



OA eBook Usage Data Analytics and Reporting Use-cases by Stakeholder

Christina Drummond
Kevin Hawkins
June 2021

ACKNOWLEDGEMENTS

This work was made possible by The Andrew W. Mellon Foundation through the *Developing a Data Trust for Open Access Ebook Usage* grant,¹ and by the dozens of community contributors who participated in workshops, anonymous online activities, and discussion through the Open Access eBook Usage (OAeBU) Data Trust’s communities of practice. Christina Drummond designed and facilitated the virtual processes that generated the input for this document. Ms. Drummond and Kevin Hawkins served as Editors. Philemon Eniola designed the graphics and layout herein.

ABOUT THE OAEBU DATA TRUST EFFORT

Originating in 2015, the OAeBU Data Trust effort has brought over 100 individuals across five continents together to surface and address the issues that complicate the analysis and use of book usage metrics for decision making and open access advocacy. To date, the project has documented the complex OAeBU data supply chain,² launched pilots of open-source infrastructure for a usage data trust, and facilitated workshops to understand the ways in which scholars and specific staff roles at libraries, publishers, and publishing platforms and services rely on OAeBU data. While these use cases were documented to inform data trust service development, the project team recognizes that they provide broader insights into the evolving analytics and linked data demands across the scholarly communications ecosystem.

METHODOLOGY

This document evolved from the outputs of multiple virtual ideation sessions and workshops among peer stakeholder groups. These community-oriented feedback mechanisms were advertised via the OAeBU Data Trust effort’s communities of practice. Co-Editors Drummond and Hawkins translated meeting and asynchronous ideation outputs into a preliminary use case document, which was shared back for public comment via the project’s communities of practice and social media.³ Public comment was incorporated with notes offered by community members included at the end of each section.

EXECUTIVE SUMMARY

Book publishers, publishing platforms and services, and libraries all rely on book usage data to inform their operations. Across these industries, staff must address the burden of managing and curating usage data provisioned in COUNTER-compliant and non-compliant reports,

1 https://educopia.org/data_trust/

2 Clarke, Michael, & Ricci, Laura. (2021, April 9). OA Books Supply Chain Mapping Report. Zenodo. <http://doi.org/10.5281/zenodo.4681725>

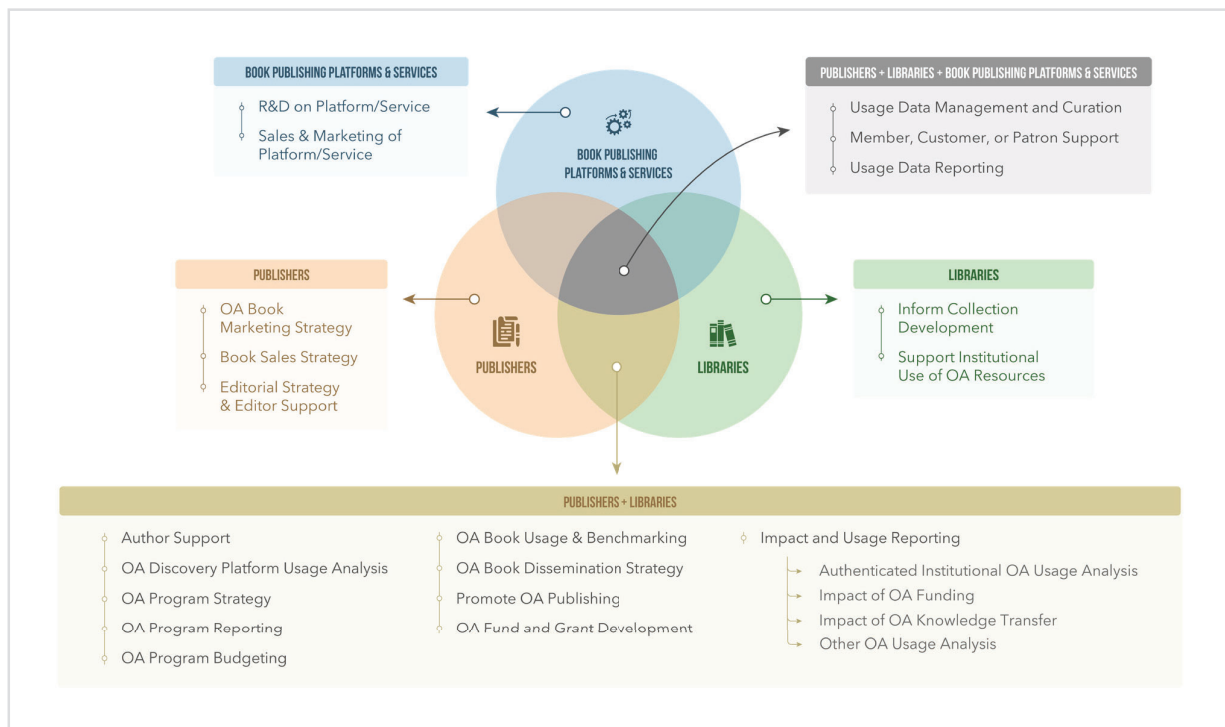
3 The OAeBU project’s processes for communities of practice development and use case development are published in Drummond, Christina (2020). “Engaging Stakeholder Networks to Support Global OA Monograph Usage Analytics,” Collaborative Librarianship: Vol. 12 : Iss. 2 , Article 9. <https://digitalcommons.du.edu/collaborativelibrarianship/vol12/iss2/9>

APIs, dashboards, and spreadsheets. These institutions individually manage, compile, and link usage data metrics, some of which are provisioned via dashboard services for users varying from scholars, research offices and funding agencies, to editors, collections and acquisitions managers, and administrative decision makers.

While the term “usage data” is often used today to refer to web analytic reports that tally page visits and file downloads, the use cases herein document a near future where linked usage data analytics regularly inform book publishing and scholarly communications operations. Publishers and libraries expressed interest in using OAeBU data analytics to inform overall OA program investment, strategy and fundraising. They noted how data on the location and context of OA ebook access can inform book marketing and dissemination strategy, acquisitions and collections development, and scholars’ promotion and tenure. OAeBU data also surfaced as vital to supporting the promotion of OA publication among scholars and to illustrating institutional impact on the local and global stage. Linked contextual usage data can illustrate how OA books reach targeted audiences in classrooms, among scholars, within industry, and among policy-makers.

OAeBU data is poised to inform customer-service relationships as publishers and libraries seek to appreciate service provider niches, evaluate book hosting and dissemination arrangements, and understand what is returned from their OA investments. Simultaneously, publishing platforms and services leverage usage data to inform the research, development, sales, and marketing of their own infrastructure.

Diverse OA book publisher models, from university or library-based publishers to commercial publisher OA teams, surfaced many common functional needs of OAeBU data as shown for publishers in the graphic below. However, their structures resulted in unique needs tied to administrative and sales reporting, as is reflected in their individual use cases.



SCHOLAR OAEBU DATA USE CASES

Scholars may interact with OAEBU data as authors; as acquisition, volume, or series editors; when under promotion and tenure review; or when serving on review committees. In these roles they hope to understand and report on how particular works are used, recognizing usage data as a complementary indicator to citations and sales data. OAEBU data can illuminate reader demographics. It also holds the promise of surfacing impacts and opportunities related to scholarly, media, policy, or classroom use. It can help authors to evaluate the effectiveness of their book marketing efforts, navigate book dissemination options, and weigh possible publishers for future work. For editors, OAEBU data may inform acquisitions and recruiting by illustrating potential audiences, niches, or demand for new editions. It also can provide supporting data for author recognition while surfacing opportunities to market scholarship across disciplines. In terms of peer review, OAEBU data may inform both the assessment of and advocacy for a candidate's scholarship, providing context for the use of impact factors and citation metrics. It also may be useful for authors wanting to understand usage patterns when defending against misinformation or attacks on scholarly freedom.

Like other stakeholders, scholars are challenged to understand book usage across reports provided by multiple sources. In addition, they may have ethical and privacy concerns over who has access to their personally identifiable OAEBU data in the absence of notice and control mechanisms that grant them the ability to understand and authorize third-party viewing and use of such information.



Author



Editor



Hiring/P&T

Use Case 1.

Understand discovery channels

How do individuals find and access the content?

Personas



- a. Inform planning for new book to frame book pitch.
 - i. How do people discover, download, purchase similar books
- b. Evaluate potential ways to host/disseminate the book.
 - i. Understand where readers are downloading OA versions from
 - ii. Understand how people discover the OA versions
 1. Social media
 2. Marketing campaigns
 - iii. Understand whether individuals know an OA version exists
- c. Understand the relationship between OA downloads and print sales
 - i. Does OA version drive purchasing
 - ii. How are people interacting with sample pages

Use Case 2.

Understand audience demographics

Personas



- a. Understand current audiences
 1. By geographical region
 2. By institution
 3. By discipline / field
 4. Academic vs. non-academic use
 - a. Academic institutions
 - i. Within my institution
 - ii. Outside of my institution
 - b. News media
 - c. General public
 - d. Government / Public policy
 5. By academic role
 - a. Students
 - b. Scholars
- b. Surface, assess potential new audiences based on similar/prior book performance
 - i. New regional audiences, translations
 - ii. New field audiences

Note: Download data is very informative - more granular the better. Citations and reviews reflect a "tiny cross-section" of readers.

Use Case 3.

Understand scholarship impact

Support varied metrics to complement traditional sales metrics

Personas



- a. Understand use of OA eBooks
 - i. Academic use
 1. Scholarship
 2. Class syllabi
 3. Student online projects
 4. Blogs
 - ii. Non-academic / public use
 1. News media
 2. Social Media
 3. Wikipedia
 4. Public policy
 - a. Nonprofits / grey literature
- b. Compare / show how OA eBook impacts relate to print impacts
- c. Understand how online usage over time compares to sales and citation activity
 - i. See changes over time
 - ii. Is OA becoming more mainstream
- d. Understand differences in book branding
 - i. in different fields / disciplines
 - ii. In different geographic communities

Use Case 4.

Evaluate book marketing and outreach

Personas



- a. Evaluate impact of time spent to connect with readers
 - i. Where do page visits or pageviews come from - FB post, twitter post, etc.
 - ii. Understand relationship between actions and reading.
 1. E.g. Does timing of promotional Facebook post, tweet, or speaking engagement by the author result in a near term increase in downloads or pageviews
 2. What OA eBook activity do actions result in? Downloads, citations, mentions
 - a. What, where, quantity
- b. Understand the impact of advertising and marketing campaigns
 - i. Are marketing campaigns effective?
 - ii. Inform whether to change promotions tactics / strategy

Use Case 5.

Understand financial investment in OA

Personas



Understand financial investment in OA

Use Case 6.

Inform editorial decisions on what to publish

Personas



- a. Explore whether something is niche or potentially broader
- b. Inform market analysis
- c. Determine need for new editions

Use Case 7.

Author identification and recognition

Personas



- a. Leverage impact and audience measures to inform recognition

Use Case 8.

Surface ways to increase scholarship accessibility

Personas



- a. Diversify audiences
- b. Public dissemination
- c. Plural / translational scholarship (across disciplines)

Use Case 9.

Assess / advocate for candidate scholarship for hiring, promotion, and tenure

Personas



- a. Inform evaluation of candidate merits
 - i. Illustrate disciplinary impacts
 - ii. Illustrate international usage
 - iii. Illustrate non-academic impacts of scholarship
 - iv. Inform qualitative peer/self-assessment
- b. Surface candidate community engagement / outreach activities
 - i. Provide data that connects with university wide innovation or engagement initiatives
- c. Advise authors on decision to publish OA
- d. Provide data and context for tenure file/job application/grant application

Use Case 10.

Guide and inform impact factor use for evaluation

Personas



- a. Understand how OA eBook usage data complements other impact factor metrics
- b. Understand and manage disparate impacts
 - i. "Don't exacerbate existing inequalities"
- c. Understand what author-affiliated data is available to whom, to inform privacy and ethical guidelines over access and use of such data
 - i. "People currently have agency over what they put in their file and can choose to omit information or provide context."
- d. Provide context for low numbers, e.g. high quality or influence based on type of usage.

Use Case 11.

Support defense against threats to scholarly freedom

Personas



- a. Fight online targeting and academic harassment
 - i. Provide evidence of activity tied to events
 - ii. Show imbalance of activity compared to similar titles or other authors
- b. Understand connections between online presence and dark-web activity

COMMERCIAL PUBLISHER OAEBU DATA USE CASES

OA-focused commercial publishing teams apply OAeBU data to support marketing, sales, and OA program strategy. They also provide OAeBU data to support editorial processes and authors. OAeBU data may surface access trends and discovery platform audience niches while also allowing editors and authors to benchmark OA eBooks against comparable books by leveraging usage data as a digital counterpart to traditional sales data. Staff roles that may interact with OAeBU data span OA management, strategy, market intelligence, marketing, sales, editorial and acquisitions, and IT.

Multiple issues make it difficult to realize the benefits of working with OAeBU data. Practices for analyzing OA usage data alongside non-OA and print sales data are developing. Situating internal OAeBU data alongside aggregator reports is a manual, time-intensive process riddled with technical challenges stemming from evolving standards and varying levels of data compatibility and existing standards adoption. Notably, there is also no existing mechanism to contextualize OAeBU against competitor benchmarks or across the OAeBU data landscape.



OA Program



Marketing



Sales



Editorial

Use Case 1.

Inform internal OA program strategy

Personas



a. Provide usage reports

- i. to demonstrate OA strategy, importance, and impact to internal audiences
- ii. to show eBook usage by audience and discipline over book lifespan
 1. Answer “which books in which editorial departments could benefit the most from efforts to diversify and increase readership?”
- iii. to show how operational changes shift engagement patterns
 1. involving usage by platform, country, discipline
 2. across known existing and anonymous customers
- iv. to track and demonstrate expanding readership benefits of OA
 1. by illustrating whether an OA version increases book readership and access over time.
 - a. If a book is written about a specific region, did the OA version help those in that region to access it?
 - b. Did the OA version help those who’d benefit from using the book to access the book?
 2. by adding context for books in smaller topic areas based on comparable usage activity in relevant topic areas

3. by surfacing potential disciplinary crossover based on usage of comparable titles
 - a. across disciplines
 - b. within target audience communities
- b. Understand eBook audiences**
 - i. to establish OA program profile or niche.
 1. "Show what someone can expect from working with us."
 - ii. to understand how OA readership patterns differ from non-OA
 1. for public access to publicly funded research outputs
 2. for specific geographic regions
 - iii. to support regional office strategy development
 1. with internal-facing usage reports for subjects by region and platform
 2. by surfacing areas that merit attention
 - iv. to identify matches between funder priorities for specific audience usage and OA program's strengths, i.e. historical usage patterns
 1. e.g. historical usage of chemistry titles in certain geographic region

Note: Ability to drill down into the data is important!
- c. Compare OA program activity against other OA publishing programs i.e. industry benchmarks**
 - i. Benchmarking against peer publishing programs (e.g. publishers of roughly the same size and type) to provide direct comparison
 - ii. Benchmark against different types of publishers
- d. Compare OA book performance against comparable usage benchmarks for a given audience and discipline**
 - i. to show authors what OA book "success" looks like when working with our OA program
 1. for an individual title based on similar titles
 2. in a discipline for target audiences
 - a. e.g. for books on topic X with a target audience in region Y, this is what OA usage looked like.

Note: each author/book has its own niche audience
 - ii. to answer "when can we expect peak engagement" based on usage trends
 - iii. to understand duration of book lifespan
 1. illustrate when peak usage is expected for a given discipline and audience. e.g. "The trend for books like yours is to see a bump in activity after x years and y years."

- iv. to inform stakeholders of expected engagement patterns and book impacts
 - 1. e.g. "Here's how we expect the OA content to be accessed"

Note: stakeholders include: authors, OA funders such as national funding bodies, library consortia, community-sourced funding, learned societies, etc.

Use Case 2.

Generate external funding impact reports

Notes: Frequency and granularity of data are important. Work is needed to define terms and metrics for lifespan, measures, and peak activity

Personas



- a. Communicate impacts to stakeholders i.e. "Move the dialogue from 'we published it' to 'here's how the OA version affected the conversation'".

- i. Report usage by:

- 1. Author institutional affiliation
- 2. Title
- 3. Source of download (platform)
- 4. Geographic region of access
- 5. For a publisher's core or traditional audiences
- 6. By author or funder targeted audiences

Note: each book may have its own success metrics in terms of what its author wants to achieve by going OA.

- ii. Generate reports for

- 1. Research funders
 - a. Usage and engagement tied to:
 - i. Research area / field
 - ii. Mission impact:
 - 1. Public policy references
 - 2. Other "broader impacts"
 - iii. Geographic / regional usage
- 2. Groups that fund OA books or whole book series to establish ROI
 - a. Learned series
 - b. Research institutes
 - c. Charities/NGOs
 - d. Corporations
 - e. Libraries and library consortia
- 3. Authors and their research offices
 - a. To show compliance with funding mandates
 - i. e.g. Support reporting requirement for UK Research Excellence Framework, "universities have to show how they achieve impact"
 - b. To show "real world" / broader impacts

Use Case 3.

Support marketing campaigns and strategy

Personas



- a. Support marketing of OA program
 - i. Encourage OA adoption by showing advantages to publishing OA compared to non-OA
 1. Compare OA vs. non-OA usage
 - a. How OA books perform compared to non-OA counterparts in subject areas
 2. Compare OA vs. non-OA discoverability
 3. Illustrate what happens when a book is “flipped to OA”
 - ii. Illustrate global profile, reach, niche of OA program
 1. Show “what one can expect from working with us”
 2. By institution type, region, usage activity
 - iii. Use data to develop white-papers for stakeholders
 1. Industry
 2. Higher-education
 3. Book series partners/funders
 4. Authors
 5. Funders
- b. Support promotions for OA book content and OA authors
 - i. Show how promotions efforts affect targeted audience usage.
 1. Examine activity by
 - a. Date or date range by institution type and region split by subject
 - i. Institution type examples: libraries, schools
 - ii. Goal is to connect usage activity to promotions campaign
 - ii. Understand content usage and engagement over time
 1. Is usage ongoing, i.e. is something “adopted” or “continuously read”
 - iii. Provide authors with usage reports
 1. Generate/send authors personalized book performance reports showing specific usage on platforms
 - a. Following promotion campaigns or launch events
 - b. Related to social media
 2. Provide self-service / easy access for authors to check their own usage directly
 - a. Author-facing dashboards
 - b. Automated “personalized book performance” emails
 - c. Support author recognition
 - i. Support “Congratulations - your book is one of the most downloaded” notifications
 - ii. Suggest marketing campaigns based on:
 1. Top downloaded books
 2. Most cited books

Note: campaigns may include email campaigns and social media campaigns

Use Case 4.

Inform sales strategy

Personas



- a. Provide data on what to expect when flipping a title OA
 - i. Understand changes in usage after making content OA on a limited basis
 1. e.g. impact of making books OA in response to COVID-19
- b. Illuminate OA impact (+/-) on sales revenue over time
 - i. View OA related sales trends for the time spanning before and after going OA to track how sales fluctuate
 1. Answer questions like:
 - a. "Is there a trend of print sales going up or down after becoming OA?"
 - b. "Does the trend vary by discipline?"
 - c. "Does the trend hold over time?"
 - ii. Understand follow-on print sales related to OA access
 - iii. Understand follow-on ePub sales related to OA access
- c. Inform pricing strategy
 - i. by understanding interactions between OA usage, book margins, and pricing
 - ii. by analyzing traditional sales revenue alongside OA revenue streams
 - iii. by understanding usage trends for OA content bundles and eBook packages
 1. to support sales team offering bundled OA access
 - a. "If there is an ebook package - librarians want to know if they can include access for a discovery POV."

Note: While participants contributed these ideas, others noted they are not yet getting these data requests from their sales teams. Some anticipated these requests would be forthcoming as OA book usage data became more accessible. Another noted, "Packages for OA books are a small part in the context of transformative agreement impacts on journals and hybrid journals."
- d. Inform print edition compensation strategy based on OA usage
 - i. if OA access drives print purchases (e.g. at libraries)
- e. Inform sales strategy for target audiences
 - i. Determine discovery platform fit
 1. e.g. "For a Latin American studies monograph, how did comparable OA titles best reach their intended audience?"

Use Case 5.

Support editors

Personas



- a. Provide usage data for editorial decision making
 - i. To understand average usage for comparable titles within a discipline or subject area
 - ii. to inform expansion strategy by identifying disciplines where interest or readership is increasing
- b. Inform book commissioning
 - i. Provide benchmarked usage data for a book's audience and topic to understand if gauge usage
 1. Is it "going well" or is it higher or lower than expected
- c. Support acquisitions by providing authors with usage reports
 - i. to establish reach of publisher's OA book dissemination
 1. as proxy for providing sales feedback for print
 - ii. to support author requests for data demonstrating OA benefits

Use Case 6.

Track OA discovery platform usage activity

Personas



- a. Ensure OA versions are being discovered
 - i. Understand whether OA version is being accessed
 - ii. Understand how OA version is being accessed by target audiences

Note: It is important to show that access by the target audience access was successful, not just that content was available OA.
- b. Assess ROI for discovery and hosting platforms and services
 - i. Compare usage, impact, and citations across platforms, "in the wild"
 1. to illuminate platform-related usage by target audiences
 - a. by region, usage "size", discipline
 - ii. Surface benefits of working with specific dissemination platforms
 1. "How well does platform X reach our target audience?"
 2. "Which platform could do the best job if we have to prioritize?"

Use Case 7.

Work towards usage data standards compliance

- a. to ensure reporting meets applicable standards for usage data (e.g. COUNTER)
- b. Navigate unclear standards
- c. Compile and aggregate platform provided usage data that is non-COUNTER compliant, legacy (e.g. COUNTER 4), and COUNTER 5

Comments re: COUNTER usage data:

- Not all 'platforms' as defined by COUNTER provide COUNTER-compliant usage reports.
- Non-COUNTER reporting makes the aggregation of usage data labour-intensive, open to institutional interpretation as to what should be measured, with greater risk of data handling error.

UNIVERSITY PRESS OAEBU DATA USE CASES

University presses and library publishers leverage OAeBU data to support marketing, sales, and editorial strategy. They share OAeBU data to support their editors in attracting prospective OA authors and collaborating with current authors. In addition, they use OAeBU data to describe OA activity for fundraising and institutional reporting. OAeBU data may also surface trends and niches among discovery platforms, signal potential markets for print distribution or translated editions, and illuminate the impacts of OA investments. Staff roles that can benefit from OAeBU data include press directors and their editorial, sales, marketing, and grant writing teams.

Challenges to benefiting from OAeBU data stem from the need to manage, curate, and normalize inconsistent usage data provided by publishing platforms and services. Such data wrangling is time-intensive, requiring expertise in data analytics, bibliometrics, and book publishing metadata. These resource requirements apply as well to the provisioning of COUNTER-compliant reports to authors and other stakeholders, and may be beyond what's available to smaller presses. Balancing data curation and analytics to inform internal press operations against report and visualization development to meet demand is a resourcing choice smaller presses in particular must confront.



Press
Director



Sales and
Marketing



Editorial and
Acquisition



Research
Admin.

Use Case 1.

Inform marketing
strategy

Personas



a. Understand eBook audiences

i. Geographic distribution

1. Where is there niche content interest?
2. How are books used in parts of the world?
3. What is usage like in places where OA is the only way to access?

ii. Niche usage communities

1. of scholars
 - a. Disciplinary access/use
 2. of institutions
 - a. Usage by institutions interested in particular subjects
 - b. Discover which institutions are accessing what content
 3. of public
 - a. advocates/enthusiasts
 - b. Policy makers
 - c. journalists
- Issue: How to normalize/present data to ensure fair comparisons
 1. Between universities, maybe normalize by
 - a. Faculty count

- b. Full-time enrollment
- c. International classification systems
 - i. [World Bank country classifications by income level](#)
- d. National classification systems
 - i. [Carnegie classifications](#) - intra-USA comparison
- 2. Between locations: maybe normalize by per capita / population
 - a. E.g. [World Bank population estimates and projections](#)
- iii. Understand end-users of content: e.g. scholars, students, public, industry
 - 1. Inter-institutional access vs. non-institutional access
 - 2. Volume of usage
 - 3. Seasonality trends of access/usage
 - 4. Industry use, i.e. a measure of access via company URLs,
 - 5. Government use to measure public policy impact
- iv. Understand user-preferred access behavior
 - 1. Investigate the demand for print copies of OA books
 - 2. Understand format access patterns/frequency for OA eBook content
 - a. e.g. *HTML or PDF*
 - b. e.g. with or without comments by other readers, by reviewers, etc
- b. Understand OA content usage patterns**
 - i. General: usage patterns and trends
 - 1. Reading cover to cover
 - 2. Dipping in to look for citation/passage
 - 3. Accessing one chapter, all chapters, >1 chapter, etc.
 - 4. Volume and **seasonality of usage**
 - a. e.g. relation to academic calendars
 - ii. by university audiences
 - 1. faculty
 - a. teaching
 - i. Accessing one chapter, all chapters, etc.
 - ii. Work with [opensyllabus.org](#)
 - b. scholarship / research
 - c. personal interest
 - d. Tenured vs. non-tenured usage differences
 - 2. Grad students
 - a. In class
 - b. In scholarship
 - 3. Undergrad students
 - a. In class

4. Librarians
 - a. in reference, advisory, and support capacity to faculty and students
 - b. as book purchasers
5. university administrators
6. Broader campus community
 - a. alumni
- iii. By primary school (K-12) audiences
 1. Students
 2. Teachers
- iv. By public audiences
 1. Policymakers
 - a. National level
 - b. Local/state level
 2. Media
 3. Advocates/enthusiasts for a given community
 - a. E.g. gamers, environmental campaigners
 4. Local public in the UPs region
 - a. To show local impact on state and community where publisher is based
- v. AI/Non-human users
 1. Text / data mining
 2. Reposting
 - a. Portals
 - b. Pirate sites
 3. Abstracting
 4. Indexing
 5. Other active uses

Note: less inference is better

c. Understand OA as a discoverability tool

- i. to generate a transaction
 1. a print transaction
 2. a premium e-transaction
- ii. To understand how OA books are presented on the sites librarians use to acquire books
 1. GOBI
 2. ProQuest
 3. OverDrive
 4. BorrowBox
- iii. To understand how OA books are presented on sites that readers use to discover scholarly content
 1. Google Scholar

2. Google
 3. Academia.edu
 4. ResearchGate
- d. Understand what a reader might want from a product to then support/drive further customer interaction. (Analog to index)

Use Case 2.

Inform sales strategy and understand OA impact

Personas



- a. Suggest/understand titles that should become OA
 - i. Based on past sales data (sales trends inform future potential)
 - ii. Identify geographic niches for future OA publication where there is a lack of retail distribution chain but past demand/use of previous editions
- b. Suggest/understand OA impact (+/-) on sales revenue
 - i. Test theories of OA impact on sales
 - ii. Surface OA usage in Global North vs. Global South
- c. Understand amount of OA funding required to meet anticipated sales
- d. Surface local edition or translation opportunities based on geographic use
- e. Inform marketing/sales resource allocation by incorporating usage data alongside print/eBook sales data

Use Case 3.

Provide basis for exploring assumptions about OA impact

Personas



- a. Impact on gated content
 - i. Is paid content out in the world in the same way?
- b. Impact on impeding access to knowledge

Use Case 4.

Editor relations

Personas



- a. Support/inform toolkits for editors to be sophisticated in their use of data
- b. Provide reports to help track impact-based editor acquisitions/goal performance. E.g. progress on acquiring books with x amount of impact

Use Case 5.

Analyze/benchmark
OA books

Note: picking the right benchmarks is going to be very important. Clarity around standard benchmarks will be needed given many different ways to look at quantitative data.

Personas



- a. **Geographic impact/use of** OA books compared to similar non-OA
 - i. Geographic range of OA readers vs. non-OA readers
 - ii. Geographic regions OA reaches that non-OA doesn't
 - iii. Percentage of total use in different regions
- b. **Qualitative / alt-metrics** for OA compared to non-OA
 - i. Soft measures of disciplinary impact
 1. Prizes/awards
 2. Reviews
 3. Media attention
 4. References in tenure/promotion
 5. References in disciplinary blogs/list-servs
 6. References in public policy documents / grey literature
 7. Statements from readers/users on why they're using the book, what they'd going to do with what they've learned, etc.
- c. **Quantitative metrics**
 - i. Readership
 1. Quantity of readers
 - ii. Scholarly Citations
 1. Quantity of citations (count)
 2. Frequency
 - iii. Downloads
 - iv. In-browser views
 1. e.g. Internet Archive tracks
- d. **Online Engagement and Reference**
 - i. Social media
 1. e.g. Twitter
 - ii. Discussion forums
 1. e.g. Reddit
 - iii. Crowdsourced reference
 1. e.g. wikipedia
 - iv. Websites / blogs
 - Examples of data sources include: Crossref Event Data, Altmetrics
- e. **Comparative metrics (apples to apples)**
 - i. How book performs compares to books in similar disciplinary area
 1. e.g. defined by BISAC subject code
 2. e.g. defined by internal subject taxonomy

Use Case 6.

Inform Editorial strategy

Personas



- a. Inform printed version based on OA usage
 - i. Whether to do a print version
 - ii. Inform format/audience for print version (e.g. academic or crossover trade book)
- b. Understand course potential
- c. Inform which lists/series/subject areas could be made OA
- d. Understand usage to identify prospective authors for future projects
 - i. at the chapter level, especially for types of monographs (anthologies, readers, edited volumes)
 - ii. for a given author
 - iii. in discipline(s)
 - iv. of particular components/resources in titles with multimedia components
 1. e.g. supplemental multimedia components used separately from the book like a song recording in an open textbook
- e. Identify thematic areas for acquisitions based on usage data
 - i. Things that are valuable to small number of scholars could be overlooked or negatively impacted if editors focus on greatest number
- f. Identify potential authors for similar books by surfacing institutions where there is a strength in the subject area

Use Case 7.

Understand impacts of OA funding

Personas



- a. Provide funders with usage information / View impacts of funding
 - i. Show compliance with federal/national OA dissemination mandates
 - ii. Impact/usage by specific funder
 1. To support ROI conversations
 - iii. To show advancement/impact on funder/grant program's particular focus
 1. e.g. advancing K-12
 2. e.g. contributing to an informed citizenry/public
 3. Other "broader impacts"
 - iv. Impact/usage by specific grant
 - v. Calculate / understand impacts of up-front OA investment downstream
 1. with OER use
 - a. passing savings on to students
 - b. Improving equitable access to educational resources; removing barriers to low-income populations
- b. Surface where OA publishing funding is more available
 - i. To inform publishing of lists

- ii. To point authors interested in OA to specific places for OA funding
- c. Understand whether books with greater OA funding potential tend to be more viable compared to non-OA funded projects?
 - 1. This may inform / steer list direction
- d. Understand the editorial impact (think authors) of the CC licenses required by different funders
- e. Understand/show changes in usage (impact) for a title after it goes OA (before/after comparison)
- f. Understand impacts of OA policies / investments on issues
 - i. of scholarly publishing practices
 - 1. Understand the degree to which public subsidies mitigate publisher risk / cost to making books OA
 - ii. of equity and social justice
 - 1. Understand impacts (e.g. potential negative effects) of OA publishing requirements on nonWestern European or North American researchers
 - 2. Help understand whether OA funding models push OA countercurrent to equity/justice concerns - i.e. can you surface negative network externalities
 - iii. On moving forward / making progress on [UN Sustainable Development Goals](#)
 - Note: Qualitative data especially important*
 - iv. of access to information
 - 1. Is valuable research being ignored or overlooked by those who can only afford "free"
 - 2. Does usage of OA resources vs. paid-for-access in every instance reflect the value, validity, and significance of the research?

Use Case 8.

Author relations

Personas



- a. Provide prospective authors with information on
 - i. the advantages / disadvantages of OA options
 - ii. data to help authors understand OA impacts and value / ROI
 - iii. Specific places for OA funding alongside multiple sources of income
 - 1. putting OA sources alongside sales treating each book project with holistic sources of income
 - iv. Impact of citations (how much of usage comes from citations from other works available online)
- b. Support authors participating in OA initiatives
 - i. Provide impact information for their individual books
 - 1. General usage stats/reports (who is using and how)
 - 2. How books are used globally

3. Usage by the subjects of their books
 - a. Specific communities studied
4. Usage in underserved communities
5. Crossover
 - a. Between disciplines
6. Engagement by professionals and advocates outside the academy:
 - a. Policy-makers
 - b. Industry researchers
 - c. Not-for-profit organizations
7. Data on their book's impact
 - a. Author defined "impact" may include:
 - i. on discipline
 - ii. on subject community
 - iii. on policy / public affairs

Use Case 9.

Research / Grant development and administration

Personas



- a. **Support proposal development**
 - i. for digital publishing grants
 - ii. for host university institutional support/matching requests for OA initiative grants
 - iii. Data needed
 1. Percentage of eBooks used compared to respective hard copies
 2. Subject matters of eBooks (including BISAC codes)
 3. Who is using the eBooks (see above)
 4. How eBooks are used (see above)
 5. Broader impact information
 - a. How OA affects sales (+/-)
 - i. Since they're also trying to figure out how much it costs to publish OA and what's scalable, sustainable, subvention amount
- b. **Inform impact stories for advancement / donors**
 - i. Go beyond the numbers to tell a compelling case that OA makes a difference
 - ii. Types of donors:
 1. University-level donors
 2. Organizational donors (e.g. corporate foundations)
 3. Individual donors
 4. New and repeat donors
- c. **Support and inform partnership development for new initiatives**
 - i. With other local entities
 - ii. With other platforms

Use Case 10.University Administration
Reporting and Relations**Personas**

- a. Support career advancement, promotion, or tenure cases
- b. Understand / Report on ROI: Impacts of institutional funding of OA
 - i. Demonstrate global reach of university press imprint / OA activities
 1. Show distribution beyond where university press normally disseminates
 - ii. Demonstrate connections to university priorities based on usage
 1. e.g. advancing diversity, equity, and inclusion
 2. e.g. influencing public policy-makers
- c. Demonstrate value-for-money of university financial investment in the press.

Reviewer Comments re: COUNTER usage data:

- You would need a document which defines all relevant keywords/terms used for collecting/reporting data.
- Going purely by platform ('platform' as defined by COUNTER), just under half of our platforms do not provide COUNTER-compliant usage reports. These vendors are generally legal publishers, and smaller publishers who may/may not know to contact organisations such as HighWire Press or EBSCO to host and report on usage. E.g. for our (Australian) university - 53 platforms/78 reports not available COUNTER-compliant, for various reasons.
- Non-COUNTER reporting makes the aggregation of usage data labour-intensive, open to institutional interpretation as to what should be measured, with greater risk of data handling error.
- Note that COUNTER5 vendors are required to provide both COUNTER platform reports and SUSHI harvested usage reporting. COUNTER4 must only provide platform reports.
- For COUNTER5 reports which need to be downloaded from vendor platforms, vendors don't provide the option to pre-filter reports at the vendor platform, though COUNTER is currently looking at opportunities to improve reporting.
- COUNTER must be involved in these discussions, as COUNTER5 can resolve much of the issues surrounding inconsistent usage data across platforms. I've asked COUNTER to also investigate potential for OER usage reporting.
- COUNTER4 identifies HTML and PDF. COUNTER5 reports do not.
- Possible opportunity to collaborate with COUNTER and [CAUL](#) around COUNTER and SUSHI usage data, and Australian Bureau of Statistics (ABS) who have experience collecting and aggregating data in secure online environments (depending on what data is required).

LIBRARY OAEBU DATA USE CASES

Libraries use OAeBU data to report OA book usage and impacts to authors and administrators at their institution. When leveraged for analytics, OAeBU data can support library collections development, strategic planning, and OA resource promotion, in addition to OA program strategy, budgeting, and fund development. Access to OAeBU data can also inform book benchmarking and dissemination strategy for librarians, and it can illuminate the global and local impacts of OA investments while clarifying how readers access OA books through varied discovery platforms. Library staff roles that may interact with OAeBU data go beyond a library's administrators to roles such as subject librarians (liaisons), collections assessment staff and collections development managers, electronic resources librarians, and research support staff. Library IT staff supporting e-resources and systems may be responsible for OAeBU data management, curation, and visualization.

Multiple challenges face libraries working with OAeBU data. Reader privacy protections make it difficult to know if unauthenticated e-Book usage is related to affiliated patrons. Non-standardized approaches to processing data for chapters and compilations cloud the reporting of OAeBU. OAeBU data management, curation, and linking are complex and time intensive. In addition to compiling COUNTER and non-COUNTER compliant data from library management systems, from publishers, and from book dissemination or aggregation services, staff may be asked to link OAeBU data to other institutional research datasets. Such time-intensive activities require expertise in data analytics, bibliometrics, and book publishing metadata that may be beyond what's available to smaller libraries.



Dean, University Librarian



Scholarly Communication



Collections Development



Liaison Librarian



IT

Use Case 1.

Promote OA publishing opportunities

Personas



- a. Provide data to support OA evangelism and advocacy aimed at encouraging people to publish OA
 - i. Show examples of audience reach and platform-based access usage for
 1. prior OA publications by authors in similar field
 2. open educational resources
 - ii. Support outreach to
 1. PhD students
 - a. Use faculty data to inspire students to publish
 2. Faculty
 3. State Library Programs (U.S.)
 4. Alumni / Patrons
 - a. Use data to support fundraising

Use Case 2.

Support OA resource and initiative budget planning

Personas



- a. Articulate the value proposition for OA
 - i. on campus
 1. impacts on student outcomes
 2. impacts on faculty outcomes
 - ii. in the local community
 1. impacts of regional access
 - a. by alumni
 - i. Degree programs
 - ii. Continued education / professional certification programs
 - b. by businesses
 - c. by policy makers
 - b. Report OA usage by academic unit, college or school
 - i. To support library budgeting conducted via a Responsibility Center Management model, i.e. when individual faculties set library budgets
 - ii. To support cost-sharing among units
 - c. Understand OA investment as a % of collections budget
- Note: Multiple concerns were raised around the potential for OA costs to cut into collections budgets. Context was noted as very important.*
- d. Inform discussion of what merits OA budget investment
 - i. Establish Return on Investment (ROI) for an OA initiative
 1. To support ongoing funding of OA initiative(s)
 2. To strategically add content
 3. To justify library staffing to support OA publishing and OA resource discovery
 4. To support communications regarding incremental OA publishing costs
 5. To articulate public good, broader impacts of OA publishing
 - ii. Understand OA book usage across different websites (i.e. platforms, services, and publishers)
 1. To inform decisions of which platforms to support, i.e. to evaluate paying for memberships or services
 - a. "If no one is using it, would dollars be better allocated somewhere else?"
 - b. Understand volume of faculty publishing OA
 - i. Generally
 - ii. Via specific OA book access point - hosting or aggregation platform

2. To articulate cost per use per platform or service
 - a. By authenticated institutional users (patrons)
 - b. Within local geographic region or city

Note: Quarterly basis is helpful to see trends, annual supports overall evaluation

Note: Multiple participants stated that in their institution OA initiative support wasn't strictly tied to their own faculty, patron, or institutional usage but instead was linked to broader impacts and public good.

Use Case 3.

Understand and support OA resource use within the institution / on campus

Personas



- a. Provide eBook usage reporting to complement reports for non-OA titles
 - i. Aggregate report for institution
 - ii. **Aggregate** report for author
 1. Overall cross-platform usage by author
 2. Chapter level reporting for authors in compilations
 - iii. Aggregate report for specific book
 - iv. Aggregate report for specific chapter
- b. Understand whether OA material is being referenced in class reading lists
 - i. e.g. via a learning management system like Canvas, TELUS or Moodle
- c. Provide usage data reports to campus units
 - i. Faculty Senate: to illustrate breadth and absence of activity
 - ii. Academic Affairs: to establish levels of OA activity by discipline
 - iii. Graduate School: to establish levels of OA publishing activity
 1. among advising faculty
 2. among graduate students
 - iv. Research Office: to indicate impacts of OA publications
 1. Story-telling by discipline or domain
- d. Support or train faculty and/or students creating their own OA impact reports
 - i. Self-service
 - ii. Supporting workshops
- e. Promote scholarship by local authors
 - i. Illustrate local and regional impact
 1. For books
 2. For chapters
- f. Support collaboration and/or contributions with university press
- g. Differentiate between access to full text or just TOC or Abstract

Use Case 4.

Inform collections development strategy

Personas



- a. Understand how library patrons use OA collections: i.e. which titles are being used and from which providers
 - i. Track demand in a collection to determine new things to support (e.g. understand user needs)
 1. Understand need for new acquisitions
 2. Support analysis of emerging models of acquisitions
 - ii. Inform what is cataloged
 1. e.g. if some of your users are accessing ebooks through a given provider that is not included in your library's discovery systems, you will want to include that provider in your systems so that your other users can also discover ebooks that way.
 - iii. Inform acquisitions
 - iv. Inform expenditures
- b. Inform decisions on who to pay for access (direct with publisher or via aggregator)
 - i. Evaluate OA content access overlaps and niches across vendors
 - ii. Understand access on specific platforms for content from a given press/publisher
- c. Inform acquisition strategy with usage as evidence
 - i. Identify items for purchase
 - ii. Inform disciplinary collection development based on access patterns

Use Case 5.

Collect, aggregate, and prepare data for reporting

Personas



- a. Gather usage data from multiple sources
- b. Input usage data into library's own data warehouse or research management system
- c. Create data visualizations
- d. Set up APIs to display usage data on websites, Tableau, or other data compilations
- e. Create usage data dashboards for campus audiences
 - i. Library
 - ii. Research office

BOOK PUBLISHING PLATFORMS AND SERVICES OAEBU DATA USE CASES

A variety of book publishing platforms and services apply OAeBU data in support of internal operations and technical development while also provisioning it to meet their customers' demand for usage data reporting. Staff roles that may interact with OAeBU data span sales and marketing, business and data analysts, IT specialists, publisher relations and community managers.

Challenges faced when working with OAeBU data arise from OAeBU data variability and the multiple processing approaches in use across the OAeBU data supply chain used to manage issues relating to book vs. chapter level data, bot vs. human access, and duplicate downloads.



IT, Data Analysts
& Product Dev.



Customer Member Support
Sales & Business Dev.

Use Case 1.

Inform publishing platform/service marketing strategy and PR

Personas



- a. Contextualize search engine traffic
- b. Evaluate SEO efforts
- c. Understand usage tied to current events. i.e. when x happened, what happened to usage data
 - i. Provide custom usage reports tied to staff provided keyword search, e.g. "immigration"
 - ii. Identify usage spikes in real-time
Note: need to view reports
 - iii. for full OA collection
 - iv. For specific sub-collections
- d. Report to major funders, stakeholders

Use Case 2.

Inform publishing platform/service sales strategy

Personas



- a. Inform differential pricing based on usage levels
- b. Identify leads for particular services based on usage, e.g. specific levels of preservation service
 - i. Publishers
 - ii. Titles
 - iii. Chapters
- c. Use OA book usage trends to inform for other parts of the business
 - i. Use country/type usage data to suggest relevant new titles (discipline based)
 - ii. journals
 - iii. chapter traction could inform edited/special editions

Use Case 3.

Manage and curate usage data

Personas



- a. Support usage data provision
 - i. Combine usage data from aggregators, customer/member publisher, and social media
 1. Find and address problems in usage data streams
 - a. Clean data real-time according to community-set rules/standards (e.g. COUNTER)
 - i. Support AI that reviews incoming data
 - b. Clean data after the fact
 2. Data versioning: Track changes made during “cleaning” process
 3. Apply APIs from third party systems
 4. Manage non-standardized data
 5. Manage “bad-behavior” / inconsistent approaches in the numbers while recognizing inconsistency might result if non-standardized approaches are applied to data cleaning
 - ii. Incorporate and apply usage data related definitions, attributes, and standards
 - iii. Develop open “clean” data, i.e. apply COUNTER processing data rules as defined at: <https://www.projectcounter.org/code-of-practice-five-sections/7-processing-rules-underlying-counter-reporting-data/>
 - iv. Identify missing data definitions to create data definitions in line with COUNTER extending code practice (<https://www.projectcounter.org/code-of-practice-five-sections/11-extending-code-practice/>) and notify COUNTER of missing data in its current definitions for consideration for inclusion in the next release.
- b. Implement APIs with third-party system delegates of member/customers to support analysis with non-usage data sources
- c. Implement security and privacy controls for personal data within usage datasets
 - i. IP addresses
 - ii. Individual identifiers
- d. Document all steps, definitions, preservation plan in full Data Management Plan
 - i. Make data provenance transparent for system users

Use Case 4.

Facilitate technical development of platform/service

Personas



- a. Support/develop usage data dashboards or visualizations
 - i. Templates for commonly requested views
 1. Growth of library (OA book count over time)
 2. Growth of publishers (# of members using usage service over time)
 3. Languages in OA corpus over time
 4. License versions in corpus over time
 5. Geographic usage
- b. Produce customized usage data reports
 - i. When x happened, what was the impact on usage
- c. Manage tool use on the platform to collect / manage usage-related data
- d. Inform design and development of features, widgets for usage data
 - i. Update APIs to support member/customer demand
 - ii. Manage situation if members define/treat chapter/book level info differently
 - iii. Manage situation if book and chapters/parts/supplemental materials are hosted individually
 - iv. Understand whether users want to see reports rolled up at a certain level.
- e. Evaluate how technical development impacts book usage and traffic
 - i. E.g. does moving download button improve usage
- f. Inform future development by understanding usage data service/product demand by member/customers
 - i. Develop potential financial use cases
 - ii. Reveal previously unknown uses or users, new or unexpected audiences
- g. Inform/prompt evaluation of why a service isn't being used
 - i. E.g. no one queries a specific API
- h. Support risk management planning tied to demand surges
 - i. Develop action plans for traffic surges
 1. Support increased demand for member/customer usage data if demand surges
 - ii. Develop strategies to manage/reduce costs associated with publication file types

Use Case 5.

Member/customer support, account management

Personas



- a. Support member/customer use of usage information
 - i. For publishers
 1. Inform publisher interfaces for usage information (direct)
 2. Support publisher's author interfaces for usage information (pass-through / white-label)
 3. Help publishers understand how to use usage data once they obtain it (see 4.c)
 4. For discovery platforms, illustrate:
 - a. how adding books to discovery platform impacts traffic back,
 - b. demand on servers,
 - c. how customer/members/libraries are getting to books e.g. directly, via aggregators
 - d. Impact on usage if book isn't in a certain directory
 - ii. For authors
 1. Inform/provide author interfaces for usage information (direct)
 - iii. Libraries
 1. Support library interfaces for usage information (direct)
 - a. Surface local use and impact
 - b. Show local connections
 - i. Author affiliation
 - ii. Local geographic focus
 2. Support librarian role in informing/driving local usage
- b. Help member/customer evaluate services/products based on usage
- c. Help member/customer place usage in historical/relative contexts
 - i. Help them address "how does Report A saying X compare to Report B saying Y"
 - ii. Explain limitations to data and aid data interpretation: e.g. analyst explaining what data means and does not mean
 1. Longitudinal usage
 2. Usage relative to other media/content
 3. Usage relative to other platforms
- d. Use usage data to help member/customer improve metadata
 - i. To get better value
- e. Support member/customers if usage surges
 - i. Track usage surges by country
 - ii. Track usage surges by member/customer
- f. Understand where to focus/develop additional member/customer support based on usage
 - i. Identify regions for internationalization based on usage
 1. New language support
 2. Region specific member/customer support

Notes: Common definitions tied to standards (e.g. COUNTER, ONIX, SUSHI) should be clear for DOI, book, chapter, audience, customer.