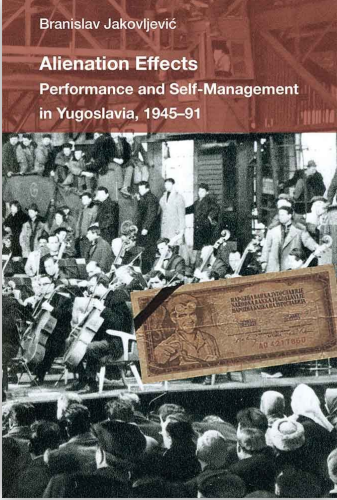



Building a Trusted Framework for Coordinating OA Monograph Usage Data

@KevinSHawkins
University of North Texas Libraries

Our Team

- Charles Watkinson (Associate University Librarian for Publishing, University of Michigan Library)
- Kevin Hawkins (Assistant Dean for Scholarly Communication, University of North Texas Libraries)
- Brian O'Leary (Executive Director, Book Industry Study Group)
- Lucy Montgomery (Director of Research, KU Research)
- Cameron Neylon (Executive Director, KU Research)
- Katherine Skinner (Executive Director, Educopia Institute)





- Published in 2016, “unlatched” by KU supporters in 2016
- “Promotion book” for faculty author
- Author interested in OA for reaching scholars abroad with interest in how dictators exert soft power through art and culture
- Interdisciplinary reach: East European history, performance studies, Jewish studies
- 2017 winner of Outstanding Book Award from Association for Theatre in Higher Education!

To help explain the issues we are trying to address, I want to share with you an example chosen by my colleague Charles Watkinson, AUL for publishing at the U-M Library and director of the U-M Press. This university press publishes about 15% of its books in OA, mostly through funding from Knowledge Unlatched.

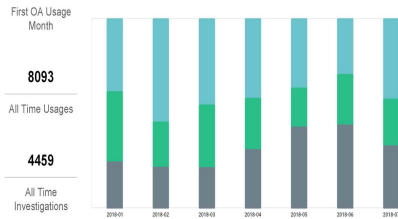
The book shown on the slide was published by the U-M Press in 2016 (and “unlatched” then as well) by an author aiming for promotion to full professor at Stanford. The author was interested in OA especially for reaching readers abroad, and given the book’s interdisciplinary subject matter, OA also seemed especially important. But the author wanted to understand the use of the book and be able to make a case about its impact beyond it winning a book award and quoting sales figures, and U-M Press wanted to reassure the author that he had chosen the right publisher and made a good decision in going with OA.



Alienation Effects: Performance and Self-Management in Yugoslavia, 1945-91

University of Michigan Press
Title usage data report from 2018-01 until 2018-09

2016-07 Usage by Time



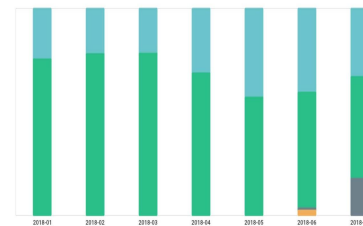
3634 Usage by Country

All Time Requests	Europe	North America	South America	Asia
102	United Kingdom 88	United States 332	Brazil 10	India 26
	Germany 76	Canada 70	Peru 4	Indonesia 22
	Ukraine 43	Mexico 3	Chile 2	China 20
	# of Unique Countries All Time		Oceania	Africa
Australia 31	South Africa 12	New Zealand 5	Nigeria 7	
Fiji 4	Uganda 5			

Alienation Effects: Performance and Self-Management in Yugoslavia, 1945-91

University of Michigan Press
Title usage data report from 2018-01 until 2018-09

Usage by Platform

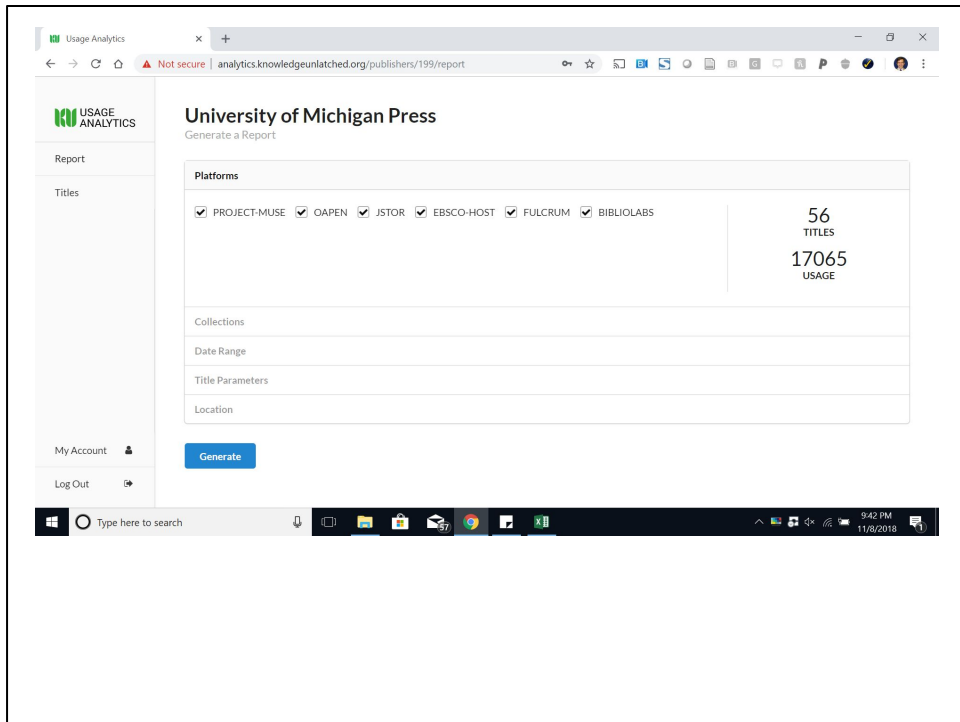


Usage by Metric / Platform

Total Item Investigations	Total Item Requests	Unique Title Requests
jstor 629	jstor 406	oapen 476
project-muse 82	project-muse 64	fulcrum 8
	fulcrum 12	bibliolabs 3
		ebcco-host 1

U-M Press has been struggling to find a way to tell such stories to its authors. U-M Press participated in the KU Open Analytics pilot effort, which aims to provide insights into the usage of OA books included in Knowledge Unlatched. The reports provided by the pilot, such as the examples shown on the slide, can be shared with authors to give insight, for example, into which countries have readers of the book. The report on the right also provides insights for the publisher, such as which platforms the users are coming through. But there are lots of other things that we'd like to know:

- Is the book really reach interdisciplinary audiences?
- What are the institutional affiliations of readers?
- Are students reading this book? Faculty members? Non-academic readers?



The Knowledge Unlatched staff say that they have been frustrated trying to form relationships with all of the platforms that their content is available on in order to get data for this pilot. And with openly licensed content, it may well be available in other locations as well, sometimes without your knowledge, such as in the Internet Archive, on Unglue.it, or in the author's institutional repository.

Title	ISBN	Platform	Metric	Compliance	Total
Alienation Effects	9780472053148, 9780472900589, 9780472121984, 9780472073146	oapen	unique_title_requests	counter-4	38
Alienation Effects	9780472053148, 9780472900589, 9780472121984, 9780472073146	jstor	total_item_investigations	counter-4	43
Alienation Effects	9780472053148, 9780472900589, 9780472121984, 9780472073146	jstor	total_item_requests	counter-4	19
Alienation Effects	9780472053148, 9780472900589, 9780472121984, 9780472073146	project-muse	total_item_investigations	counter-4	35
Alienation Effects	9780472053148, 9780472900589, 9780472121984, 9780472073146	project-muse	total_item_requests	counter-4	33
Alienation Effects	9780472053148, 9780472900589, 9780472121984, 9780472073146	fulcrum	total item requests	counter-5	3
Alienation Effects	9780472053148, 9780472900589, 9780472121984, 9780472073146	fulcrum	unique title requests	counter-5	2

The KU staff has also been disappointed in the quality of the data that they get from the platforms. *Alienation Effects* was assigned four ISBNs by U-M Press for different product formats, but the different platforms have used different ones in their reports. Platforms produce different kinds of COUNTER reports (more on that in a bit), sometimes for books as a whole and sometimes on a chapter level. These are the frustrations that led to our Mellon-funded research project.

Objectives

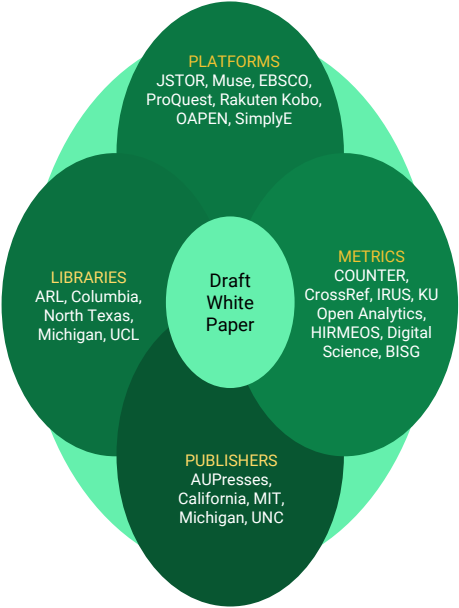
Project aims to . . .

- Create a structured conversation around usage tracking for open access ebooks
- Understand implementation challenges
- Define opportunities for collaboration
- Define a framework for moving forward
- Ultimately: Make a compelling case for investment in OA book publishing by authors, publishers, funders, and libraries.

Timeline

- Spring 2018: Andrew W. Mellon Foundation funded planning exercise.
- Summer 2018 / Fall 2018: KU Research prepared a discussion document providing landscape review and outlining data trust concept.
- Fall 2018 / Winter 2019: Community consultation around the discussion document culminating in an invitation-only summit in New York (December 3–4, 2018).
- Spring 2019: BISG will publish a white paper including recommendations for next steps.

NYC Summit



We've explicitly tried to bring together a number of communities, especially between Europe and North America and between the for-profit and not-for-profit communities.

How does OA ebook usage data work today?

With that overview of the motivations, objectives, and timeline for our work, I want to go into a bit more detail about how usage data for OA ebooks works (and doesn't work) today.

You may be familiar with usage data for online journals and databases: for these types of resources, generally speaking, views and downloads are reported according to one of the formats specified by the COUNTER Code of Practice as a standard way to compare usage between products. The Code of Practice was expanded in recent versions to address the specific needs of books and of OA content, which certainly helps with comparing usage of ebooks across platforms.

Besides COUNTER reports, there are other types of usage data. Notably, some web platforms use tools like Google Analytics to provide a richer look at how website users engage with content. These tools provide more information than the simple tallies of searches and views found in a COUNTER report, instead assisting with path analysis (how the user navigated through the web).

However, information about the usage of academic ebooks, especially OA books, is much more difficult to gather, analyze, and communicate than comparable information about electronic journals.

E-journals vs. ebooks

E-journals are usually delivered through the publisher's website.

Ebooks are usually delivered through website of intermediaries:

- Library providers: *EBSCO, ProQuest, JSTOR, Project MUSE*
- Direct-to-consumer retailers: *Rakuten Kobo, Amazon Kindle, Google Play*

E-journals are usually delivered through the publisher's website even if that website in turn uses a third-party platform like Silverchair. But ebooks are instead usually delivered through intermediary channels. The various intermediaries for ebook distribution, such as those listed in italics on the right, compete for market share and tend to view the usage data that they collect as proprietary. What they do share with publishers and with libraries is inconsistent and in formats that can't be easily compared. For example, library aggregators may report chapter downloads or whole book downloads, which indicate different types of usage.

Other sources of usage data

- Authors or their publishers sometimes make one or more versions of an ebook available online separate from the usual distribution channels. There might be usage data from:
 - author's personal website
 - a self-service platform like Figshare, Academia.edu, or ResearchGate
 - an institutional repository
 - a digital library system
- The publisher might choose to distribute the work through channels specifically for OA content, such as the OAPEN Library, and if the work is openly licensed, third parties such as Unglue.it and the Internet Archive might distribute their own copies of the work.

But there are other sources of usage data besides that generated by ebook distributors.

(read slide)

Any attempt to represent the usage of an OA ebook needs to take into account these various channels, plus storytelling indicators such as altmetrics and Crossref Event Data. As Lucy Montgomery, Cameron Neylon, Alkin Ozaygen, and Tama Leaver argue in a recent article in *Learned Publishing*, we want to know not just the quantity of usage of a book but also about the audiences who are engaging with it and, if possible, how and why they use the content.

How are ebooks used?

But it's not just publishers and libraries that need a full picture of ebook usage. Authors and funders of research want to know more about how their books are used, but they are usually reliant on publishers sharing the appropriate data with them. But while journal publishing is concentrated among a few large publishers with a fairly stable revenue stream, monograph publishing is highly distributed among publishers on shoestring budgets. Few monograph publishers have staff with time and expertise to examine usage data closely.

To put it another way, advocates of OA often say that an OA book will be more often downloaded, used, and cited than a comparable restricted-access title. However, all stakeholders in scholarly communication want proof of this:

- Publishers need to demonstrate such impact to receive support for their open access programs
- Funders look for usage data to demonstrate return on their investments
- Authors are eager to show evidence of additional reach and influence for their work

This study is aimed at figuring out how to establish a mechanism to do this.

Collective action problem

Our discussion document argues that the problems here are less technical and more social—a *collective action problem*. We need to establish a trusted framework for coordinated action between all relevant stakeholders that will allow data to be shared in an appropriate way that guards against misuse. We call this a “data trust”.

What do we mean by a “data trust”

cooperative organization to address the needs of scholarly publishers, libraries, and other stakeholders in the collection, validation, aggregation, normalization, and dissemination of usage data

Features of the data trust

- Repository of data about OA books and their use
 - Framework for collecting, normalizing, aggregating, and maintaining
 - Well documented pathways and workflows for data access and use
- Organization with legal responsibility for data
 - Community/member-governed entity
 - Multi-stakeholder empowerment
 - Oversight responsibility
- Standards-making (and keeping) responsibility
 - Fair/equitable access
 - Data ethics laboratory

Questions the data trust will resolve

How can information be collected?

How can data be contributed?

How can data be normalized?

How can data be stored?

How can data be curated and preserved?

How can information be shared and used?

What are the terms of use (and are they different for different members)?

Who can see which parts of the data?

How can aggregated data be accessed and used?

Who can build services on the data (and how)?

A sneak peek at our findings in the forthcoming white paper (1)

- A good deal of data is already available to those who want to study the impact of OA monographs but is sometimes in closed environments.
- Certain types of data are of interest to stakeholders but have never been compiled.
- The number of available data sets, whether in closed or open environments, dwarfs the data that is not yet available.
- The data of greatest interest varies by audience (authors, publishers, funders, vendors, libraries, and readers). Across the several audiences, relatively little of the available data is being used widely or consistently.
- There are marketplace and ethical concerns about use of certain data points.
- COUNTER does not provide some of the qualitative information about OA ebook usage that stakeholders want, but its governance group is willing and eager to adapt the standard to be more useful for OA ebooks.

A sneak peek at our findings in the forthcoming white paper (2)

- Use cases for OA monograph discovery, access, consumption, and engagement have not been widely or fully developed. Relevant use cases, when developed, must be mapped against the needs of audiences identified above.
- Significant work is being done outside of North America, but coordination with European and other international efforts has been inadequate.
- There is significant debate about how to build such a trust—specifically, whether its governance and operation should be centralized, federated, or distributed.
- For creating a data trust, we need agreements in at least three areas:
 - standards for data exchange
 - where and how data is stored and managed
 - how analytics will be built on top of that data

A sneak peek at recommendations for next steps

1. Define the governance and architecture for the data trust and articulate priorities
2. Create a pilot service that implements the defined governance and architecture
3. Implement and extend relevant open-source technologies across a base of stakeholders in the US
4. Develop personas and use cases that demonstrate who benefits from OA monograph usage information and how a data trust can better serve their needs
5. Build engagement across markets beyond North America and Europe
6. Better document the supply chain for OA monographs

#3 builds on #2, using work by HIRMEOS in Europe.

Final whitepaper to be published at
<https://bisg.org/page/WhitePapers>

Questions?

@KevinSHawkins