinson)
- New York summit date and venue confirmed (O’Leary)
- KU Research starts research work (Montgomery)
- Invitations issued to potential summit attendees (Hawkins, Watkinson)

September – November 2018

- KU Research finishes research work. First draft of “landscape survey” and “action plan provocation” by end of September. Second public discussion draft by end of October. (Montgomery)
- PI team refines KU Research draft documents (Hawkins, Watkinson, O’Leary)
- Circulation of draft for comment along with survey for response (Hawkins)
- Travel arrangements finalized with attendees (Watkinson)
- Facilitation plan for the summit agreed (Skinner)

December 2018 – February 2019

- Logistical plan for summit held in New York City executed; early December (O’Leary)
- Facilitation of the summit (Skinner)
- Travel reimbursements issued (Watkinson)
- Discussions at summit captured and collated (BISG staff)

- KU Research report, survey feedback, and summit conversations collated into draft white paper. End of February deadline. (Hawkins, Watkinson)

March 2019 – May 2019

- White paper refined by PI Team during March. Revision by end of March (Hawkins, Watkinson)
- Submitted to BISG by end of March. White paper issued as BISG publication in May 2019 (O’Leary)

June 2019 – August 2019

- Write-up and submit final report to Foundation (Watkinson, Hawkins)

Expected Outcome and Benefits

Three main outcomes of the project are anticipated to be as follows:

- Cross-sector, transatlantic appreciation of the importance of OA ebook usage information in demonstrating the return on investment of OA book publishing.
- Shared understanding of the challenges of aggregating, analyzing, and communication OA ebook usage, informed by international studies and experiences.
- Agreement on the next steps necessary to establish a system to aggregate, analyze, and communicate OA ebook usage and the funding options to build the necessary technological and governance infrastructure.

The tangible output of the project is a White Paper incorporating the landscape survey and “provocation” prepared by KU Research with the results of the Summit and comments of online respondents. It is anticipated that the findings of project will inform the future implementation of measures of OA ebook impact and engagement and the project team has had several explicit conversations with participants who are applying for funding. These include John Sherer, submitting a proposal for the “Open Access History Monograph” project at the invitation of the Andrew W. Mellon Foundation, and Pierre Mounier, working on an application to the European Commission for a next phase of funding for the HIRMEOS project and explicitly looking for international collaboration.

List of Invitees to Summit

North American Non-Profits (five representatives from the list below)

[Redacted list of North American Non-Profits]

European Non-Profits (five representatives from the list below)

[Redacted list of European Non-Profits]

Multinational Aggregators and Platforms (five representatives from the list below)

[Redacted list of Multinational Aggregators and Platforms]



Organizers

1. Kevin Hawkins, Assistant Dean for Scholarly Communication, University of North Texas Libraries <Kevin.Hawkins@unt.edu> (USA)
2. Lucy Montgomery, Director of Research, KU Research <lucy@kuresearch.org> (Australia)
3. Cameron Neylon, Executive Director, KU Research <cn@cameronneylon.net> (Australia)
4. Brian O’Leary, Executive Director, Book Industry Study Group <brian@bisg.org> (USA)
5. Charles Watkinson, Director, University of Michigan Press <watkinc@umich.edu> (USA)
6. Rebecca Welzenbach, Research Impact Librarian, University of Michigan Library <rwelzenb@umich.edu> (USA)