

## Plato's Cave Revisited

Bruce Heterick  
*JSTOR, Portico*, [bruce.heterick@ithaka.org](mailto:bruce.heterick@ithaka.org)

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>



Part of the [Library and Information Science Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

---

Bruce Heterick, "Plato's Cave Revisited" (2013). *Proceedings of the Charleston Library Conference*.  
<http://dx.doi.org/10.5703/1288284315238>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

## Plato's Cave Revisited

*Bruce Heterick, Vice President, JSTOR, Portico*

*The following is a transcription of a live presentation at the 2013 Charleston Conference. Slides are available online at <http://slidesha.re/1gTZk84> and video at <http://youtu.be/ZkGSQIFZ0BI>.*

Good afternoon, everyone. My name is Bruce Heterick. I work with JSTOR and with Portico, and thanks, everyone, for postponing your happy hour for a little to come and talk a bit about discovery. It probably does not seem like a real good tradeoff, but I will try not to disappoint.

I am happy to be here today talking about some of the work that I personally have been engaged in for the past year and a half. At JSTOR, we have really been diving into this topic for the last 9 months, and I am really excited today to talk about some of the results of those efforts. I did a talk about a year and a half ago at the Fiesole Retreat, entitled "Plato's Cave"—talking about discovery, the lack of transparency that existed at that point in time, and understanding how things work from a content provider perspective. So the people at Charleston asked me to do a reprise of that presentation which I am calling "Revisiting Plato's Cave." Now that we have a little more information and a little more data, I thought it would be really interesting to share what we are seeing as a content provider on this front (although, admittedly, a fairly unique content provider). I want to make sure we state that upfront. Some of the things I am going to talk about today certainly should not be applied across all of publishing. It is important to recognize that JSTOR is primarily backfile content. We have some current journals, and we have some books, but what people primarily know us for is the backfile content of the 2,000-plus journals that we have on the platform.

I am going to try to go through this information relatively quickly so we leave a lot of time for questions, because I think there will be a lot of questions at the end of this, and to me that is the most interesting part. So we will do our best; and

if there are not a lot of questions to answer then happy hour can start sooner than we expected.

The first thing that I want to do is I want to give thanks to my colleagues because this has been a lot of effort; a lot of work. Many of you know Jenny Walker; she was actually going to present with me today. She is in South Africa, and it was very difficult for her to get away from what she was doing to join us today, so I am going to be doing this solo. But she did a tremendous amount of work on this project, particularly with the deep dives we did with each of the web-scale discovery partners. And then two folks at JSTOR who also were instrumental, Teddy Hein and Ross Houseright. Just to give you a sense of the importance of these things, we actually have built an analytics team in JSTOR to help us start to understand this data in a much better way; a much more robust way. This is the beginning of some of this work, particularly some of the stuff that Ross Housewright, our senior data analyst, is doing. Many of you know Ross's name from the Ithaca S+R research and faculty surveys that he has been involved with in the past. Both Teddy and Ross were instrumental in getting us to this point.

I alluded to this in a presentation this morning (Figure 1). This chart gives you a sense of where we see content accesses happen at JSTOR; the origin of those content accesses. And for our purposes today, when I say content access, I am talking about an article download or an article view. This chart gives you a sense of where people are starting before they actually come to JSTOR; before we see a content access on our platform. As you can see, a large percentage of people start at JSTOR—about 33%. We see a lot of content accesses coming from Google, Google Scholar, and institutional resources (things that usually have the ".edu" domain). We also have a number of linking partners: organizations like Research Papers in Economics (RePEc) or AMS (MathSciNet). We have 50-plus formal linking

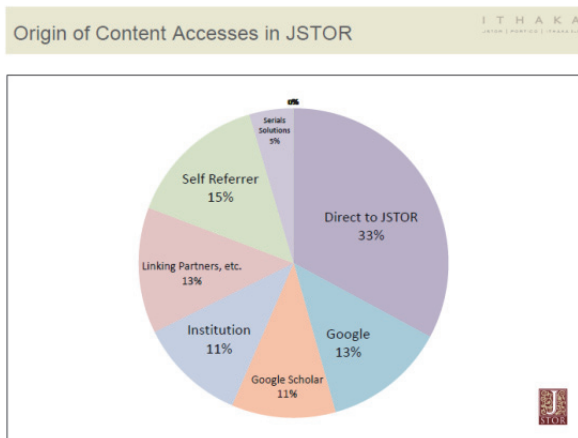


Figure 1.

partners<sup>1</sup> that link to JSTOR directly from that particular resource. The self-referrers listed in the chart are typically people who are either linking within JSTOR, or they came to JSTOR, perhaps from another place (e.g., Google), their session ended, but they restarted the next session in JSTOR. And then, we see 5% of our traffic driven through the Serials Solutions's domain name, which is primarily their link resolver.

Now one of the really interesting things that has come out of reviewing this over the past year is that we really cannot tell with any precision which content accesses originate from a discovery service versus the content accesses that come to JSTOR through a link resolver. Some significant percentage of discovery service traffic (with the exception of the direct linking traffic) is hidden by the link resolver. And one of the very clear things that came out of this research, and one of the things we are going to immediately begin to address with the discovery providers, is establishing persistent origins through the linking process, so we can see when something starts at EDS/Primo/Summon versus a link starting at another origin but passed through the link resolver. We have had a good collaboration with all of these discovery providers during this process, and we have learned a lot on both sides. That is important as we move forward. I talked earlier about the high percentage of people who start at JSTOR, and it was brought up earlier today that perhaps that is a function of JSTOR's "unique"

<sup>1</sup> "formal" as in having executed a formal linking agreement

place in academia: it has brand recognition among faculty and among students. There is something to be said about that, for sure, but let me reiterate that the research we are looking at today is from one content provider's point of view, so it is really important that we do not try to make too many assumptions for all content providers from this research, nor carry forward too many assumptions from this work to other areas of research.

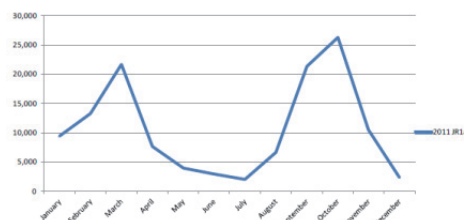
JSTOR sees a great deal of use, including a high number of referrals from Google. We see a lot less traffic than we used to, and many of you in this room know, particularly the publishers, that about a year and half ago Google changed their ranking algorithm which really disproportionately affected publishers that have pay walls. In essence, Google began promoting free content first in their search results. At JSTOR, like other content providers, we have seen our Google referrals drop quite a bit since that change. Now we have been working with Google to make them understand that there is actually a hole in this "free first" philosophy; because, if you are at an institution that has licensed this content, it is, in essence, "free" to your constituents. Yet, when they search in Google, they are not going to see that content because many times that content falls way down in the search result, or out of the search results altogether. We are trying to figure out a way to sort of close that "loophole," as much as you can do with the folks at Google, but often times it is hard to find someone to talk to on the Google Web Search side of the house. Thus far, we have been fortunate. We have been able to get through to some people, and we are trying to see if we can actually get them to do something about this in the coming months.

We also see a lot of traffic from Google Scholar. Many of you have connected your link resolvers with Google Scholar, and we see a high number of links coming from that connection. I also wanted to highlight—based on the discussion we just had about link resolvers—that the Google Scholar number here is probably underrepresented. What we are picking up here are links that come directly to JSTOR from Google Scholar; not those links that pass through a link resolver. If it came through a

link resolver, that content access is represented in Serials Solutions, or SFX, or Link Source, or something of that nature. So the Google Scholar number here is underrepresented for sure. Okay, so this is to say to everybody here that the amount of content accesses that we are seeing come through web discovery services is pretty small compared to the other things that we have to deal with, and, frankly, are not our highest priority. Based on the percentages reflected in the chart, the priority for our organization is improving the discovery experience for those individuals that start their research at JSTOR. How do we help that 48%? Next, we are focused on Google. How do we affect that experience in a more positive way? So it is not that the web discovery services are not important. They are, but they are important for a smaller percentage of the people coming to JSTOR in our particular situation.

So over the past several years, I have done specific JSTOR usage analysis for hundreds of institutions. We have been looking at the trends of how usage has changed over time. I am bringing this next set of slides up as an illustrative example. This is a small institution in the US (Figure 2). This was their COUNTER usage for JSTOR in 2011. This was their COUNTER usage in 2012 (Figure 3). At this particular institution, their usage dropped 24%. Now, this change in usage did raise a red flag, and I noticed a handful of similar situations over the next year. What is happening here? Why are we starting to see some of these anomalies? This change in usage is probably not explained by the drop in content accesses coming from Google Web Search. In fact, it is probably not all explained by other factors we can control for. But, it was important—at least at the surface level—to try and understand what was happening in situations such as these. In this particular institution, in June 2012, these folks implemented a discovery service. You can see that their 2012 usage was climbing along at a bit higher rate than 2011 and then it started to drop in mid-2012. The drop continued into 2013 (Figure 4); it stayed low through the first 6 months of the year (which is when this particular analysis was first run).

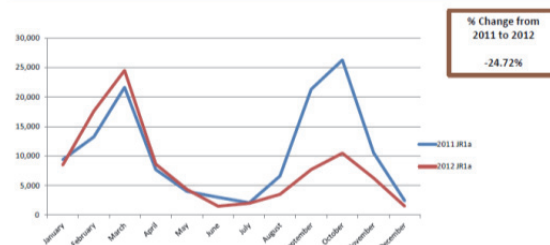
#### 2011 Usage – JSTOR Small



COUNTER Journal Report 1a (JR1a) counts the number of successful full-text article requests by month and journal from an archive. The metrics that drive into this are Article Views and PDF downloads, excluding Article Views and PDF Downloads of the same item in the same session if occurring within 30 minutes of a previous View of the same item, or 30 seconds of a previous Download of the same item.

Figure 2. 2011 Usage—JSTOR Small

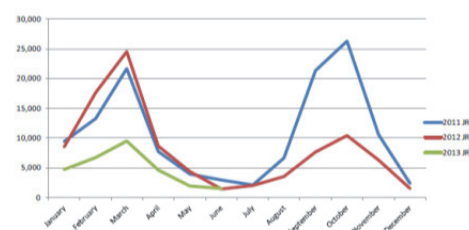
#### 2012 Usage – JSTOR Small



COUNTER Journal Report 1a (JR1a) counts the number of successful full-text article requests by month and journal from an archive. The metrics that drive into this are Article Views and PDF downloads, excluding Article Views and PDF Downloads of the same item in the same session if occurring within 30 minutes of a previous View of the same item, or 30 seconds of a previous Download of the same item.

Figure 3. 2012 Usage—JSTOR Small

#### 2013 Usage (YTD) – JSTOR Small



COUNTER Journal Report 1a (JR1a) counts the number of successful full-text article requests by month and journal from an archive. The metrics that drive into this are Article Views and PDF downloads, excluding Article Views and PDF Downloads of the same item in the same session if occurring within 30 minutes of a previous View of the same item, or 30 seconds of a previous Download of the same item.

Figure 4. 2013 Usage (YTD)-JSTOR Small

What is actually causing this? We are fortunate at JSTOR in that, for most of the institutions represented here today, the cost per use for JSTOR is extremely low; so if we see a big drop in usage, the impact is probably smaller than another publisher where the cost per use is higher. However, where we are seeing drops in usage like what is reflected in this example, that

would become a worrying trend. And so what we wanted to see from our perspective was what is happening here? Are we doing something wrong? And, if we are doing something wrong, what can/should we do to take some corrective actions to make it better?

In order to attempt to understand what was happening in these situations, we decided to get some data. The first thing we did back in May was to send out a simple little survey to all of our participating institutions asking (1) what discovery service are you using and (2) when did you implement the service. Considering that we have almost 9,000 participating institutions, we received a pretty small number of responses: 422. And to make matters worse, there were no consistent implementation dates. There was nothing that we could actually attach a date to and actually begin to measure usage before and after said date (Figure 5).

Then we went to Marshall Breeding's database (Figure 6), because we knew that he was tracking which institutions had which discovery services and which link resolvers, and we were able to increase the pool of institutions by quite. Still, getting consistent implementation dates from that database also proved to be a challenge. The export from that database certainly increased the number of products that people said was their "discovery service," and that was interesting, but the lack of consistent implementation dates made any potential usage analysis almost impossible.

We finally decided to go direct to the source. We went to EBSCO, Proquest, Ex Libris, and OCLC and asked them directly: tell us who your customers are, tell us when they implemented your service. We will then begin to look at how usage looked before and after implementation, and we will see if there is any correlation that we can make between implementation of a web-scale discovery service and changes in usage patterns at JSTOR.

So we worked directly with EBSCO, Ex Libris, OCLC, and ProQuest to get their customer database and implementation dates (Figure 7). Those organizations were supportive of this endeavor, but it was also very challenging. The companies are not used to just handing that sensitive information over

to somebody, and so while they were supportive, it did take time.

Discovery service "A" here, gave us almost 5,000 institutions. We were able to match about 3,100 of those with institutions in our data warehouse. About 1,700 or 1,800 of those were current JSTOR participants, about 925 of that pool were in higher education, and that is the pool of institutions that we decided would be most useful for our initial data analysis.

#### Survey of JSTOR participating institutions (May 2013)

- o 422 responses
- o No consistent implementation dates (< 100)
- o Too few responses across institutional archetypes to be statistically relevant

| Discovery Service               | # responses |
|---------------------------------|-------------|
| AquaBrowser (Serials Solutions) | 1           |
| EDS (EBSCO)                     | 154         |
| Encore (Innovative Interfaces)  | 10          |
| Primo (Ex Libris)               | 69          |
| WorldCat Local (OCLC)           | 36          |
| Other <sup>1</sup>              | 48          |
| Summon (Serials Solutions)      | 102         |
| Vufind (Villanova University)   | 2           |

Figure 5.

#### Supplemented with data from lib-web-cats database (Marshall Breeding)

- o Increased # of institutions with data to 1,480
- o Again, no consistent implementation dates

| Discovery Service                  | Number |
|------------------------------------|--------|
| AquaBrowser (Serials Solutions)    | 45     |
| EDS (EBSCO)                        | 379    |
| Encore (Innovative Interfaces)     | 121    |
| Primo (Ex Libris)                  | 410    |
| WorldCat Local (OCLC)              | 117    |
| Other <sup>1</sup>                 | 73     |
| Summon (Serials Solutions)         | 259    |
| Vufind (Villanova University)      | 41     |
| Backlight (University of Virginia) | 5      |
| Enterprise (SirsiDynix)            | 20     |
| Locally-developed                  | 10     |

Figure 6.

- Worked directly with EBSCO, Ex Libris, OCLC, and ProQuest to get customer list and implementation dates (July - Sept. 2013)
  - Supportive of effort
  - Confidentiality required

| Discovery Service | Institutions provided | Matched in CRM | JSTOR participants | Higher Ed | % JSTOR participants | % higher education |
|-------------------|-----------------------|----------------|--------------------|-----------|----------------------|--------------------|
| A                 | 4,992                 | 3,149          | 1,781              | 925       | 36%                  | 19%                |
| B                 | 760                   | 645            | 576                | 417       | 76%                  | 55%                |
| C                 | 63                    | 57             | 53                 | 48        | 84%                  | 76%                |
| D                 | 623                   | 540            | 397                | 308       | 64%                  | 49%                |

Figure 7.



Let me say one thing before I start showing you numbers, all right? This is really important. Statistics are like bathing suits. What they reveal is interesting, but what they conceal is essential. We are going to see some interesting things here, but really the “essential” stuff is the work that has not yet been done—the next set of statistics that we start to look into in the next phase of this work.

Let us look at the data: There are any number of variables that we could/should look at when evaluating the change in usage at a particular institution. How have they established their administrative setting in their discovery service or in their link resolver? What type of institution are they? Are they a research library? Are they a liberal arts school? Are they public or are they private? Are they in the United States or in the UK? We are going to start looking at trends in all of these different archetypes to really try to understand if certain combinations of discovery service and link resolver behave differently in different types of institutions, and I think that is the thing that all of us in this ecosystem—libraries, content providers and discovery service providers—are really interested to begin to find out.

I do not know how many of you attended Michael Levine-Clark and Jason Price’s presentation yesterday morning, but this is dovetailing a little bit with what they are doing. I have talked a lot with Michael, and we are providing them some data for what they are doing, and we are going to see about how we can continue to work together in a way that begins to provide a holistic picture. They are taking an institutional perspective with their research, while we are trying to look at potential usage impacts from a content provider perspective. We really have not heard a lot about what the impact of these implementations might be vis-à-vis the content providers, and we believe it is an important perspective to bring to the discussion.

So we took these lists of JSTOR participants in higher education and we then took the implementation date (Figure 8). The reason I have the implementation date highlighted on the slide is to call out an important caveat: this is the implementation date that the discovery provider

provided to us. That does not necessarily mean, I have come to find out, that this is actually when the system was turned on. This could have been when they decided to activate the JSTOR collections in the system, but it may have taken 6 more months (or more) for the service to actually go into full production. These are the dates that we are working from, but they are imperfect, and so I want to make that clear to everybody. What we have tried to do, much like Michael Levine-Clark and Jason Price have tried to do, is to look at the average monthly usage at JSTOR for 12 months before that implementation date and compare that to the average monthly usage for the 12 months after the implementation date.

As a comparative baseline, we looked at the change in JSTOR usage for all higher education institutions that participate in JSTOR over that same time frame (August 2009–September 2012). Whether they had a discovery service implemented or not, the change in average monthly usage change was a decline of 3.2%.

When you look at the discovery providers, and looking at those same set of institutions in U.S. higher education over that same time period, you can see that average monthly usage after implementation of a discovery service dropped across the board: one service at a smaller percentage than the baseline change in JSTOR use (-3.2%), one just above that percentage, and one at a much higher percentage. The one thing I want to say here is that discovery service “C” is too small of a sample to provide any statistical relevance. I wanted to make sure that we included it in this discussion, but we are going to have to

Culled customer lists to JSTOR participants in U.S. higher education for EDS (EBSCO), Primo (Ex Libris), WorldCat Local (OCLC), and Summon (Serials Solutions/ProQuest)

- o Looked at average content access per month for each JSTOR Class for 12 months prior/post implementation date
- o JSTOR average usage change for all U.S. higher education (August 2009 – September 2012): -3.2%

| Discovery Service | Usage Change Post-Implementation |
|-------------------|----------------------------------|
| A (218)           | -8.7%                            |
| B (100)           | -0.4%                            |
| C (13)            | -13.3%                           |
| D (117)           | -4.4%                            |

Figure 8.

figure out how to actually work with a broader set of data as we move forward

Now, when we look at JSTOR institutions in higher education worldwide (including US), the average monthly change in JSTOR usage change during that time period was about a 1% decline (Figure 9). When that average monthly usage change is broken down by discovery service implementations, the differentials are smaller (as compared to the U.S. results versus the baseline). Again, we cannot make any judgments from Discovery Service “C” because that sample is too small.

While this data is interesting at the macro level, what we really wanted to start to look at was whether there were differences by type of JSTOR institution (e.g., research universities versus a small liberal arts colleges) and so in the limited time that we had, we decided to let us use the JSTOR Classifications as a relative proxy for different institution types. For the 1,600 JSTOR institutions that are in this sample, when you look at the discovery service implementations by JSTOR Class, you can see how the variances do not show much of a pattern (Figure 10). Some went up, some went down, and to me this begs the really interesting question(s): Why?

For instance, if you look at Discovery Service “B”, you will see that the average monthly JSTOR usage increased an average of 7.1% for the 15 institutions in that pool after a discovery service implementation. I really want to dig into those 15 institutions. Were there any commonalities amongst those 15 institutions? Were they all a particular type of institution? Do they all have a particular type of link resolver? As you saw earlier this morning from the Clark-Price presentation, there are always some real outliers in these analyses. We need to take those into account and dig into these numbers a lot deeper. But, at the very least, you can see that there are really some interesting (perhaps important) differences in these data.

Seeing these data is really interesting because, up to this point, we (JSTOR) have treated these discovery services equally. We provide each of them with the same metadata for indexing, and

Culled customer lists to JSTOR participants worldwide for EDS (EBSCO), Primo (Ex Libris), WorldCat Local (OCLC), and Summon (Serials Solutions/ProQuest)

- o Looked at average content access per month for each JSTOR Class for 12 months prior/post implementation date
- o JSTOR average usage change for all higher education (August 2009 – September 2012): **-0.7%**

| Discovery Service | Usage Change Post-Implementation |
|-------------------|----------------------------------|
| A (541)           | <b>-4.6%</b>                     |
| B (340)           | <b>-1.3%</b>                     |
| C (18)            | <b>7.1%</b>                      |
| D (238)           | <b>-1.3%</b>                     |

Figure 9.

- There are significant differences by JSTOR Class within each discovery service. Why?

| Discovery Service | Very Large            | Large                 | Medium                | Small                 | Very Small            | Any           |
|-------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|---------------|
| A (218)           | <b>-3.1%</b><br>(7)   | <b>-11.1%</b><br>(16) | <b>-9.9%</b><br>(70)  | <b>-11.1%</b><br>(54) | <b>-12.1%</b><br>(65) | <b>-8.7%</b>  |
| B (100)           | <b>1.1%</b><br>(21)   | <b>7.1%</b><br>(15)   | <b>-11.6%</b><br>(32) | <b>-7.4%</b><br>(11)  | <b>-3.9%</b><br>(20)  | <b>-0.4%</b>  |
| C (13)            | NA<br>(0)             | NA<br>(0)             | <b>-8.6%</b><br>(3)   | <b>-19.2%</b><br>(5)  | <b>5.7%</b><br>(5)    | <b>-13.3%</b> |
| D (117)           | <b>-10.0%</b><br>(21) | <b>2.1%</b><br>(13)   | <b>5.6%</b><br>(46)   | <b>-5.7%</b><br>(20)  | <b>9.3%</b><br>(17)   | <b>-4.4%</b>  |
| JSTOR (1,599)     | <b>-2.9%</b><br>(86)  | <b>-3.2%</b><br>(87)  | <b>-2.7%</b><br>(390) | <b>-3.2%</b><br>(363) | <b>-5.9%</b><br>(638) | <b>-3.2%</b>  |

Figure 10.

that metadata is syndicated at the same time and in the same format. JSTOR has not provided these discovery services with full text for indexing for a myriad of reasons (a different presentation topic for a different day). These data raise all sorts of questions as to whether we should be differentiating our efforts and, if so, on what basis. More importantly, it raises questions as to what we could/should be doing better—across all the discovery services—to ensure that we are helping libraries leverage those substantial investments as effectively as possible. Looking at the changes in JSTOR usage worldwide in higher education (Figure 11), you can see that the usage changes differ a little bit from the U.S. numbers we saw previously and, actually, are even less consistent. You can see again there are differentials (above and below the baseline of -0.7%) in almost every JSTOR class. We have got good sample sizes for these data, so it will be very helpful—in the next phase of this analysis—to have those sample sizes in trying to ascertain the commonalities and differentials amongst those institution sets.

- There are significant differences by JSTOR Class within each discovery service. Why?

| Discovery Service | Very Large     | Large         | Medium         | Small          | Very Small      | Any   |
|-------------------|----------------|---------------|----------------|----------------|-----------------|-------|
| A (541)           | 2.1%<br>(11)   | -4.9%<br>(51) | -4.5%<br>(134) | -9.7%<br>(109) | -4.9%<br>(220)  | -4.6% |
| B (340)           | -1.2%<br>(26)  | -0.3%<br>(80) | -2.8%<br>(114) | -4.0%<br>(53)  | 5.2%<br>(62)    | -1.3% |
| C (18)            | NA<br>(0)      | 15.3%<br>(1)  | -10.8%<br>(4)  | -19.2%<br>(5)  | 30.6%<br>(8)    | 7.1%  |
| D (238)           | -7.3%<br>(24)  | 2.5%<br>(30)  | 4.6%<br>(89)   | -3.4%<br>(36)  | -2.6%<br>(58)   | -1.3% |
| JSTOR<br>(5,258)  | -2.5%<br>(111) | 0.2%<br>(359) | -0.9%<br>(917) | -2.0%<br>(864) | 0.9%<br>(2,637) | -0.7% |

**Figure 11.**

Okay, this is all very interesting and head scratching, but what did we really pull from this work? This effort was really less about the discovery service providers, and more about JSTOR. What is happening here?

While we were identifying and evaluating this usage data, we were also doing deep dives with each of the providers looking at (1) how relevancy ranking works in these systems, (2) how each of the providers use the metadata that we send to them, (3) what, from their perspective, could we be doing different to make content in the JSTOR archive work better within these systems, and (4) what could/should libraries know about the setups of these systems that impact when/how content is actually surfaced for discovery. These are more relevant “lessons learned” from that research.

First, subject metadata matters a lot. A lot. In terms of relevancy—for all of the services—subject metadata is critical. Now you have to remember, JSTOR was started with a mission to be the long-term preservation home for the journals on the platform. We were building our metadata for preservation purposes, not for access purposes, so we have very limited subject metadata in the corpus we are providing to the discovery services for indexing today. That has to change. We have to put much more robust metadata in our data syndication, and we have started working on methodologies to accomplish that. This probably means going back through 10 million articles and populating them with metadata that is going to be more useful for people when they are doing discovery within these services. That is not an inexpensive proposition to do., but it is important.

It is really important. And to us, it is actually much more important than providing full text to these services for indexing. It is a bit of a proposition that I am testing. I think there is little doubt that full text will give someone more results. The proposition that I am testing is will it give someone more relevant results. We are going to be doing some testing by giving full text of the Archive Collections to one of the providers and looking at usage to see if it actually provides more relevant results within the discovery service and then whether it drives more content accesses, too. We are going to start running that test at the beginning of next year, and if it works, if it is important, if full text proves to be substantially more relevant and a substantially better vehicle for driving content accesses on the JSTOR platform, then we are going to figure out a way to give full text to these providers. Since we are not the publisher, and because we do not own the rights to the full text, we are going to need to talk to the 1,000-plus publishers who are participating in JSTOR and get their buy in. I do not think it will be a problem with most of the participating publishers, but I want to have some data to back up why we want to do this.

The second point is that it became very clear to us from our interviews with dozens of libraries that libraries do not spend enough time configuring these systems before (or after) implementation. Most of them come out of the box in the default setting, and that is how libraries leave them. Now, I realize there are a lot of institutions that do not have staff to stay on top of these systems day after day (nor should they, really), but some of these systems have default settings that are really disadvantageous to surfacing the content on the JSTOR platform. If you have not adjusted those default settings, you may not see any content from the JSTOR platform in your results set. I am sure this is an issue that every content provider needs to be aware of. Link resolver configuration is critical, as well. These two things work hand-in-hand, and your link resolver and your discovery service have to be in sync. There is not a broad enough understanding in many of the libraries that we spoke with that the link resolver was overriding some of the discovery service settings. I think we definitely need some better education on these topics to enable librarians to act more purposefully.



And that is why one of the first things that we are working on at JSTOR, based on what we discovered here, is to create a set of guides that we can provide to libraries that say, "If you are using Summon, and you are using SFX as your link resolver, here are the things that you would need to do in those systems if you actually want your JSTOR content to show up and show up more prominently." Some libraries will say, "I do not want to bother with that." That is fine, but a lot of libraries in this room have made significant investments in JSTOR over the past 15 years, and they think it is an important resource for their campuses. I want to make sure—at the very least—that we have put that information out there so librarians will have a choice and will have a better understanding of the impact of the default settings. In the next few weeks, we are going to begin working on that series of guides, and I hope we will be able to begin releasing them in March 2014. Finally, we have learned that we must change the way that we syndicate content out to these services. The way we were doing it before was completely insufficient. We were sending the same data to everyone on a quarterly basis. No anomalies. We have now rebuilt a system that is going to be able to differentiate whether we are sending backfile journal content from the archive collections or current journal content from our current scholarship program. Those content types have very different syndication needs, and the syndication needs of the journal content may be different from that of the book content on the platform. We also need to better differentiate what we send to discovery providers depending on what works best in their particular system. Building a syndication platform to offer this type of flexibility is going to be important as we move forward. We have done a good deal of work on this already, and it has been a significant financial investment for this organization, but it is a necessary cost of doing business as we move forward.

Connected to the syndication conundrum is the importance for maintaining good, substantial KBART files for the link resolvers. Giving institutions up-to-date, robust files of titles they have licensed from JSTOR for loading into their link resolvers is essential to ensure proper linking from the discovery services. Nobody wants to go into their

link resolver and "turn on" access to hundreds (or thousands) of individual book titles or journals. We have to make this much easier for libraries or they just will not do it at all.

So syndication of metadata and getting our subject metadata much more up to speed for access purposes are the two big takeaways from this first set of analysis. Also, building the library guides so that we can better inform libraries on how the combination(s) of discovery services and link resolvers work with the content indexed from the JSTOR platform is another important takeaway.

Our next steps here are really quite simple. Besides doing the work I just mentioned, we are going to continue to dig into this usage, and we are going to continue to look at these different institutional archetypes by different discovery services so that we can start to see if there any commonalities about which we can inform the community. The discovery providers are very interested to get this, as well.

I have to admit that I sometimes butt heads with these folks, but they were wonderful to work with on this project. We went through these presentations with many of them, and they had incredibly good feedback to offer. I really wanted this to be a collaborative effort, and they provided us some very useful things for us to think about in our next round of research. I did not want this to be a "we" versus "them" finger pointing exercise. These things are still relatively new, and whatever we can learn from them for the benefit of the community is going to be good. We need to do that together. It cannot be something where we are putting stuff out and people are saying "well, that is wrong because of this" or "that is wrong because of that." We need to actually build on this together. Just like the guy said yesterday morning, this is a start, but we really want to understand what is happening and why because, in the type of organization that we are, a small not-for-profit organization, we have to think hard about where those investments are going. I want to make sure we are doing things that are smart and not just doing them based on some set of anecdotal evidence that we may have. So I am going to stop there. I am happy to take any questions from the audience.