

# An Open Approach to Scholarly Reading and Knowledge Management



# AN OPEN APPROACH TO SCHOLARLY READING AND KNOWLEDGE MANAGEMENT

Exploring Scholarly Reading through Publisher,  
Librarian, and Reader Perspectives

Hugh McGuire, Boris Anthony, Zoe Wake Hyde, Apurva  
Ashok, Baldur Bjarnason, and Elizabeth Mays

**The Rebus Foundation**

**Montreal, Canada**



An Open Approach to Scholarly Reading and Knowledge Management by The Rebus Foundation is licensed under a Creative Commons Attribution 4.0 International License, except where otherwise noted.

Published March 1, 2018.

# Contents

Introduction	1
<i>Research Goals</i>	1
<i>Our Process</i>	2
The Scholarly Publishing Ecosystem	5
<i>Background</i>	5
<i>Publishers</i>	5
<i>Aggregators/Distributors</i>	6
<i>Libraries</i>	7
<i>Readers</i>	8
<i>The Scholarly Publishing Cycle</i>	9
Considering Publishers	10
<i>Background</i>	10
<i>The Challenges of Monographs</i>	10
<i>Print vs. Digital</i>	11
<i>PDF vs EPUB</i>	12
<i>Digital Rights Management (DRM)</i>	12
<i>The Relationship With Libraries</i>	14
<i>Possibilities for the Future</i>	14
<i>Conclusions</i>	15

Considering Librarians	17
<i>Background</i>	17
<i>Print vs. Digital</i>	17
<i>Publisher Tools and Digital Rights Management (DRM)</i>	18
<i>Scholarly Reading Practices</i>	19
<i>Possibilities for the Future</i>	20
<i>Conclusions</i>	21
Considering Scholarly Readers	22
<i>Background</i>	22
<i>Print vs. Digital</i>	22
<i>Not Just Monographs</i>	23
<i>Access to Content</i>	24
<i>Workflows and Workarounds</i>	26
<i>Conclusions</i>	31
Online Survey	32
<i>Methodology</i>	32
<i>Summary of Results</i>	32
<i>Conclusions</i>	38
The Rebus Approach	40
<i>Open, Open, Open</i>	40
The Way Forward	43
<i>A Sustainable Business Model</i>	43
<i>From Prototype to Product</i>	44
Conclusion	47
Appendices	49
<i>i. The Rebus Personal Library Prototype</i>	49
<i>ii. Interviewees</i>	54
<i>iii. Online Survey Questionnaire</i>	55
<i>iv. Summary of Rebus's W3C strategy</i>	55
<i>v. Acknowledgements</i>	57
<i>vi. About the Rebus Foundation</i>	57

# Introduction

Throughout 2017, the Rebus Foundation undertook a research and development project to investigate both the ecosystem of monograph publication and people’s approaches to “deep” reading (i.e. deliberate, critical, and reflective reading for comprehension). This work was made possible through the support of a USD\$75,000 planning grant from the Andrew W. Mellon Foundation.

The driving thesis behind the project was to focus on how the Open Web—its technologies, its business models, and its ethos—can improve scholars’ access to, and interaction with, scholarly monographs, and whether these approaches could transform scholarly communications for the better.

We began by developing a simple prototype of a scholarly reading management system (see Appendix i for a detailed overview of this prototype), working on our existing understanding of readers’ needs. In parallel we engaged with the major stakeholders in the scholarly publishing ecosystem (scholarly readers, academic libraries, and university presses) to understand whether a tool such as ours might find a place among them, and to learn how it could be improved to better serve their needs. In total, twenty-two interviews were undertaken (see Appendix ii for list of interviewees) as well as an online survey with more than 100 respondents, all of which serve as the basis for the conclusions in this report.

This report explores the perspectives of these major stakeholders in the scholarly publishing ecosystem, and provides direction for future work by the Rebus Foundation on scholarly reading.

## **Research Goals**

We set out on the research portion of this project to engage with the major stakeholders identified in the previous section and achieve the following:

1. **Learn more about the scholarly publishing ecosystem from those most closely involved and invested in it.**

While our research team had some experience in the context

in which our stakeholders operate, the experience of the stakeholders themselves, on the front lines of scholarly publishing, proved indispensable in shaping our understanding of what could be possible in the future.

**2. Learn more about reader needs.**

As with any (potential) product development process, an intimate understanding of user needs is critical. By engaging with scholarly readers, we could explore their current workflows, challenges, preferences, and pain points.

**3. Solicit input and feedback on product direction.**

Along with reader needs, we sought to record specific feedback on the prototype in development and gather ideas for future feature development.

**4. Secure buy-in from key stakeholders for future work.**

In addition to general observations on the existing system, we also hoped to secure commitments from organizations interviewed to work with the Rebus Foundation as we continue to develop our product and a sustainable business model.

**5. Develop a feasible business model for future work.**

The long-term viability of a project to develop new technology for reading scholarly material is only possible with a sustainable business model. The development of such, informed by conversations with publishers and librarians, was a key goal for this project.

## **Our Process**

We conducted a number of interviews with a wide variety of people who engage with scholarly books: publishers and librarians, as well as professors and graduate students. The interviews were free-form conversations around the topic of scholarly books and digital reading, and each culminated with a demonstration of a prototype.

The prototype consisted of a web-based “personal e-library” software system, which evolved over the course of the research project. The prototype was based on a number of our hunches, with ideas for further development gleaned from interviews.

Interviews were recorded, transcribed, and synthesized into summary overviews, which were then excerpted into the present report. Each of the groups we spoke with—librarians, publishers, and readers—encompasses a diverse range of experiences and perspectives. Consequently, interviewees were expected to speak only to their individual professional experience. Effort was made to ensure the group of participants represented as wide a range of roles, organizational structures, and years of experience as possible. Participants included:

- scholarly communications, subject, copyright, and data librarians



- multiple directors and editors of university presses of varying sizes and models
- graduate students, adjunct professors, assistant professors, and department chairs in a range of Humanities disciplines, representing the spectrum of research experience

All interviewees were based at North American institutions. Please see Appendix ii for the full list of interviewees, who we gratefully thank for their time and insights.



# The Scholarly Publishing Ecosystem

## Background

Scholarly publishing is a broad term that describes the range of activities related to the publication of academic research. For the purposes of this report, we define scholarly publishing as a system, or series of interactions, through which academic writing or research is disseminated and consumed within scholarly communities.

Throughout our research and this report, we focus on one type of scholarly publishing output: the scholarly monograph. While other forms of publishing (namely journals) were touched upon in our interviews, the focus of our research and prototyping is on the monograph ecosystem.

The nominally “closed system” of digital academic publishing has four main actors: publishers, aggregators, libraries, and readers.

Note that there is a fifth actor in the ecosystem, ebook retailers (Amazon, Apple, Nook/Barnes & Noble, Google, Kobo, among others) that we have purposefully excluded from our research. Our focus is on exploring how technologies can help academic libraries improve scholarly reading.

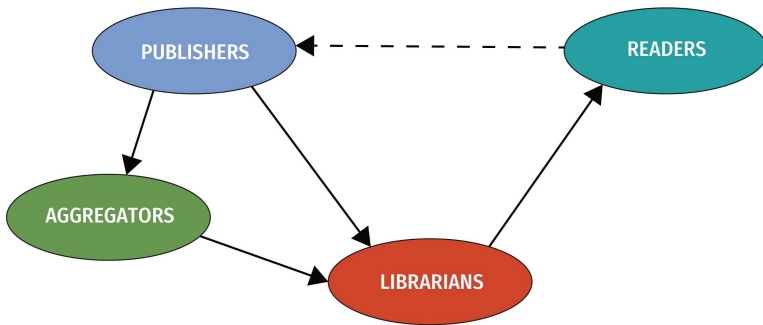
Below, we describe the role that each group plays in the scholarly publishing ecosystem, and trace their interactions with the other actors. We posit that the interactions between actors are not linear, with content simply flowing from publisher through aggregator to library and finally to reader. Rather, the reader in a research context is also a creator, feeding input back to publishers.

The consumption of content flows directly into creation of more content, which is published and distributed to other readers. This cycle is modelled in the following figure.

An overview of each actor’s role in this ecosystem is described below.

## Publishers

Publishers work to “make public” scholarly work in the form of textbooks, journals, and monographs, and represent a wide range of pub-



lishing approaches, business models, budgets, and institutional affiliations. With our focus on monographs, the two most significant groups are large commercial publishers and university presses. These publish the vast majority of monographs in circulation, although in recent years, smaller open access publishers have also begun to emerge.

The role of publishers includes (among other things):

- acquisitions and list curation
- editorial work and coordinating peer review
- design and production (for various formats, typically: print, digital PDF, and EPUB)
- distribution and marketing of finished products into various channels (libraries, aggregators, stores) where readers can access books

The last of these—distribution—involves direct relationships between publishers and either distributors/aggregators or libraries, which is typically a commercial relationship. Most publishers of monographs now provide digital files (PDF and/or EPUB) into academic libraries.

The overarching problem for publishers of monographs is one of revenues. In general, publishers report that revenues for scholarly monograph titles are in a downward slope.

It's important to note that publishers tend to be two removes from the readers, since content/digital files usually go from publishers to aggregators for distribution, and then from aggregators into libraries, who in turn serve the readers. This means that structurally, publishers have little interaction with readers.

However, as publishers acquire content from the same people who consume their published works, there is a separate relationship between the two which we will revisit in the Readers section.

### **Aggregators/Distributors**

While publishers create the marketable product (monographs), distrib-

utors (known as aggregators in the context of digital distribution) are usually responsible for getting monographs to the places where readers access them: libraries and retailers. (In this report, we focus only on the library ecosystem, and ignore the role of retailers).

The aggregator role involves:

- acquiring books (digital files + metadata) from publishers
- delivery of books (both digital and print) and associated meta-data to libraries
- managing commercial relationships with and between libraries and publishers
- building tools to improve discoverability of books
- (sometimes) providing reading platforms for digital books to library patrons

Most commercial aggregators (Proquest, Ebsco, and Overdrive) provide complementary digital reading platforms to libraries. However, these are closed-silo systems, which do not allow other tools or systems to build on them. In general our interviews, particularly with readers and librarians, suggest that these digital reading platforms tend to pose significant frustrations to users.

Non-profit aggregators (such as Project Muse, JSTOR, and ACLS Humanities e-Book), have not provided digital reading platforms to date.

## **Libraries**

Libraries are pivotal to the scholarly publishing ecosystem. They are significant purchasers of content from publishers and aggregators, and they are a primary source of published research, archival resources, and other materials for researchers.

Their role in the academic ecosystem is:

- curating collections of books (print and digital) by purchasing from publishers and aggregators
- making books discoverable or searchable to readers
- providing readers access to print books, digital files, and aggregator-created reading platforms
- providing patrons with services, tools, and other research support

Libraries are uniquely poised in that they interact with publishers, aggregators, and readers in various capacities, and they are ultimately responsible for placing books in readers' hands. Unlike other relationships in the ecosystem, the one between libraries and readers is not a commercial one, as a library's parent institution typically finances its activities.

However, library budgets have been dwindling over recent years, which has impacted their purchasing decisions. This has been a factor in a strong preference for providing digital access (via publisher or aggregator platforms) over print, since the overhead cost for digital files is far less than print, which requires checkout, shelving etc.

Despite these budget cuts, libraries continue to offer important services to students and faculty at their institutions, providing essential support to the research community.

## Readers

Readers are, of course, the driving force behind the whole scholarly reading ecosystem.

The most obvious reason is that readers are consumers of content. Without demand for content from readers, publishers (as well as aggregators and libraries) would have little reason to exist.

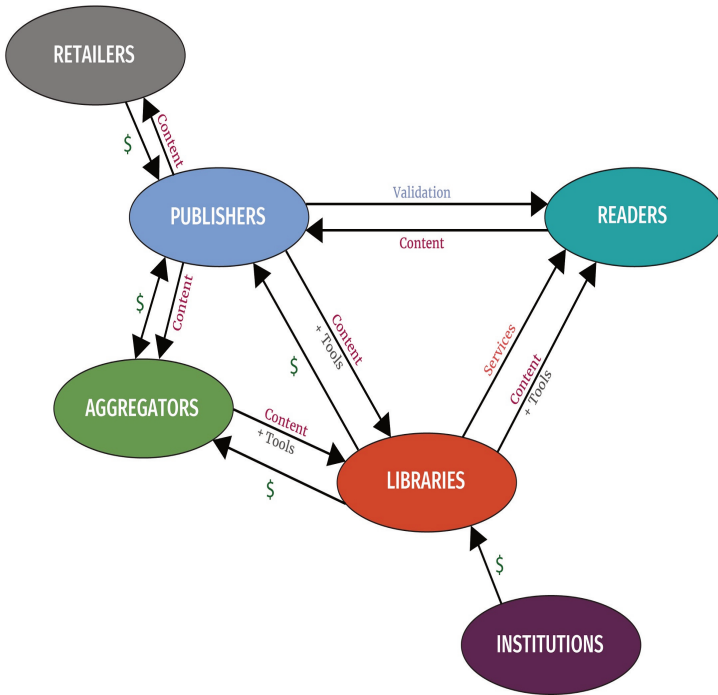
However, in the cycle of scholarly publishing as described above, the reader has two critical roles:

- to access and consume content
- to produce more scholarly research

Readers are also authors, and they are authors who are driven to publish not for direct financial reasons, but because published original research is the central mechanism of career advancement in academia. The connection between consuming (reading) and creating (writing) is critical to the academic ecosystem.

Publishers acquire (or commission) manuscripts, which allows them to publish and sell books. In exchange, authors may receive royalties (which range from very small to very large amounts—in the current climate of low sales, it is more often the former than the latter), but they also receive professional validation. Publishing a monograph is critical to career advancement for scholars in the Humanities, and a publisher offers not only an avenue to publication, but also quality assurance, via editorial and peer review, as well as prestige value through their reputation. This relationship is the critical last piece of the cycle introduced earlier.

However, while readers form the linchpin of the scholarly publishing ecosystem, their needs — beyond basic “access to content” — are rarely considered a priority. Readers are indirectly impacted by publishers and aggregators who supply reading and discovery tools, that are often restricted in various ways by digital rights management (DRM). In many cases these DRM-enforced restrictions are in conflict with the readers’ needs. This mismatch of interests has emerged in large part due to the evolution of publishing technologies, the considerable challenges of the shifting digital landscape, and increasing financial pressures on publishers and libraries.



## The Scholarly Publishing Cycle

Having explored the scholarly publishing ecosystem and its primary relationships, we can update the cycle as follows:

Our project set out to explore and address the shortfall in serving the scholarly reader identified in this section. This shortfall is made clear in two connected points:

- Scholarly readers are not just content consumers; scholarly reading is an act of creation as well.
- Publishers and aggregators are not incentivized to create better tools to support scholarly reading.

From here, this report will consider the experiences of publishers, librarians and readers through a synthesis of interviews conducted with several members of each group, as well as a short online survey aimed at readers. We will then share some of our own philosophy on the future of scholarly reading, then detail the path forward we see for our own work in the area.

# Considering Publishers

## Background

We started with the following hunches about publishers in the scholarly monograph ecosystem, born largely of the experience from members of our team who have worked in digital publishing technologies:

- Scholarly monographs have low unit sales, and sales are declining.
- Scholarly monographs are nonetheless seen as important to the mission of university presses, even if their economics are increasingly difficult.
- Libraries are the main buyers of scholarly monographs.
- EPUB is the lingua franca of publishing digital versions of monographs.
- Publishers are not willing to publish digital versions without DRM.
- Working with university presses on new digital initiatives is difficult.

It is important to note that not all of these hunches are correct, as we discuss below, and as highlighted in the conclusions to this section.

We have synthesized the results of our interviews with academic publishers into the following themes.

## The Challenges of Monographs

Perhaps unsurprisingly, the first theme to emerge from our interviews was the current challenges in monograph publishing. While monographs continue to be important to the mission of the university press, the economics around them are increasingly difficult. As one publisher of a smaller press says:

“Monograph sales have been dropping my entire career. I’ve been in the



business about 24 years, when I started the Chicken Little moment was, ‘Oh my God, monograph sales are falling below 2,000 copies a year, what are we going to do?’ Today we’re lucky if we sell 300.”

This has impacts across the press, both in economics of the whole publishing operation, but also on decisions about acquisitions:

“Books that ... have not really done anything for us in terms of list, building a reputation and also we’re losing money, those are the ones that we’re thinking we should just not publish.”

Still, while sales numbers for monographs are declining, in aggregate, backlist sales can be important:

“A monograph will move onto the backlist after a couple of years, and will still continue to sell, in both print and ebook formats, albeit in relatively small numbers. But that long tail of the backlist contributes to a little bit of an annuity almost, that the press has every year, that they can depend on.”

This means that while any given monograph might have questionable economics, for some publishers, the overall collection backlist of monographs can contribute significantly to revenues. Given the tenuous nature of university press finances, any threats to existing revenue streams will be approached with caution.

Sales from monographs represent a relatively small portion of revenues for a university press, but they are still significant enough that some care has to be taken around any project that might disrupt revenues, small as they might be, from monograph sales.

### **Print vs. Digital**

Print remains the primary revenue stream for monographs, with reports from our interviewees ranging from 70-90% revenues from print and 10-30% revenues from digital. This has an impact on where publishers can focus their innovation/development resources:

“If you look at individual book sales...the market is still very much interested in the print book and so, just the fact of where the money goes obviously has had an impact on what we invest in developing, right?”

While print remains a much larger revenue stream for university presses (often selling through brick-and-mortar and online bookstores), libraries drive digital acquisitions:

“The primary sales for the digital edition of a scholarly monograph are also to academic libraries. Those are generally done through third-party vendors also like ProQuest or EBSCO.”

Indeed, while print is a revenue driver for university presses, many libraries are moving to a digital-only acquisitions policy, which is shifting the terrain:

“There are a lot of libraries now who are on electronic-only, ebook-only approval models, where if it’s available as an ebook they won’t even source the print book.”

This poses a challenge for publishers, because while print remains the economic driver, many libraries require digital format, to the exclusion of print. And so, because scholars wish for their books to be purchased by libraries, they value digital publishing highly when selecting a press for publication:

“Total digital sales are well south of 10% of our total sales. Our cost for creating these are north of 10% of our total cost. So, the thing is, from a pure ROI investment, it’s not yet making a whole lot of sense to do everything in digital, but if we don’t, then there is a danger that that could even hurt the print sales, because we could be perceived as not really being engaged in the process, in the dissemination of scholarly research.”

### **PDF vs EPUB**

One of the surprises in our discussions with university presses (as well as scholarly readers) is the importance of PDF as the digital format of choice. This is driven partly by cost.

“And frankly, because it costs more to convert to EPUB, there’s no reason to spend the extra cost to create, that I’ve discerned, to create an EPUB edition to sell that content to an academic library, because you’re not going to increase the sales numbers by having increased the production cost.”

While our research has been driven by particular interest in developing a networked, web-based reading platform (with HTML as the base format), it is clear that PDF is currently the preferred digital format for many scholarly readers, largely because of two factors: portability and searchability.

“That’s why PDF is working better, maybe not ultimately, but is working better, because it’s searchable and digital has helped scholars because they can say, ‘Okay, I’m working, this is the next project I’m working on, I want to know if this book has something about this.’”

This is one of the challenges we are likely to face as we build out a digital reading platform — to build tools that will surface the value of content that is much more accessible and usable, in the form of web-based reading (or even EPUB), and demonstrate clearly the value — to readers, and to publishers — of moving beyond PDF.

### **Digital Rights Management (DRM)**

Another potential blocker for innovation is Digital Rights Management (DRM) or digital-locking technologies that limit what a reader can or

cannot do with a given work, and further limit what a reading system can do.

Typically DRM might block a reader's ability to download a work, so that research can be done at a later date; or to copy-paste from the work, for instance to support quotations that will go into a research paper.

Digital Rights Management continues to be contentious, but embrace of DRM was far from universal and many publishers see it as an unnecessary blocker to legitimate use. Still, for books that might make significant contributions to sales numbers, one publisher had this to say:

"I will never put front list into a DRM-free platform again."

The reality, however, is that most monographs are not front list, and it's very unclear that DRM has any impact on business numbers of monographs at all.

Our own interviews with readers and librarians suggest that DRM causes significant friction in acquiring and using digital content (more on this in the following sections), and one publisher admits,

"We're not necessarily protecting a whole lot by slapping DRM on these titles."

The conflicted views of DRM were prevalent, even among those publishers we spoke with who are generally supportive of DRM in at least parts of their business:

"I'm coming to the point where for the standard scholarly monograph, I'm less concerned about DRM and tightly controlling that, but I will admit, because that's a genie that can't get put back in the bottle, I am proceeding cautiously."

One large press has taken a very pragmatic view about the actual (versus perceived) value of DRM in protecting content from unauthorized use:

"Pretty much anything that we publish is, for lack of a better word, pirated within a matter of weeks. And that's just the reality of it, and so anything that we do has to take that into account. We also feel very strongly about in most cases, unless it's a trade author or the author is concerned, we don't pursue takedown actively. We're moving away, pretty much across our platform except for textbooks... away from DRM altogether. "

In general the support for DRM was present, but lukewarm, and especially unconvinced when it comes to monographs. In some cases, there is direct evidence that content without DRM can be even more commercially valuable:

"To give you an example, our best-selling book of the past season is a textbook on deep learning, [that sold] multiple tens of thousands of copies. It's an openly available book in HTML on the author's site, which is great. But that's the kind of experience that we want to be able to provide [for our readers]."

Our expectation is that as tools are built to show the real value of non-DRMed content — the ability to allow scholars to use their digital texts in new ways — support for DRM among publishers will continue to fall, driven largely by demands from readers.

We heard this antipathy to DRM reflected in the desires of librarians, and from those publishers who have the resources to be thinking more deeply about the digital future of monographs. The cost of the friction caused by DRM in acquisition and use of digital content by readers may, on balance, outweigh the benefits of “protection” of content, especially for publishers starting to think about selling collections, rather than individual units, to libraries:

“If you look at what libraries rightfully expect in terms of features and functionality of digital books: no DRM, unlimited multiple users, perpetual access, the line between those principles of ebooks, ebook subscriptions, and persistent access is really blurry.”

### **The Relationship With Libraries**

Some presses, as they think about shifting thinking driven by digital access to books, are beginning to rethink the traditional supply chain and commercial arrangements. As established earlier, publishers give their content to aggregators, who in turn sell content into libraries. But again, this removes publishers from their true customers (libraries), and puts them even farther from the people actually using their books: the readers. So, for publishers considering their lists as digital collections, the need for an aggregator in the middle of the transaction with libraries starts to be questionable:

“We’re saying, ‘Yeah, you’re going to pay us a subscription to put this content pretty much in the wild.’ And so, that’s another reason why I feel very strongly about owning that relationship [with libraries]. And I have no problem whatsoever with the openness of [the content].”

### **Possibilities for the Future**

The publishers were all given an opportunity to look at the Rebus open web-based reading prototype, and all of them expressed interest in engaging and exploring the potential increased value to readers, scholarship, and perhaps presses themselves:

“I think there’s a need to create these levels of research or tools, and it’s interesting, this is bringing some of the stuff that’s happening in open science into the discussion. There’s a lot of this kind of stuff that’s been happening in scientific publishing [that] hasn’t been happening in [Humanities and Social Sciences] publishing.”

In some cases the interest was tied specifically to the possibility that a system like the Rebus Reader prototype can help surface works (and open access works) by a particular press:

“So, this seems to give an option to actually show an institutional view of the books that were published by scholars at our institution, through the same platform, right? You could have a press-type view, then you have an institution view from multiple publishers. That’d be really cool.”

Further, the power of the reading system links directly with the possibility of increased innovation on the back of open access publishing, by giving readers and scholars new ways to interact with, and use, monographs:

“How would open access help push forward innovation? It’s only fairly recently, through things like [the Rebus Reader Prototype], that I’ve realized what the connection is.”

The range of publishers we spoke to includes those fighting for financial survival in the face of declining sales (partly related to the digital transition, partly related to other factors), to those whose whole mission is directly related to digital:

“I don’t worry about open access or digital gleaning away sales, because my purpose is the digital. My purpose is to say to authors ‘You are most likely in the 97% of scholarly authors who are not going to realize large amounts of personal income for writing your book. And therefore, your objective should be getting people to read your ideas, and that is what we can help you do.’”

## Conclusions

We started off with a number of hunches about how university presses view the current, and future ecosystem for scholarly monographs. Some of these hunches were confirmed by our interviews, namely:

1. Scholarly monographs have low unit sales, and these are declining.
2. Scholarly monographs are nonetheless seen as important to the mission of university presses; even if their economics are increasingly difficult.

However, our findings suggest that things are more complicated. Monograph sales are indeed declining, but in aggregate, the “long tail” sales of monographs remain important to university presses. And, they are still seen as important to the mission of university presses, even if their increasingly perilous economics make publishing them harder.

Libraries are indeed the biggest buyers of digital monographs. However, according to our interviews, print sales make up a larger percent of revenues (70-90%), and most print sales are happening through brick-and-mortar and online stores.

In other areas, we were surprised to learn that PDF (not EPUB) is the lingua franca of publishing digital versions of monographs. EPUBs are produced by some presses, but not seen as necessary for digital pub-

lishing, and PDF is the baseline digital format for monographs. EPUB is seen as more of a luxury, and indeed comes with added costs for publishers to produce and distribute.

Finally, as a positive and encouraging surprise, we found that while there is still some nervousness about removals of Digital Rights Management (and the possibility of lost sales through pirating), there is an increasing openness to publish non-DRM versions of digital files into libraries. Aggregators including Project MUSE, JSTOR, and ACLS Humanities E-Book are now routinely offering their content with no DRM. One publisher reported a significant growth in library purchases and use once DRM was removed. Indeed, every university press we spoke to was interested in working together on new DRM-free digital initiatives, on projects such as the Rebus Reader system.

The overall impression from the interviews with university presses was very positive for new models and new experiments. In general, presses are cognizant that the existing business model for monographs (unit sales) will continue to be challenging, and increasingly so. They recognize that a new approach to digital could offer new value to readers, libraries, and indeed the presses themselves; and further that DRM may be hampering the ability to develop these approaches.

There is a willingness among university presses — indeed an excitement — to experiment with new digital models with the right partners.

# Considering Librarians

## Background

Librarians are vital members of the research community and are uniquely placed to engage with both sides of the publishing equation: publishers and readers. Their role involves engaging with and supporting researchers, both faculty and students, who are accessing all kinds of materials through libraries. They're also privy to the system of content delivery comprising publishers and aggregators and were thus able to offer valuable insight into the current landscape. In doing so, librarians expressed frustration with limitations of the current digital ecosystem, as it often adversely impacts readers, but in many cases librarians felt poorly-positioned to provide solutions to these problems.

Setting out, we had several hunches about current problems in the digital reading ecosystems in the library context, namely:

- Unsatisfactory user experience for delivery of digital versions of monographs to readers
- Digital Rights Management (DRM)/digital locks and usage restrictions, which make it hard for readers to get and use books as they wish
- Closed-silo reading systems provided by distributors/aggregators, that restrict development of innovative tools for scholarly readers

These hunches were largely confirmed through our interviews, and interviewees shared their vision for what the future of scholarly reading could and should look like.

## Print vs. Digital

The first issue we explored was how researcher preferences for digital or print are playing out in the current library landscape. The interviewees unanimously reported an entrenched preference among patrons for print when it comes to long-form texts, evident in everything from the

ongoing popularity of printing, to requests to be able to filter catalogues by print availability. While search and discovery are acknowledged as having improved in a digital context, and the utility of “skimming” large sets of resources to get a broad/shallow understanding of a topic were noted, there was no question that researchers consider print the better option for deep reading.

While some of this is likely influenced by habit, particularly for established researchers, almost all the interviewees noted that it’s difficult to say whether the bias comes from something inherent to print, or if it is a response to the inadequacies of existing digital books and reading systems. In short, they are,

“not sure whether [the preference for print is] because paper is amazing, or whether it’s because the digital tools aren’t good enough”.

### **Publisher Tools and Digital Rights Management (DRM)**

The issues cited with available tools were significant, starting with user experience. In the words of one librarian:

“The platforms are terrible. The user interface is bad. You get into a system, you have to click five or six times to get into it... [Once] you get into the system, even if you have a wide-screen monitor... the viewing of the content itself is limited to half the screen... It looks like it’s from the 90s. There’s limitations on printing, if you can print at all. The annotation is either non-existent or terrible, and it’s not hierarchical or smart in any way, and getting those annotations back out (because they have locked it down, because they view it... as a workaround to get the content out...) is near impossible and not functional. So you feel like you’re sort of looking at this thing through a digital glass box that you can’t touch or interact with.”

The proprietary publisher/aggregator content delivery model has also led to wide variations in what’s permitted on different texts, creating frustration for users, and for librarians who are left unable to effectively advise their patrons.

“We can’t encourage people to use things that are incredibly frustrating to use; they just won’t use them.”

Of particular importance in reflecting on user experiences of publisher/aggregator reading platforms is the fact that many of these limitations (on access, printing, and annotations, for example) are deliberately put in place due to fears of content being removed from their platform. These limitations exist under the umbrella of DRM, and our interview subjects unanimously named it the most frustrating feature of the current digital content ecosystem.

DRM is particularly counterintuitive for users when it comes single-seat or limited-seat licenses for digital content, where an artificial restriction placed on access removes one of the big advantages of digital over



print. As one interviewee put it, “It’s actually a value decrease, except for the accessibility, not a value add in most cases, to the print [version].”

“The DRM. [The biggest issue] has to be the DRM... [print is] easy to understand. You walk into the stacks, and you look for a book and it’s not there, or you look in the catalogue, and it says this book is checked out until May 11<sup>th</sup> 2017, right? That’s easy to conceptualize. You can recall books from people, but when you find it in digital format, and you’re thinking digital? That should be limitless and it’s not. It says the book is checked out, and it’s not clear whether it’s checked out for an hour or for a month, or do I just keep checking back? There’s no way to get notification when you’ve been accessing. I think that is the biggest frustration from a direct user perspective.”

In addition, in the case of aggregators, the most cautious publishers dictate what happens for all content in a platform, holding back others who may be more willing to allow certain features:

“When ebooks first arrived in libraries, there was a lot of hand wringing from the publishers, and they really wanted to restrict how much you could do anything with them. So the model that has survived for that is this aggregated platform, where a vendor has created a platform to host ebooks on for libraries... And because that vendor is having to work with so many different publishers with so many different levels of risk that they’re willing to take, they end up going with the lowest common denominator.”

### Scholarly Reading Practices

At the core of librarians’ frustration with the current system—poor tools, publisher variation and DRM—is the simple fact that existing digital reading systems do not support the kind of reading researchers are actually doing. A consistent theme from our interviewees was the idea that digital books—ebooks in particular—were only useful for the kind of linear reading associated with novels or other non-academic texts.

The issue when it comes to scholarly reading is that it is a much more dynamic, active, and interactive process that is simply not supported by existing platforms.

“If you’re not just sitting there with your Kindle on an airplane, reading cover to cover, not stopping, not making notes, maybe occasionally stopping to highlight something... but not really interacting with the text... That’s not how you work with scholarly books. You really need to interact with the text a lot...”

Scholarly reading involves not only making annotations, underlining, highlighting, and adding other marginalia, but also being able to flip back and forward within a text, or break it apart, comparing and rearranging pages and sections. All of these actions, while possible with print, are either limited or non-existent when working with texts restricted by DRM in various forms. Several interviewees said the kinds

of actions they take for granted in print might be made possible, and even improved, with digital texts:

“I would like to be able to take different sections of the book and lay them all out so that I can reference between the pieces. Just replicating the whole idea of being able to flip back and forth between pages pretty seamlessly. To be able to leave marginalia for myself that I could read later on, which is a problem if it’s a library book because you return the electronic copy, and all of your notes are gone too, so something that I could retain from usage to usage.”

“I think smart annotations that allow you to replicate things like indexing and tagging, and commenting, and highlighting. All of the things that we can physically do with the text and maybe envisioning more things. But annotation is not a simple experience. We can mark pages and so I think there’s that, but there’s also that ability to then take that and do something with it.”

Beyond existing reader actions, one of our interviewees also posited that as other kinds of reading become more interactive and readers’ ways of consuming texts change across the board, scholarly reader preferences are shifting in a similar direction, reiterating the need for innovative approaches to digital reading in the academic context.

Another added to this developing idea of scholarly reading by sharing the concept of a text as a dataset, on and from which a reader/researcher can extract data, annotate, layer, compare, and contrast and analyze.

“I think if you’re looking at text as a dataset, it’s about, ‘How do you visualize? How do you enable an integration for visualization or re-combination?’... This is how I read, and the digital environment doesn’t allow me to underline and tag and mark up, and then also export text into a note tool or any of those things.”

## **Possibilities for the Future**

Increasingly, as the availability of content at scale becomes the norm, the value lies in the experience of accessing that content. As those on the front lines of providing that experience, librarians can see that the current system of providing digital texts through proprietary publisher and aggregator platforms is not delivering what their patrons want and need.

The tools, with their embedded restrictions and limited features, are unable to deliver a useful reading experience for scholars, to the extent that digital reading is viewed as vastly less desirable and effective than print. This is disappointing, considering the innovations that digital technologies have offered in other areas of the scholarly publishing ecosystem, and should not be taken as the result of something inherent to digital reading. The digital environment can be dynamic, flexible, and connected, not siloed and restricted, and so too could digital scholarly reading.

Ultimately, the biggest barrier to improvement that librarians identify is publisher restrictions and DRM that not only limit access to and inter-

action with a single text, but also put walls between platforms, texts, and tools that might otherwise work together to serve readers' needs.

Any future innovations must move past this paradigm in order to be able to offer true value to readers. Tools and formats should be interoperable and agnostic to accommodate the full range of activities involved in scholarly work, and the librarians interviewed seem poised and ready to embrace such a change. They share our belief that academia is an ideal environment to pursue this kind of approach, and that once the value proposition is undeniable, scholars are poised for a big shift in their content consumption.

## **Conclusions**

The overall impression from the interviews with librarians is that there is a clear need not being met for scholarly readers. Our hunch around unsatisfactory user experiences with existing reading tools was confirmed, as was our expectation that DRM and usage restrictions are an ongoing source of frustration, limiting user interaction within texts. While we didn't dive deeply into the issue of closed-silo reading systems, our general impression from librarian experience is that publishers and commercial aggregators are not likely to be the source of new, innovative approaches to digital reading.

These interviews have also confirmed that libraries are not only enthusiastic about the possible emergence of alternative tools for digital reading, but they are ideally placed, and indeed critical, to supporting their development and adoption.

# Considering Scholarly Readers

Moving to our interviews with readers, our main objective was to listen to them free of assumptions, understand how they work, hear their concerns and frustrations, and identify ways that digital reading technology might effectively support their academic pursuits.

## **Background**

Scholarly readers are those whose reading is done in pursuit of new knowledge, primarily—but not exclusively—in professional academic settings. We focused primarily on professors and graduate students who engage in research, teaching, and publishing.

We were quickly reminded that scholarly readers don't read or annotate for a single purpose. An interviewee explained,

“For me, it's more a question of the function the reading that I'm currently doing has for me, and I have sort of three layers of engagement. There is—in print—there's a reading for teaching, there is a reading for understanding and things that I need to use for my research. And then there is reading digitally, which I usually actually use if I really need a text for a research project.”

Scholars engage in a wide range of activities with regard to scholarly materials, most of which are, according to our interviews, poorly addressed by existing digital solutions, and to some extent, even hindered by them.

## **Print vs. Digital**

We discovered that scholars don't favour print or digital per se, but rather use a mix of formats depending on the type of reading they are doing. For deep reading, most interviewees chose print. For one interviewee, this was just a matter of habit:

“I never embraced the digital note-taking phenomenon. It may well be that I was just too old when it came along. Sure it looks cool, but I never saw the value added. So, I guess the theory is you have a PDF, you can write notes

on the PDF, you can save those written notes. I guess it does make sense, I just never adopted it. And I think it's just because when that came around, by that time, I was at least thirty-five years old, and I had been a graduate student for ten years. And you continue with something you feel comfortable with, right?"

However, others would choose digital in certain contexts. For instance, digital was preferred when traveling, as scholars could take their collections with them; for skimming content when conducting a shallow read; to search for keywords to quickly assess the relevance of a text; or to locate quotations when writing; and most importantly, to retain a copy for quick referral while writing. While there was a consensus that "there is something about the printed text that's really wonderful," there was also agreement on the value of digital formats, during the writing process in particular. One interviewee described their process, saying,

"So, up until this point, up until the Ph.D., I really sided with electronic forms of printed materials, or of literature. And that was specifically because I was really mining from various sources, and I needed to— I realized that how I put together my ideas was in kind of gathering quotations from here and there, and the easiest way to do that without completely transcribing a book was to be able to copy and paste."

Surprisingly to us, PDF emerged as the favoured choice of digital formats among all the interviewees. One explained:

"When it comes to digital, I use PDFs. I never use EPUBs, I never use, you know, Kindle, and so on. Largely, because this problem that I think then, that the location gets saved in some digital page number. But I do read quite a bit, in PDF form."

The preference for PDF might stem from the similarity of layout in the digital PDF with the printed book, or the undeniable fact that the majority of digital and digitized monographs are only available in this format. The advantages of EPUB, let alone networked HTML books, are not apparent, partly, we surmise, because the tools for reading digital texts are so limited.

### **Not Just Monographs**

While our research was focused primarily on scholarly monographs, when it came to investigating reading behaviors, the conversations and responses encompassed reading of all sorts of formats, not just books. Monographs certainly figured, but so did primary sources, essays, criticism, reviews, articles, and journals, and often, only individual chapters of books were used.

"We have materials that are very hard to find, manuscripts. Manuscripts, bits of manuscripts, letters, or something like that."

What's more, these sources aren't always in a typical typeset format,

such as PDF. They may be flat PDFs from scans and photocopies, or even photos.

“People tend to like scan or photocopy or take [photos] with their iPhone [of] thousands of primary sources and then, figure out what is important when you get back home.”

The nature of these materials can also influence a researcher’s choice between print and digital formats. One researcher explained their shift from digital-first to a more hybrid approach once they began their Ph.D.:

“So if there were [digital] scans of books, then already in circulation on the Internet, then I would run it through an OCR [Optical Character Recognition] and be able to kind of use that in a more dynamic way than as an image... Now, I’m at a point where the amount of material that I must contend with—that model that I’ve just described is no longer appropriate. The amount of work, the amount of hands-on labor that I would need to do for myself is just— it puts that model to shame. It makes it obsolete (emphasis added).”

Readers we spoke to did not make much, if any, distinction between monographs and other materials, in contrast to how we had intended to frame the conversations. Any system of scholarly knowledge management would then naturally need to handle texts of all sorts and all formats (at the very least PDF, EPUB/Web, and images). This was a critical observation and has helped to shape our understanding of user needs.

### **Access to Content**

Curious to see the interplay between various groups in the publishing ecosystem, we asked scholars how they accessed the content needed for their research. It became apparent that access to content was by no means equitable or uniform.

For nearly all interviewees, the library was the primary point of access to content, with the Internet and search engines being the next:

“Immediately I go to the library. I go the library, I go to Google Books, right? ... I would do it normally in that order, library, Google Books, I see whatever. I just Google the author, and see where copies are available online and again, it’s that. And in 95% of the cases I’ll be able to get it.”

Others face the challenge of essential materials to their field being difficult to access for other reasons. They may have to travel to access archives, be allowed a limited time with them, or even be restricted from taking anything but handwritten notes. Yet, these materials are still a vital part of the researcher’s corpus.

“Let’s say that I go to a library in [city redacted], right? I go to a library in [city redacted], they have a manuscript, they don’t allow me to even to take a photo of it, nothing. No citations. I just copy out passages by hand. Never-

theless, I want to be able to cite it, I want to know how much of my notes I've used, you know, something, that's a very extreme example. But there can be many examples in the middle, of books that if they ever become digitized, it'd take five years or whatever."

"I'm now having to contend with texts that are really old, or are [in] archives that I cannot scan, or they're not accessible to me at all times. I only have them for a certain period of time, like a couple of hours sometimes, right?"

Our interviewees were quick to remind us that, even for more typical published materials, not everything they need is accessible through the libraries at their institutions:

"[The] most recent scholarship is generally available in digital form, through the library system. That having been said... a lot of the texts that I do work on... they haven't been digitized at all, and sometimes I have to use the ILL [inter-library loan system] to actually request a physical book from somewhere else, just because it just doesn't exist in digital form."

As a result, scholars in particular disciplines have resorted to sharing materials, such as manuscripts, letters, and photos of archived texts, in private groups. They are wary of copyright infringement but keen to help their peers.

Others pointed out that some libraries have subscriptions and access to larger databases. An interviewee explained:

"I've always been in an elite university, and that's relevant because I have had access to about as wide a range of secondary sources to read and then, later, I've been in universities that are so big and wealthy that they can buy the whole range of consortium. So, when the digital age came about, I could have access to any journal, any journal article and then later, when ebooks started to be included, I could have digital access to more or less any monograph that was a secondary source."

Adjunct scholars affiliated with elite universities may have the challenge of being located far from campus, meaning that they cannot browse the library's physical collections. Instead, they have to resort to other means to get materials – such as requesting library staff scan relevant works and send these via email.

Inequities in access to content are further evident in the case of scholars who are affiliated with universities that have less comprehensive access to resources. They must resort to paying exorbitant rates to access content. In many cases, these scholars will contact their peers with institutional access to "restricted" materials behind paywalls, and request a copy. For those academics who are unable to access materials critical to a research project, the effect can be detrimental to their careers.

"[It's a] convoluted system, where you can't make yourself attractive on the tenure-track market unless you have an affiliation....If not, you'll have no access to the journals that you need....It's a vicious, vicious cycle."

## Workflows and Workarounds

Among the scholars interviewed and surveyed, none made reference to any standard, formalized, or taught process, or workflow for managing their reading-related activities. However, a pattern emerged of improvised systems, combining different tools and developed over years of research, each remarkably unique from the others.

There were some commonalities, with four major areas emerging: file management, annotations and other text markup, note taking, and finally, writing. Beyond this, however, each researcher we spoke to had created their own system from scratch, piecing together different tools and methods available to them to create something functional.

### File Management

For file management, some use cloud storage services such as Dropbox or Google Drive for digital files. Others stored files on their computers, developing an often complicated, idiosyncratic folder structure. Some used a spreadsheet or other document to keep track of readings related to a given project, including things like reading statuses, priorities, availability, and format (print from library, need inter-library loan, PDF on computer, etc.):

“What I do is when I start a project, I usually write a bibliography of possible sources and I also use that to track what I’ve seen and what I haven’t seen. What I’ve ordered through inter-library loan, and what I haven’t. What I’ve gotten a PDF of, and what I haven’t. And I simply— I make a link in that as a Word file. I make a link in the Word file to the PDF, if I have it.... And if it’s a physical text, I put in the call number, the library, or I mean, if it’s— if I have it on my overflowing shelf, I have a lot of books, I really need to buy more shelves. Then, I know of course that I have it, usually. It’s really getting to the point where I don’t know what’s there.”

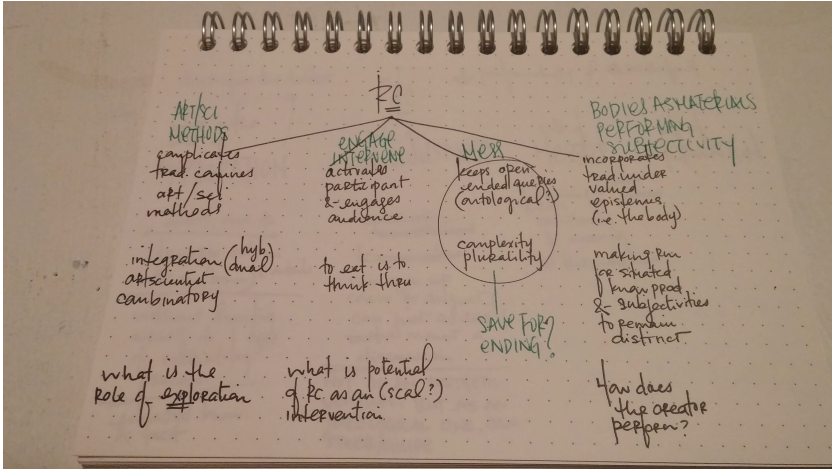
In one case, a researcher also included a system of accountability:

“I have a separate list, actually, for while I’m on project holding myself accountable to actually reading these texts. So, I make a subsection [of] which texts I have to read through, which project, I put them in parentheses, and these are just like plain-text files... where [I] just put in parentheses the final number, the total number of pages. And track myself, where I am in that text, because it forces me to read it, but the same time, where I have a subsection at the bottom, where I just keep accumulating the texts that I’ve read. Sometimes for a specific project but also overall for the texts that I’ve read in the year, just because it gives me a sense of what I’ve read, what I’ve achieved.”

### Annotations and Notes

Annotations and other text markup were done either on the page, in the case of print, or with a number of available tools for ebook and PDF





An example of a conceptual map created by one of our interviewees

annotation. We spoke to scholars to find out different ways that they write, annotate, or interact with texts. These included: notes or comments on paper, marginalia, index cards, blurbs, quote bubbles, highlights, sticky notes, conceptual maps, drawings, and conversation. More interesting was the language of colours, symbols, and other markers that each researcher developed to capture their thoughts as they read, and communicate significance, connections, project associations, and myriad other nuances of their interpretation of a text. Our interviewees explained:

“[I have] three levels of notations of one that I’m reacting to, one that’s like straight author’s points, main points. And then, my own reaction, or sorry, my own kind of question or discussion point that I want to bring up to the class.”

“I think in pictures and I think in lists. I think in like little blurbs and quotes, quote bubbles, key words, and then I’ll write where that thing is. ...This is a key word that is really kind of coming up. And so, then, the process of me drawing all of this out, that becomes the conceptual map within which I operate to write....drawing becomes my way of pre-writing.”

“I wish I could annotate and draw at the same time. I wish that the iPad, for example, screen were maybe five times this size, right? So like have four, right? What if it was this big? With the text still that big, right? And I could use the extra margins in a way where I could start to do this more...”

It seemed at times that the remarkable freedom of writing freeform allowed these languages to form, but it was difficult, if not impossible, to replicate that freedom on available digital tools. Printing out articles or chapters of interest and annotating them with pen or pencil is still seen as the way to go by many. Having physical copies on hand also means easier management as this benefits from the very natural use of space for arranging things, e.g.: “The pile on the right contains my primary

sources; on the left are things I've flagged as potentially interesting and to revisit." Often mentioned was the use of digital editions for quick consultation and search, but print versions for in-depth reading and annotation. Most collect important works in print.

While some note taking did take place alongside annotation, each of our researchers would reach a point where they needed to take the texts they had read and turn the notes, quotes, and other takeaways into something they could then begin to incorporate into their writing. Again, the approaches to this varied widely, and depended on the tools used initially. Some would take handwritten annotations and highlighting and type them into a word processor. Others would export annotations from tools in whatever way is available to them, though an arduous copy-paste process was more common. One interviewee wrote his own script to extract his notes from a PDF annotation tool:

"What I do, is in any PDF I mark it up, using iAnnotate, usually using the note function, I... type a note into iAnnotate, and then, what I do is I programmed a script essentially, which grabs all of the notes from the PDF file and puts them in order. And then, it produces my Word file for me, so for that that's very useful."

At this point, there is no clear end to the reading and annotating process as writing begins. Rather, the two coexist and continue in tandem.

"So, the problem is, if you leave it at the end, and you write, and you're like, 'Oh, I should have looked at this and that.' And then it's too late, right? So, you write very early in the process. And the writing process and the research process is iterative."

### Reading to Write

Over the course of the interviews it was clear that the acts of reading and writing were inseparable in the minds of researchers. Every journal article, chapter in an edited volume, or monograph is read with the purpose of a specific output.

"The writing process is not something that's done at the end... The writing process is something that starts, and this is something I emphasize with my doctoral students, is that the writing process literally starts right at the beginning. You write drafts, with the expectation that those are going to be revised, maybe thirty or forty times. But you start off with a draft and it's the process of writing, which itself is an intellectual activity, which exposes other things you need to research."

Annotations and highlights make their ways into notes and indexes, eventually to be massaged into whole new paragraphs, quotations, citations, and bibliographies. Scholarly reading is the continuous transformation of read material into newly written material.

This lack of separation between reading and writing is unique to scholarly reading, making it very different from leisure reading. One of our

interviewees explained the reason for this inseparability between reading and writing:

“You’re providing an original perspective, original research. But you need to engage with an existing body of knowledge [to do so].”

We learned that, for academics, reading is not simply deciphering the words in a single text. Rather, scholarly reading is reading writ large: a set of activities that includes simple consumption and comprehension of a text, and also a number of related activities, including collecting and managing a large corpus of texts, annotating, note-taking, and writing.

Academics read for a number of purposes, and toward multiple goals – reading to write a journal article or monograph, reading to include a text in a course, reading to prepare a conference presentation, or reading to stay informed about a particular topic. And as they read, as they write, they discover newer avenues:

“It’s just completely a snowballing effect, it really is. So, every article that I read itself has let’s say fifty references ... [T]he first thing we do, as researchers, is actually not even look at the content, as much. We’re looking at the sources that they’re using. Not just the primary sources, but the second, everyone’s sort of like, ‘Hey, I haven’t read that.’ I haven’t read that article.”

What’s more, any given reading project is likely to have multiple written outputs offshooting from a central focus. One interviewee explained,

“Authors will work through their ideas in journal articles, before they lead up to the big honking book. And so, it’s those are usually like the breadcrumbs that we follow until the book comes out, and then we all switch over to citing that book.”

Ultimately, scholarly reading cannot be separated from scholarly writing. This means, with our focus on the former, we must expand our understanding of what it means to read for research. Scholarly reading is not passive, or linear. As one interviewee put it, very succinctly,

“It’s a creative process itself.”

### Avenues for Improvement

Having spoken to the range of readers we did, it appears that while they had largely developed a system that worked for them, it was a case of making the best of what is available. All ran up against limitations in the tools they chose, or were frustrated by not being able to do something quite how they’d like. Where an action is not supported—and in some cases not “permitted”—scholars resorted to manual workarounds which could potentially be automated if the technology were more open.

PDF clutter and convoluted file folder arrangements were common complaints, as was the difficulty of managing both print and digital resources.

“What I find increasingly frustrating is when you have different research projects, but at the same time, sort of a general repository of texts. Sometimes, I end up saving new texts that I should be adding to the repository to a specific subject folder for a research project. And then, a month later, I’m incapable of finding it without putting fifty units of research on my computer, restricting around or using the search function....But it would be useful if there was a general system where you have a double access to the same file, through the folders.”

“And so I’m now straddling these two worlds, where I either have to keep renewing books because I’m not quite done with them yet. Or if they get recalled, then I just have to click on that button two weeks from now, so that I can get it back, right? Or everything is on my computer and I just have to keep a mental inventory of where that document is. And I really wish I didn’t have to expend my mental energy on that.”

Many were also frustrated by the cost of stringing together all the different tools they needed. While some are offered through libraries, others are not, and paying for cloud storage, for example, was cited by many as an annoyance, and is certain to be a barrier for others. But, given the sheer volume of content a researcher accumulates, storage is essential.

“In my academic career in speaking with different people, attending conferences and liaising, I very frequently get PDF texts like books, monographs, in email attachments. And I can’t just search for them in my email and delete only those things. And so, it became a, ‘Okay, do I spend a week clearing out all thirteen years of my email? Or do I just buy my sanity and just shell out?’ And so that’s what I did. I probably will have to do something similar to Dropbox also.”

Overall, it is clear that workflows require time and patience to figure out and hone over years, largely without support or guidance, even from institutions.

“[There is] very minimal support on the institution’s side, mostly because the assumption is that at this point in your career, you should have figured that out by now. That is the baseline is, in order to have gotten to this point you have to have had figured that out.”

And the end result is generally still a very manual, sometimes clunky process. And yet, the development of a personal workflow is considered in many ways a critical part of developing your skills as a researcher over time.

“It’s something that I’m developing that just continues to develop. I think this also might be interesting, how the methods for people who are still doing graduate studies or undergraduate studies with reading and writing are very different from somebody who’s finished their graduate education [and is] now [a] professor. Because you have five to ten years, to slowly develop that process and come to something, arrive at something that you’re comfortable with. And that’s a part and parcel of the academic training and the reading that you do.”

## Conclusions

The overall impression from our interviews with readers is that their methodologies of reading, writing, managing collections, and finding content are deeply personalized. While there are clear commonalities, there are no agreed-upon standards, best practices, or widely accepted models that a researcher could adopt or adapt.

As mentioned at the start of this report, readers are the driving force behind the scholarly ecosystem, so any software and tools must work for them if they are to have any value at all. These interviews made clear that a personal library or collections management software must meet certain requirements in order to be a meaningful addition to the array of tools available to them. These include the ability to:

- manage a scholar's entire corpus or collection of texts
- accommodate multiple formats and materials (e.g. books, images, videos, audio, journal articles, reviews, archival material, letters, reviews, personal notes, or documents)
- remain flexible enough to accommodate disparate workflows
- support the development of personalized workflows, without restricting or mandating behaviours

Ultimately, the value our tool could offer will lie in managing an entire collection of texts, not just a single reading experience, and in the capacity to accommodate user-driven workflows. Neither of these criteria are met by any software currently on the market, and there is a real opportunity for one to emerge and, perhaps for the first time, properly support readers in their academic pursuits.

# Online Survey

The needs of the scholarly readers are rarely prioritized over those of other stakeholders in existing digital tools. While our in-person interviews allowed us to focus on some readers' workflows, behaviours, and needs, we could not assume these to be representative of all researchers, or all readers, around the world. To better understand the varying needs of scholarly readers, we conducted a short online survey.

## **Methodology**

Our online survey was modeled after our interviews, with similar questions regarding researchers' approaches towards print and digital monographs, preferred citation and annotation tools, and workflows (please see Appendix iii for the full list of survey questions). We created a thirty-five-question survey using Survey Monkey. We distributed the survey in the Rebus Community and Rebus Foundation newsletters, as well in social media channels (Twitter and Facebook). The survey was open for roughly one month.

## **Summary of Results**

We received a total of 104 responses to the online survey. Respondents were mostly from North America, despite our efforts to promote the survey to scholars in other parts of the world. To our surprise, the survey was answered not only by scholars (including undergraduate and graduate students, tenured and non-tenured faculty, independent researchers) but also by librarians, publishers, editors, instructional designers, experts in higher education policy, and university staff such as deans. This range of responses has reaffirmed the notion that a viable business model must consider the needs of a variety of stakeholders.

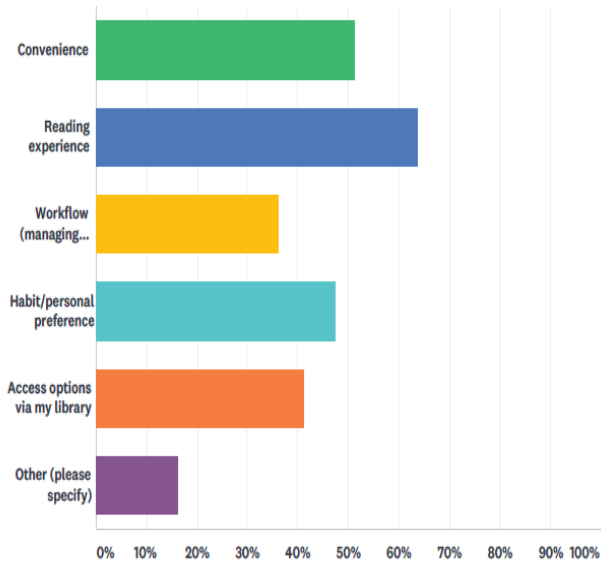
We've organized the results of the survey into four broad themes below.

## Print vs. Digital

Why do some researchers abhor digital and favor print, or vice-versa? The classic print vs. digital debate was necessary for us to understand readers' preferences with each format.

### Q11 What factors influence your choice of print? (select all that apply)

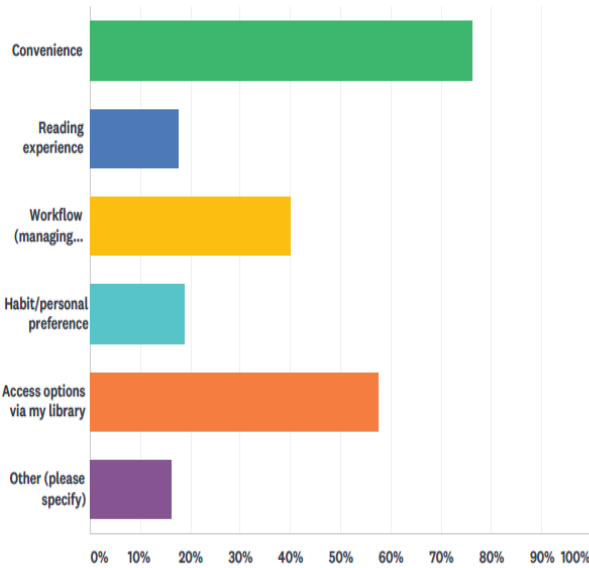
Answered: 80 Skipped: 24





## Q12 What factors influence your choice of digital? (select all that apply)

Answered: 80 Skipped: 24



Unsurprisingly, there was no clear winner, with 46.25% saying “it depends.” Their choice of format depends on the task at hand:

“For cover-to-cover reading I much prefer print; for annotating, searching, and referencing during writing, I much prefer the digital.”

When prompted about why print was the preferred format for “immersive” reading, a respondent said, “Reading online simply does not allow for a deep, reflective, thoughtful, and active reading experience.” According to some, print works allow readers to conceptualize the work in its entirety, in a way that is not possible online. Others noted that the kinesthetic experience of a print book contributed to their preference of this format:

“The embodied reading experience is, I think, the biggest factor. Reading something long on a device feels disembodied; it’s hard to remember where I was last; it’s hard to remember reading it at all, later on. Digital works great for searching and scanning, but for multiple serious paragraphs, it feels abstracted.”

As the respondent above noted, deep reading using digital formats is still a challenge. Many respondents admitted to skimming content in digital form, and printing sections out later to read more thoroughly.

While print was a choice for many for an initial thorough read, a number of respondents turned to digital formats during the writing stages of

their research projects. They praised the quick, easy searchability of digital formats, the ease of copying text into their working documents, as opposed to typing this in manually, and the portability of digital files.

Others noted that digital files, if prepared properly, were more accessible than print books:

“Since I am blind, I have scanned print books, making my own digital files. When monographs are available digitally, this saves me a labor-intensive step. However, it only helps if the files are designed with accessibility in mind. I often have to print out an image-only PDF and scan the print pages to make an accessible file.”

Overall, respondents concluded that print and digital formats are “complementary tools; one is better for retention while the other is better on grounds of searchability.” Along those lines, respondents made requests to publishers and librarians:

“We need to lobby bookstore chains for discounted ebook version[s] when purchasing hard copy. They should make a bundle available.”

### Managing Collections, or Organized Chaos

For researchers who are juggling a multitude of projects at a given time, keeping track of the books in their collection is crucial. One respondent shared their frustration with managing ebook collections:

“I now have hundreds of Kindle books. Amazon provides exactly ZERO useful tools for managing them. I fought with them a couple of weeks ago to ask them if there was a way I could just download a complete list of what I have, print it, and sort through discussing it with my wife, who shares the account, and then delete many obsolete ones. THERE IS NOT A WAY TO DOWNLOAD OR PRINT A COMPLETE LIST OF WHAT YOU’VE GOT.”

Some admitted to forgetting which ebooks they owned, resulting in the repurchase of these books. Overall, we found that most respondents used a number of organizational tools to manage their collections. Some kept track of print books in an online bibliography, which they continuously updated. Another respondent noted that they use a “combination of apps which have varying degrees of features and accessibility—Zotero, Onenote, Foxit, Docear, Qiqqa.”

This combination of tools was fairly common for managing digital PDFs, EPUBs, citations, primary sources, and other materials. Respondents shared the following tools: Google Drive, Dropbox, Zotero, Mendeley, Papership, Endnote, Onenote, Foxit, Docear, Qiqqa, as well as flash drives, and storage space on their personal computers.

These mechanisms are similar to those described by our interviewees, and appear to be a system that each reader has honed over a number of years. A new collections management system would need to not only accommodate the variety of materials, in different formats, that a scholar has gathered over the years, but also be simple to use and cus-

tomizable to fit the needs of the scholar. Looking toward this type of system, a respondent wished,

“It would be nice to have a software to collect materials from multiple databases in one place and to link the relevant materials together using my own organization structure.”

### Interoperability of Citation and Annotation Tools

As evident from the section on collections management, researchers use a variety of tools for citation and annotation. These include: Zotero, Google Scholar plugin, EndNote, RefWorks, Mendeley, BibTeX, Bookends, Docear, Qiqqa, Hypothesis, Kindle built-in tools, Adobe built-in tools, MS Word built-in tools, GoodReader, PDF Expert, Weave, Diigo, Papership, iAnnotate, and DocuWiki.

Respondents expressed their frustration at having to resort to multiple tools to work with different digital formats.

“No one has yet invented a native-to-digital format that facilitates scholarly reading, e.g., endnotes, cross-references, annotation to an external standard format. Drives me absolutely nuts and is the main reason I’m doing this survey, because you might be thinking the same way. The ebook formats we have are functional if you are reading a murder mystery — as functional as a papyrus scroll. They have not yet advanced to incorporate the advantages of the codex book.”

The lack of “interoperability” between the tools was brought up by other respondents. The “lack of integration” and “absence of systematic annotation” systems that aggregate easily with one another poses a challenge for readers, who simply want to be “centralizing everything and [accessing] it all in the same place.” One reader notes that there’s also a challenge in keeping up with newer versions of tools:

“Citation management tools keep changing, which makes it hard to keep training materials up to date.”

In addition, readers also noted the challenge of accessing files from different computers, or in different formats, where page numbers might not correctly align with their existing notes.

Respondents also noted the challenge of implementing a single standard or tool, as it could not fit all needs (ideal though this might sound):

“For instance, Hypothesis can be a great way to annotate all sorts of things (and, somehow, keep ad hoc collections through tagging) but the Open Annotation standard has yet to be widely adopted. In an ideal world, all the full-text content from my research materials, notes, annotations, Open Data repositories, and personal files would be part of the same database structure, with proper mechanisms for classification, linking, and reuse.”

## Restrictive Access: Lending Periods and DRM

The survey results revealed that access to monographs also greatly impacted readers' behaviours. Lending periods and DRM stood out as the key factors.

Respondents observed that lending periods varied greatly for print vs. digital books.

“For a print monograph a faculty checkout period at my library is one full semester. For a digital copy it's one to fourteen days depending on vendor.”

Readers find themselves forced to check out digital versions, as fewer print copies are held in libraries. But they are made to “return” the digital copy far sooner than they would be expected to return the print copy. These “artificial limits on borrowing periods” prove challenging for library patrons, who are simply trying to complete their reading.

“I don't like having to ‘check out’ an ebook repeatedly to finish a close reading.”

For researchers, being able to transfer content easily across devices, or sync a number of annotation tools, is critical. DRM can prevent this ease of transfer, and even prevent readers from copying and pasting quotes from books into their research papers/working documents. One respondent said,

“DRM is the No. 1 reason digital reading isn't my default. I want to be able to put the text on whichever devices I want and have a common set of annotation tools that sync annotations across devices. Kindle's environment is getting decent at this, but I'm still locked into Kindle's platform, and many/most scholarly monographs I have access to through my library are only available on platforms like Ebook Library.”

Readers wanted to remove DRM to bridge the gap between print and digital formats. A respondent said, “I would love it if print books were sold with instant, DRM-free digital access (or, conversely, if digital copies sold with very affordable access to print-on-demand).”

## Conclusions

This survey has provided additional insight into the needs of readers, and, significantly, other members of the publishing ecosystem. It has shown that researchers use a range of different formats in the course of their work, all of which need to be accounted for in the tools they use. In the case of citation and annotation tools, we must work towards integrating with existing software, as opposed to creating our own. The results reinforce the idea that digital tools cannot be developed with a view to replacing print formats, but rather, must be built to supplement the same.

As one respondent put it, open publishing ecosystems are the way of the future:

“Closed ecosystems are a deal-breaker for me, as they should be for all academics, since they are prima facie violations of academic freedom and the general project of innovative scholarship – which should not be enclosed in high walls!”

# The Rebus Approach

Having explored multiple external perspectives in the scholarly reading ecosystem, we turn now to sharing ours as well. We believe, based on our experience, values and the issues raised by our interviewees, that open, web-based technical solutions are essential for addressing the needs of scholarly deep readers. It is our view that only in an open technical ecosystem can deliver on the promise of digital reading technology, an ecosystem built for interoperability of reading platforms with external tools, a reading system that can be built upon and innovated upon. Reading is the starting point of scholarly activity, but it is only the starting point, and tools need to be developed to support all the other things scholars wish to do: from reading and annotating, through the management of collections, notes and reading contexts, to their reuse and recrafting into whole new bodies of work.

To date, the commercial platforms supporting digital reading are not answering this need. We believe there is space for a new, open model of digital scholarly reading, that is expressly built to encourage, rather than discourage, interoperability.

## **Open, Open, Open**

The history of human knowledge is written in, on, and with open, available, and accessible technologies: language, writing, paper, pens, typewriters, printing presses. No proprietary or closed technology has survived the tests of time to preserve and propagate the continuous progress of scholarly investigation. In the digital world, the movements of Open Web, Open Content Licenses, and Open Source Software tools continue with this imperative today.

The advent of digital information technologies has made possible entirely unprecedented deep engagement and manipulation of information and knowledge; and yet so far, for much of scholarly content, we have locked it up in closed technologies and business models.

The future of scholarly work, and the human progress that depends on it, lies in Open.

## Open Content

For scholarly endeavors to thrive and continue their contributions to human progress, source materials must be available not just for looking at and reading, but manipulating, extracting, annotating, quoting, copying, reworking, recontextualizing, and pushing further forward. This means doing away with DRM and enabling direct access, manipulation, and reuse of content. At its most radical, this imperative suggests Open Licences. But, we needn't be so radical to hugely improve the scholarly reading ecosystem. Simply letting scholars who have legitimate access to digital content (through library or private purchase) use and manipulate that content as they wish to support their scholarship is radical enough.

## Open Standards / Open Web Platform

Unfettered innovation in new ways of interacting with scholarly materials depends on interoperable, open standards-based approaches. The Web, whose very fabric and existence depends on such, is the natural model and platform.

Publishers must be involved in defining the Web Publication standard now developing in the W3C's Digital Publishing Working Group. (We have been involved in preliminary discussions with the group since mid-2017. See Appendix iv for more details of this process). Libraries should seek out and support web-based systems which incorporate such things as collection and book APIs and portability.

New interactions and models can best emerge when we agree on interoperable standards and protocols, which allow new innovations to build on existing technologies.

## Open Source Software

Along the same path, the software that supports the systems and services used in scholarly publishing need to be open to use, modification, repair, and innovation by anyone. Ideally, they should not be controlled nor limited by any one entity, so as to allow their development without the pressure of (only) commercial imperatives.

## Values, Not Only Business Models

Finally, all of the above are driven by economic factors: the flow of value through society. If the motivations are solely financial, then much non-accountable value is lost. Human and social values should be given at least equal weight, if not primacy. when considering the systems and technologies we have underpinning scholarly inquiry. The library is the perfect example of this model; technologies that support the work of libraries should share their values.

In our context, that means valuing readers, their research, and their teaching as more than just activities from which we can profit, instead recognizing and supporting the value that they contribute back to soci-

ety. Mission-driven businesses and organizations should continue to be given a chance to emerge, through funding—be it venture or grant based—and support.

We at the Rebus Foundation believe in and pursue open standards in web publishing, open source software and collaboratively-sourced open content, and we are grateful for the Andrew W. Mellon Foundation's planning grant allowing us to undertake this initial research project. Gathering such insight into the situation and needs of the scholarly publishing ecosystem is invaluable for our mission: to build open web-based tools for scholarly publishing and reading.



# The Way Forward

## **A Sustainable Business Model**

A central goal of the research done under this grant was to develop the outline for a feasible business model for the technology to be developed, and begin getting feedback from potential partners/stakeholders.

While experimental digital projects are important, the long-term sustainability — and ultimate value — of any technology developed is only possible with a sustainable revenue and business model. A sustainable financial model allows the technology to continue to grow and improve, and enables proper support (customer support, marketing, and sales) of the software.

We approached the development of a feasible business model with two central ideas in mind:

- The focus of software development should always be the needs of the scholarly reader.
- We should try to develop a business model that works in partnership with existing players in the academic ecosystem, rather than in competition.

As described above, the existing players in the ecosystem are:

- university presses, who produce and publish monographs
- aggregators, who serve as commercial and delivery intermediaries between publishers and libraries
- academic libraries, who are the buyers of content and offer support to readers and the research community
- scholarly readers, who are the “end-customer” of both the libraries and publishers, and the creators of content

While we did not interview aggregators officially as part of this report, we did have several conversations with some aggregators (namely Project Muse, and ACLS Humanities E-Book), and it was clear that their technology focus is on discoverability. The reading platform is either

not the focus (in the case of non-profit aggregators), or is part of an explicitly closed-silo system (in the case of commercial aggregators, such as Proquest, Ebsco, and Overdrive). In the case where there is a reading system associated with an aggregator, that system comes for “free” as part of an acquisitions and delivery platform.

Further, the central insights gleaned from interviews with libraries and publishers are:

- Libraries are indeed the ultimate buyers of digital monograph content.
- Libraries are generally unsatisfied with the digital platforms for reading made available by aggregators.

This opens up the possibility to develop a business model, whereby a hosted version of open source Rebus Reader software can be provided to libraries, and purchased through the “tools”/IT infrastructure budget, rather than the “content acquisitions” budget, which is allocated to aggregators, who usually provide their reading platforms for “free”.

This allows the reading system to be decoupled from the acquisitions process and platform, and allows Rebus to offer an added value service, which can work in partnership with aggregators, particularly those non-profit aggregators (such as ACLS Humanities e-Book, Project MUSE and JSTOR), whose missions align more closely with The Rebus Foundation.

In general this model has been received positively by the libraries we have talked to, and with openness from aggregators we have interacted with (ACLS Humanities e-Book, and Project Muse).

While the details of the business model — marketing, pricing, service, support — need to be fleshed out, there is enough indication from the critical players in the ecosystem, especially the potential buyers, that this model could well be successful, and allow us to truly begin to build digital tools for the scholarly reader, with an expectation of a sustainable future.

## **From Prototype to Product**

The other central goal of this project was to gain a better understanding of the needs of scholarly readers, and to use that understanding to shape the future development of the Rebus Reader/Personal Library Software. The prototype of this tool helped to illustrate what it might look like to stakeholders, and generated significant (positive) feedback. We are now equipped with more insight into our potential users and stakeholders, we can begin moving from prototype to a fully functional and fleshed-out product.

The core purpose of our proposed tool is to improve the reading and research experience for scholarly readers, leveraging the combined power of digital and the open web.

The technology will be based on a set of general technical principles:

- A digital reading platform should make many things possible

that are not possible with print.

- An open, web-based digital reading platform should make many things possible that are not possible with “closed” digital reading platforms (e.g. Kindle, Overdrive, Ebsco, Proquest).
- An open, web-based digital reading platform should allow third parties to build services, tools, and innovations upon the basic framework of the underlying system.
- Integrations with web-based digital scholarship projects (such as Fulcrum and Manifold Scholarship) will improve both the reading platform, and those projects.
- Content and user data should not be locked into the platform.
- Data generated by readers and users should belong to them; users should control how data about their usage is used.
- Readers should be allowed to do what they want with the texts they are reading (as long as they don’t break laws): search, copy, modify, annotate, etc.
- Scholarly reading is (usually) part of a broader research activity, and any tools must fit within this reality.
- While features must be developed to improve the reading experience for readers, libraries, university presses, and aggregators must be key partners in developing technology that answers the needs of whole ecosystem.

### Insights on Product Development

The interviews conducted as part of this project, in particular with readers, have been indispensable in shaping our approach to development. The following are the four most significant takeaways:

#### **Supporting scholarly reading means supporting multiple activities and workflows.**

If we are to build a tool that effectively supports the reading of academics and students, it will have to truly support the multitude of activities associated with that reading. It is not simply a question of consuming a single text, but interacting with that text, connecting it with others, and using it to create potentially multiple written works. What’s more, each scholarly reader has their own approach to these activities, which must be accommodated.

#### **Value lies in managing a reader’s entire collection of texts.**

Related to the previous point, it is clear that in order to be a truly valuable contribution to the current array of tools, ours must be able to cover a reader’s full collection of texts, not just some. This means offering the ability to create library entries not just for books available through a

library, or just books available in web or EPUB. While some features may only be available for a subset texts, depending on format and availability, others must still be represented in the collection. Otherwise, this risks becoming one more tool in a chain, further complicating, rather than simplifying, users' workflows. This also means the tool must also allow for the inclusion of all kinds of materials (from full books, to book chapters, journal articles, primary sources, images, audio, and perhaps others), and it should remain agnostic about where these materials are sourced from.

### **Print and digital coexist.**

Print's continuing popularity seems to come down to two drivers: habit and poor support of digital reading. While the latter of these is clearly something we seek to mitigate, that is far from an easy task, and the former is likely to remain entrenched for many years to come. As a result, the goal cannot be to try to replace print reading, but should instead be to complement it intelligently, and ensure that wherever possible, hybrid print/digital scholarly reading should be supported.

### **Access is critical.**

This last point is perhaps less clear-cut than the others here, and emerged in a less explicit way from our interviews. The readers we interviewed represented a wide range of positions, even within a small group, and some clear differences in needs stood out. Many of the group touched on various issues relating to access, as did librarians, and our own philosophy means we take access and inclusive design seriously. As a result, to our minds, the development of our tool must take into consideration who will access it and how. This cannot be limited to permanent faculty with access to institutional resources. Students and contingent faculty tend to have less secure access to institutional resources, are more likely to change institutions (or leave them, temporarily or permanently), and often have very different working conditions from permanent faculty, yet make up a significant proportion of the users we envision benefitting from our tools. These user profiles will remain top of mind throughout development, and must be served just as well as the model of a "traditional" researcher.

These insights will continue to guide us as we move forward with the product development process, and influence the measures of success we establish along the way.

# Conclusion

Reflecting on what we have achieved in the past year, the Rebus Foundation team are confident we stand in good stead to move forward with the next phase of this project. Our discussions with publishers suggest a strong interest in, and incentive for, new models for digital reading. We've also confirmed that (non-profit) aggregators are interested in partnering with us to deliver better reading experiences to readers, who are the ultimate end-users of their platforms. Libraries are key potential partners in developing a new digital reading ecosystem for their patrons and are well placed to drive discussions on readers' needs, and responded positively both to our prototype and proposed business models. Libraries also have, collectively, the potential financial leverage in the ecosystem to effect change. And finally, we know that readers, the end users of any software we develop, represent the part of the ecosystem with the most to gain from a new approach to digital scholarly reading. Focusing on development of the highest priority features for readers will be critical, and we have a clear vision for what those features should be.

Over the course of our research, we have had some hunches confirmed, some concerns allayed, and gained some significant insights into reader behaviour that we could not have without going through this process. We are deeply indebted to all those who took the time to share their thoughts and ideas with us over the course of the project and are excited to continue working with them. As this report is part of what we hope to be ongoing work, it is difficult to find a single, summarizing note to end on, except to say that there is clearly enormous potential for innovation in the support of digital scholarly reading. We at the Rebus Foundation believe that, through an open, collaborative approach to the problem at hand, we can help lead the way in shaping that future.



# Appendices

## **i. The Rebus Personal Library Prototype**

### About

The Rebus Personal Library Prototype is an exploration of a web-based personal reading and annotation management platform. The goal of the prototyping process is to figure out and help us imagine what meaningful and valuable aspects of academic deep reading need to be supported and made available to academics in their research and writing work.

From a user experience perspective, the prototype takes a very specific line of inquiry: “Can we build rudimentary, web-based digital experiences which return to the user some of the tangible, spatial, temporal, and social experiences of having books in one’s physical environment?” This means the focus was not on the details of reading and annotating books, but the management, collecting, and “living with” of books. Future additions would be some of the more specific reading management needs of academic deep readers, such as grouping by reading intent, prioritization, and scheduling, etc.

In terms of technology, we are driven by a strong belief in the Open Web—with its pillars of open source software and openly licensed content—as the driver of innovation and democratic access to knowledge and tools. To this end, the prototype is built using open web technologies and open source software, and the monographs used for demonstration purposes are openly licensed (Creative Commons). The codebase for the prototype itself is provided as is on GitHub.

### Core Elements

#### **Library Overview**

The Library Overview gives the user an immediate visual sense of the scale of their collections, as well as hinting at the heft of the books it contains. Book cover thumbnails are sized according to their page counts, giving the user an idea of how big or thick a book is. Grouping and sort-

The screenshot shows a web interface for "My Rebus Library — Prototype B." At the top, there are navigation options: "Stacks", "By Read State", and "By Date Added". Below this, the date "March, 2017" is displayed, along with a count of "16 books". The books are arranged in a grid. The first row includes titles like "THE INDIGENOUS STATE" by Nancy Postero, "IMPERIAL MATTER" by Lori Khatchadourian, "EQUITY, GROWTH, and COMMUNITY" by Chris Benner & Manuel Pastor, "LUXURY and RUBBLE" by Erik Harms, "FINDING JERUSALEM" by Katharina Galor, "KEYS TO PLAY" by Roger Mosley, "HINDU PLURALISM" by Elaine M. Fisher, and "HOKUM!" by Rob King. The second row includes "The Multi-Agent Transport Simulation Method" by Andreas Horni, Kai Nagel & Kay W., "ISSUES IN OPEN RESEARCH DATA" by Samuel A. Moore, "LEARNING FROM WEST AFRICAN FAMILIAL CHAIRMAN" by Alan Brylan, "WRITING AS MATERIAL" by Kathryn E. Piquette & Ruth, "GOVERNANCE DU SECTEUR DE PALETTE CHAIRMAN" by Alan Brylan & Louise Chauveau, "RECONSIDERING CULTURAL" by Akiru Matsuda & Luisa Elena, "HOGSKOLANS ANSVAR" by Andrew Casson, and "EUROPEAN HANDBOOK OF CRISTINA CAJINERI".

Below the March section, the date "February, 2017" is displayed, along with a count of "18 books". The books in this section include "Imperial Genus", "INSTRUMENTS FOR NEW MUSIC", "MAKING THINGS", "Scale", "Polemics and", "MIGRATING INTO FINANCIAL MARKETS", "Precarious Claims", and "PRECARIOUS CREATIVITY".

ing can be done according to aspects such as Date Added, “Read State,” and “Stack.” More dimensions can be added.

View the library overview.

## Stacks

Stacks are user-definable groupings of books. They are meant to represent the stacks of books related to specific research projects or any other activity or context the user may wish to define. Stacks are non-exclusive, meaning a book can be “in” multiple stacks at once.

A Stack of books is presented with its relevant metadata, as well as all of a reader’s behavioral traces, e.g.: annotations, notes, etc. This provides “one place” for all context-related (research project, etc.) reading data.

View the “Open Science” stack.

## Book and Reading Details

Individual book detail views include relevant metadata, table of contents as well as user annotations and notes. Also displayed are the Stacks the book is part of, as well as a hint of similar books—based on keyword intersections in this case. Easy citation functionality is an example of app-like developments that can be added.

View the book and reading details.



“Open Science” Stack

4 books



Read: HTML — EPUB

Issues in Open Research Data

Samuel A. Moore

DOI: 10.5334/ban

172 pages

My Annotations

Under the Directive, rights are granted to the one who makes a substantial investment in obtaining, verifying or presenting the contents of the database.

12:51 pm, Friday, December 2nd, 2016

*science subsumed to capitalism and its next evolution, vectorialism.*

12:51 pm, Friday, December 2nd, 2016

The United States, for example, does not impose any sui generis rights. Copyright ownership belongs to the creator or his employer, but may be transferred to another (such as a publisher) hence copyright ownership can be difficult to ascertain, particularly where multiple researchers have contributed to the whole.

12:54 pm, Friday, December 2nd, 2016

*ditto, more often than not, "employer" ("who is paying for the labor") and publisher ("who provides distribution infrastructure") will demand rights, this becomes increasingly true when they can no longer capture revenue from the selling of atoms and must lock down the bits.*

12:55 pm, Friday, December 2nd, 2016

A focus solely on the symptoms of dysfunction in research, rather than the underlying causes, will fail to deliver meaningful positive change.

1:41 am, Friday, December 2nd, 2016

*true of all things*

1:45 am, Friday, December 2nd, 2016

Open Context, the data sharing system I direct, has adopted a model of “data sharing as publication.” Open Context publishes a wide variety of archaeological data, ranging from archaeological survey datasets to excavation documentation, artifact analyses, chemical analyses of artifacts, and detailed descriptions of bones and other biological remains found in archaeological contexts.

9:53 am, Friday, December 2nd, 2016

*interesting, in a way it is "secondary" or "side" publishing: not the glamour and prestige of the full report, but the important, necessary and value of the data, both of findings themselves and the setting/context*

5:57 am, Friday, December 2nd, 2016



Read: HTML — EPUB

Turn on the Light on Science

Antonio Tintori, Rossella Palomba

DOI: 10.5334/bba

120 pages

My Annotations

Stereotypes help us in the complex task of simplifying our world by sorting everyone and everything into tidy categories. It is an abstract but clear and simple mental process. In his work on the theory of social identity, Tajfel highlighted the close relationship between simplification and distortion of reality.

11:02 pm, Thursday, December 1st, 2016

Stereotypes are basically generalizations concerning social groups, aimed at binding the cognitive process to the cultural context. In order to do that, stereotypes emphasize and overestimate the characteristics of a social group that make it different from the others

11:02 pm, Thursday, December 1st, 2016

A stereotype is a rigid perspective on the world. It is based on bias. It represents beliefs that are not necessarily negative but certainly irrational, and may result in very negative attitudes and behaviours, as for example in the case of racism and xenophobia.

# 52 An Open Approach to Scholarly Reading and Knowledge Management



Read: HTML — Epub

## Open

The Philosophy and Practices That Are Revolutionizing Education and Science

Rajiv S. Jhangiani, Robert Biswas-Diener

March, 2017  
DOI: 10.5195/bbc  
ISBN: 97809129006  
Goodreads  
Education, Open education, Open access, Science, Open science, Social science, Pedagogy, Psychology, Open textbooks, Open pedagogy, Open educational resources.

704 pages  
reading  
Cite this book

### Stacks

#### Technology & Politics



#### Open Science



#### Similar books



### Table of Contents

- Cover
- Title
- Copyright
- Table of Contents
- Acknowledgments
- Contributors
- Introduction
  - Introduction to Open (Robert Biswas-Diener and Rajiv S. Jhangiani)
  - A Brief History of Open Educational Resources (T. J. Bliss and M. Smith)
  - Open Licensing and Open Education Licensing Policy (Caleb Green)
  - Openness and the Transformation of Education and Schooling (William C. Hitt and David M. Monette)
  - What Can OER Do for Me? Evaluating the Claims for OER (Martin Welser, Beatriz de los Angeles, Rob Farrow, Rebecca Pitt and Franck McAndrew)
  - Are OER Resources High Quality? (Regan A. R. Gurung)
- Open Practices
  - Opening Science (Brian A. Nosek)
  - Open Course Development at the OERs (Wayne Mackintosh)
  - From OER to Open Pedagogy: Harnessing the Power of Open (Robin DeRosa and Scott Robinson)
  - Opening Up Higher Education with Screenscasts (David B. Miller and Addison Zhao)
  - Librarians in the Pursuit of Open Practices (Quill West)
  - A Library Viewpoint: Exploring Open Educational Practices (Anita Walsh)
- Case Studies
  - The International Journal of Wellbeing: An Open Access Success Story (Dan Weijers and Aaron Jardine)
  - Heading Toward Openness: Lessons Learned on a Personal Journey (David Wiley)
  - Open-Source for Educational Materials Making Textbooks Cheaper and Better (El Diener, Carol Diener and Robert Biswas-Diener)
  - Free is Not Enough (Richard Barzantini, Nicole Finckelstein, David Harris, Dan Nicholson and Daniel Williamson)
  - The BC Open Textbook Project (Mary Burge)
  - TeachPsychScience.org: Sharing to Improve the Teaching of Research Methods (David B. Stoshinetz, Natalie J. Ciurro and Gary W. Lewandowski, Jr.)
  - DIY Open Pedagogy: Freely Sharing Teaching Resources in Psychology (Jenica Hartnett)
- Conclusions
  - You Can't Sell Free, and Other OER Problems (Robert Biswas-Diener)
  - Open as Default: The Future of Education and Scholarship (Rajiv S. Jhangiani)

### Write-up

Affordable education. Transparent science. Accessible scholarship.

These ideals are slowly becoming a reality thanks to the open education, open science, and open access movements. Raining separate—if parallel—courses, they all share a philosophy of equity, progress, and justice. This book shares the stories, motives, insights, and practical tips from global leaders in the open movement.

### My Annotations

have enabled the marginal cost of education, such as the sharing of resources, to approach zero.

11:29 pm, Thursday, December 07, 2016

*the language of capitalism is deeply pervasive...*

11:29 pm, Thursday, December 07, 2016

not only job skills-related education but also a liberal arts education

11:30 pm, Thursday, December 07, 2016

*what a strange choice of examples.*

11:31 pm, Thursday, December 07, 2016

The real tragedy is the loss of every scientific discovery, artistic work, invention, new business, and cultivated mind because of lack of opportunity according to random chance at birth.

11:31 pm, Thursday, December 07, 2016

This means that the entire blue-collar labor force will continue to be functionally illiterate for another generation?

11:31 pm, Thursday, December 07, 2016

*along with property ownership, this is the bedrock of exploitation capitalism*

11:31 pm, Thursday, December 07, 2016



This notion that education can unlock a person's potential is relatively non-controversial. More controversial is the related notion that education should be made available to all. Indeed, despite the fact that it is specifically mentioned in Article 26 of the United Nations Universal Declaration of Human rights,<sup>2</sup> this 'education as a fundamental human right' idea is a relatively recent phenomenon. Historically, learning was an activity reserved for privileged citizens: the priestly class, the wealthy, men. In *The Theory of the Leisure Class*, Veblen (1899/2009) argues that the 'learned class' has therefore long been associated with the aspects of higher education that have no immediate industrial use, such as the study of the classics or music. However, over the last couple of decades advances in technology **have enabled** the marginal cost of education, such as the sharing of resources, to approach zero.<sup>3</sup> This means that barriers to education are now being removed for a broader swath of humanity than at any time in history. What is more, that all people should have equal access to not only job skills-related education but also a liberal arts education is a case that can be made more strongly on a human rights platform than on an economic empowerment platform. The real tragedy of poverty is not that the poor need more opportunities to be factory foremen, office managers, or stock traders. The real tragedy is the loss of every scientific discovery, artistic work, invention, new business, and cultivated mind because of lack of opportunity according to random chance at birth.

Unfortunately, educational inequality abounds. Braun (2014) states that as many as 57 million school-aged children worldwide do not attend primary school. In 1999 the ambitious *Voices of the Poor* project was launched in order to conduct field interviews with tens of thousands of poor people from around the globe.<sup>4</sup> With regard to education, poor people identified the strain of costly and distant schools, lack of quality education, feelings of low self-worth, and competition with child labor as widespread educational problems. In his study of poverty in Kolkata

## Reading and Annotation

Texts and annotations can be consulted, searched, and read online. Annotation itself would be possible with more development work.

View annotations in a book.

## Screen Saver

When the system is not in use, or set to a "display" mode, this screen-saver displays a user's own or popular highlights from books in one's collection. A new highlight fades into view every five minutes. As a glanceable space, this can be used to surface opportunities for serendipitous discoveries in books one has collected but perhaps not read yet.

View the screen saver.

## Feedback and Future Directions

In order to solicit feedback on the prototype and input on the future direction of development, we shared this prototype with our interview participants. Beyond specific features, we sought more broadly an understanding of what function the product must serve for it to be useful to readers, but also to complement the work of publishers and aggregators, rather than conflict with them.

For more on the future direction of the prototype, see From Prototype to Product.



## ii. Interviewees

We would like to thank the following interviewees for their time and insights:

- Heather Coates, Data Librarian, IUPUI University Library
- Anna Creech, Head, Resource Acquisition and Delivery, Boatwright Memorial Library, University of Richmond
- Will Cross, Director, Copyright & Digital Scholarship Center, NCSU Libraries
- April Hathcock, Scholarly Communications Librarian, NYU Libraries
- Stacy Konkiel, Director of Research & Education, Altmetric
- Margaret Smith, Physical Sciences Librarian, NYU Libraries
- Megan Wacha, Scholarly Communications Librarian, CUNY Library
- Amy Brand, Director, MIT Press
- Mark Eddington, Director, Amherst College Press
- Bridget Flannery-McCoy, Editor, Columbia University Press
- Michael Jon Jensen, Director, Technology, Westchester Publishing Services

- Dennis Lloyd, Director, University of Wisconsin Press
- Tony Sanfilippo, Director, Ohio State University Press
- Erich van Rijn, Director of Publishing Operations, University of California Press
- Charles Watkinson, Associate University Librarian for Publishing, Director, University of Michigan Press
- Felix Fuchs, PhD Candidate, Department of English, McGill University
- Mona Gupta, Assistant Professor, Department of Psychiatry, Université de Montréal
- Maya Hey, 2<sup>nd</sup> year PhD student, Department of Communications, Concordia University
- Pasha Mohammed Khan, Chair in Urdu Language and Culture, Assistant Professor, Institute of Islamic Studies, McGill University
- Katie Kirakosian, Adjunct Lecturer, Department of Anthropology, University of Massachusetts, Amherst
- Zain Rashid Mian, MA student, Department of English, McGill University
- David Wright, Professor and Canada Research Chair, Department Chair, Department of History, McGill University

### **iii. Online Survey Questionnaire**

We would like to thank all those who participated in our online survey. View the questionnaire.

### **iv. Summary of Rebus's W3C strategy**

#### Context

A key requirement for the wide adoption and use of digital publications (both in education and trade) is device compatibility. Both EPUB and PDFs have serious limitations here.

- Reading an EPUB requires bespoke applications which may or may not be pre-installed on your device—all of which suffer from significant compatibility issues.
- PDFs are fixed in size and format. Choosing one size for your PDF excludes those using incompatible device sizes.

The platform that offers both significant distribution and pre-established — tested — tactics for cross-platform compatibility is the web. But essential to the use of the web for publications is making sure that it, as a platform, supports the needs of those making publications. To that

end we have been participation in the W3C standardization process as they are trying to put together a set of specifications designed to ease the transition from EPUB and PDF to web publications for publishers.

The W3C is only one of many standards organizations responsible for the web. Others include: WHATWG (markup and APIs), IETF (networking), IANA (naming), and ECMA (JavaScript). It is currently the only one of these organizations that has explicitly reached out to the publishing community but is by no means the most important.

Their initial effort is focused on a manifest format (Web Publication Manifest) that adds support for the document structures and metadata that publishers consider essential for their participation in the open web stack.

## Our Goals

The fundamental goal of our participation in the W3C is to increase the likelihood of the W3C ratifying Web Publication specifications that both benefit open education and publishing and increase the odds of those specifications being implemented.

The latter is more important than the former. It is more important that we have free and open publication-oriented technologies implemented as a part of the open web stack than to have those technologies standardized specifically by the W3C.

So far, the W3C is the best bet to accomplish these goals but we plan on regularly pausing to investigate if other options would be more effective. Possible alternatives in the long run would be direct participation in the WHATWG (Web Hypertext Application Technology Working Group) who now have de facto control over HTML and many related specifications or a greater focus on open source projects.

## Progress so Far

Most of our work so far has been focused on providing a counterweight to traditional publishing interests in the W3C Publishing Working Group. This is not out of some sort of adversity towards the industry but to increase the odds of the Working Group's Web Publication specifications being broadly implemented. Many of the strategies and needs that drove the development of EPUB directly hindered the adoption of that specification adoption outside of the very limited sphere of a small portion of the market for ebook reading systems.

We are by no means alone in our efforts here. Many of the participants are very keen on putting together a format that is more widely adopted than just in the current publishing industry. This was especially apparent when we contributed to the metadata sections of the proposed specification as our suggestions were met with much broader approval than we had dared to hope.

Currently, the Working Group is evaluating several proposals for a publication manifest format, and there isn't much consensus on the pros and cons of each.

## Next Steps

The next step for the Working Group is to conduct more experiments with the proposed specifications to see if they can be made to fulfil requirements. To answer many of the questions still lingering about the specifications, several assumptions need to be put to test and that requires some of them to be actually implemented.

Most of these experiments can and should be lightweight but the more extensively these formats are tested in this early phase of their life when they can still be changed, the more robust and—hopefully—popular they will be later on.

## **v. Acknowledgements**

The researchers would like to thank the support of the Andrew W. Mellon Foundation, and in particular Donald Waters and Molly McGrane-Cleary.

## **vi. About the Rebus Foundation**

The Rebus Foundation is a Canadian not-for-profit, established in April 2016, with a mandate to foster the development of the infrastructure and ecosystem for open webbooks. The Rebus Foundation received grants from the Hewlett Foundation in August 2016, to promote the creation and consumption of Open Textbooks; and from the Andrew W. Mellon Foundation in Jan 2017 to explore the digital scholarly reading ecosystem. More information can be found at: <https://rebus.foundation>