



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

Christina Drummond

Kevin Hawkins

January 2022

ACKNOWLEDGEMENTS

This work was made possible by The Andrew W. Mellon Foundation through the *Developing a Data Trust for Open Access Ebook Usage* grant¹ 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 created this report by translating the meeting and asynchronous ideation outputs generated in 42 virtual boards through 499 individual participation sessions. A draft was shared with participants for public comment via the project's communities of practice and social media.³ Public comment was incorporated into this document, with notes and quotes offered by community members highlighted. The use cases for usage data that follow, grouped by stakeholder type, describe why usage data is important and how it is applied for a given use case. Specific questions or queries of usage data that surfaced in discussion are noted.

OA EBOOK USAGE DATA USE CASES TABLE OF CONTENTS

i. Scholar: Author, Editor, Hiring/P&T	3-6
ii. Commercial Publisher: OA Program, Marketing, Sales, Editorial	7-12
iii. Library: Dean/University Librarian, Scholarly Communication, Collections Development, Liaison Librarian, IT	13-16
iv. University Press: Director, Sales, Marketing, Editorial, Acquisition, Research Administration	17-24
v. Book Publishing Platforms & Services: IT, Data Analysts, Product Development, Support, Sales	25-28

1 https://educopia.org/data_trust/ and <https://zenodo.org/communities/oaebu>

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

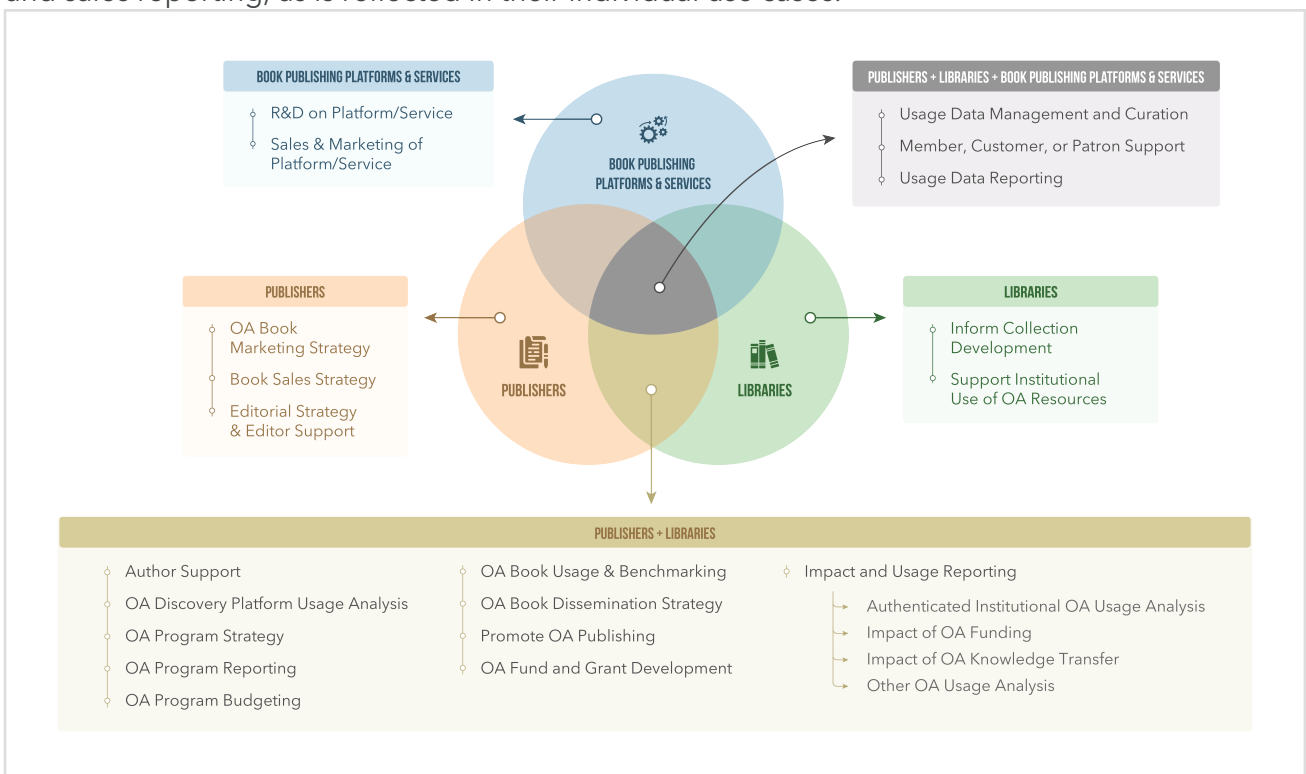
3 The OAEBU project's processes for communities of practice development and use case development are published in Drummond, Christina (2020). "Engaging Stakeholder Networks to Support Global OA Monograph Usage Analytics," Collaborative Librarianship: Vol. 12 : Iss. 2 , Article 9. <https://digitalcommons.du.edu/collaborativelibrarianship/vol12/iss2/9>

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, APIs, dashboards, and spreadsheets. These institutions individually manage, compile, and link usage-data metrics, some of which are provisioned via dashboard services for users ranging 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 to refer to web analytics 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 investment, strategy and fundraising in OA programs. They noted how data on the location and context of OA ebook access can inform book marketing and dissemination, acquisitions and collections development, and scholars’ promotion and tenure. OAeBU data 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 models for OA book publishers, from university or library-based to commercial OA teams, surfaced common publisher functional needs of OAeBU data as shown 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 OPEN ACCESS EBOOK USAGE (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 promote 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 or 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



Why | How

a. Inform planning for a new book to frame book pitch

- i. to understand how people discover, download, purchase similar books

b. Evaluate potential ways to host/disseminate the book

- i. by seeing where readers are downloading OA versions from
- ii. to understand how people discover OA versions via
 1. social media
 2. marketing campaigns
- iii. to understand whether individuals know an OA version exists

c. Understand the relationship between OA downloads and print sales

- i. to know if an OA version drives purchasing
- ii. to understand how people interact with sample pages

Use Case 2.

Understand audience demographics

Personas



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

a. Understand a current audience for an eBook

- i. by geographical region
- ii. by institution
- iii. by discipline or field
- iv. academic vs. non-academic use
 1. by academic institutions
 - a. use within my institution
 - b. use outside of my institution
 2. by news media
 3. by the general public
 4. by government or public policy stakeholders
- v. by academic role
 1. students
 2. scholars

b. Surface and assess potential new audiences based on similar or prior book performance

- i. to identify new regional audiences, translation opportunities
- ii. to identify new field audiences

Use Case 3.

Understand scholarship impact

Support varied metrics to complement traditional sales metrics

Personas



a. Understand the use of OA eBooks

- i. for academic use
 1. in scholarship
 2. within class syllabi
 3. for student online projects
 4. within blog posts
- ii. in non-academic and public use
 1. by news media
 2. in social media
 3. on Wikipedia
 4. for public policy
 - a. by nonprofits or in grey literature

b. Compare the impact of eBooks to print editions

c. Understand how online usage over time compares to sales and citation activity

- i. to see changes over time
- ii. to track if OA is becoming more mainstream

d. Understand differences in book branding

- i. in different fields or disciplines
- ii. in different geographic communities

Use Case 4.

Evaluate book marketing and outreach

Personas



a. Evaluate impact of effort spent to connect with readers

- i. to understand where page visits or page views originate, e.g. a Facebook post, tweet, etc.
- ii. to understand the impact of promotion activities
 1. e.g. *"Does timing of a promotional social media or speaking engagement by the author result in a near-term increase in downloads or page views?"*
 2. e.g. *"What OA eBook activity do promotions result in, e.g. downloads, citations, mentions, etc.?"*
 - a. understand what occurred, where, in what quantity

b. Understand advertising and marketing campaign impacts

- i. to evaluate if marketing campaigns are effective
- ii. to inform promotional tactics and strategy

Use Case 5.

Understand financial investment in OA

Personas



a. Understand who is investing in OA

Use Case 6.

Inform editorial decisions on what to publish

Personas



a. Explore whether something is niche or of broader interest

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 author recognition

Use Case 8.

Surface ways to increase scholarship accessibility

Personas



- a. Identify more diverse audiences
- b. Inform public dissemination
- c. Support plural or translational scholarship across disciplines

Use Case 9.

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

Personas



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

Use Case 10.

Guide and inform impact factor use for evaluation

Personas



"Don't exacerbate existing inequalities."

- a. Understand how OA eBook usage data complements other impact factor metrics
- b. Understand and manage disparate impacts
- c. Understand what author-affiliated data is available to whom, to inform privacy and ethical guidelines over access and use of such data
"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. by providing evidence of activity tied to events
 - ii. by showing imbalance of activity compared to similar titles or other authors
- b. Understand connections between online presence and dark-web activity

COMMERCIAL PUBLISHER OA EBOOK USAGE (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 data against competitor benchmarks or across the OAeBU data landscape.



OA Program



Marketing



Sales



Editorial

Use Case 1.

Inform internal OA program strategy

Personas



"Which books in which editorial departments could benefit the most from efforts to diversity and increase readership?"

"Did the OA version help those who'd benefit from using the book access it?"

a. Provide usage reports

Why | How

- i. to demonstrate OA strategy, importance, and impact to internal audiences
- ii. to show eBook usage by audience and discipline over book lifespan
- 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 Open Access impact on expanding readership
 1. by illustrating whether an OA version increases book readership and access over time
 2. by illustrating if a book about a region was used within that region
 3. by adding context for books in smaller topic areas based on comparable usage activity in relevant topic areas
 4. by surfacing potential disciplinary crossover based on usage of comparable titles
 - a. across disciplines
 - b. within target audience communities

"Show what someone can expect from working with us."

"The ability to drill down into the data is important!"

"Each author and book has its own niche audience."

"I'd like to be able to say here's how we expect the OA content to be accessed..."

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

b. Understand eBook audiences

- i. to establish OA program profile or niche
- 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 the OA program's strengths and historical usage patterns
e.g. "What is the historical usage of chemistry titles like this in this specific geographic region?"

c. Compare OA program activity against other OA publishing programs via industry benchmarks

- i. for peer publishing programs, i.e. publishers of roughly the same size and type
- ii. for 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
e.g. "For books on topic X with a target audience in region Y, this is what OA usage looked like?"
- ii. to project usage based on past trends for similar books
 1. by responding to authors asking, "When can we expect peak engagement based on usage trends?"
- iii. to understand duration of book lifespan
 1. by illustrating 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 book stakeholders of expected engagement patterns and book impacts

Use Case 2.

Generate external funding impact reports

Personas



"Move the dialogue from 'we published it' to 'here's how the OA version affected the impact.'"

"Support reporting requirements that require universities to show how they achieve impact."

Note: Frequency and granularity of data are important.

Work is needed to define terms and metrics for lifespan, measures, and peak activity.

a. Communicate impacts to stakeholders

- i. by reporting usage by
 1. author institutional affiliation
 2. title
 3. source of download (platform)
 4. geographic region of access
 5. a publisher's core or traditional audiences
 6. author or funder targeted audiences
 7. Note: Each book may have its own success metrics in terms of what its author wants to achieve by going OA.
- ii. by generating reports for
 1. research funders
 - a. usage and engagement tied to
 - i. research area or 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 societies
 - b. research institutes
 - c. charities/NGOs
 - d. corporations
 - e. libraries and library consortia
 3. authors and their research offices
 - a. to establish compliance with funding mandates, e.g. "Support reporting requirement for UK Research Excellence Framework"
 - b. to show 'real world' or broader impacts

Use Case 3.

Support marketing campaigns and strategy

Personas



"The goal is to connect usage activity to promotional campaigns."

Note: campaign may refer to email and/or social media campaigns.

a. Support marketing of OA program

- i. to encourage OA adoption by showing advantages to publishing OA compared to non-OA
 1. compare OA vs. non-OA usage, e.g. *"How do OA books perform compared to non-OA counterparts in this subject area?"*
 2. compare OA vs. non-OA discoverability
 3. illustrate what happens when a book is "flipped to OA"
- ii. to illustrate the global profile, reach, niche of an OA program
 1. show "what one can expect from working with us"
 2. by institution type, region, usage activity
- iii. to develop white papers for stakeholders
 1. industry
 2. higher education
 3. book-series partners and/or funders
 4. authors
 5. funders

b. Support promotions for OA book content and OA authors

- i. to show how promotions affect targeted audience usage by examining activity by date or date range by institution type (e.g. libraries, schools) and regional split by subject
- ii. by understanding content usage and engagement over time, e.g. *"Is usage ongoing? Is something "adopted" or "continuously read?"*
- iii. by providing authors with usage reporting
 1. generating custom, personalized book performance reports showing specific usage on platforms
 - a. following promotion campaigns or launch events
 - b. related to social media activity
 2. providing self-service reporting access so authors can check their own usage directly via
 - a. author-facing dashboards
 - b. automated "personalized book performance" emails

c. Support author recognition

- i. to trigger *"Congratulations - your book is one of the most downloaded"* notifications
- ii. to suggest marketing campaigns based on
 1. top downloaded books
 2. most cited books

Use Case 4.

Inform sales strategy

Personas



"Is there a trend of print sales going up or down after becoming OA?"

Does the trend vary by discipline ?

a. Provide data on what to expect when flipping a title OA

- i. to understand changes in usage after making content OA on a limited basis, e.g. *"What was the impact of making books OA in response to COVID-19 from date X to date Y?"*

b. Illuminate OA impact (+/-) on sales revenue over time

- i. by viewing OA-related sales trends for the time spanning before and after going OA to answer questions
- ii. to understand follow-on print sales related to OA access
- iii. to understand follow-on eBook 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 to support sales team offering bundled OA access e.g. *"If there is an eBook package, librarians want to know if they can include access for a discovery POV."*

d. Inform print-edition compensation strategy based on OA usage

- i. to understand if OA access drives print purchases, e.g. at libraries

e. Inform sales strategy for target audiences

- i. to determine discovery platform fit, 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. by providing benchmarked usage data for a book's audience and topic to understand and gauge usage, e.g. *"Is it 'going well' or is it higher or lower than expected?"*

c. Support acquisitions by providing authors with usage reports

- i. to establish the reach of a publisher's OA book dissemination
- ii. to complement print edition sales data
- iii. to support author requests for data demonstrating OA benefits

Use Case 6.

Track OA discovery platform usage activity

Personas



"It is important to show that the target audience successfully accessed the eBook, not just that content was available OA."

a. Ensure OA versions are being discovered

- i. to understand whether OA version is being accessed
- ii. to understand how OA version is being accessed by target audiences

b. Assess ROI for discovery and hosting platforms and services

- i. by comparing usage, impact, and citations across platforms "in the wild"
- ii. to illuminate platform-related usage by target audiences segmented by region, usage "size", discipline
- iii. by surfacing benefits of working with specific dissemination platforms
 1. e.g. "How well does platform X reach our target audience?"
 2. e.g. "Which platform could do the best job if we have to prioritize?"

Use Case 7.

Work towards usage data standards compliance

Personas



a. 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

Notes regarding COUNTER usage data

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

LIBRARY OPEN ACCESS EBOOK USAGE (OAEBU) DATA USE CASES

Libraries use OAeBU data to report the impacts and usage of OA books 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 inform book benchmarking and dissemination strategy for librarians. 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 eBook 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, publishers, and 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



"Faculty data can inspire students to publish OA."

Why | How

- a. **Provide data to support OA evangelism and advocacy aimed at encouraging people to publish OA**
 - i. to show examples of audience reach and platform-specific access for
 1. prior OA publications by authors in similar field
 2. open educational resources.
 - ii. to support outreach to library patrons and constituents
 1. PhD students
 2. faculty
 3. state or provincial library programs
 4. alumni or public patrons
 - iii. to support fundraising

Use Case 2.

Support budget planning for OA resources and initiatives

Concerns exist over the potential for OA costs to cut into collections budgets. Context is very important.

Quarterly data can surface trends while annual data is important for reporting.

Personas



"If no one is using it, would funding be better allocated somewhere else?"

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

a. Articulate the value proposition for OA

- i. on campus
 1. as impacts on student outcomes
 2. as impacts on faculty outcomes
- ii. in the local community
 1. as impacts of regional access
 - a. by alumni
 - i. for degree programs
 - ii. for continuing education or 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 percentage of the collections budget

d. Inform discussion of what merits OA budget investment

- i. by establishing the 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 or broader impacts of OA publishing

e. Understand OAEBU across different websites, i.e. platforms, services, and publishers

- i. to inform discussion of which platforms to support, i.e. to evaluate paying for memberships or services
- ii. to understand the volume of faculty OA publishing
 1. generally, across platforms
 2. through a specific access point, hosting or aggregation platform for OA books
 3. to articulate the cost per use per platform or service
 - a. by authenticated institutional users (patrons)
 - b. within a local geographic region or city

Use Case 3.

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

Personas



- a. **Provide eBook usage reporting to complement non-OA title reports**
 - i. by reporting aggregated usage by institution
 - ii. by reporting aggregated usage by author
 1. with overall cross-platform usage by author
 2. with chapter level reporting for compilations
 - iii. by reporting aggregated usage for a specific book
 - iv. by reporting aggregated usage for a specific chapter
- b. **Understand whether OA material is being referenced in class reading lists**

e.g. "Is an OA eBook being used on 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 the impacts of OA publications through discipline or domain storytelling
- d. **Support or train faculty and/or students on how to create their own OA impact reports**
 - i. by providing self-service OA usage reports
 - ii. by supporting workshops for faculty or students
- e. **Promote scholarship by local authors**
 - i. by illustrating local and regional impact
 1. of books
 2. of chapters
- f. **Support collaboration and/or contributions with university press**
- g. **Understand what elements were accessed**
 - i. by differentiating between usage of Abstract, Table of Contents, and/or full text

Use Case 4.

Inform collections development strategy

Personas



"Which titles are being used from which providers?"

a. Understand how library patrons use OA collections

- i. by tracking 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. to inform what is catalogued.
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. to inform acquisitions
- iv. to inform expenditures

b. Inform decisions on who to pay for access, e.g. direct with publisher or via aggregator

- i. by evaluating OA content access overlaps and niches across vendors
- ii. by understanding access on specific platforms for content from a given press or publisher

c. Inform acquisition strategy with usage as evidence

- i. to identify items for purchase
- ii. to 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, in Tableau, or other in data compilations

e. Create usage data dashboards for campus audiences, e.g. Library, and Research office

UNIVERSITY PRESS OA EBOOK USAGE (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



"I'd like to know what a reader might want from a product to then support and drive further customer interaction."

a. Understand eBook audiences and end uses

Why | How

- i. by geographic distribution
 1. to identify where there is niche content interest
 2. to understand how books are used in specific areas of the world
 3. to view what usage is like in places where OA eBooks are the only accessible option
- ii. by niche usage communities
 1. within universities
 - a. to report disciplinary access and use
 - i. by faculty
 1. in teaching, e.g. Accessing one or all chapters, or cited in a syllabus on opensyllabus.org
 2. in scholarship / research
 3. for personal interest
 4. to understand tenured vs. non-tenured usage differences
 - ii. by graduate-level students
 1. in class
 2. for scholarship
 - iii. by undergraduate students in class
 - iv. by librarians

1. in reference, advisory, and support capacity to faculty and students
2. as book purchasers
- b. to report use to university administration
- c. to report use among alumni and the broader campus community
2. within primary/grade schools (K-12)
 - a. by students
 - b. by educators
3. within industry
 - e.g. measure of access via company URLs
4. within news media and by journalists
5. within the government
 - a. impacts and references for policy-making
 - b. at multiple levels, e.g. national, state or province, city or locality
6. within other institutions
 - a. by understanding usage at institutions with an interest or specialty in particular subjects
 - b. by viewing which institutions are accessing what content
 - c. by understanding intra-institutional vs. non-institutional access
7. within the 'public' such as
 - a. local public in the university press' region
 - b. *"To show local impacts on the state and community where the press is based."*
 - c. advocates or enthusiasts for a defined subject, e.g. *"Such as gamers or environmental campaigners."*

b. Ensure fair comparisons of institutional usage

- i. by normalizing institutional data by
 1. faculty count
 2. full-time enrollment
 3. international classification systems, e.g. [World Bank country classifications by income level](#)
 4. National classification systems, e.g. [Carnegie classifications](#) - intra-USA comparison
- ii. by normalizing geographic data by
 1. per capita / population, e.g. [World Bank population estimates and projections](#)

c. Understand usage patterns and trends

- i. to investigate the demand for print copies of OA books
- ii. to understand format access patterns and frequency for OA eBook content to support reading behavior
 1. e.g. *"Do readers appear to prefer HTML or PDF?"*
 2. e.g. *"Do readers prefer versions with or without comments by other readers, by reviewers, etc.?"*

"Practices for normalizing and presenting data are still in development, but are required to ensure fair comparisons across institutions."

"Less inference is better"

3. e.g. "Do seasons or academic calendars impact usage?"
4. e.g. "Do readers prefer to read books like this cover to cover, or do they dip in to look for citation/passage?"
5. e.g. "Do readers tend to access one chapter, all chapters, >1 chapter, etc.?"

d. Understand AI and non-human usage

- i. for text or data mining
- ii. for reposting
- iii. for portal inclusion or use
- iv. for pirate site inclusion or use
- v. for abstracting
- vi. for indexing
- vii. other active uses

e. Understand OA as a discoverability tool

- i. that generates additional activity
 1. print transactions
 2. premium e-transactions
- ii. by understanding how OA books are presented on the sites librarians use to acquire books such as GOBI, ProQuest, OverDrive, BorrowBox
- iii. by understanding how OA books are presented on sites that readers use to discover scholarly content such as Google Scholar, Google.com, Academia.edu, ResearchGate
- iv. to understand what a reader might want from a product to then support or drive further customer interaction, e.g. "analog to index"

Use Case 2.

Inform sales strategy and understand OA impact

Personas



a. Suggest titles that should become OA

- i. by using comps similar to decisions based on past sales data
e.g. "Similar to how sales trends inform future potential."
- ii. by identifying geographic niches for future OA publication
e.g. "Where there is a lack of retail distribution chain but there was past demand or use of previous editions."

b. Understand the OA impact (+/-) on sales revenue

- i. by testing theories of OA impact on sales
- ii. by surfacing OA usage in Global North vs. Global South

c. Understand the amount of OA funding required to meet anticipated sales figures

d. Surface local edition or translation opportunities based on geographic use

e. Inform marketing and sales resource allocation by incorporating usage data alongside print or eBook sales data

Use Case 3.

Provide basis for exploring assumptions about OA impact

Personas



- a. Study how OA impacts gated content access
e.g. "Is paid content out in the world in the same way?"
- b. Study how OA impacts access to knowledge

Use Case 4.

Editor relations

Personas



- a. Support development of tool-kits for editors so they can be sophisticated in their use of data
- b. Provide reports to help track impact-based editor acquisitions and goal performance.
e.g. "Show the progress made on acquiring books with x amount of impact."

Use Case 5.

Analyze or benchmark OA books

Personas



- a. Examine geographic impact or use of OA books compared to similar non-OA books
 - i. by geographic range of OA readers vs. non-OA readers
 - ii. for geographic regions that OA reaches but non-OA does not
 - iii. using the percentage of total use in different regions
- b. Compile qualitative or alt-metrics for OA compared to non-OA
 - i. to provide soft measures of disciplinary impact
 1. with related prizes or awards
 2. with related reviews
 3. with related media attention
 4. with related references in tenure or promotion files
 5. with related references in disciplinary blogs or listservs
 6. with related references in public policy or grey literature
 7. with related statements from readers or users on why they're using the book, what they'd going to do with what they've learned, etc.
- c. Compile quantitative metrics for OA compared to non-OA
 - i. for readership, e.g. "Count or quantity of readers"
 - ii. for scholarly citations
 1. as quantity of citations, i.e. count
 2. by frequency of citation

"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."

"Examples of complementary data sources that relate to usage data to illustrate impact include Crossref Event Data and Altmetrics."

- iii. for downloads
- iv. for in-browser views, e.g. "Internet Archive tracks"

d. Understand online reader engagement and references

- i. on social media, e.g. Twitter
- ii. on discussion forums, e.g. Reddit
- iii. in crowd-sourced reference resources, e.g. wikipedia
- iv. on websites or blogs

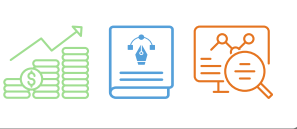
e. Provide comparative metrics (i.e. apples to apples)

- i. to illustrate how a book performs compares to books in similar disciplinary area
 - 1. e.g. "As defined by BISAC subject code."
 - 2. e.g. "As defined by internal subject taxonomy."

Use Case 6.

Inform editorial strategy

Personas



a. Inform printed book version based on OA usage

- i. by providing data to influence decision to make a print version
- ii. by understanding the format and audience for a print version
e.g. "Is this an academic or crossover trade book?"

b. Understand course potential

c. Inform consideration of which lists, series, and subject areas that could be made OA

d. Understand usage to identify prospective authors for future projects

- i. at the chapter level, especially for certain types of monographs, e.g. anthologies, readers, edited volumes
- ii. for a given author
- iii. in a given discipline(s)
- iv. for particular components or resources in titles with multimedia components

e.g. "How are supplemental multimedia components being used separately from the book like a song recording in an open textbook?"

e. Identify thematic areas for acquisitions based on usage data

"Things that are valuable to small number of scholars could be overlooked or negatively impacted if editors focus on the 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



"Qualitative data is especially important to illustrate impact alongside usage metrics."

- a. **Provide funders with usage information**
 - i. to show compliance with federal or national OA dissemination mandates
 - ii. to illustrate a funder's return on investment for OA initiatives through their grantees' impact or usage
 1. across all grants
 2. for a specific grant
 - iii. to show advancement or impacts on a funder or grant program's particular objective, such as advancing K-12 or contributing to an informed citizenry/public
 - iv. to illustrate other 'broader impacts'
 - v. to calculate or understand the impacts of up-front OA investment downstream
 1. e.g. for OA investments in Open Educational Resources
 - a. by passing savings on to students
 - b. by improving equitable access to educational resources and 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?**

"This may inform or steer list direction over time."
- d. **Understand the editorial impact on authors of the CC licenses required by different funders**
- e. **Show the changes in usage or impact for a title when it goes OA, i.e. "Before and after comparison."**
- f. **Show the impacts of OA policies and investments on**
 - i. scholarly publishing practices by understanding the degree to which public subsidies mitigate publisher risk and costs related to making books OA
 - ii. equity and social justice
 1. by understanding the potential negative or positive impacts of OA publishing requirements on non-Western European or North American researchers
 2. by understanding whether OA funding models push OA in opposition to equity/justice objectives
"Can you surface negative network externalities?"
 - iii. the [UN Sustainable Development Goals](#)
 - iv. access to information
 1. e.g. *"Is valuable research being ignored or overlooked by those who can only afford 'free'?"*
 2. e.g. *"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

- i. on the advantages and disadvantages of OA options in their discipline
- ii. to help them understand OA impacts, value, and ROI
- iii. to surface sources of OA funding alongside other sources of income, e.g. *"Putting OA sources alongside sales, treating each book project with holistic sources of income."*
- iv. to show the impact of citations on usage, e.g. *"This is how much usage of book x comes from citations within other online works."*

b. Support authors participating in OA initiatives

- i. by providing impact information for their individual books
 1. on trends for how the books are used around the globe, e.g. *"Who is using the book, where, and how?"*
 2. on usage by the subjects of their books, such as use by the specific communities or populations studied
 3. on usage in under-served communities
 4. on crossover usage across and between disciplines
 5. on engagement by professionals and advocates outside the academy
 - a. within government or public policy-makers
 - b. by industry researchers
 - c. within not-for-profit organizations
 6. on author defined "impact" metrics, such as those specific to
 - a. the author's discipline
 - b. the book's subject community
 - c. relevant policy or public affairs

Use Case 9.

Research or grant development and administration

Personas



a. Support proposal development

- i. for digital publishing grants
- ii. for host university institutional support or matching requests for OA initiative related grants
- iii. to provide requested data such as
 1. percentage of eBooks used compared to respective hard copies
 2. eBook subject matters (including BISAC codes)
 3. who is using the eBooks
 4. how are the eBooks being used
 5. how OA affects sales (+/-), e.g. *"They're trying to figure out how much it costs to publish OA and what's scalable, sustainable, subvention amount."*

"Go beyond the numbers to tell a compelling case that OA makes a difference."

Use Case 10.

University administration reporting and relations

Personas



b. Inform impact stories for advancement or donors

- i. by creating reports for types of donors such as
 1. intra-university donors or sponsors
 2. external organizational donors, e.g. corporate foundations
 3. individual donors, both new and repeat

c. Support and inform partnership development for new initiatives

- i. with other local entities and partners
- ii. with other publishing platforms and services

a. Support career advancement, promotion, or tenure cases

b. Report on the impacts of institutional OA funding

- i. by demonstrating the global reach of the university press imprint and OA activities
- ii. to show distribution beyond where the university press normally disseminates
- iii. to 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 the value provided through 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)."

BOOK PUBLISHING PLATFORMS AND SERVICES

OPEN ACCESS EBOOK USAGE (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 staff, business and data analysts, IT specialists, publisher relations staff, 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 to manage issues relating to book vs. chapter level data, bot vs. human access, and duplicate downloads.



IT, Data Analysts, and Product Development



Customer or Member Support, Sales, and Business Development

Use Case 1.

Inform publishing platform or service marketing strategy and communications

Personas



Why | How

- a. Contextualize search engine traffic
- b. Evaluate Search Engine Optimization (SEO) efforts
- c. Understand usage tied to current events.
 - i.e. "When x happened, what happened to usage data?"
 - i. to provide custom usage reports based on keyword search, e.g. "immigration"
 - ii. to identify usage spikes in real time
 1. for the full OA collection
 2. for specific sub-collections
- d. Report to major funders, stakeholders

Use Case 2.

Inform publishing platform or service sales strategy

Personas



- a. Inform differential pricing based on usage levels
- b. Identify potential customers for particular services based on usage, e.g. specific levels of preservation service
 - i. by reporting aggregate usage by publisher
 - ii. by reporting aggregate usage for titles
 - iii. by reporting aggregate usage for chapters
- c. Report usage trends to other parts of the business, e.g. journals
 - i. by country and type of usage for a given discipline, to suggest relevant new titles
 - ii. by using chapter data to inform edited/special editions

Use Case 3.

Manage and curate usage data

Personas



a. Support usage data provision

- i. by combining usage data received from aggregators, customer or member publishers, and social media
- ii. by finding and addressing problems in usage data streams
 - a. when cleaning data in real time according to community established rules or standards, e.g. COUNTER
 - b. when supporting or developing AI that reviews incoming data
 - c. to clean data after the fact

b. Support data versioning, i.e. tracking the changes made during the “data cleaning” process

- i. by applying APIs from third party systems
- ii. to manage non-standardized usage data
- iii. to manage “bad behavior” or inconsistent approaches to usage data processing, recognizing inconsistency might result if non-standardized approaches are applied to data cleaning

c. Incorporate and apply usage data related definitions, attributes, and standards

- i. to develop open “clean” data, i.e. apply COUNTER processing data rules such as <https://www.projectcounter.org/code-of-practice-five-sections/7-processing-rules-underlying-counter-reporting-data/>
- ii. to identify missing data definitions and suggest data definitions extending the COUNTER code practice such as <https://www.projectcounter.org/code-of-practice-five-sections/11-extending-code-practice/>, i.e. notify COUNTER of missing data in current definitions for consideration in the next release

d. Implement APIs with third-party system delegates of customers or members to support analysis with non-usage data sources

e. Implement security and privacy controls for personal data within usage datasets such as IP addresses or individual identifiers

f. Document all steps, definitions, and preservation plan in usage data management plan to make data provenance transparent for system users

Use Case 4.

Facilitate technical development of platform or service

Personas



a. Develop usage data dashboards or visualizations

- i. to provide templates for commonly requested views
 1. of changes in library usage, i.e. OA book count represented on platform or service over time
 2. of changes in publisher usage of platform or service for OA, e.g. number of members using platform or service over time
 3. of languages represented in OA corpus over time
 4. of license versions represented in corpus over time
 5. of geographic usage
- ii. to provide custom usage data reports for customers or members
e.g. *"When x happened, what was the impact on usage?"*

b. Design, develop, and support usage-related platform or service features

- i. e.g. APIs that support member or customer demand
- ii. e.g. widgets or tools that collect usage-related data
- iii. to manage situation when members or customers define or process chapter or book level data differently
- iv. to manage situation when a book and its chapters, parts, or supplemental materials are hosted separately
- v. by understanding how users want to see reports rolled up

c. Evaluate how technical platform or service developments impact book usage and traffic

e.g. *"Did moving the download button improve usage?"*

d. Inform future development by understanding usage data related service or product demand by members or customers

- i. to develop potential financial use cases
- ii. by revealing previously unknown uses, users, and/or audiences

e. Inform or prompt an evaluation of why a service isn't being used

e.g. *"No one queries this specific API"*

f. Support risk management planning tied to demand surges by informing action plans to support temporary increase in demand for usage data

g. Inform strategies to manage and/or reduce costs associated with usage related to specific publication file types

Use Case 5.

Customer or member support and account management

Personas



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

a. Support customer or member use of usage information

- i. for publishers
 1. to inform publisher staff interface for usage data, i.e. direct
 2. to support publishers' author interfaces for usage data, i.e. pass-through or white-label websites
 3. to help publishers understand how to use usage data once they obtain it
- ii. for discovery platforms
 1. to show how adding books to discovery platform impacts traffic back
 2. to manage demand on servers
 3. to understand how customers, members, and/or libraries are getting to books directly or via aggregators
 4. to understand impact if a book isn't on a certain platform or service
- iii. for authors seeking self-service
 1. by informing or providing author interfaces for usage data, i.e. direct
- iv. for libraries
 1. to support library interfaces for usage data, i.e. direct
 - a. by surfacing local use and impact
 - b. by showing local connections
 - i. by author affiliation
 - ii. to local geographic region
 2. to support librarians informing or driving local usage

b. Support customer or member evaluation of services or products based on usage

c. Support customer or member placing usage in historical or relative contexts

- i. to help them answer "How does Report A saying X compare to Report B saying Y?"
- ii. to explain limitations to data and aid data interpretation, i.e. "What does data mean and what does it not mean?"
 1. for longitudinal usage
 2. for usage relative to other media/content
 3. for usage relative to other platforms

d. Use usage data to help customer or member improve their own metadata

e. Support customer or member if or when usage surges

- i. by tracking usage surges by country
- ii. by tracking usage surges by customer or member

f. Understand how and where to grow additional customer or member support based on usage

- i. by identifying regions for internationalization
 1. of new, additional language support
 2. for increased regional customer or member support