Rationale

The Need

The A.W Mellon Foundation funded project Developing a Pilot Data Trust for Open Access Ebook Usage (2020-2022, the "OAeBU" project) identified a growing analytics capability gap which places the diversity of the scholarly book publishing system at risk. Ensuring diversity among authors, perspectives, topics, audiences and languages in OA scholarly books depends on a publishing ecosystem that includes both large and small presses, operating in different contexts, with diverse missions and stakeholders. In turn this means that *all* presses need access to data that can support informed strategic decisions in pursuit of their varied missions.

Without this data, small and medium sized presses with fewer resources will simply be left out of a process of digital transition that is reshaping scholarly communication landscapes globally. There is a danger that OA book publishing, which is currently notable for the role that small publishers play, will undergo the extreme concentration that has already occurred in the journal space: becoming dominated by a handful of commercial, multinational players.[1]

A print sale is comparable to a print sale through a different vendor, but digital use is recorded in radically different ways by different platforms. This makes the task of aggregating and interpreting usage data for OA books time consuming and technically complex.

The vast majority of scholarly book publishers lack the technical and staffing capacity to engage with a growing tide of potentially valuable information relating to how OA books are used.[2] This limits publishers' ability to improve the provision of OA books to the communities most likely to benefit from them; as well as to advocate for the value of OA scholarly books to authors and funders.

For diverse presses to develop and advocate for OA books, and to effectively represent and amplify underrepresented and marginalized voices, information on usage needs to be collated and presented in ways that speak to the needs of the many stakeholders engaged in OA book publishing.

Through the OAeBU project we verified a compelling demand for shared services to address this analytics gap. This demand is global in its reach, and while the most immediate calls for support arise in Europe and North America, and from presses that publish primarily in English, we believe it is incumbent on any digital infrastructure project to seek to address wider geographies

[1] For a more extensive discussion of the importance of diversity in the data landscapes that support scholarly communication and research see: Montgomery, L., Neylon, C., Wilson, K., Huang, C-K., Hosking, R., Ozaygen, A., and Handcock, R. (2020). *Who puts the 'open' in open knowledge?* Cultural Science Journal. 12(1), 13–22. https://sciendo.com/article/10.5334/csci.136

[2] For a discussion of the challenges facing small and medium sized presses in relation to usage data; as well as the opportunity and potential benefits that usage data presents see: Montgomery, L. , Neylon, C. , Ozaygen, A. and Leaver, T. (2018), Getting the best out of data for open access monograph presses: A case study of UCL Press. *Learned Publishing*. doi:10.1002/leap.1168

and user bases so as to mitigate the risk of increasing existing and historical inequities in scholarship. The analytics gap is not merely a technical gap, but also one of capacities and skills to interrogate information, including its completeness and quality, in making strategic decisions. We also developed the initial technical infrastructure and workflows needed to underpin these services.

An opportunity now exists to build on the OAeBU project's insights, community, and codebase in order to support the development of a sustainable community infrastructure capable of creating a step-change in the analytics capabilities of small and medium sized OA publishers. Doing so will have a powerful impact on the capacity of these publishers to leverage the potential of OA to address the needs of diverse communities.

Previous Work

The OAeBU project developed pipelines capable of efficiently capturing and bringing together data relating to OA Books. It also implemented structured approaches to the management and aggregation of OA Book data. Together, these approaches make it possible to deliver high-quality data analytics for OA Books, including the creation of interactive visualization dashboards that can be tailored to the needs of individual members of the OA book community One example of a dashboard can be seen live on the University of Michigan Press site.[3]

During the OAeBU project stakeholders in OA book usage data were engaged in an extensive process of consultation and requirements scoping. This included the development of detailed use cases focussing on the needs of commercial OA book publishers;[4] university presses;[5] scholars;[6] libraries[7] and dissemination platforms.[8] An additional process of requirements scoping and prototype development was also undertaken with a smaller group of Dashboard Pilot Partners: the University of Michigan Press, Wits University Press, UCL Press, ANU Press, SpringerNature and OAPEN. By working with Dashboard Pilot Partners the project's data science team was able to gain deep insight into the technical systems, workflows and book related data used by a diverse group of OA book players in a variety of contexts. Collaboration with Dashboard Pilot

[3] Michigan University Press has already chosen to deploy its prototype dashboard on its public website: https://ebc.press.umich.edu/impact/

[4] Drummond, Christina, & Hawkins, Kevin. (2022). Commercial Publisher OA eBook Usage (OAEBU) Data Use Cases. In OA eBook Usage Data Analytics and Reporting Use-Cases by Stakeholder. Zenodo. https://doi.org/10.5281/zenodo.5879727

[5] Drummond, Christina, & Hawkins, Kevin. (2022). University Press OA eBook Usage (OAEBU) Data Use Cases. In OA eBook Usage Data Analytics and Reporting Use-Cases by Stakeholder. Zenodo. https://doi.org/10.5281/zenodo.5879450

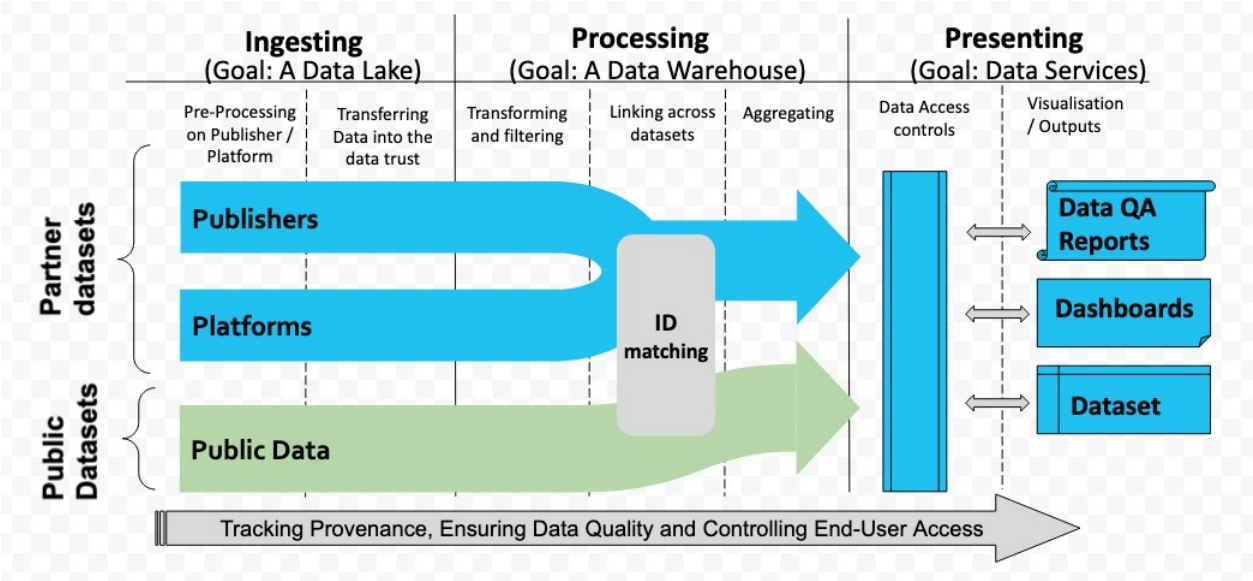
[6] Drummond, Christina, & Hawkins, Kevin. (2022). Scholar Open Access eBook Usage (OAEBU) Data Use Cases. In OA eBook Usage Data Analytics and Reporting Use-Cases by Stakeholder. Zenodo. https://doi.org/10.5281/zenodo.5879666

[7] Drummond, Christina, & Hawkins, Kevin. (2022). Library Open Access eBook Usage (OAEBU) Data Use Cases. In OA eBook Usage Data Analytics and Reporting Use-Cases by Stakeholder. Zenodo. https://doi.org/10.5281/zenodo.5879678

[8] Drummond, Christina. (2020). Engaging Stakeholder Networks to Support Global OA Monograph Usage Analytics. Collaborative Librarianship, 12(2). https://doi.org/10.5281/zenodo.5567206

Partners supported the development of fully operational working prototype dashboards for each partner.

As part of the project the Curtin team developed a scalable technology platform that integrates book usage data to support visualizations and analytics. Data is collected from multiple sources including general bibliographic data (eg Crossref, Microsoft Academic) and usage data from multiple platforms (OAPEN, JSTOR, Google Books) and integrated using publisher data via the book-specific metadata standard, ONIX.



ONIX is a book-industry standard metadata exchange system which is used throughout the book publishing industry to facilitate the exchange of information about books.[9] This technical approach has many advantages in terms of scaling (given common platform usage data a new publisher need only provide their ONIX feed to on-board data), flexibility (the separation of usage data from core metadata allows for complex book-objects and usage analysis, and inclusion (standards-based approaches have allowed for the inclusion of non-Latin scripts in multiple languages).

To our knowledge the system developed during the OAeBU project is the only large-scale open source usage data integration platform that is built from the ground up for books specifically, and which treats books as a first-class object. This is important because systems built to capture information relating to journal articles are not capable of capturing or processing a significant proportion of the data that relates to books.

The difficulties associated with infrastructure that is not specific to books are highlighted in the *OA Book Supply Chain Mapping Report*,[10] which was produced by the OAeBU project. As the

[9] See: Clarke, Michael, & Ricci, Laura. (2021). OA Books Supply Chain Mapping Report. Zenodo. https://doi.org/10.5281/zenodo.4681725

[10] *ibid*

report details, important differences exist between the standards, metadata, persistent identifiers, and information workflows that support the publication of OA books and those associated with journal articles.

Although many journal-focussed systems (eg. OpenAlex, OpenAIRE, Pubmed, Crossref) capture *some* data about *some* books, none are currently able to provide OA book publishers with the accurate, comprehensive, and high-quality data that is needed to support strategic decision making. As a result, OA book publishers generally have little choice but to manually capture and compile usage data relating to their books from a number of different sources as best they can. Almost none have the ability to use data to support deeper knowledge of the ways in which diverse communities engage with this form of publication.

In addition to developing technical infrastructure, the OAeBU project worked closely with stakeholders to gather information about the analytics needs of different types of dashboard users. Community consultation and engagement, led by Educopia, took place in parallel with deep-dive exploration of the technical realities, workflows and analytics requirements of 6 Dashboard Pilot Partners: the University of Michigan Press; Wits University Press; ANU Press; UCL Press; SpringerNature and OAPEN. Each of these Dashboard Pilot Partners is operating in a different geographic context: Africa, Australia, the US and Europe. Dashboard Pilot Partners were also chosen to maximize diversity in scale of operations: SpringerNature has published more than 1000 OA books, while Wits University Press has published just 16; as well as diversity in business models (for profit/not for profit) and institutional context (traditional university press/library based fully OA press/OA book dissemination platform engaging with partners from across the globe).

With the support of a full-time Project Officer, the project team worked with OA book stakeholder communities to explore and document the data related products and services most likely to deliver value to this community; and the capacity and willingness of community members to contribute to the long-term sustainability of an OA eBook usage data aggregation and dashboarding service. This process involved the development of a volunteer advisory board (invited by the project team with the goal of geographic and stakeholder representation) and a set of facilitated focus and interest groups that gained ongoing stakeholder engagement from over 100 people.[11]

Through the consultation processes described above the OAeBU project identified two key areas for further development: a *data exchange* and a *dashboarding service.* The community consultation process also allowed for areas of tension and concern to be explored with the

[11] See https://educopia.org/data_trust/ for a list of advisors and for links to each of the seven stakeholder-specific groups; see also Drummond, Christina (2020) "Engaging Stakeholder Networks to Support Global OA Monograph Usage Analytics" *Collaborative Librarianship*: Vol. 12 : Iss. 2 , Article 9. https://digitalcommons.du.edu/collaborativelibrarianship/vol12/iss2/9/ for a summary of the engagement process.

community. Two key issues identified during the 2018 OA book stakeholder workshop[12] emerged as particularly relevant to planning the next phase of the project. These were:

1. The potential for conflict between the interests of large, commercial, OA book publishers and platforms, and those of small and not for profit presses;
2. The danger that incentives for participating in shared industry standards exercises would be conflated with those of accessing metrics analytics services (such as dashboards).

In the words of one response to the 2018 Data Trust consultation document[13] shared with the community for comment:

"We are concerned that the present proposal for a "Data Trust" integrates the fundamental role of defining and managing industry-wide standards with a variety of data metric services - and as such risks conflating incentives and having a damaging rather than enabling impact on OA publishing in the longer term."[14]

To address these dangers and maximize the positive impacts of future initiatives on the wider OA book publishing landscape, we identified two strands of activity that should be pursued independently.

First, there is a need for a trusted and standards based approach to the sharing of potentially sensitive usage and other proprietary data. This *data exchange* function would provide a trusted mechanism and set of standards, rules, and expectations for the pooling and use of data. It would create incentives for participating by commercial presses and platforms; as well as small and not-for-profit players. The *data exchange* is addressed in a separate grant.

Secondly, in order to ensure that stakeholders are able to act on and use the growing volume of data that is becoming available in relation to OA books, it is also necessary to address the analytics gap. There is a need for a *data analytics service* which can provide translation, comparison and integration of data in a form which is useful for stakeholder decision making.

Demonstrating that data, standards and best practice data handling guidelines provided by the *data exchange* can be used by a community infrastructure, and translated for specific use cases, is also critical to the long-term success and impact of the exchange. Our goal is that the *dashboarding service* will help to lead the way for other services that draw on the *data exchange*

[12] Attendee agenda associated with an invitation-only workshop held in Tribecca, NY on December 3-4 2018. The workshop involved key stakeholders in OA eBooks and formed part of the Andrew W. Mellon Foundation funded project, Understanding OA Ebook Usage: Toward a Common Framework. http://dx.doi.org/10.17613/fpcz-gp24

[13] Neylon et al (2018), Building a Trusted Framework for Coordinating OA Monograph Usage Data [online] https://docs.google.com/document/d/1PjeRaz6XRvOiGY2GsmY7J5LHUqwFiUI9v-iC9ZsZaj0/edit#heading=h.wiqkn3xdk73u

[14] Gatti et al (2018) Thoughts on "Building a Trusted Framework for Coordinating OA Monograph Usage Data" [online] https://docs.google.com/document/d/1bk8ljGq0ZWHq3HvuVXfkvZTcdEUJuxS3QvzVgsiKXKw/edit#

to meet the needs of diverse OA publishing communities. **This document focuses on the data analytics service.**

The decision to split the *data exchange* and *dashboarding service* efforts was validated via the OAeBU project board, and further developed through a round of mission-vision-values work with invited stakeholders which defined the specific requirements for each strand of work.

The OAeBU community and project team worked together to articulate a shared Guiding Principles[15] statement to guide both the data exchange and data analytics service. The community and project team then separated into two task-force groups, one focused on the data exchange[16] and the other on the data analytics service,[17] and each developed and formalized a mission and vision statement, as well as initial service plans and impact models, for these as separate entities. This work has positioned both the data exchange and the data analytics service to now pursue specific aims and engage in new partnerships in the next phase of development work.

The OAeBU project has developed in close and conscious alignment with other key initiatives underway globally in scholarly communication data analytics, including those involved with standards, tools, and aggregation/workflow development. In particular, the OAeBU project has complemented and resonated with OAPEN's focus on delivering better information about the uses of OA books to a diverse and highly international stakeholder community. OAPEN (short for Open Access Publishing in European Networks) was from 2008-2010 an EU co-funded project initiated by six European university presses.[18] From 2010 OAPEN became a non-profit Dutch foundation. It now hosts, distributes, and preserves books from over 370 international OA academic publishers via the OAPEN Library.

OAPEN has a long history of working with publishers of all types and sizes and a track-record of successfully identifying the means to provide affordable infrastructures that can support diversity in OA book publishing landscapes.[19] This includes finding approaches that can support publishers operating at a range of scales, in different national contexts and languages (including non-Latin scripts), and with diverse missions. OAPEN's success in supporting diversity in OA book publishing has been achieved through a focus on providing user friendly services that build on established book-specific standards and infrastructures. This approach makes it possible for diverse OA book publishers to engage with technologies and capacities that would otherwise be the preserve of the largest players.

---

[15] Kemp, Jennifer, Potter, Peter, Scrivener, Brian, Watkinson, Charles, & Skinner, Katherine. (2021). Guiding Principles for OA Book Usage Data Services. Zenodo. https://doi.org/10.5281/zenodo.5842414

[16] Drummond, Christina, Elwell, Jon, Emery, Jill, Hawkins, Kevin, Kemp, Jennifer, O'Leary, Brian, Skinner, Katherine, & Watkinson, Charles. (2021). OA eBook Usage (OAeBU) Data Trust Impact-driven Service Model. Zenodo. https://doi.org/10.5281/zenodo.5579257

[17] Drummond, Christina, Kroes, Kara, Montgomery, Lucy, Neylon, Cameron, Potter, Peter, Scrivener, Brian, Skinner, Katherine, Snijder, Ronald, Queen, Wendy, Watkinson, Charles, & Welzenbach, Rebecca. (2021). (SAMPLE) OA eBook Usage Data Dashboarding Service Model. Zenodo. https://doi.org/10.5281/zenodo.5567231

[18] Further details of the history of OAPEN can be found in this personal tale "A brief saga of open access books"

[19] A recent example is Ferwerda, E., Mosterd, T., Snijder, R., & Mounier, P. (2021). *UKRI Gap Analysis of Open Access Monographs Infrastructure*. UKRI. https://doi.org/10.5281/zenodo.5735021

In addition to providing a high quality, cost-effective platform for the dissemination and preservation of OA books, OAPEN is a key player in the development of OA book discovery and visibility infrastructures. The Directory of Open Access Books (DOAB), hosted by OAPEN, is one example. Another is the long-standing collaboration between OAPEN and IRUS-UK (Institutional Repository Usage Statistics, United Kingdom). IRUS-UK is a service for capturing and processing institutional repository usage data, making it possible for institutional repositories to generate COUNTER compliant usage data. In 2013 OAPEN entered into a groundbreaking collaboration with IRUS-UK: developing new workflows and processes to ensure that OAPEN could provide COUNTER compliant usage reporting to its members.

OAPEN's collaboration with IRUS-UK was, in turn, leveraged during the OAeBU project. The IRUS-UK and OAPEN collaboration made it possible for OAPEN to provide the OAeBU project with consistent and scalable COUNTER data. The availability of this high-quality usage data played an important role in the OAeBU project's ability to develop efficient down-stream data workflows and visualizations for OAeBU Dashboard Pilot Partners[20] that participated in the project.

The OAPEN Foundation is a not-for-profit organization based at the National Library in the Hague, Netherlands. The organization is dedicated to supporting peer reviewed OA books; and operates 3 internationally significant services: 1. The OAPEN Library: a central repository for hosting and disseminating OA books; 2. The OAPEN OA Books Toolkit; and 3. The Directory of Open Access Books (DOAB): A discovery and indexing service for OA books.

The OAPEN Library and DOAB both operate on open source DSpace 6 platforms allowing for seamless integration with research library catalogs (according to our data more than 1,200 libraries worldwide integrate with the OAPEN Library). With monthly OAPEN usage of around 1 million COUNTER compliant downloads, publishers acknowledge OAPEN as a central and international publication platform for books that keeps growing with new publishers joining regularly. Furthermore, OAPEN provides services for a number of research funders and initiatives like the CERN SCOAP3 for books and ScholarLed. More information is provided in the 2020 annual stakeholder report.

OAPEN's commitment to developing infrastructures capable of supporting diversity in OA book publishing, as well as their experience developing and implementing financial sustainability and governance models that are inclusive and transparent, make them a natural partner for this follow-on project.

The proposed project *OA Book Usage Analytics for Diverse Communities - A Demonstration* builds on all of this work in order to deliver and scale a sustainable OA Book focused analytics service capable of supporting continued diversity, experimentation, and innovation in OA Book publishing. It invites OAeBU project (2020-2022) partners to be among the first users of a dashboarding service; scales workflows, infrastructure and customer support in collaboration

[20] OAeBU Dashboard Pilot Partners comprised: OAPEN, Springer-Nature, University of Michigan Press, ANU Press, UCL Press, and Wits University Press. See: https://educopia.org/data_trust/

with community engagement partners; and develops a long-term plan for housing, maintenance and funding of the analytics service as sustainable community infrastructure.

What needs to be done?

The OAeBU project developed a detailed set of user stories and requirements for varying stakeholders, and has published these via a 2021 report, OA eBook Usage Data Analytics and Reporting Use-cases by Stakeholder.[21] Reliable, relevant, and transparently handled usage data is needed to support a diverse OA book publishing ecosystem. Examples of the types of data that publishers would like to access include information about the number of times a book has been downloaded; the countries and cities readers are located in; and whether readers are accessing books via an institutional library system or the open web. If they are accessing books via an institution, is it a university library that has helped to pay the costs of publishing a book in open access? Are there particular platforms that are more likely to help books to connect with readers in specific locations. For example, JSTOR may be very heavily used by academic communities in North America; but Worldreader might be a dissemination channel more likely to connect with readers in countries and communities that have low bandwidth and where mobile phones are the main way that people access the internet. Information about the timing of usage can also be helpful: allowing publishers, libraries and authors to understand the factors that help drive readership, as well as synergies between different open scholarship initiatives. Examples might include usage of an OA book that spikes when a MOOC based on its content is launched; or when a particular community is in urgent need of information about the topic that it addresses.

OA book publishers need access to usage data that is comprehensive, accurate and trustworthy so that they are able to make better informed choices about the books that they choose to publish; as well as the dissemination channels and strategies that they use to make their books visible and accessible to the communities that they serve. Funders and libraries need data that can support decisions about how to invest limited resources, as well as the evaluation of investments in the context of an organization's mission. All stakeholders need data that can support conversations with authors, funders, institutions, and communities about the benefits of investments in OA for scholarly books, including greater usage, and wider reach.[22]

The costs, including time and expertise, of collating and managing data related to the use of scholarly books means that currently only the very largest publishers are able to gather, integrate, analyze and interpret usage data to support strategic decision making. Smaller players are dependent on data from providers and intermediaries who have different, and sometimes conflicting interests, including aggregators and supply chain actors who are paid based on volume, platforms that seek to maximize apparent usage at the cost of accuracy, and peer publishers that may inflate usage for promotional purposes. Mission driven publishers, those that are less resourced and those from underrepresented regions or publishing in underrepresented

[21] Drummond, Christina, & Hawkins, Kevin. (2021). OA eBook Usage Data Analytics and Reporting Use-cases by Stakeholder (Draft). Zenodo. https://doi.org/10.5281/zenodo.5572841

[22] Neylon, Cameron, Alkim Ozaygen, Lucy Montgomery, Chun-Kai (Karl) Huang, Ros Pyne, Mithu Lucraft, and Christina Emery. 2021. "More Readers in More Places: The Benefits of Open Access for Scholarly Books". *Insights* 34 (1): 27. DOI: http://doi.org/10.1629/uksg.558, see also the Use Cases documented above

languages are most at risk from being unable to identify opportunities or maintain their position and unique characteristics. Scholarly hegemonies are built on self-reinforcing power structures that align with capital and historical advantage. Supporting the power of underrepresented groups to make their own choices, about what scholarship matters, what questions can be asked, what are the important qualities of scholarly work and how it is represented, distributed and critiqued is not merely a market intervention but also an attention to reparative justice in scholarship.

As an example, incomplete and non-comparable data is currently contributing to a lack of transparency among the platforms distributing OA books. This makes it impossible for small and medium sized OA book publishers to make appropriately informed choices about the distribution platforms they use. It also creates risks that high quality platforms will be penalized for seeking to provide the cleanest possible data[23] (Snijder, 2021), at the cost of reducing apparent usage counts. High quality information and its use requires scale and capacities that are unavailable to smaller and less well-resourced presses.

This combination of interlinked technical and community trust challenges makes a community-focused, shared service model an attractive solution. Such a solution will need to:

- Be trusted and trustworthy for presses, providing reliable, timely and audited data, alongside well-designed visualization and analysis tools that are usable for presses, trusted and seen to support good decisions;
- Be responsive to and working in the interests of the diverse range of publishers, with a remit to provide and demand best practice across that community in terms of data comparisons, quality and validity;
- Have a sustainability model that provides confidence that services and systems will remain available, and therefore represent a useful investment of time and effort for users, while also being affordable for a diverse range of publishers.

What resources are required?

The resources required to create a service that is financially and technically sustainable are two-fold. From a technical perspective it is necessary to undertake standardization in order to ensure that onboarding new users and partners at scale is both possible and practical. Ensuring that technical systems also provide the flexibility to support a diversity of potential users will be key to maximizing the positive impact of the technical infrastructure on the OA book publishing ecosystem. From a management perspective, customer relationship and community engagement systems also need to be shifted towards a footing that is both professional and scalable. In order to ensure long-term financial sustainability, working with OA book stakeholders to develop a financial model that strikes a balance between the need for partner contributions; and the importance of affordability for small and medium sized university presses is important.

[23] Snijder, R. (2021). Open access book usage data – how close is COUNTER to the other kind? Insights, 34(1), 9. https://doi.org/10.1629/uksg.539

A grant of $AUD1.05M over 3-years provided by the Andrew W. Mellon Foundation will allow this project to move beyond proof of concept and towards becoming a financially viable service embedded within a larger ecosystem of infrastructures that support OA book publishing. During the first year the budget incorporates funding from COARD (Collaborative Open Access Research and Development previously known as Knowledge Unlatched Research), a not-for-profit organization that partnered on the first two phases of this project and which Montgomery and Neylon lead. The revenues generated through COARD's work as a research consultancy and as a provider of tools an analysis relating to OA book usage[24] have generated an operating surplus that COARD will re-invest as a contribution to the next phase of this project. This is consistent with COARD's goal of creating tools and services for understanding data related to OA books in order to support and sustain a diversity in OA scholarly communication landscapes.[25]

$502,286 is requested from the Foundation in Year 1 tapering to $363,231 in Year 2 and $183,836 in Year 3. Additional funds will be provided by COARD in Year 1 ($200k). In years 2 and 3 user contributions will be invited, and additional funding opportunities will be explored, informing a long-term sustainability plan for the project.

In Year 1: Resources will be mostly focused on technical developments to support scaling, legal support to ensure that data agreements are robust and initial work on grounding and scaling community engagement.

In Year 2: The focus moves to expanding the user community, and continuing to work with that community to co-create a sustainability model. Technical work will focus on addressing any specific technical issues identified in Year 1, with priority given to ensuring that capturing data for new partners and generating larger numbers of dashboards can be scaled smoothly.

In Year 3: Resourcing is focused on scaling up and diversifying the user community, including potential engagement with wider user groups (funders and libraries). There is also a focus on the development of a long-term plan for financial sustainability; as well as the transition of service maintenance and management to OAPEN, which is anticipated to function as the long term organizational host.

How will results be made available to the community?

This project, like the initial OAeBU project, will be grounded in and responsive to the community it serves. Our team works to include a broad range of stakeholder voices, not just as potential clients but as connected partners and active voices in the work ahead. The work undertaken in the initial project included ongoing two-way communications between the project team and seven stakeholder working groups, in addition to the 19-member Advisory Board,

[24] See for example the two packages of work described in Montgomery et al. (2018) Getting the best out of data for open access monograph presses: A case study of UCL Press, Learned Publishing, 31: 335-344. https://doi.org/10.1002/leap.1168 ; Neylon et al. (2021) More Readers in More Places: The Benefits of Open Access for Scholarly Books. Insights 34 (1): 27. DOI: http://doi.org/10.1629/uksg.558

[25] See: https://coard.community/about

seven-member Technical Advisor group, and the six Dashboard Pilot Partners. We begin the second project with more than 100 invested stakeholders, representing the interests of libraries, authors and scholars, university presses, library publishers, societies and associations, commercial publishers, and publishing platforms, and we will continue to build upon these relationships via structured communication pathways that include working and advisory groups.

In alignment with the Guiding Principles[26] that drive our work, and the mission and vision established by the OAeBU community in 2021, this project will develop with these stakeholders a robust, user-governed and community-focused infrastructure that maintains cost and pricing transparency and demonstrates its affordability for diverse publishers. In doing so it will help to ensure that publishers are able to continue supporting author communities; and that important processes of business model experimentation, innovation and development are able to continue (see e.g., the OA Books Supply Chain Mapping Report).[27] The continued business model development will be undertaken in lockstep with stakeholder representatives.

In addition to the organizational and governance work, the code and workflows developed by the project will be made open source and will be accompanied by high quality technical documentation. The Curtin team has a strong record on software release and documentation as demonstrated through our code release,[28] associated documentation,[29] and connection to publications.[30] This will help to ensure technical accountability, as well as the ability of a wider open research community to build on and apply the work of the project in diverse local contexts.

Project Organization

The project will be led from Curtin University, which will hold fiscal and final reporting responsibility. The project team leads (Montgomery, Neylon, Stern, Snijder, Skinner) will meet regularly (generally every two weeks) via video call to manage coordination and progress. The collaborating organizations will have budgetary responsibility within their agreed budget envelope. The project will require the coordination of work across all three collaborating organizations as well as the engagement of the participating publishers and other contributing organizations. Achieving this will require a tightly integrated team and means of operating. This will be made easier by the fact that all of the organizations involved in the next phase of this project worked together on the OAeBU pilot; and have a strong sense of mutual trust, as well as established approaches to working together across geographies and time zones.

A specific challenge with mapping the detailed responsibilities through this project is that the process of establishing an infrastructure that is sustainable in the long-term requires a project in

[26] Kemp, Jennifer, Potter, Peter, Scrivener, Brian, Watkinson, Charles, & Skinner, Katherine. (2021). Guiding Principles for OA Book Usage Data Services. Zenodo. https://doi.org/10.5281/zenodo.5842414
[27] Clarke, Michael, & Ricci, Laura. (2021). OA Books Supply Chain Mapping Report. Zenodo. https://doi.org/10.5281/zenodo.4681725
[28] See for example repositories at https://github.com/The-Academic-Observatory/ and https://github.com/Curtin-Open-Knowledge-Initiative/
[29] See documentation of code at ReadTheDocs e.g. https://oaebu-workflows.readthedocs.io/en/latest/ and https://github.com/The-Academic-Observatory/academic-observatory-workflows
[30] See for example resources associated with Neylon et al 2021 - https://doi.org/10.5281/zenodo.4014904, Huang et al 2021 - https://doi.org/10.5281/zenodo.4040403

which long-term maintenance and management move from within a University to a long-term hosting organization. The 3 partners involved in this project will thus need to work together to manage the appropriate transfer of responsibility. In order to achieve this, we have developed 3 broad types of work packages: 1) Work packages which involve investigating models (for revenue, governance, and technical prioritization) which then move into implementation; 2) Work packages which involve technical research and development which then needs to transition into a maintenance mode; and 3) Transition of overall hosting and management of this project to a long-term organizational host, with appropriate governance structures and community oversight.

A PI and Partner Organisation are nominated as the Lead for each Work Package. Success will require close coordination across the PI team and active collaboration between partner organizations throughout. In order to facilitate this, almost all work packages will actively involve the whole collaborating team. The PI and Partner Organization nominated as the Lead for each work package will take responsibility for coordinating activities and ensuring that the goals of the work package are achieved. Other PIs and Partner Organizations will be closely involved in each work package in collaborating and supporting roles.

Curtin:

Montgomery (Lead PI) will have management responsibility for the project as a whole, and will hold the overall roadmap and strategic plan. **Neylon** (CI) will take a strategic lead on technical issues and be responsible for the technical roadmap. Kathryn **Napier** (Curtin-based Lead Data Scientist) will continue to have operational management responsibilities throughout the project. The listed **Software Engineer** (to be appointed) will focus on code development and maintenance, specifically the development of software systems providing pipelines for the ingest of metadata and usage data. The **Data Scientist** (to be appointed, we have a candidate in post currently funded on another project) will focus on dashboard development and deployment as well as user analytics and reporting. The project will also be able to draw on the expertise of the Curtin Institute for Computation,[31] which Neylon and Montgomery are both members of.

The Curtin Institute for Computation (CIC) employs a substantial pool of data scientists and software developers in order to allow for flexible support for research projects across the university. This ensures that individual projects benefit from the combined expertise of a dynamic, interdisciplinary, and highly skilled group of data scientists, and that the need for career development opportunities and job stability can be addressed in a context in which individual projects are generally supported by 'soft money' and the uncertainty that surrounds it.

Educopia:

Skinner (CI) will lead on community cultivation and facilitation and be deeply involved in the development and community testing of the governance, sustainability, and pricing models. This work will include facilitating the processes needed to develop a robust, user governed and community-focused governance model that provides appropriate and well-documented mechanisms and processes for decision-making and oversight. Skinner will also provide a range

[31] See: https://computation.curtin.edu.au/

of "Validation" lifecycle stage tools and facilitated work sessions to assist the dashboard service in 1) building and refining its financial plan and organizational model, 2) developing its initial leadership group and governance procedures, 3) structuring and refining community input to guide formalization of data sharing and data transfer agreements, and 4) building and refining its communications plan and relationship management infrastructure. Skinner will contribute to the health and stability of this venture, ensuring that its organizational and business models adhere to the "Guiding Principles" and serve the community's mission and vision.

OAPEN:

Stern (PI) will take the strategic lead on community engagement, based on the experience of OAPEN in running a service for a diversity of publishers and other stakeholders, as well as having responsibility for development of the onboarding workflows and opportunities for shared systems for the service within OAPEN. **Snijder** (PI) will have responsibility for the technical implementation of onboarding processes, quality assurance and integration of the demonstration project systems with OAPEN systems where appropriate. Snijder will also contribute significantly to design and implementation of technical workflows by the Curtin team, providing an external technical perspective based on their extensive experience. Additional resources at OAPEN will be dedicated to specific development tasks involved in integration and to management of the community engagement process including customer support when the service is launched.

Partner Publishers:

As partners are onboarded, both those engaged in the existing pilot and new cohorts, we will refine the requirements and expectations for engagement, as well as the service guarantees that can be provided at each stage. Specifically, partners will have the responsibility to audit and agree to data transfer agreements, to provide data to a specified quality level and to engage with the developing governance and user requirements systems. Partners will also be invited and encouraged to participate in the development of financial sustainability models, including through contributing financially to the further developing and sustainability of the service.

Technology

Within the Demonstration Project we will make full use of the technical capability that was developed within the OAeBU Pilot Dashboard project. This documented[32] open-source code and framework[33] is capable of gathering, integrating and visualizing ebook usage data and has been tested and piloted with the current Dashboard Pilot Partners. This has the joint benefit of providing for future compatibility with the partner Data Trust project as well as reducing the risk of technical lock-in for dashboard users.

The technology stack uses the book-industry standard metadata interchange standard ONIX, combined with open bibliographic metadata (Crossref, MAG – which will move to OpenAlex –

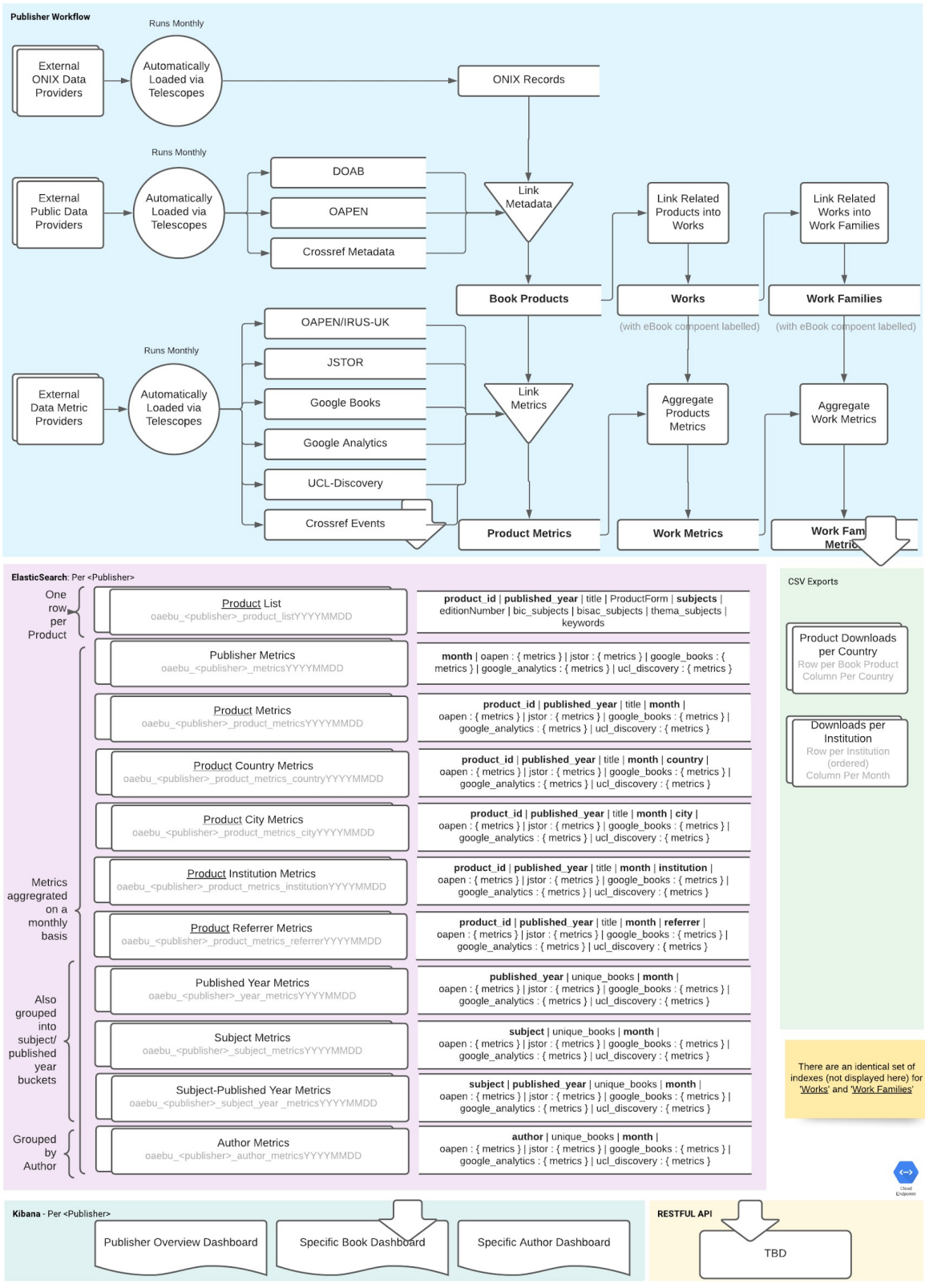
[32] OAeBU Team (2018-2022) OAeBU Workflows Code Documentation, ReadTheDocs, https://oaebu-workflows.readthedocs.io/en/latest/

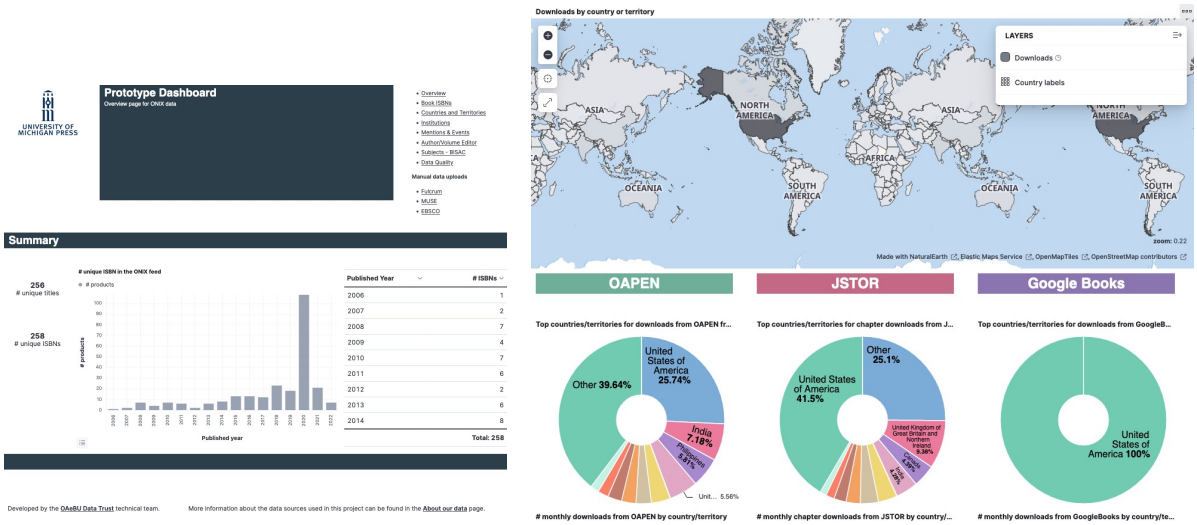
[33] OAeBU Team (2018-2022) OAeBU Workflows Code Repository, github, https://github.com/The-Academic-Observatory/oaebu-workflows

Directory of Open Access Books and OAPEN) to integrate usage data from multiple sources. Data integration through the OAeBU Workflows code base is built on an open source workflow system[34] written in Python and built on AirFlow which ingests and combines data into a cloud database (Google BigQuery) which uses standard SQL for data operations. The OAeBU workflows disambiguate and combine data from multiple sources (e.g. OAPEN, COUNTER, JStor, Google Books) which may use differing identifier systems.

Data is pushed from the cloud datastore into ElasticSearch where it is accessible to Kibana, providing an open source data analytics and dashboarding system. Dashboards in Kibana can be developed through a visual interface in collaboration with users, and are specified in JSON so they can be maintained and versioned in a code repository. Data access limitations flow through from the underlying sources into the cloud database and Elastic/Kibana where stakeholder data is sandboxed into separate areas with no cross access, providing strong security and privacy.

[34] COKI Team (2017-2022) Observatory Platform, github, https://github.com/The-Academic-Observatory/observatory-platform see also documentation, https://docs.observatory.academy/en/latest/


Publisher Workflow
Runs Monthly
External ONIX Data Providers
Automatically Loaded via Telescopes
ONIX Records
Runs Monthly
External Public Data Providers
Automatically Loaded via Telescopes
DOAB
OAPEN
Crossref Metadata
Link Metadata
Link Related Products into Works
Link Related Works into Work Families
Book Products
Works
Work Families
(with eBook compoent labelled)
(with eBook compoent labelled)
Runs Monthly
External Data Metric Providers
Automatically Loaded via Telescopes
OAPEN/IRUS-UK
JSTOR
Google Books
Google Analytics
UCL-Discovery
Crossref Events
Link Metrics
Aggregate Products Metrics
Aggregate Work Metrics
Product Metrics
Work Metrics
Work Fam
Metric
ElasticSearch: Per <Publisher>
One row per Product
Product List
oaebu_<publisher>_product_listYYYYMMDD
product_id | published_year | title | ProductForm | subjects | editionNumber | bic_subjects | bisac_subjects | thema_subjects | keywords
Metrics aggregrated on a monthly basis
Publisher Metrics
oaebu_<publisher>_metricsYYYYMMDD
month | oapen : { metrics } | jstor : { metrics } | google_books : { metrics } | google_analytics : { metrics } | ucl_discovery : { metrics }
Product Metrics
oaebu_<publisher>_product_metricsYYYYMMDD
product_id | published_year | title | month | oapen : { metrics } | jstor : { metrics } | google_books : { metrics } | google_analytics : { metrics } | ucl_discovery : { metrics }
Product Country Metrics
oaebu_<publisher>_product_metrics_countryYYYYMMDD
product_id | published_year | title | month | country | oapen : { metrics } | jstor : { metrics } | google_books : { metrics } | google_analytics : { metrics } | ucl_discovery : { metrics }
Product City Metrics
oaebu_<publisher>_product_metrics_cityYYYYMMDD
product_id | published_year | title | month | city | oapen : { metrics } | jstor : { metrics } | google_books : { metrics } | google_analytics : { metrics } | ucl_discovery : { metrics }
Product Institution Metrics
oaebu_<publisher>_product_metrics_institutionYYYYMMDD
product_id | published_year | title | month | institution | oapen : { metrics } | jstor : { metrics } | google_books : { metrics } | google_analytics : { metrics } | ucl_discovery : { metrics }
Product Referrer Metrics
oaebu_<publisher>_product_metrics_referrerYYYYMMDD
product_id | published_year | title | month | referrer | oapen : { metrics } | jstor : { metrics } | google_books : { metrics } | google_analytics : { metrics } | ucl_discovery : { metrics }
Also grouped into subject/ published year buckets
Published Year Metrics
oaebu_<publisher>_year_metricsYYYYMMDD
published_year | unique_books | month | oapen : { metrics } | jstor : { metrics } | google_books : { metrics } | google_analytics : { metrics } | ucl_discovery : { metrics }
Subject Metrics
oaebu_<publisher>_subject_metricsYYYYMMDD
subject | unique_books | month | oapen : { metrics } | jstor : { metrics } | google_books : { metrics } | google_analytics : { metrics } | ucl_discovery : { metrics }
Subject-Published Year Metrics
oaebu_<publisher>_subject_year _metricsYYYYMMDD
subject | published_year | unique_books | month | oapen : { metrics } | jstor : { metrics } | google_books : { metrics } | google_analytics : { metrics } | ucl_discovery : { metrics }
Grouped by Author
Author Metrics
oaebu_<publisher>_author_metricsYYYYMMDD
author | unique_books | month | oapen : { metrics } | jstor : { metrics } | google_books : { metrics } | google_analytics : { metrics } | ucl_discovery : { metrics }
CSV Exports
Product Downloads per Country
Row per Book Product
Column Per Country
Downloads per Institution
Row per Institution (ordered)
Column Per Month
There are an identical set of indexes (not displayed here) for 'Works' and 'Work Families'
Kibana - Per <Publisher>
Publisher Overview Dashboard
Specific Book Dashboard
Specific Author Dashboard
RESTFUL API
TBD



A live example of an existing (prototype) [dashboard can be seen on the University of Michigan Press website] (see the embedded dashboard under 'Open Access Book Usage').

The entire stack from workflow system, through data ingest, database queries, data movement, and dashboard specification is maintained as an open source codebase under an MIT License (https://github.com/The-Academic-Observatory/oaebu-workflows) and documented through the industry standard ReadtheDocs (https://oaebu-workflows.readthedocs.io/en/latest/).

To our knowledge this system is the only large-scale data integration system that provides scalable, open-source technology that is designed specifically for books and focuses on book industry standards such as ONIX. It is built on high-quality underlying technologies and standards (i.e. UTF-8 compliant technologies including SQL, Airflow, Python, ElasticSearch etc) that are capable of supporting multiple languages, including non-Latin scripts such as Chinese characters[35]. Because the system focuses on provision of metadata from publishers, and using standards built for books, it is not limited by indexing or other visibility issues that are often problematic for scholarly books.

Major Activities

Project activities are organized into 10 Work Packages covering technical, engagement and sustainability work. Across the work packages there is a core focus for each year of the project on standardization, scaling, and consolidation, as described below.

Year 1 Standardization and Business Scaffolding: The focus of the first year of the project will be on the standardization of technical workflows, the formalization of processes for prioritizing technical development goals in collaboration with users, and the refinement of legal arrangements necessary to support data management by an analytics service. The Curtin team will run and maintain the existing service stack (data ingest and integration workflows, cloud

[35] From a technical perspective the workflow system requires an ONIX feed with standards-compliant character encoding declared, in which book products are identified by ISBN13.

data store in BigQuery and transfer and visualization via ElasticSearch/Kibana), based on the OAeBU technology for the course of the current project. A Technical Development Roadmap based on community consultation undertaken during the OAeBU project will be shared as an openly accessible document with public commenting facilities and refined in collaboration with partners. Initial implementation of the Technical Development Roadmap will be carried out in order to support the first in a series of annual functionality releases. This will involve moving from our current continuous development process to defining specific functionality for each release.

The first year will also include development of critical business scaffolding that will support scaling up of this initiative in year 2. This work centers on implementation plans for governance and community engagement, building on the work conducted within the OAeBU project. In Year 1 this will include formation of a Project Advisory Board[36], an evaluation of options for organizational form[37], and the development of the Organizational Form Implementation Plan for implementation over Years 2-3. Aligned with this will be a work package clearly defining the value proposition to potential service users that will elaborate on the initial mission, vision, and service and impact model developed in 2021, and will guide our ongoing work and communications with these potential users. While our core users at launch will be small and medium-sized publishers and their associations specifically, we will also explore the potential value of our services to academic libraries and research funders. The work package will be crucial for identifying ways to avert the risks involved in sustainability while deploying pricing models that ensure diversity and inclusion. Sustainability and a sound financial footing are crucial to support diversity and inclusion amongst users and membership. The plan developed through consultation with existing and potential partners will propose a model or models including details around membership, sponsorship, consulting, and other revenue/relationship streams.

Year 2 Scaling: Year 2 will focus primarily on introducing the service to user communities through the networks we have been developing (see Drummond, 2020);[38] and on developing and testing the business models developed in Year 1 (see e.g., the data use cases series published January 2022[39] and the mission, vision, and impact and service model developed in 2021).[40] The

[36] See pp 198-199 in Drummond, "Engaging Stakeholder Networks" https://digitalcommons.du.edu/cgi/viewcontent.cgi?article=1467&context=collaborativelibrarianship for a summary of this process in our 2019 and 2020-21 projects. As previously, we will select individuals based on competencies, skills, stakeholder identities, and geographies to ensure a deliberate cross-section of knowledge and perspectives.
[37] This process will involve our Advisory Board and will be structured to help to guarantee fit-for-purpose. We have done extensive research already (forthcoming, Skinner and Drummond, Scholarly Kitchen 2022) and will build on the knowledge and relationships we have built to date both within and beyond this project.
[38] Drummond, Christina (2020) "Engaging Stakeholder Networks to Support Global OA Monograph Usage Analytics," Collaborative Librarianship: Vol. 12 : Iss. 2 , Article 9. Available at: https://digitalcommons.du.edu/collaborativelibrarianship/vol12/iss2/9
[39] All of the use case documents are available via the OAeBU 'community' collection in Zenodo: https://zenodo.org/communities/oaebu/?page=1&size=20
[40] Drummond, Christina, Kroes, Kara, Montgomery, Lucy, Neylon, Cameron, Potter, Peter, Scrivener, Brian, Skinner, Katherine, Snijder, Ronald, Queen, Wendy, Watkinson, Charles, & Welzenbach, Rebecca. (2021). (SAMPLE) OA eBook Usage Data Dashboarding Service Model. Zenodo. https://doi.org/10.5281/zenodo.5567231

precise form of this will depend on the experience of the initial Year 1 cohort.[41] Our expected plan is to conduct an outreach campaign and solicit responses from a second cohort of publishers through an Expression of Interest process and selected qualified presses will then be brought on board to test and refine the onboarding process. An alternative is to move directly to continuous onboarding. We expect initial responses to be predominantly from Europe and North America. Year 2 will therefore also include development of a focused strategy for connecting with broader networks (eg OAPEN, ScELO Books) to target geographic and linguistic inclusion.

As the number of institutions engaging with the service grows we will test technical systems and processes as well as refine the community oversight functions for larger numbers and wider engagement. This will feed into the development of a long-term sustainability plan for the project in Year 3. Following the technical prioritization and user-needs processes implemented in Year 1 we will shift technical development towards a more stable and less development focussed technical phase, which will be focussed on ensuring the stability of existing systems. This will be the focus of the second annual update.

Year 3 Consolidation: Year 3 will involve a further round of scaling and outreach to grow the uptake of the service towards identifying a financially sustainable service model; as well as exploring options for additional funding. Year 3 focuses on a further cycle of outreach, the implementation of the Organizational Form Roadmap and consolidation of onboarding and technical development processes onto a steady-state footing to allow additional refinement of the sustainability model before moving into the next phase from April 2025.

While the technical provision will be run out of Curtin over the course of the project, a university is not likely to be the best long-term host for the service. There are a number of reasons for this with the most important being that most universities are unable to support the accumulation of capital reserves within a project, can be slow at issuing contracts and invoices, and create unnecessary management overheads for small, service-oriented initiatives. OAPEN is well placed to offer a long-term home for the service and is willing to take carriage of the service, if this is an outcome that the stakeholder community is happy with. The Organizational Form Implementation Plan will provide an opportunity to validate a decision about the long-term organizational host both within the project and with the broader interested community.

Work Packages

WP1: Technical Roadmap Consultation: Developing a Process and Model

April - December 2022
Lead: Curtin
Involved: All collaborators

[41] The year 1 cohort will consist of six small to medium sized university presses. Interest in participating in the next stage of the dashboarding service has been expressed by Michigan University Press; UCL Press; ANU Press; UNC Press. Early conversations are now underway with a number of other small and medium sized presses, in Europe, including CEU Press and members of the OPERAS network.

A draft technical roadmap will be created and shared as a commentable public document with engaged publishers and the wider community groups assembled through the OAeBU project, and feedback from these stakeholders will be invited. Following a feedback and community consultation process, technical priorities for the 3-year project period will be finalized and made public using our Zenodo repository and announced through both project and community listservs. We will also develop and announce an ongoing process by which the team will continue to engage the community with its technical roadmap from 2025 forward.

This WP will underpin core technical work throughout the 3-year project and signal the project's commitment to transparency and accountability in its technical development processes. These processes will give the community a formal record of the development priorities and commitments made over time. It will draw on community consultation and technical development work carried out during the previous project (Developing a Pilot Data Trust for OA eBook Usage) in order to support the articulation of technical development priorities during the current Demonstration Project. We will use this time to also formalize an ongoing public technical roadmap and user feedback process that will be refined as needed throughout the scaling up of this service. Implementing these processes in detail will move the project from signaling accountability through publishing the roadmap through to demonstrating it via the delivery and community modification of that roadmap over time.

WP2: Workflow Standardization

April 2022 - January 2023
Lead: Curtin
Involved: OAPEN

This work package focuses on an initial standardization of the process of bringing on new partners. It will develop and implement an updated Data Sharing Agreement with existing partners and refine processes for onboarding of presses. It will ensure that the technical and legal capacity needed to scale new members in WP3, WP7 and WP9 are in place.

The previous project developed and piloted code, workflows, and data sharing agreements to support an OA book focused analytics service. Curtin will continue to maintain and run the existing service stack during the course of the current project. In order to scale these systems and create a sustainable OA book analytics service capable of operating at scale it will be necessary to develop standard approaches to the legal and technical aspects of bringing new partners into the system.

During the previous project significant time was spent tailoring data sharing agreements to address the concerns of individual partners. In order to scale successfully further work to balance the needs of diverse partners with the importance of simplicity will be carried out. Technical onboarding workflow standardization will also be carried out. This will include the implementation of standard workflows for: auditing partner metadata for completeness; the activation of data capture pipelines; account setup; and dashboard deployment.

WP3: Partner Onboarding Phase 1

April 2022 - March 2023
Lead: Curtin
Involved: OAPEN

This work package focuses on the practical tasks associated with quality assurance and onboarding partners. This includes: 1) ensuring that existing partners are comfortable with the Demonstration Project and have signed updated data sharing agreements; 2) undertaking rigorous quality assurance processes with partners; and 3) testing and iterating onboarding processes with a small number of new partners.

Our existing workflows and data ingest procedures have been through three cycles of testing and user acceptance with the current Pilot Dashboard Partners. Workflows for capturing and cleaning new partner data will continue to be tested, along with processes for creating new partner dashboards, providing appropriate access control, and providing partners with an appropriate level of customization of dashboards, within the framework identified in the technical development roadmap. Lessons learned during this work package about the process of onboarding a limited number of new Dashboard Partners will feed into the workflows used to support onboarding of additional partners later in the project. There will be limited options for customisation of processes and dashboards through the core service offering at that point (customisation does offer a source of future revenue as a fee-for-service offering eg for developing an ingest pathway for a custom data source).

A primary output will be a guide for new partners that describe the Dashboarding Service's requirements for data quality and completeness; provides practical information about steps that need to be taken to make data available to the Dashboard Service. A key part of OAPEN's experience in providing a cost-effective service has been to create a balance between the need for consistent data from presses and partners, and the value of supporting diverse partners interested in engaging with a service. This experience will play a key role in helping the Dashboard Service to develop sensitive approaches that manage the costs of data handling while supporting a diverse OA book publishing community.

WP4: Community Cultivation

October 2022 - March 2025
Lead: Educopia
Involved: All collaborators

A process for developing an annual technical roadmap that reflects community priorities will be established and implemented.

The success of our project rests on developing an engaged community of partners that feel they have a stake in the success, sustainability and value created by the service. Using the initial Technical Roadmap as a provocation this work package will engage first with the existing partners and interested community as an example of community input with the goal of designing and implementing the process of an annual cycle of reporting, community prioritization, technical development and release. This work package will also establish a project volunteer

advisory board function in the first year as part of the community engagement process. As in the previous two projects, we will select individuals based on competencies, skills, stakeholder identities, and geographies to ensure a deliberate cross-section of knowledge and perspectives.

WP5: Sustainability Model and Community Consultation

April 2022 - June 2023
Lead: OAPEN
Involved: All collaborators

A range of sustainability models will be investigated and co-created with community stakeholders, leading to a call for Expressions of Interest for Phase 2 partners which includes an expectation of financial contribution.

Drawing on the experience of both Educopia and OAPEN in providing services through a mixed revenue model this work-package will document sustainability options and consult with the wider community to solidify our Phase 2 partnership expectations. This work is crucial in achieving our goals of supporting diversity of service users, and therefore of supporting the diversity of scholarly publishers.

Consultation will be deliberately wider than the initially engaged group of partners and will also benefit from the work in WP8 on identifying networks to provide input. The work of our PI group in multiple settings has emphasized how the resource limitations that prevent engagement of smaller publishers are not limited to financial issues but include the capacity of staff to dedicate time to engage with new initiatives. This makes engaging through networks (such as OPERAS, SciElO and EIFL) the most effective way to gain diverse input on what sustainability models will work for the widest diversity of contributors and supporters.

WP6: Onboarding of Phase 2 Partners

April - December 2023
Lead: Curtin
Involved: OAPEN

During Year 2 the second phase of partners selected from the EOIs in WP3 will be brought onto the platform. Onboarding processes will be stress tested and issues identified to inform a second scheduled round of technical work in WP9.

This work package will stress test on-boarding procedures and documentation, refined and updated through WP3, with a larger number of publishers (10-20). Lessons learned during this process will feed into Year 3 Technical Development priorities, specifically the updates to technical capacities in WP9. The core outputs will be a further revision of standardized workflows and onboarding systems. In particular it will focus on refining the onboarding guidance documentation for new partners.

WP7: Sustainability Development Plan

July 2023 - December 2024
Lead: OAPEN
Involved: All collaborators

A year 3 membership drive, with relevant adjustments to the product offering based on feedback gathered in WP5 and WP6.

Based on the results of WP5 OAPEN as lead on market engagement will refine the offering based on the experience of the second phase of onboarding. This will include developing the initial membership revenue models and service inclusions, testing these with the existing partners and (following any further refinement) leading engagement to bring on new partners. The goal will be to identify 20-40 additional partners (dependent in part on community input into the developing sustainability model).

In addition, this work package will identify other potential users and funding sources that can diversify the revenue streams. Offerings to libraries is one area OAPEN is already exploring. The Curtin team is engaged with OPERAS as part of a potential metrics offering through the European Open Science Cloud and as a potential contributor to other expected calls for proposals from the European Commission. Providing a service which is accessible to diverse publishers will rest on our ability to identify opportunities to diversify revenue and this will be an ongoing process.

WP8 Regional Strategy Development

Jan 2023 - March 2025 (and beyond)
Lead: Educopia
Involved: All collaborators

This work package will focus on building relationships with appropriate regional (and language-based) networks. We will pilot data analytics services in collaboration with network partners in order to explore the potential of the analytics service to meet the needs of specific communities.

The technical infrastructure that underpins this project is capable of supporting publications in diverse languages, including those using non-Latin scripts. A strategy for ensuring the geographic and linguistic diversity of partners engaging with the services developed in this project will be key to maximizing its benefit for a diverse global OA publishing community.

A tension at the heart of this project exists between our desire to create a sustainable service with viable revenue streams, and our commitment to ensuring the diversity of the stakeholders that benefit from the service we are creating. This tension is particularly evident in relation to our goals of supporting OA publishers in different regions across the globe. In purely financial terms the largest markets for the service are likely to be concentrated in Europe and North America and to be dominated by English language publications.

Working with regional networks (including, but not limited to, Europe and North America) will allow this project to efficiently engage with and respond to the needs of diverse local contexts. Partnering with regional networks of publishers is also a strategy that will help us to maximize transparency and community involvement in the project, while respecting the resource constraints of smaller publishers.

WP9: Implementing Processes for Ongoing Technical Operations

Jan 2024 - March 2025
Lead: Curtin
Involved: OAPEN

Refine the form, structure and documentation for ongoing operations and onboard partners through Year 3.

In the final year of the funded project we will implement the working processes for ongoing provision of the service into the medium-long term with ongoing maintenance work (e.g. underlying system updates, security management and dealing with upstream changes in data formats and flows, ideally the latter will reduce over time through the work of the Data Trust partner project). This work package includes dedicated time to refine systems and the documentation needed to support their long-term viability. The processes and systems for ongoing data onboarding, customer relations management, as well as long-term technical maintenance and development will be put in place. The planned handover to OAPEN (or a different long term host if that is the conclusion of WP10) will provide an additional point within the project to test and address the robustness and scalability of business processes and documentation; as well as additional funding opportunities, for example via European Commission sources.

WP10: Organizational Form Development and Implementation

October 2023 - March 2025
Lead: OAPEN
Involved: All collaborators

Identify preferred organizational model for long-term operations, develop transition plan, and move towards implementation.

As discussed above, a university is unlikely to provide an ideal environment for the long-term provision of the service. As discussed elsewhere in this proposal, OAPEN provides a strong candidate, and is willing to host the service. It is nonetheless incumbent on the project to consider alternatives and to consult with the community of users and stakeholders developed through the project. A number of possible options are available. These include transitioning the service to another existing entity, founding a new organization and/or some combination (e.g., a new organization operating the service on behalf of a partner that provides community relations). There are benefits and challenges for all of these approaches. Most specifically there are relatively few appropriate organisations which have both community trust and long term sustainability plans, and developing new organisations is challenging and expensive. Educopia

will lead on a validation of OAPEN as a future host. A strategy will be developed in consultation with the engaged community and implemented in Year 3 of the project with the final deliverable being the implementation plan for the future transition to a sustainable ongoing operation of the data analytics service.

Expected Outcomes

The goal of the project is to develop and implement a sustainable, community governed, usage analytics service offering, based on the technical and community work undertaken in the OAeBU project. The key outcome is therefore an inclusive and sustainable service for usage data analytics capable of supporting a diverse range of scholarly book publishers. Diversity includes considerations of geography, mission, language and scale. The benefits to the broader publishing and scholarly community is the preservation of diversity amongst scholarly book publishers, and therefore support for a diversity of topics, voices, writers and readers of scholarly work in book and book-like form.

Central to the work packages within the project is identifying a viable sustainability model that enables engagement with a diversity of small and less well-resourced publishers. There is a clear business need for high quality data analytics to support strategic decisions making within scholarly publishers. There is also a range of viable business options to develop such a service. However, only a subset of these models will enable participation by a diversity of publishers. Many potential models will exclude smaller and less well-resourced publishers. Identifying a combination of sustainability model, organizational host, and pricing models that provides both confidence in sustainability and is inclusive of a diversity of publishers is therefore critical.

Through the services that COARD offers (and charges for) we have identified that there are medium-scale scholarly book publishers able and willing to pay for services at a level (around $AUD10-15k per annum) where around 40-60 contributors would cover costs. However, these levels will exclude many smaller presses. OAPEN has substantial experience of navigating this tension with their service offering. The current OAPEN revenue model involves a mix of (low) membership fees, grant-funded development, sponsorship and subsidy, alongside seeking to diversify the membership base.[42] In initial brain-storming, alongside the Business Model Canvas developed within the OAeBU project (see Appendix B) have identified a number of options meriting deeper testing including freemium models (a free version with limited functionality; as well as a premium service with much more functionality); cross subsidy via tiered pricing; infrastructural contributions from libraries and funders and other subsidies.

[42] For some background and history see on OAPEN see this personal perspective "A brief saga of open access books"