



BIROn - Birkbeck Institutional Research Online

Eve, Martin Paul (2019) Open Metrics for Monographs: Background Contexts and Literature Review. Jisc Open Metrics Lab ,

Downloaded from: <http://eprints.bbk.ac.uk/27807/>

Usage Guidelines:

Please refer to usage guidelines at <http://eprints.bbk.ac.uk/policies.html> or alternatively contact lib-eprints@bbk.ac.uk.

Jisc Open Metrics Lab

Open Metrics for Monographs: Background Contexts and Literature Review

Professor Martin Paul Eve

Birkbeck, University of London

All content published can be shared (CC BY 4.0)

<https://creativecommons.org/licenses/by/4.0/>



Table of Contents

Background Contexts.....	3
Existing OA Monograph Metric Initiatives.....	7
The Jisc Open Metrics Lab Monograph Experiment in Context	10
Bibliography	15

Background Contexts

The Jisc Open Metrics Lab's Monograph Experiment takes place at a time of transition for the academic monograph in the United Kingdom. Amid debate over whether a "crisis of the academic monograph even exists", in the past few years there have nonetheless been several signals that open access for academic monographs is becoming a reality.¹ The 2015 Crossick report to HEFCE, for instance, noted that this was a time for experiment for these important media in the Humanities and Social Scientific disciplines,² a time in which "[o]utside the framework of any policies, funders should play a role in facilitating through pilots and the formulation of standards those developments that will help digital open access realise its potential for innovation in research communication, collaboration and practice".³

The true impetus, though, for a move towards open access for academic monographs was given in the 2018 announcement by the global coalition of funders, known as cOAlition S, that their uncompromising mandate for open access was to extend to monographs at some point in

-
- 1 On the debate about the crisis in conventional monograph production, see Ronald Snijder, 'Measuring Monographs: A Quantitative Method to Assess Scientific Impact and Societal Relevance', *First Monday*, 18.5 (2013) <<https://firstmonday.org/ojs/index.php/fm/article/view/4250>> [accessed 12 May 2019]; Marilyn Deegan, 'Academic Book of the Future Project Report', 2017 <https://academicbookfuture.files.wordpress.com/2017/06/project-report_academic-book-of-the-future_deegan3.pdf>; Michael Jubb, 'Academic Books and Their Future', 2017 <https://academicbookfuture.files.wordpress.com/2017/06/academic-books-and-their-futures_jubb1.pdf>; For more on the definitions of open access, see Peter Suber, *Open Access*, Essential Knowledge Series (Cambridge, MA: MIT Press, 2012) <<http://bit.ly/oa-book>>; Martin Paul Eve, *Open Access and the Humanities: Contexts, Controversies and the Future* (Cambridge: Cambridge University Press, 2014) <<http://dx.doi.org/10.1017/CBO9781316161012>>.
 - 2 While it is true that monographs exist in the natural scientific spaces, and indeed many of the most important works in the history of the natural sciences have been published in book form, such as Darwin's *The Origin of the Species*, this report confines its remit to the humanities and social sciences. Indeed, the fact that the Wellcome Trust's open-access mandate includes monographs indicates that even a scientific funder with only a small cohort of medical humanities authors takes seriously this media form. Further, there have been recent attempts to appraise the bibliometrics of health monographs. See Pamela Royle and Norman Waugh, 'Bibliometrics of NIHR HTA Monographs and Their Related Journal Articles', *BMJ Open*, 5.2 (2015), e006595 <<https://doi.org/10.1136/bmjopen-2014-006595>>.
 - 3 Geoffrey Crossick, 'Monographs and Open Access: A Report for the Higher Education Funding Council for England', *Higher Education Funding Council for England*, 2015, p. 68 <<http://www.hefce.ac.uk/pubs/rereports/year/2015/monographs/>> [accessed 24 May 2015].

the near future, although the precise timeframe was left unspecified.⁴ Given that Crossick's calls for experiment have barely begun, this has come as a shock to many in the humanities and social sciences as no single business model for open access has yet been developed (or may be desirable).

That said, there have been many advances since Crossick. Knowledge Unlatched continues to be the largest and most successful open-access monograph initiative,⁵ facilitating the opening of hundreds of academic monographs.⁶ The transfer of Knowledge Unlatched to a for-profit structure in 2018, however, has prompted some hand-wringing among libraries and

-
- 4 cOAlition S, 'Plan S', *Plan S and COAlition S*, 2018 <<https://www.coalition-s.org/>> [accessed 12 May 2019]; Funder mandates have often driven the uptake of open access. For more on mandates, see Ulrich Herb, 'Recommendations, Statements, Declarations And Activities Of Science Policy Actors On Shaping The Scholarly Communication System', *Zenodo*, 2017 <<https://doi.org/10.5281/zenodo.1003229>>; David Sweeney and Ben Johnson, 'Seeking a Fresh Perspective: A Research Funder's View of Open Access', *Insights: The UKSG Journal*, 27.1 (2014), 51–57 <<https://doi.org/10.1629/2048-7754.114>>; José Carvalho and others, 'Monitoring a National Open Access Funder Mandate', *Procedia Computer Science*, 13th International Conference on Current Research Information Systems, CRIS2016, Communicating and Measuring Research Responsibly: Profiling, Metrics, Impact, Interoperability, 106 (2017), 283–90 <<https://doi.org/10.1016/j.procs.2017.03.027>>; Philippe Vincent-Lamarre, Jade Boivin, Yassine Gargouri, Vincent Larivière, and others, 'The Effect of Open Access Mandate Strength on Deposit Rate and Latency', 2014 <<http://eprints.soton.ac.uk/366815/>> [accessed 23 July 2014]; Philippe Vincent-Lamarre, Jade Boivin, Yassine Gargouri, Vincent Larivière, and others, 'Estimating Open Access Mandate Effectiveness: The MELIBEA Score', *ArXiv:1410.2926 [Cs]*, 2014 <<http://arxiv.org/abs/1410.2926>> [accessed 12 May 2019]; Jingfeng Xia and others, 'A Review of Open Access Self-Archiving Mandate Policies', *Portal: Libraries and the Academy*, 12.1 (2012), 85–102; Alma Swan, 'Open Access Policy Effectiveness: A Briefing Paper for Research Institutions' (Pasteur4OA) <<http://www.pasteur4oa.eu/sites/pasteur4oa/files/resource/Policy%20effectiveness%20-%20institutions%20final.pdf>>; Vincent Larivière and Cassidy R. Sugimoto, 'Do Authors Comply When Funders Enforce Open Access to Research?', *Nature*, 562.7728 (2018), 483 <<https://doi.org/10.1038/d41586-018-07101-w>>; it can be difficult to see, though, how mandates will translate from the journal to the monographic space Martin Paul Eve and others, 'Cost Estimates of an Open Access Mandate for Monographs in the UK's Third Research Excellence Framework', *Insights*, 30.3 (2017) <<https://doi.org/10.1629/uksg.392>>.
- 5 In this report, "open access" is hyphenated as "open-access" when used as a prepositive adjective.
- 6 See Frances Pinter and Christopher Kenneally, 'Publishing Pioneer Seeks Knowledge Unlatched', 2013 <<http://beyondthebookcast.com/transcripts/publishing-pioneer-seeks-knowledge-unlatched/>>; Higher Education Funding Council for England, 'Knowledge Unlatched Pilot given HEFCE Backing', 2013 <<https://www.hefce.ac.uk/news/newsarchive/2013/news85263.html>> [accessed 21 December 2013]; Lucy Montgomery, 'Knowledge Unlatched: A Global Library Consortium Model for Funding Open Access Scholarly Books', *Cultural Science*, 7.2 (2014), 1–66; Knowledge Unlatched, 'How It Works', 2013 <<http://www.knowledgounlatched.org/about/how-it-works/>> [accessed 5 December 2013].

criticisms from other presses around some of its activities.⁷ Other smaller initiatives – often working under the banner of the ScholarLed coalition – have also shown early success though, particularly punctum books, Open Humanities Press, and Open Book Publishers. There has also been a rise of the “new” university press, specialising in open-access monographs, among which number UCL Press, Goldsmiths Press, Luminos Press, Lever Press, Calvary Press, and many many others. (Notably, these new university presses are a diverse and heterogeneous grouping, with some operating out of the library on a budget close to zero, while others receive substantial budgetary subsidy. Some, also, such as Lever Press span multiple academic institutions.⁸) On the whole, the rise of open-access monographs appears set to continue.⁹

An important part of the debate around open-access monographs, though, has been usage, situated within a broader context of developing bibliometric indicators that are sensitive towards, and can work in, the humanities and social sciences.¹⁰ Indeed, the Crossick report stressed that a “clear articulation of the opportunities and benefits of open access for monographs will be an essential component of policymaking in this area”; an articulation that can only be made when backed by evidence.¹¹ Some publishers, such as Springer-Nature, have already made moves in this direction, demonstrating and publicising a seven-fold increase in general usage among their open-access monographic titles, by various measures.¹² However, in addition to the

-
- 7 Springer Nature, ‘Open Research Library’, *Open Research*, 2019 <<https://www.springernature.com/gp/open-research/journals-books/books/orl>> [accessed 26 May 2019].
- 8 For more on this theme, see Janneke Adema, Graham Stone, and Chris Keene, ‘Changing Publishing Ecologies: A Landscape Study of New University Presses and Academic-Led Publishing’ (Jisc, 2017).
- 9 Simba Information, ‘Open Access Book Publishing 2016-2020’, 2016 <<https://www.simbainformation.com/Open-Access-Book-10410716/>> [accessed 12 May 2019]; Eelco Ferwerda, Frances Pinter, and Niels Stern, *A Landscape Study On Open Access And Monographs: Policies, Funding And Publishing In Eight European Countries* (Zenodo, 1 August 2017) <<https://doi.org/10.5281/zenodo.815932>>.
- 10 Björn Hammarfelt, ‘Beyond Coverage: Toward a Bibliometrics for the Humanities’, in *Research Assessment in the Humanities: Towards Criteria and Procedures*, ed. by Michael Ochsner, Sven E. Hug, and Hans-Dieter Daniel (Cham: Springer International Publishing, 2016), pp. 115–31 <https://doi.org/10.1007/978-3-319-29016-4_10>.
- 11 Crossick, p. 68.
- 12 Christina Emery and others, ‘The OA Effect: How Does Open Access Affect the Usage of Scholarly Books?’ (Springer-Nature, 2017) <<https://media.springernature.com/full/springer-cms/rest/v1/content/15176744/data/v3>> [accessed 12 May 2019].

regular complexities of citation and reference analysis, undertaking such analysis in the humanities and social sciences presents specific difficulties.¹³ Not least of these is the issue of coverage of these disciplines' outputs within the conventional databases that are used for bibliometric analyses.¹⁴ This is usually attributed, threefold, to “diverse publication channels, the importance of ‘local’ languages as well as the wide-ranging audience of research”.¹⁵

Yet there is also a problem of disciplinary definition at work here. When it is claimed, for instance, that books are more frequently cited in the humanities than journal articles (and vice versa in the social sciences), this is a generalization too far. For, in aggregating up to “the humanities” and “the social sciences”, this elides the fact that the citation of journals plays a central role in, for instance, history and linguistics, while sociology and library information sciences hold the monograph in high citation regard (although it is always worth noting that the lack of semantic value in citation metrics means that it is impossible to bestow a positive characteristic upon what is merely attention – a citation).¹⁶ There are also, though, serious problems of obtaining accurate, centralized data for OA monographs, often by the publishers themselves.¹⁷ For one, the permissive distribution clauses of the Creative Commons licenses – a feature, not a bug, of open-access dissemination in that copies can end up distributed in different locations – means that statistics must be aggregated and are unlikely, even where such figures are available, to be

13 Cameron Neylon, ‘The Complexities of Citation: How Theory Can Support Effective Policy and Implementation’, 2016 <<http://repository.jisc.ac.uk/6553/>> [accessed 12 May 2019].

14 See, for just a selection Jordi Ardanuy, ‘Sixty Years of Citation Analysis Studies in the Humanities (1951–2010)’, *Journal of the American Society for Information Science and Technology*, 64.8 (2013), 1751–55 <<https://doi.org/10.1002/asi.22835>>; Anton J. Nederhof, ‘Bibliometric Monitoring of Research Performance in the Social Sciences and the Humanities: A Review’, *Scientometrics*, 66.1 (2006), 81–100 <<https://doi.org/10.1007/s11192-006-0007-2>>; Maria Teresa Biagetti, Antonella Iacono, and Antonella Trombone, ‘Testing Library Catalog Analysis as a Bibliometric Indicator for Research Evaluation in Social Sciences and Humanities’, in *Challenges and Opportunities for Knowledge Organization in the Digital Age*, ed. by Fernanda Ribeiro and Maria Elisa Cerveira (Berlin: Ergon Verlag, 2018), pp. 892–99 <<https://doi.org/10.5771/9783956504211-892>>.

15 Hammarfelt, ‘Beyond Coverage’, p. 117.

16 Björn Hammarfelt, ‘Following the Footnotes: A Bibliometric Analysis of Citation Patterns in Literary Studies’ (unpublished Doctoral, Uppsala University, 2012), p. 31.

17 Charles Watkinson, Rebecca Welzenbach, and others, ‘Mapping the Free Ebook Supply Chain: Final Report to the Andrew W. Mellon Foundation’, 2017, p. 4 <<https://deepblue.lib.umich.edu/handle/2027.42/137638>>.

collected in standardised ways in all instances across multiple platforms (e.g. COUNTER compliance).¹⁸

Existing OA Monograph Metric Initiatives

Following in these difficult methodological footsteps, there have been several projects that have, nonetheless, tried to gauge the impact and measure the changes to usage that open access has had on academic monographs. Of particular note are the OAPEN-NL and OAPEN-UK projects, which attempted to measure usage and sales figures for matching controlled sets of monographs.¹⁹ The aforementioned Springer-Nature report has also attempted to provide a comparative measure of usage between the company's OA and non-OA books.²⁰ Knowledge Unlatched Research – the non-commercial research arm of KU – has also conducted a comparative analysis of usage within the JSTOR ecosystem of the first four publishers to begin distributing their OA books through this channel.²¹ This approach, as with the Springer-Nature study, has the advantage of isolating its analysis to one particular context, thereby avoiding the above noted problems of statistical aggregation. The difficulty, of course, is that such an analysis is more likely to favour OA books, as the non-availability of a title is less likely to lead to a download.

This highlights the important interrelationship between measuring “usage” (be this citations, references, or views and downloads) and understanding discoverability (how users

18 'Project COUNTER - Consistent, Credible, Comparable', *Project Counter* <<https://www.projectcounter.org/>> [accessed 12 May 2019].

19 See OAPEN-UK, 'The Pilot', 2013 <<http://oopen-uk.jiscebooks.org/pilot/>> [accessed 25 March 2014]; Eelco Ferwerda, Ronald Snijder, and Janneke Adema, 'OAPEN-NL: A Project Exploring Open Access Monograph Publishing in the Netherlands Final Report', 2013 <<http://www.oopen.nl/images/attachments/article/58/OAPEN-NL-final-report.pdf>> [accessed 24 March 2014]; Janneke Adema, 'Overview of Open Access Models for Ebooks in the Humanities and Social Sciences', *OAPEN*, 2010 <<https://curve.coventry.ac.uk/open/file/a976330e-ed7a-4bd5-b0ed-47cab90e9a5e/1/ademaoopen2comb.pdf>> [accessed 12 August 2014].

20 Emery and others.

21 Lucy Montgomery and others, 'Exploring Usage of Open Access Books via the JSTOR Platform' (Knowledge Unlatched Research, 2017) <http://kuresearch.org/PDF/jstor_report.pdf>.

come to find material). However, as Neylon *et al.* note, “[t]he question of visibility is [...] a complex one”.²² The problems that they identify for monographs – a print-centric discoverability system, intermediaries rather than direct reader interactions, lack of persistent identifier redirects, unexpected audience groups, poor quality assurance on data that is collected, small presses with little capacity for data collection, and inconsistent metadata – appear pervasive and will take many years to address.

Nonetheless, and despite the statistical problems encountered in the OAPEN-UK project, a subsequent and more recent OAPEN-CH project in Switzerland has managed to find some statistically significant differences between OA and non-OA books.²³ Namely that:

- “Open access had a statistically significant positive influence on the trackability and visibility of the monographs”
- “Placing open access monographs in the OAPEN Library increased international reach”
- “Open access had a statistically significant influence on the use of monographs (number of book visits, page views and downloads). Monographs in the experimental group were used more frequently than books in the control group.”
- “Statistically, open access did not have a negative influence on the sales figures for printed books. The average number of monographs sold in the experimental group was only negligibly lower than the number in the control group. In fact, more copies overall were sold in the experimental group. However, the reverse conclusion – open access has a positive impact on sales figures – does not hold statistically either since there were hardly any differences between the two groups.”

22 Cameron Neylon and others, ‘The Visibility of Open Access Monographs in a European Context: Full Report’ (Knowledge Unlatched Research, 2018), p. 7
<<https://hcommons.org/deposits/objects/hc:18270/datastreams/CONTENT/content>>.

23 Eelco Ferwerda and others, *OAPEN-CH – The Impact Of Open Access On Scientific Monographs In Switzerland. A Project Conducted By The Swiss National Science Foundation (SNSF)* (Zenodo, 23 April 2018) <<https://doi.org/10.5281/zenodo.1220607>>.

These findings are clearly of interest to those piloting business models for open-access monographs. Amid existing debates over whether the monograph is sustainable, the knowledge that OA appears not to have damaged sales figures is a potentially heartening finding, although there are multiple explanations for why this may be the case (poor discoverability of OA editions, unawareness of OA editions etc.). That said, there are also convincing rationales for how this finding should have come about (people favour reading in print, libraries buying print to support OA etc.).

The HIRMEOS project (High Integration of Research Monographs in the European Open Science Infrastructure) also has a work package devoted to metrics and monographs. Led by Ubiquity Press, this has resulted in the development of a metrics standard that includes DOI scraping, altmetrics (attention scores), and geolocation data on readers.²⁴ This project also convened a workshop in Paris on metrics for open-access monographs.²⁵ Of note, perhaps, here and stemming from the phrasing of the HIRMEOS workshop is the ambiguity in the term “open metrics”, and whether this refers to metrics that are, themselves, open, or metrics pertaining to research objects that are open access.

Thinking more around the values of metrics and the ways in which such facilities are often abused in the research evaluation process, the Stateside Humane Metrics Initiative has turned its attention to developing “an initiative for rethinking humane indicators of excellence in academia, focused particularly on the humanities and social sciences (HSS)”. These centre

24 HIRMEOS and Ubiquity Press, ‘Deliverable D6.1: Metrics Services Specification’, *High Integration of Research Monographs in the European Open Science Infrastructure*, 2019
<https://www.hirmeos.eu/wp-content/uploads/2017/11/Hi61-Metrics_Service_technical_specification-final.pdf> [accessed 3 June 2019].

25 HIRMEOS, ‘HIRMEOS Workshops on Annotation and Metrics for OA Monographs, 10-11 Jan 2019, Paris’, *High Integration of Research Monographs in the European Open Science Infrastructure*, 2019
<<https://www.hirmeos.eu/2018/11/05/hirmeos-workshops-on-annotation-and-metrics-for-oa-monographs-10-11jan-2019-paris/>> [accessed 3 June 2019].

around collegiality, quality, equity, openness, and community.²⁶ That said, the very idea that metrics for scholarship *can even be humane* has been contested and opened for debate.²⁷

Finally, there is at least one other major project in train that aims to investigate the usage of ebooks more broadly – and understanding that many of the above problems are not just specific to OA books, but simply digital books in general remains key. The Book Industry Study Group, working with Knowledge Unlatched Research, is currently undertaking a study, funded by the Andrew W. Mellon Foundation, that aims to convene “a structured community conversation around usage tracking for OA ebooks”.²⁸ Using both conventional (that is, quantitative citation) metrics and altmetrics as its basis, this project offers a potentially promising route to addressing the systemic problems with metrics for monographs, and thereby yielding convincing rationales for any transition to open access.

The Jisc Open Metrics Lab Monograph Experiment in Context

Bibliometrics for monographs – and open-access monographs – remain extremely difficult to do well. There are a range of basic issues and problems when profiling books that are simply not as far along as they are in the world of journals. This is, in part, due to the lateness of books to come to the digital format compared to journals. That said, of course, coming late to a field also

26 HuMetricsHSS, ‘About HuMetricsHSS’, *Humane Metrics Initiative*, 2018

<<http://webcache.googleusercontent.com/search?q=cache:A8FHPfPBuC8J:humetricshss.org/about/+&cd=1&hl=en&ct=clnk&gl=uk>> [accessed 3 June 2019].

27 Stacy Konkiel, ‘Approaches to Creating “Humane” Research Evaluation Metrics for the Humanities’, *Insights the UKSG Journal*, 31 (2018) <<https://doi.org/10.1629/uksg.445>>; Martina Franzen, Eileen Joy, and Chris Long, *Humane Metrics/Metrics Noir* (Coventry UK: Post Office Press / meson press, 2018) <<https://hcommons.org/deposits/item/hc:19823/>> [accessed 30 May 2019].

28 Charles Watkinson, Kevin Hawkins, and others, ‘Understanding OA Ebook Usage: Toward a Common Framework’ (Knowledge Unlatched Research, 2018), p. 2 <https://deepblue.lib.umich.edu/bitstream/handle/2027.42/143840/Redacted%20Grant%20Narrative%20-%20OA%20Ebook%20Usage_FINAL%20SUBMISSION_042718.pdf?sequence=1&isAllowed=y> [accessed 13 May 2019].

confers concomitant advantages in that one can learn from the difficulties faced in other contexts.

One of the overarching challenges, though, remains the fact that bibliometrics are inextricably associated with research assessment. In the eyes of many humanities and social scientific researchers, the only use to which the development of accurate citation metrics for books could be put is to develop ever-more coercive evaluation procedures, which is an undesirable outcome for most academics. Coupled with the extremely long citation half life in many HSS disciplines,²⁹ the lack of a convincing rationale for citation metrics for monographs (beyond evaluation) is hindering the uptake of open access.

It is within this context that the Jisc Open Metrics Lab Open Monographs Experiment places itself. For, if those in the humanities and social sciences do not want bibliometrics to be used for assessment, they are all actually already used to using the citation graph in another type of utilitarian exercise: cross-referencing in order to gain an understanding of a field. As outlined in the launch blog post for this project, one discovery technique used at present is to travel to a national deposit library and order ten or so books that appear to have pertinent titles. The researcher can then cross-reference the bibliographies of these books in order to ascertain what they cite in common. This allows the researcher to quickly understand a new field: the most-cited items in common will be good pieces to read in order to rapidly understand a new disciplinary space.

This is a labour intensive process. It involves the move to a physical space in the first place – a physical research library – which on its own has implications for accessibility for those with mobility conditions or long-term health problems. This is then followed by a search of the

29 See Nigel Vincent and Chris Wickham, 'Debating Open Access: Introduction', in *Debating Open Access*, ed. by Nigel Vincent and Chris Wickham (London: British Academy, 2013), pp. 4–12 and the accompanying volume.

catalogue, a wait for the delivery of the items, and then a laborious process of note taking, observation and cross-referencing across hundreds of permutations of bibliographic entries.

For the experiment that we are undertaking, we decided to implement a digital system to perform this task using open-access monographs. For what if, in the contemporary digital publishing landscape, there were a better way than this manual searching? The project has three components:

1. This literature review of existing material on bibliometrics for open-access monographs and bibliographic intersection tools;
2. A tool that will allow people to download a corpus from the DOAB;
3. A tool that will parse references from open-access monographs and tell the user which items are cited in common among the selected titles.

There are existing tools in this space, but none, so far as we know, for monographs. The most well-known of these is CitationGecko, which acts as a visualization aide for CrossRef's repository of interlinked citations. This relies on the publisher having deposited semantically rich citation metadata with CrossRef, which we suspect many open-access book publishers are not doing. There are also attempts at referencing mining that are variable in their success rates. For instance CERMINE uses a visual approach to PDF parsing to attempt to identify references and to parse them into uniquely identifiable data objects in the JATS XML format. Unfortunately, as with many PDF parsing solutions, there are serious problems with generically identifying the visual styling of citations and our initial attempts indicated that line breaks between entries in a test corpus of Cambridge University Press books caused serious problems. The anycite.io parser, written in Ruby, suffered from a similar problem of distinguishing references from one another. However, this latter parser also has a mode in which, if one can pass it clean, single-line, single reference plaintext, it has a high success rate for parsing the result. This makes it a viable option if one can parse the books into, at the very least, individuated references.

There are a range of strategies that we will deploy in order to convert from the free-text referencing of OA books to semantically rich and uniquely identifiable data objects that can be cross-compared. Further, the format delivery of the identifier link in the DOAB OAI feed is not consistent. Instead, when following through the OA edition of a book, we will have to be able to detect the passed format (PDF/epub/landing page with options/Dropbox-encapsulated PDF) via MIME type and select appropriately for the targeted publishers in this iteration. Existing approaches of examining the PDF files has clearly failed on projects with far more resource than this. Instead, we will use the following approaches:

- epub versions often have semantically rich TEI or HTML markup within them. Where we are presented with an epub option, we will attempt to parse this first. Many SciELO books adopt this format.
- In cases where the OAI identifier redirects the client to a URL beginning with `oapen.org/view`, we have found that often we can obtain semantically rich data from this source, which uses an XSLT transformation to render references in a standard form. This appears to be the format used by Palgrave Macmillan, thereby opening a substantial catalogue of OA books to this technique.
- In some cases, publisher platforms provide semantically rich formatted HTML versions of referencing systems. For instance, in the case of Cambridge University Press, a landing page template of <https://doi.org/10.1017/CBO<ISBN-13>> yields semantically rich formatting of references, broken down into constituent components (e.g. author names, titles etc.).

In this way, we hope to demonstrate value to humanities and social science researchers in being able to parse an open citation graph that goes beyond merely developing bibliometrics for research assessment. Certainly, the task remains difficult for all the reasons above. But we

believe that this tool could truly be of use to researchers in these disciplines and can then be used as further evidence of the value of opening access to monographs.

Bibliography

- Adema, Janneke, 'Overview of Open Access Models for Ebooks in the Humanities and Social Sciences', *OAPEN*, 2010 <<https://curve.coventry.ac.uk/open/file/a976330e-ed7a-4bd5-b0ed-47cab90e9a5e/1/ademaoopen2comb.pdf>> [accessed 12 August 2014]
- Adema, Janneke, Graham Stone, and Chris Keene, 'Changing Publishing Ecologies: A Landscape Study of New University Presses and Academic-Led Publishing' (Jisc, 2017)
- Ardanuy, Jordi, 'Sixty Years of Citation Analysis Studies in the Humanities (1951–2010)', *Journal of the American Society for Information Science and Technology*, 64.8 (2013), 1751–55 <<https://doi.org/10.1002/asi.22835>>
- Biagetti, Maria Teresa, Antonella Iacono, and Antonella Trombone, 'Testing Library Catalog Analysis as a Bibliometric Indicator for Research Evaluation in Social Sciences and Humanities', in *Challenges and Opportunities for Knowledge Organization in the Digital Age*, ed. by Fernanda Ribeiro and Maria Elisa Cerveira (Berlin: Ergon Verlag, 2018), pp. 892–99 <<https://doi.org/10.5771/9783956504211-892>>
- Carvalho, José, Cátia Laranjeira, Vasco Vaz, and João Mendes Moreira, 'Monitoring a National Open Access Funder Mandate', *Procedia Computer Science*, 13th International Conference on Current Research Information Systems, CRIS2016, Communicating and Measuring Research Responsibly: Profiling, Metrics, Impact, Interoperability, 106 (2017), 283–90 <<https://doi.org/10.1016/j.procs.2017.03.027>>
- coalition S, 'Plan S', *Plan S and COAlition S*, 2018 <<https://www.coalition-s.org/>> [accessed 12 May 2019]
- Crossick, Geoffrey, 'Monographs and Open Access: A Report for the Higher Education Funding Council for England', *Higher Education Funding Council for England*, 2015 <<http://www.hefce.ac.uk/pubs/rereports/year/2015/monographs/>> [accessed 24 May 2015]
- Deegan, Marilyn, 'Academic Book of the Future Project Report', 2017 <https://academicbookfuture.files.wordpress.com/2017/06/project-report_academic-book-of-the-future_deegan3.pdf>
- Emery, Christina, Lucraft Mithu, Agata Morka, and Ros Pyne, 'The OA Effect: How Does Open Access Affect the Usage of Scholarly Books?' (Springer-Nature, 2017) <<https://media.springernature.com/full/springer-cms/rest/v1/content/15176744/data/v3>> [accessed 12 May 2019]
- Eve, Martin Paul, *Open Access and the Humanities: Contexts, Controversies and the Future* (Cambridge: Cambridge University Press, 2014) <<http://dx.doi.org/10.1017/CBO9781316161012>>
- Eve, Martin Paul, Kitty Inglis, David Prosser, Lara Speicher, and Graham Stone, 'Cost Estimates of an Open Access Mandate for Monographs in the UK's Third Research Excellence Framework', *Insights*, 30.3 (2017) <<https://doi.org/10.1629/uksg.392>>
- Ferwerda, Eelco, Frances Pinter, and Niels Stern, *A Landscape Study On Open Access And Monographs: Policies, Funding And Publishing In Eight European Countries* (Zenodo, 1 August 2017) <<https://doi.org/10.5281/zenodo.815932>>
- Ferwerda, Eelco, Ronald Snijder, and Janneke Adema, 'OAPEN-NL: A Project Exploring Open Access Monograph Publishing in the Netherlands Final Report', 2013

- <<http://www.oapen.nl/images/attachments/article/58/OAPEN-NL-final-report.pdf>> [accessed 24 March 2014]
- Ferwerda, Eelco, Ronald Snijder, Brigitte Arpagaus, Regula Graf, Daniel Krämer, and Eva Moser, *OAPEN-CH – The Impact Of Open Access On Scientific Monographs In Switzerland. A Project Conducted By The Swiss National Science Foundation (SNSF)* (Zenodo, 23 April 2018) <<https://doi.org/10.5281/zenodo.1220607>>
- Franzen, Martina, Eileen Joy, and Chris Long, *Humane Metrics/Metrics Noir* (Coventry UK: Post Office Press / meson press, 2018) <<https://hcommons.org/deposits/item/hc:19823/>> [accessed 30 May 2019]
- Hammarfelt, Björn, 'Beyond Coverage: Toward a Bibliometrics for the Humanities', in *Research Assessment in the Humanities: Towards Criteria and Procedures*, ed. by Michael Ochsner, Sven E. Hug, and Hans-Dieter Daniel (Cham: Springer International Publishing, 2016), pp. 115–31 <https://doi.org/10.1007/978-3-319-29016-4_10>
- , 'Following the Footnotes: A Bibliometric Analysis of Citation Patterns in Literary Studies' (unpublished Doctoral, Uppsala University, 2012)
- Herb, Ulrich, 'Recommendations, Statements, Declarations And Activities Of Science Policy Actors On Shaping The Scholarly Communication System', *Zenodo*, 2017 <<https://doi.org/10.5281/zenodo.1003229>>
- Higher Education Funding Council for England, 'Knowledge Unlatched Pilot given HEFCE Backing', 2013 <<https://www.hefce.ac.uk/news/newsarchive/2013/news85263.html>> [accessed 21 December 2013]
- HIRMEOS, 'HIRMEOS Workshops on Annotation and Metrics for OA Monographs, 10-11 Jan 2019, Paris', *High Integration of Research Monographs in the European Open Science Infrastructure*, 2019 <<https://www.hirmeos.eu/2018/11/05/hirmeos-workshops-on-annotation-and-metrics-for-oa-monographs-10-11jan-2019-paris/>> [accessed 3 June 2019]
- HIRMEOS, and Ubiquity Press, 'Deliverable D6.1: Metrics Services Specification', *High Integration of Research Monographs in the European Open Science Infrastructure*, 2019 <https://www.hirmeos.eu/wp-content/uploads/2017/11/HI61-Metrics_Service_technical_specification-final.pdf> [accessed 3 June 2019]
- HuMetricsHSS, 'About HuMetricsHSS', *Humane Metrics Initiative*, 2018 <<http://webcache.googleusercontent.com/search?q=cache:A8FHPfPBu8J:humetricshss.org/about/+&cd=1&hl=en&ct=clnk&gl=uk>> [accessed 3 June 2019]
- Jubb, Michael, 'Academic Books and Their Future', 2017 <https://academicbookfuture.files.wordpress.com/2017/06/academic-books-and-their-futures_jubb1.pdf>
- Knowledge Unlatched, 'How It Works', 2013 <<http://www.knowledgeunlatched.org/about/how-it-works/>> [accessed 5 December 2013]
- Konkiel, Stacy, 'Approaches to Creating “Humane” Research Evaluation Metrics for the Humanities', *Insights the UKSG Journal*, 31 (2018) <<https://doi.org/10.1629/uksg.445>>
- Larivière, Vincent, and Cassidy R. Sugimoto, 'Do Authors Comply When Funders Enforce Open Access to Research?', *Nature*, 562.7728 (2018), 483 <<https://doi.org/10.1038/d41586-018-07101-w>>
- Montgomery, Lucy, 'Knowledge Unlatched: A Global Library Consortium Model for Funding Open Access Scholarly Books', *Cultural Science*, 7.2 (2014), 1–66

- Montgomery, Lucy, Neil Saunders, Frances Pinter, and Alkim Ozaygen, 'Exploring Usage of Open Access Books via the JSTOR Platform' (Knowledge Unlatched Research, 2017) <http://kuresearch.org/PDF/jstor_report.pdf>
- Nederhof, Anton J., 'Bibliometric Monitoring of Research Performance in the Social Sciences and the Humanities: A Review', *Scientometrics*, 66.1 (2006), 81–100 <<https://doi.org/10.1007/s11192-006-0007-2>>
- Neylon, Cameron, 'The Complexities of Citation: How Theory Can Support Effective Policy and Implementation', 2016 <<http://repository.jisc.ac.uk/6553/>> [accessed 12 May 2019]
- Neylon, Cameron, Lucy Montgomery, Alkim Ozaygen, Neil Saunders, and Frances Pinter, 'The Visibility of Open Access Monographs in a European Context: Full Report' (Knowledge Unlatched Research, 2018) <<https://hcommons.org/deposits/objects/hc:18270/datastreams/CONTENT/content>>
- OAPEN-UK, 'The Pilot', 2013 <<http://oapen-uk.jiscebooks.org/pilot/>> [accessed 25 March 2014]
- Pinter, Frances, and Christopher Kenneally, 'Publishing Pioneer Seeks Knowledge Unlatched', 2013 <<http://beyondthebookcast.com/transcripts/publishing-pioneer-seeks-knowledge-unlatched/>>
- 'Project COUNTER - Consistent, Credible, Comparable', *Project Counter* <<https://www.projectcounter.org/>> [accessed 12 May 2019]
- Royle, Pamela, and Norman Waugh, 'Bibliometrics of NIHR HTA Monographs and Their Related Journal Articles', *BMJ Open*, 5.2 (2015), e006595 <<https://doi.org/10.1136/bmjopen-2014-006595>>
- Simba Information, 'Open Access Book Publishing 2016-2020', 2016 <<https://www.simbainformation.com/Open-Access-Book-10410716/>> [accessed 12 May 2019]
- Snijder, Ronald, 'Measuring Monographs: A Quantitative Method to Assess Scientific Impact and Societal Relevance', *First Monday*, 18.5 (2013) <<https://firstmonday.org/ojs/index.php/fm/article/view/4250>> [accessed 12 May 2019]
- Springer Nature, 'Open Research Library', *Open Research*, 2019 <<https://www.springernature.com/gp/open-research/journals-books/books/orl>> [accessed 26 May 2019]
- Suber, Peter, *Open Access*, Essential Knowledge Series (Cambridge, MA: MIT Press, 2012) <<http://bit.ly/oa-book>>
- Swan, Alma, 'Open Access Policy Effectiveness: A Briefing Paper for Research Institutions' (Pasteur4OA) <<http://www.pasteur4oa.eu/sites/pasteur4oa/files/resource/Policy%20effectiveness%20-%20institutions%20final.pdf>>
- Sweeney, David, and Ben Johnson, 'Seeking a Fresh Perspective: A Research Funder's View of Open Access', *Insights: The UKSG Journal*, 27.1 (2014), 51–57 <<https://doi.org/10.1629/2048-7754.114>>
- Vincent, Nigel, and Chris Wickham, 'Debating Open Access: Introduction', in *Debating Open Access*, ed. by Nigel Vincent and Chris Wickham (London: British Academy, 2013), pp. 4–12
- Vincent-Lamarre, Philippe, Jade Boivin, Yassine Gargouri, Vincent Lariviere, and Stevan Harnad, 'Estimating Open Access Mandate Effectiveness: The MELIBEA Score', *ArXiv:1410.2926 [Cs]*, 2014 <<http://arxiv.org/abs/1410.2926>> [accessed 12 May 2019]

- Vincent-Lamarre, Philippe, Jade Boivin, Yassine Gargouri, Vincent Larivière, and Stevan Harnad, 'The Effect of Open Access Mandate Strength on Deposit Rate and Latency', 2014 <<http://eprints.soton.ac.uk/366815/>> [accessed 23 July 2014]
- Watkinson, Charles, Kevin Hawkins, Lucy Montgomery, Brian O'Leary, Cameron Neylon, and Katherine Skinner, 'Understanding OA Ebook Usage: Toward a Common Framework' (Knowledge Unlatched Research, 2018)
<https://deepblue.lib.umich.edu/bitstream/handle/2027.42/143840/Redacted%20Grant%20Narrative%20-%20OA%20Ebook%20Usage_FINAL%20SUBMISSION_042718.pdf?sequence=1&isAllowed=y> [accessed 13 May 2019]
- Watkinson, Charles, Rebecca Welzenbach, Eric Hellman, Rupert Gatti, and Kristyn Sonnenberg, 'Mapping the Free Ebook Supply Chain: Final Report to the Andrew W. Mellon Foundation', 2017 <<https://deepblue.lib.umich.edu/handle/2027.42/137638>>
- Xia, Jingfeng, Sarah B. Gilchrist, Nathaniel XP Smith, Justin A. Kingery, Jennifer R. Radecki, Marcia L. Wilhelm, and others, 'A Review of Open Access Self-Archiving Mandate Policies', *Portal: Libraries and the Academy*, 12.1 (2012), 85–102