This paper has been authored and submitted by Dr.Katherine Skinner (katherine@investinopen.org), Christina Drummond (christina.drummond@oabookusage.org), Dr. Rebecca Handcock (rebecca.handcock@curtin.edu.au), and John Sherer (jsherer@email.unc.edu).

In 2015, an SCI project, "Collecting and analyzing usage data for online scholarly publications" began exploring how to gather and analyze data about open access book usage, ideally in partnership with publishers of all kinds (including library publishers, university presses, and commercial publishers). The initial team (Kevin Hawkins, Lucy Montgomery, Sarah Melton, Joyce Chapman, and Katherine Skinner) used its SCI week to study two main problems: 1) the administrative and technical burden entailed in data collection and analysis around Open Access (OA) book usage; and 2) ethical questions and dilemmas regarding collection and usage of data in institutional (research/press) environments.

Our initial research and discussion at SCI helped us more clearly articulate that in 2015, data associated with scholarly communities was increasingly being gathered, analyzed, and converted into new products and services. This work was largely led by commercial entities that were opaque about what they were collecting/packaging for administrators in university settings and what they would be charging for these new services. We focused on open book data because this was a tangible and important, yet underserved, area of research.

Our concluding presentation in September 2015 discussed the ways that research institutions might be able to gather, analyze, and share data about their own publishing practices and impact in an ethically grounded, transparent, and community-governed "data trust" environment. In that presentation, we argued that "The value of aggregate data, the ethical implications of its use, and associated administrative burdens have created the need for a neutral point of scholarly communication data aggregation, management, and dissemination." We proposed to build this coordination framework as a global scholarly communication data cooperative in which privacy, ownership, and control over both institutional and aggregated data was carefully managed.

Since the 2015 workshop, this initiative has evolved into two core teams and six grant awards totalling more than $3.4M. Facilitating the aggregated review of book metrics for institutions through dashboards, it has resulted in the development and successful launch of the Book Analytics Service (BAS) hosted by OAPEN and COKI. Addressing the complex data governance requirements related to the exchange and management of multi-platform data connectors for granular, sensitive usage data, SCI also seeded the formation of the OA Book Usage Data Trust, which has developed its community governance and sustainability with Mellon Foundation support while fundraising for a planned 2025 technical build and launch with fiscal sponsor OPERAS. In addition, two NSF grant awards have explored how the lessons learned from these efforts may extend to other national interests in open scholarship metrics aggregation and exchange.

Timeline of Projects to Date	
2015	***Collecting and Analyzing Usage Data for Online Scholarly Publications (Scholarly Communication Institute)*** Participants: Kevin Hawkins (University of North Texas), Lucy Montgomery (Curtin University), Joyce Campbell (Duke University), Sarah Melton (Boston College), Katherine Skinner (Educopia Institute).
2018-2019 **Issue Identification**	***Understanding OA Ebook Usage: Toward a Common Framework (93k USD)*** A one year award from Mellon supported the exploration and development of international standards and practices for measuring and reporting ebook usage. Outputs included: The Exploring Open Access Ebook Usage whitepaper and The Building a Trusted Framework for Coordinating OA Monograph Usage Data paper. **Principal Investigators**: Charles Watkinson (University of Michigan), Kevin Hawkins (University of North Texas), Lucy Montgomery (Curtin University), Brian O'Leary (Book Industry Study Group) ; **Event Facilitator:** Katherine Skinner (Educopia Institute).

2020-2022 **Background research and proof of concept technical development**	***Developing a Pilot Data Trust for Open Access eBook Usage (OAEBU) (1.2M USD)*** A two-year Mellon award enacted the recommendations in the Exploring Open Access EBook Usage whitepaper, engaging over 100 contributors to document usage data supply chains and use cases, develop proof of concept open infrastructures for usage data aggregation and provision, and identify community-based governance models to support a diverse, global data trust for usage data on open access (OA) monographs. Project outputs included: OA eBook Usage (OAeBU) Data Trust Impact-driven Service Model, Governance Documentation, Supply Chain Mapping, and Analytics and Reporting Use Cases by Stakeholder. Technical deliverables related to proof of concept usage data dashboard pilots for multiple partners compiled by the project's technical team at Curtin University included: workflow automation code to fetch, process, and analyze usage data, documentation on ingesting, processing, and outputing workflows and processing workflows for mapping books products, linking metrics, and exporting results, and a 'Proof of Concept' OA Book Usage Data Dashboard. **Principal Investigators**: Kevin Hawkins (University of North Texas), Lucy Montgomery (Curtin University), Cameron Neylon (Curtin University), Brian O'Leary (Book Industry Study Group), Katherine Skinner (Educopia Institute), Rebecca Welzenbach (University of Michigan)
2022-2026 **Open scholarly infrastructure development**	***Open Access eBook Usage Data Trust (OAEBUDT) Data Exchange a.k.a.*** ***Advancing to Launch by Developing IDS Governance Building Blocks (1.2M USD)*** This ongoing three-year Mellon award develops data space governance mechanisms, quantifies participation benefits, and provides a sustainable budgetary model for an international data space for OA book usage. Project outputs to date include: An overview discussion paper, a Technical Gap Analysis Assessing Open Scholarly Infrastructures for an IDS model role, a Participant Rulebook, and multiple community consultation reports, board policies and committee charters. **Principal Investigators**: Christina Drummond (University of North Texas), Prodromos Tsiavos (2023-Current OpenAIRE); Paolo Manghi, (2022-OpenAIRE), Yannick Legré (OPERAS) --- ***Open Access E-book Usage Data Trust - Data Analytics (Book Analytics Dashboard (BAD) (750k USD)*** This ongoing three-year, 750k USD Mellon award creates a sustainable OA Book usage aggregation service from the proof of concept dashboards created in the prior OAeBU project by scaling workflows, infrastructure and customer support while transitioning the service from research project to sustainable community infrastructure hosted by OAPEN. Project outputs include: Technical documentation, Summary materials, and a Template dashboard. **Principal Investigators**: Lucy Montgomery (Curtin University), Cameron Neylon (Curtin University), Niels Stern (OAPEN), Ronald Snijder (OAPEN), Katherine Skinner (Educopia) --- ***Exploring National Infrastructure for Public Access and Impact Reporting (50k USD)*** This $50k USD 2023 conference workshop award from the National Science Foundation (NSF) (Award 2315721) gathered thought leaders in usage and impact reporting to discuss the efficacy and future of FAIR research impact reporting. The workshop, co-located with the 2023 CNI Spring Member Meeting, explored the current state and recommendations to move forward. Key outputs included: workshop proceedings, a video collection of expert recorded talks that capture the current state of scholarship usage and impact data analytics, and a peer-reviewed publication (preprint). **Principal Investigators**: Charles Watkinson (University of Michigan), Christina Drummond (University of North Texas); **Collaborators: :** Katherine Skinner (Consultant), Jennifer Kemp (Consultant) --- ***EAGER: Secure Research Impact Metric Data Exchange: Data Supply Chain and Vocabulary Development*** This 100k USD award from NSF advances recommendations from the 2023 workshop to provide foundational information about data flows and shared vocabularies. Usage data supply chains for journals and data

	(complementing the above referenced supply chain report for books), and a combined glossary and crosswalk of usage and impact related vocabularies resulted. **Principal Investigator**: Christina Drummond (University of North Texas); **Collaborating Researchers**: Laura Ricci and Michael Clarke (Clarke and Esposito); Jennifer Kemp (STRATOS)

Mellon Foundation Funded Project Details

Understanding OA Ebook Usage: Toward a Common Framework (2018) **-** As our first project began, monograph publishers were struggling to move beyond base-level reporting of financial transactions, such as print and e-book sales and fee-based licensing, and toward capturing and articulating the value of monograph investments in the context of their users across multiple sites and formats. They lacked the requisite scale to justify collecting and analyzing this usage data at a single press, particularly since such small pockets of data without broader market context would reveal little. To be useful, the data needed to have potential for benchmarking, seeing patterns and trends over time and across multiple publishers in a way that no single publisher could accomplish alone.

We wanted to explore how a community approach to usage data might be undertaken and whether such an approach could be made palatable to competitors, enabling them all to benefit from data sharing that would bring economies of scale for collection, analysis, and display of usage data for OA books, as well as richer context for that data. We were cognizant that usage data includes sensitive information about users that requires privacy safeguards, and also involves commercial information such as sales figures of print-on-demand copies. Any multi-institutional "usage data community" would have to begin with clear rules and boundaries or data sharing between publishers would be disincentivized by the amount of risk entailed. We also were aware that authors needed a say in the types of data that were collected and used so that they, too, could benefit from usage data results and to provide a crucial check-and-balance regarding data ethics.

In this one-year study, we developed a strong case for the need for action and produced a description of the landscape, and we proposed a "community data trust" as a way forward for the monograph community. We also proposed specific technical and standards-based issues that needed further study and community-based agreement in order to enable usage data to be efficiently gathered, normalized, and displayed in a set of pilot dashboards and analytic frameworks.

Developing a Pilot Data Trust for Open Access eBook Usage (OAEBU) (2019-2022) Research and development work fueled our development of two related pilot services: 1) OA Book Usage Data Dashboarding/Reporting, and 2) an OA Community Data Trust. These two elements were intertwined as we started the project. Over the first year, we recognized that we had a classic incentives problem baked into the duality of a data trust (focused on building enough structure and governance to enable trust between competitors) and a usage data services layer (dashboards and visualizations). If the Trust produced dashboards and visualizations, it would have unfair access to an "inside track" for both development and marketing/awareness building about the service. That could disincentivize others from engaging in service creation, leading to the potential creation of a "sole source" option that ultimately could undercut the Trust. This observation was well timed, as it coincided with learning about the emerging European Data Strategy impacting many global publishers, presses, and services, and the related European Data Governance Act, which would require such neutrality from data intermediation services.

In early 2021, in line with advisory board guidance, the trusted, neutral data-intermediation was separated from dashboard analytics so that both efforts could continue advancing distinct from each other. Joint planning continued in order to develop common principles to ground each of these efforts as their development trajectories diverged. Between June-October 2021, we undertook the development of three related outputs to provide this common core: 1) Guiding Principles for OAeBU, 2) a logic model/strategic framework for the Data Trust, and 3) a sample logic model/strategic framework that could be used for dashboard/analysis services. We also produced a governance model and documentation, while naming an initial Board of Trustees to govern and carry out the next phase of work with the Data

Trust. Each of these crucial pieces of documentation involved extended groups of project team members and interested parties within the broader book publishing community.

Open Access eBook Usage Data Trust (OAEBUDT) Data Exchange (2022-2025) and ***Book Analytics Data Trust (2022-2026)*** community efforts resulted in the creation of two distinct services, now in different stages of maturity.

1) **The Book Analytics Service (BAS) |** Now formally launched as a service by the OAPEN Foundation, with a technical backend hosted by COKI, this dashboarding and data analytics aggregation and visualization service was designed as a proof of concept by 2022 and began transitioning its business structure in 2024.
2) **The OA Book Usage Data Trust International Data Space (OAEBUDT-IDS)** | Through Mellon Foundation support, governance is developing to guide a technical IDS pilot by OA book usage data "creators" and data "recipients" in a managed international data space that will be conformant to the emerging Data Space Protocol. Substantial consultations documented in Zenodo and facilitated board work resulted in a participant rulebook to inform model contractual clause development. In September, we launched a Supporting Membership pilot to allow stakeholders to contribute financially to OAEBUDT-IDS development. The OAEBUDT Board of Trustees selected OPERAS-EU as the fiscal sponsor for its data space service and Think-IT as its technical data space development partner.

Lessons learned (so far) across the projects include:

- **The benefits and challenges of multi continent collaboration** (e.g. Australia, North America, Europe) for logistics and overall connectivity among principals is ongoing; so are the benefits and challenges of working across a mix of academic and nonprofit business models. Different drivers and calendar pressures impact these environments, both physically (e.g. time zone and hemisphere differences) and socially (e.g. fiscal calendars, vacation schedules and expectations, bureaucratic requirements, leadership changes). It requires ongoing cultivation of relationships and recalibration of awareness/understanding to balance these very different environments in a project team's leadership group.
- **The ongoing challenges of cultivating a (changing) market to be ready for a new product at the time of its launch.** Building a service or environment that publishers need is the easy part; engaging those same people in using and paying for that service or environment within a short timespan is incredibly difficult, especially in such a crowded market. Grant funding often enables initiatives to get started and build momentum; but even the best aims toward "sustainability" rarely provide the needed marketing/sales/user support teams that drive adoption for commercial products, which are often the biggest competitors to the grant-funded, "not-for-profit" or "community-led" efforts.
- **The importance of (and the costs associated with) hosting strategic, widespread community consultations** for getting feedback from diverse stakeholders on complex policy and technical issues. Our human networks are crucial, and when we are working across stakeholder groups, these quickly expand in complexity of vocabulary, values/principles, and needs. Staying in near-constant contact with stakeholders is often invisible work. Someone has to manage that process, track learnings, and manage their integrations into a service/environment. As with market cultivation, the essential market analysis and stakeholder mapping work that projects need to undertake take substantial time and are sometimes absent from grant-funded, "not-for-profit" or "community-led" efforts. Both BAS and OAEBUDT benefited greatly from Mellon's support for this vital community development work.
- **The importance of creating governance and sustainability models in parallel or in advance of building/assembling the technology solutions or products.** This ensures appropriate buy-in from stakeholders; it also establishes crucial checks-and-balances. Such model-building requires specialized knowledge that many grant-based teams lack, and as with other elements, this work is rarely adequately funded, even in "sustainability" oriented grants. Again, Mellon's support was vital to providing dedicated resources for this work for both BAS and OAEBUDT.
- **The strain of introducing new products/environments into a market where stakeholder ability to invest in new options is low and where common models are under scrutiny and reconsideration** (e.g., membership models, annual subscriptions). Even the best new products that result in high volumes of interest and desire may simultaneously meet with an unfortunate reality in today's scholarly communication landscape: There is a steady

decline in available capital to invest. Channeling resources for operations in such a tight environment is tough at best. For example, the full OAEBUDT - IDS technical build, quoted at 300k-500k Euro,is currently on hold pending fundraising that retains stakeholder trust, despite substantial public/private support for such an effort.

- **Challenges resulting from managing an intentional project split into two distinct-but-aligned efforts.** Via deliberate crossover in Advisory and Governance Board members, we aimed to provide enough autonomy for each project to make its own decisions but enough connectivity so that unnecessary duplication of effort was avoided. For the most part, we were successful; yet, keeping up with what *one* project doing is challenging for volunteer advisors and board members) Coordinating across time zones and business environments, especially with seasoned leaders who sit on the boards, requires dedicated and specialized facilitation time, energy, and bandwidth that may not be adequately accounted for in a project budget or work plan, despite best intentions to be globally inclusive.
- **Fundraising is an omnipresent need and resources to do it, much less be successful at it, are limited within most project teams.** A PI can never just focus on the project at hand; they must constantly lay groundwork for the next 2-3 projects to maintain research and development work once the current project expires. Even when a project is ready to advance to become an ongoing program/business, that project's readiness is only one of so many factors that influence its actual survival. The market's readiness, procurement processes, the political environment, and the shifting pace of change (and attention that goes with it) are all variables that impact and influence that transition moment. While travel support to share research findings helps to raise awareness of the innovative services that have emerged, grant writing and membership drives require additional key support.

4. Future Plans

As they secure funding, the **OAEBUDT-IDS** plans to develop an ISO standards compliant proof of concept International Data Space (IDS) to support secure multi-party access and use of sensitive data without harvesting, prior to creating a full-fledged IDS with API discovery and data pipeline controls and access transaction logs for usage data recipients and providers (pending 500k+ Euro in funding). Membership models will launch in 2025 to generate unrestricted funds to complement grants (2025, post 2024 pilot), while the data model and technical roadmap develop through technical pilots (2024-2026). Partner networks are emerging to explore how this OAEBUDT-IDS could extend to support data exchange for other types of data, as well as open knowledge graphs and the evaluation of OA funding models.

The Book Analytics Service is adding new data streams to increase the relevance and usefulness of the product; it is also actively growing and onboarding its client base and working with this group to test and refine its business, governance, and technical models. Topics both the BAS and OAEBUDT teams could explore with other SCI participants include:

1. **How should open infrastructures seek to share back office service and departments (legal, communications, marketing, IT, etc.).** Many models are used in project-to-service transitions, including hosting by academic institutions (e.g., APTrust), multi-institution partnerships (e.g., HathiTrust), fiscal sponsors (e.g., SPARC), or founding new nonprofits. Relationships with the administrative entities that host early stage research, development, and service launch shape how these emerging infrastructures can launch services and products.Are national or global pathways better suited to supporting such development and scaling of interoperable yet inclusive open infrastructure?
2. **Are there interoperable ways to measure or compare infrastructure impacts on increased operational efficiencies or carbon footprints?** For example, if we move to shared solutions, does that mean less computation or cloud-services are needed across the scholarly communications ecosystem, thereby generating an environmental impact by working together?
3. Just as the OAEBUDT-IDS is looking to translate a model from other industries to SSH and scholarly communications, **are there other opportunities out there related to emerging technologies that could make SSH more inclusive, dynamic, and human?**