

2017-11-09

Open Access Monographs: Promise or Bust

Welzenbach, Rebecca

<http://hdl.handle.net/2027.42/148277>

[slide 1]

Good morning.

[slide 2]

I'm a librarian at Michigan Publishing, which is a division of the University of Michigan Library. We publish in a variety of ways, but for the purposes of this talk, I'll be focusing specifically on the OA monographs published by the University of Michigan Press.

[Slide 3]

We have been publishing OA monographs for more more than a decade as part of our DigitalCultureBooks imprint, which now has more than 55 titles.

[slide 4]

In the last three or four years, though, we have started to see a lot of change. We're seeing increased interest in OA (and in measureable outcomes) from our authors, funding agencies, governments, and institutions. We're seeing new opportunities to participate in OA monograph publishing at scale through cross-institutional programs, such as Knowledge Unlatched and Open Library of the Humanities. And, we're seeing a proliferation of platforms--an increasing number of ways to distribute and discover OA monographs.

So how do we know if we're succeeding? Where's the promise? Where's the bust? As we just saw from Erich's presentation: data. We're looking for data about usage, about reach, and about engagement. We've been gathering usage data on our OA books for years, but we haven't been very satisfied with what we've been able to do with it.

[slide 5]

For example, on our own platform, we use Google Analytics. This allows us to gather a ton of information, but it can be overwhelming, complex, and difficult to share title-level information with authors or with others. The biggest issue, though, is that we can only use it to track things that we are in fact hosting on our own servers. We miss out on all usage that happens on other platforms. Which, as Erich showed in his presentation, is in fact probably **most** of the usage.

[Slide 6]

Here's another example: from KU/OAPEN, we receive regular COUNTER Reports, and this is useful to tell us how **frequently** the books are being used. It's also really helpful that there's a consistent standard here, making comparisons easier. But this tool is really designed for libraries to measure usage, not publishers, and it's not as rich as we'd really like these reports to be.

[slide 7]

One final example: about a year ago we were finishing up a research project called Mapping the Free Ebook Supply Chain. One element of that project was that we invited readers to take a survey telling us about how they found OA books and what they intended to do with them.

These qualitative survey results were actually one of the most interesting and promising parts of the study for us! But this approach, too, is limited: it reflects only those users who opt in, and it's time consuming to execute and analyze.

[slide 8]

This was the opportunity to conduct this research project with Frank, Erich, Dean, Lara, and the KU Research folks was so exciting to us. From the start, we saw much higher usage of our OA monographs on the JSTOR platform than we had seen elsewhere.

[slide 9]

We wanted to know more about that but, just as important, we were excited about the opportunity to compare notes with other presses, and to really try to figure out what works and what doesn't, what's useful and what isn't, in capturing and communicating usage of OA monographs.

[slide 10]

Our first task was to figure out what questions to ask. They included:

Wow that's tiny. Clearly we had a lot of questions!

- What percentage of usage comes from within JSTOR, vs. other sources?
- Where do readers come from in the world, and can institutional affiliation also be traced?
- What are the most popular subjects across the JSTOR platform?
- Are the most popular books on the OA platform similar in subject matter to the subjects already popular in JSTOR?
- What do readers do when they download, ie are they typically downloading just one chapter or multiple chapters?
- Are stats for the four initial publishers similar?
- Are there multiple downloads of the same content (eg. Chapter) from the same institution?
- What is the readers' behaviour in terms of the proportion of who downloads chapters and the proportion who just view?
- What can JSTOR data tell us about how long readers spend on each book that they view?

From my point of view, one of the most useful outcomes of this study was to show us where the surprises *weren't,* where lines of questioning didn't reveal much--and thus we can sort of let them go.

[slide 11]

Some of these were in fact really important to ask once, but can now be set aside. For example, the question that essentially launched this project: are the stats for the four publishers similar?

This graphic shows the usage of all monographs on JSTOR for the four presses in the study from August 2015 through August 2017, and it's clear that the story, the shape, is roughly the same for all the presses. UCL press is the most different in that their entire catalog is open access, so they have no usage at all until late 2016, and then their numbers go way up because they started with more OA titles than the rest of us did.

[slide 12]

This slide shows a breakdown of where the readers using our OA monographs on JSTOR are coming from, and we see essentially the same story: really similar patterns across all four presses. We saw this throughout the report: the story for all four presses generally matched quite closely with the overall picture.

That's useful and reassuring--it gives us some firm ground to walk on as we go forward with thinking about OA monograph usage reporting.

[slide 13]

Another question that we asked was, what percentage of readers download PDFs from JSTOR, vs. viewing the content within the platform. Well....ok, across all four presses, a little more than half view, rather than download.

[slide 14]

This can be broken down further: 21% of users directly downloaded, without viewing the item. The remaining 79% viewed the item. Of those who viewed the item, just under a third went on to download it, while the others did not.

We could spend a lot of time trying to parse this and dig into it further. As it is, it doesn't tell us much. We might get a more meaningful story if we associated view vs. download data with time spent in the book. Or, we could tie this behavior to geography, institutional status, or even device being used: does people's behavior differ w/r/t views and downloads if they're on a phone vs. on a desktop? Or if they're in Boston vs. in Manila?

We *could* do all these things, but the important thing for us to ask ourselves is, do we want to? How will it help? What would we do differently? Maybe at some point understanding user behavior on different devices or in different countries might inform platform decisions by JSTOR. But for us, the publishers, this question doesn't really help us. It doesn't help us tell a meaningful story about OA to our authors, institutions, or funding bodies. It won't help us make decisions differently--we'll deliver the same file formats to JSTOR until told to do differently.

So this question that really seemed important at the start of this study, to me, seems like one we can really set aside as we focus on how best to measure and communicate OA monograph usage.

[slide 15]

And, one more example--we asked whether the most popular books on JSTOR's OA platform were similar, topically, to the most popular books on "the rest" of JSTOR.

Now the sample for OA books from our four presses was not that large, but in general, yes, these books are in the same areas. Political science, history, anthropology—and *general*.

This, too, is a good foundation for us going forward. For the purposes of studying and analyzing usage, it lets us know that the books in the OA group and the non OA group are comparable. One fewer variable to worry about when trying to assess usage between open and closed monographs. It is also a reminder that--as Erich pointed out--YES, JSTOR is a good place for people to find the kinds of books we want to offer to them. So, this is useful to know as a baseline, but probably not something that needs to be part of an ongoing OA monograph usage report.

[Slide 16]

OK, so I've talked a lot about the questions that turned out to be not that useful, or that can now be set aside. But what actually was useful? What information do we want more of, and to share with our stakeholders?

Well, first of all, just as Erich showed, the vast increase in usage of our OA materials:

Are OA books used more than non-OA books? Yes, clearly, without question, by a lot. This is this is the kind of story we'd like to be able to tell. It would be even more effective with a reminder of the number of books in each area (i.e. 20 OA Michigan titles making up all of that)

[slide 17]

Next, where users come from in the world and, crucially, the version of this that's normalized to account for how many academic institutions where usage might be happening, as opposed to just raw numbers.

[slide 18]

If we just look at the number of events--views and downloads--the English speaking western world--and especially the US--dominates the picture. But when we look instead at the number of downloads per institution, now we can see a great deal more usage all around the world--especially in the Philippines, which is right at the top.

[slide 19]

We also want to know what types of readers are using the books. In the case of UM Press in the lower right hand corner, we saw usage from two institutions in Manila, one in Indonesia, a high school in California, and a community college system in Texas. Great! These are the audiences we want to reach with OA.

This does reveal, however, one of the major gaps in this study: that we don't currently have any good way at this time of accounting for unaffiliated users using OA books on the JSTOR platform. And of course, that's exactly what we'd really like to know more about.

[slide 20]

So, to return to where we started: are we asking the right questions? There are some that I think we now feel comfortable letting go of--at least I do. I think we also discovered some new questions--or new facets of questions--that we want to be thinking about: What about non-institutional users? When we measure usage around the world, are we taking into account population/number of institutions, not just raw numbers? How well are readers able to find this content on the open web?

[slide 21]

We want to use what we've learned from this study to establish good practices for reporting the usage of OA monographs that reveals the reach and the impact of OA to authors, funders, and institutions.

[slide 22]

Now that we have a better sense of what questions we want to ask, and what information is useful, we want to get to a place where reports of OA monograph usage are comprehensive--that's a bigger issue, pulling in usage from other platforms--but starting with the space where we're seeing the most usage makes sense. Consistent, customizable, and easy to share with relevant stakeholders. And, to steal another adjective that I heard yesterday from Ross McIntyre from JISC, reports that are easily consumable: visualizations like the ones in this study, as opposed to mega spreadsheets.