

## In this issue

From Keith and Denise

Text & Data Mining

The Future of the Book

Bookmetrix

Beyond the E-Book

Open Access (OA)  
Monographs

CMU Support for OA  
Monographs

Double Dipping

The Discipline of  
Organizing

Actions for Researchers

Actions for Librarians

## We want to hear from you

Send us your questions, suggestions for topics, and information to include in the SC Digest.

Contact  
Denise Troll Covey

## From Keith and Denise

This issue of the SC Digest introduces content provision for text and data mining, explores the future of the book, including support for publishing monographs open access, warns researchers about double-dipping publishing practices, and describes two important international initiatives endorsed by Carnegie Mellon University.

## Text & Data Mining

Text and data mining (TDM) are increasingly popular ways to conduct research. What is TDM? Using automated tools to process large volumes of digital text or data and extract key bits. Why? To identify relevant information and discover previously unknown patterns or connections. You can assemble the extracted bits to reveal new facts or formulate hypotheses to explore using conventional methods.

What content can you mine? US copyright law makes this tricky. If the content is copyright protected, some people believe TDM is a [fair use](#) – no permission required – because it's a transformative use, a use different from the purpose for which the author created the work. Other people believe TDM rights must be licensed from the copyright owner. The upshot is a mix. Some content is freely available for TDM. Other content requires a license to TDM.

The University Libraries is pursuing TDM rights for content we license if the content provider requires a license. Providers have different models for enabling TDM. Some maintain the content themselves and provide researcher access through an API or simple FTP. Others provide the library with a copy to maintain locally. In all cases, access is restricted to Carnegie Mellon users. A TDM license might or might not incur a one-time or recurring fee. While we explore what's available, **if there is particular content you want to mine**, send email to Denise Troll Covey, [troll@andrew.cmu.edu](mailto:troll@andrew.cmu.edu). We'll make exploring TDM rights to the content you want to mine a high priority.

In the meantime, we've created a [Text & Data Mining](#) web page of content you can TDM. Links are provided to relevant policies and instructions. For a short news item about TDM, see [Text Mining – Partnering to Open New Doors to Big Data Research](#). To help remove barriers to TDM, consider signing the [Hague Declaration](#). We did.

## The Future of the Book

What is happening with the academic book? Digital formats, devices, distractions and deluge are challenging print and changing how and what we read. The [Academic Book of the Future](#) project in the UK is studying the changing state and contexts of the book and exploring how future scholarship in the arts and humanities will be produced, read, shared and preserved.

In a recent post on the LIBLICENSE listserv, Arizona State University Librarian Jim O'Donnell observed: "the way we read, what we retain, and what we can do with what we've read will differ widely based on format and reading practice." He thinks we're not paying enough attention to what's happening or making wise choices. As media, reading practices and business models change, the books we produce and consume will change. The nature of the things written and published as "scholarly books" will change.

O'Donnell poses two questions: What are we doing when we read and write scholarly work? And what are the appropriate or supportive forms of production and consumption for those activities? He wonders if we're ready for what lies ahead, citing the new models of production underway at the University of Michigan and Stanford University as important and non-trivial. See [Beyond the E-Book](#).

## Bookmetrix

Until recently, the development of alternative metrics focused on journal article-level metrics, such as the number of citations, downloads and social media mentions. Now [Bookmetrix](#) provides alternative metrics for books.

A collaborative development of [Altmetric](#) and Springer, Bookmetrix provides book- and chapter-level metrics on reach, use and readership, including citations, downloads, book reviews, readers, and mentions in policy documents, news outlets, Wikipedia entries and social media. The metrics are displayed on book and chapter pages in [SpringerLink](#).

## Beyond the E-Book

In [The Medium is the Massage](#), Marshall McLuhan explained that the initial content of a new medium is the old medium. Over time, however, possibilities inherent in the new medium are realized, yielding entirely new kinds of work.

Today, most e-books and electronic theses and dissertations (ETDs) are essentially print books in digital form. What could these works be if the potential of the digital were realized? Let's see.

The University of Michigan Press recently received a \$899K grant from the Andrew W. Mellon Foundation to develop a [publishing platform](#) that will connect publications to datasets, interactive information, video and other digital content. Maximizing the publishing strengths of university presses and the preservation expertise of libraries, the project will create an open-source solution for born-digital complementary monograph materials.

Stanford University Press recently received a \$1.2M grant from the Andrew W. Mellon Foundation to address a problem that is stifling innovation and the recognition of innovators. To date, scholars investing in new ways to visualize and analyze data are hosting their work on personal websites or public platforms. Because these venues lack rigorous peer review processes, these scholars cannot achieve with these works the publishing pedigree that traditional book publishing provides for consideration in hiring and tenure decisions. The Stanford project will establish a [publishing methodology and distribution channel](#) for interactive scholarly works that will be held in the same high regard as traditional monograph publishing. Stanford University Press will seek interactive scholarly works, conduct peer and technical reviews, and publish and market high quality titles. Stanford Libraries will perform web and data archiving to preserve the content. The project will also develop a cost-basis for publishing digital objects, and document and disseminate best practices.

## Open Access (OA) Monographs

For roughly a decade, efforts to provide free online (open) access to scholarly work focused on journal articles. Disciplines that require a longer work to share findings and reward scholars were seemingly kept out of the fray over escalating costs and access restrictions. The landscape is changing.

Monograph publishers are adopting open access options under a variety of business models. Explore the following:

- [Cambridge University Press](#)
- [Luminos](#)
- [Open Book Publishers](#)
- [Open Humanities Press](#)
- [Open Library of Humanities](#)
- [Ubiquity Press](#)

See also the directory [Publishers of OA Books](#) and the article [Innovative approaches to publishing open access monographs](#). In addition, some traditional monograph publishers now allow you to provide open access to a chapter in your book. Ask your publisher about open access options, but beware of “double dipping.” See [Double Dipping](#) in the adjacent column.

## CMU Support for OA Monographs

The [Open Library of Humanities](#) (OLH) publishes open-access journals and monographs in the humanities. It follows an academically rigorous approach and does not levy author-side fees, providing humanities scholars with an appropriate and open outlet for their work. Initially funded by the Andrew W. Mellon Foundation, ongoing support for the OLH is provided by an international library consortium called the Library Partnership Subsidy (LPS). CMU Libraries recently joined the LPS to support open access publishing in the humanities.

CMU Libraries also supports open access publishing of humanities and social sciences monographs through [Knowledge Unlatched](#).

## Double Dipping

When traditional journal publishers cautiously adopted open access, they created the “hybrid” journal: a subscription journal that makes selected articles available open access upon payment of an Article Processing Charge (APC). When a publisher charges both a subscription fee and an APC for the same content without financially compensating subscribers who paid to access the freely available content, critics call it “double dipping.” Publisher policies against double dipping are often disingenuous.

Unfortunately, as monograph publishers are cautiously adopting open access, they too in some cases appear to be double dipping. According to a recent [blog post](#) by Martin Paul Eve, University of London, if book publishers are basing the author-side publication fee on the financial return they want from publishing the work, without factoring in the projected revenue from print sales, they are double dipping.

We encourage CMU authors to avoid the double dip. Two scoops may be good on an ice cream cone, but they’re not good on a budget.

## The Discipline of Organizing

**Robert J. Glushko**

**June 16, 2015**

**4-5:30 pm**

**Gates Hillman 6115**

**RSVP to [gbalbier@andrew.cmu.edu](mailto:gbalbier@andrew.cmu.edu)**

Library and Information Science and Computer Science view and teach information organization (IO) and information retrieval (IR) in significantly different ways. Want to learn how to bridge these differences and integrate IO and IR? Attend the BrainHub lecture by [Robert J. Glushko](#), University of California, Berkeley.

Glushko’s book [The Discipline of Organizing](#) was named Information Science Book of the Year in 2014. It provides a framework for the theory and practice of organizing and integrates IO and IR through the unifying concept of an Organizing System, an intentionally arranged collection of resources and the interactions they support.

## Elsevier Action and Backlash

On April 30, 2015, Elsevier announced changes in its policy on author rights, proclaiming it was “[unleashing the power of academic sharing](#).” The changes

- Impose an embargo of one to four years on the deposit of author manuscripts in an open-access institutional repository such as [Research Showcase @ CMU](#), regardless of whether the institution mandates deposit in the repository. (From 2012 until this recent change, Elsevier imposed an embargo on institutional repository deposits only if the institution mandated deposits.)
- Mandate that deposited manuscripts be licensed under the most restrictive Creative Commons license, [CC-BY-NC-ND](#). (Attribution is required; commercial use and creation of derivative works are prohibited.)
- Apply to “all articles previously published and those published in the future.” May 26, 2015, Elsevier clarified that no retrospective action is necessary on works already deposited, but any previously published articles deposited in the future must be embargoed.

In [Stepping Back from Sharing](#), Kevin Smith (Duke University) calls Elsevier’s announcement “a masterpiece of doublespeak.” The lengthy embargoes and restrictive license unnecessarily limit the usefulness of the articles. Authors must now wade through a [50-page list](#) to discover the embargo period that applies to their work. This is “a retreat from open access,” “a retreat from sharing and an effort to hamstring the movement toward more open scholarship.” Smith suggests it’s time to consider another [boycott of Elsevier](#).

On May 20, a global coalition of organizations released a [statement](#) denouncing the new policy and urging Elsevier to revise it. According to an analysis by the [Scholarly Publishing and Academic Resources Coalition](#) (SPARC) and the [Confederation of Open Access Repositories](#) (COAR), Elsevier’s new policy presents a significant obstacle to the dissemination and use of research results, and creates unnecessary barriers for authors who need to comply with funders’ public access policies. As of May 28, the [statement](#) had been signed by 94 organizations and 619 individuals. **Carnegie Mellon University signed the statement. We encourage you to read and sign it.**

A recent webinar in the UK, [Getting the Rights Right, or When Policies Collide](#), predicted the day is coming when compliance with funder policies will force institutions to constrain where their researchers can publish.

## The Hague Declaration

Significant barriers to the analysis of facts, data and ideas must be removed to reap the economic and societal benefits of knowledge discovery in the digital age. [The Hague Declaration](#) aims to foster agreement about the best way to remove these barriers and enable access. Signatories endorse five principles they believe can help shape ethical research practice, legislative reform and the development of open access policies and infrastructure. Briefly stated, the principles articulated in the Declaration are:

1. Intellectual property (IP) was designed to promote research activity, not to regulate the free flow of facts, data or ideas.
2. IP law should not restrict innovative or commercial use of facts, data or ideas.
3. Licenses should not restrict use of facts, data or ideas.
4. People should be free to exercise intellectual and analytic curiosity without fear of monitoring or repercussions.
5. For the benefit of society, ethical norms for content mining must evolve as technologies evolve.

**Carnegie Mellon University signed the Declaration. We encourage you to read and sign it [here](#).**

## Actions for Researchers

### Push back if a publisher claims your article is a work for hire

Ever publish or consider publishing with the American Society of Mechanical Engineers (ASME)? Perhaps it's time to stop. The ASME recently revised its [publication agreement](#). The new agreement requires authors to

- Stipulate that their article was a work made for hire, commissioned by the ASME
- Waive all moral rights to the article, including the right to be attributed as the author

According to the blog post [From control to contempt](#) by Kevin Smith (Duke University), this is an "outlandish" contract and the "epitome of disrespect" for authors. Authors who respect themselves and academic freedom should refuse these terms. If you agree, tell the [ASME](#).

### Compare your workflow with your peers

Do researchers use Google Drive instead of Word? Google Scholar instead of Scopus? Megajournals instead of topical journals? ResearchGate instead of repositories? New scholarly communication tools abound, changing the research workflow. (See this [spreadsheet](#) of 400+ tools.) Who is using what and why?

Utrecht University is studying and [charting](#) the changing workflows of faculty members and researchers worldwide. The work includes an international survey investigating how tool usage varies by discipline, country and position.

We invite and encourage you to take the [graphical survey](#). Yes, a fun graphical survey. Invest ten minutes and get a nice graphical representation of your workflow compared with the workflow of your peers.

Feel free to share the link to the survey with your CMU colleagues (<http://tinyurl.com/SCtoolSurvey>). The survey will run until February 2016.

See the [Innovations in Scholarly Communication](#) website for background information. Preliminary results from the survey will be posted on the site.

### Get ready, get recognized with ORCID

Over 1.3 million researchers worldwide have registered for an ORCID ID. Why? Because ORCID uniquely and persistently identifies them and enables publishers, funders and other researchers to recognize and reward their work.

Journal publishers – including AIP Publishing, IEEE, Nature Publishing Group, the National Academy of Sciences, Springer, and Thomson Reuters – are capturing and embedding author ORCID IDs in their articles. Wiley asks authors for their ORCID ID and, if they don't have one, encourages them to get one. Hindawi captures and includes in article metadata the ORCID IDs of authors, editors and peer reviewers. Capturing the ORCID IDs of peer reviewers enhances the discoverability of peer reviewers and makes peer review activities measurable.

Funders, altmetrics platforms, and data repositories are also integrating ORCID IDs into their workflows. For example, Autism Speaks, National Institutes of Health, Wellcome Trust, Dryad, figshare, Plum Analytics and ImpactStory request ORCID IDs. A graph of completed and underway ORCID integrations is available [here](#).

Carnegie Mellon wants your ORCID ID to be linked to your Andrew ID in CMU's identity management system. Why? To recognize and reward your work and facilitate communication with your research sponsors. If you have not yet used the **ORCID @ CMU** web app, please go to <https://orcid.library.cmu.edu> and follow the brief instructions. See the [ORCID @ CMU FAQ](#) for more information.

### Read it. Sign it.

Carnegie Mellon University has endorsed the following initiatives. We encourage you to read these brief documents, think about the issues, and if you agree with the positions taken, sign your name to show your support.

- [The Hague Declaration on Knowledge Discovery in the Digital Age](#)
- [Statement against Elsevier's sharing policy](#)

## Actions for Librarians

### Read *The Decolonized Librarian*

[The Decolonized Librarian](#) blog explores how libraries can be freed of the political control entrenched in the hierarchical, Western, rational systems that name, organize and provide access to materials in our collections. Why would we want to do this? To enable other ways of knowing and being in the world and to improve service to the diverse community that is CMU. Recent posts examined the role libraries play – through biased collection development and pejorative classification and subject headings – in suppressing [genocide](#) in North America and [skepticism](#) over the theory that Shakespeare was Edward de Vere, 17th Earl of Oxford.

The Decolonized Librarian is Michael Dudley, Indigenous & Urban Services Librarian at the University of Winnipeg.

### Read *The Feral Librarian*

[The Feral Librarian](#) blog addresses many topics. The recent post [Never neutral: Libraries, technology, and inclusion](#) explains that our lack of neutrality is the result of (a) our “absorb[ing] and reflect[ing] the inequalities, biases, ethnocentrism, and power imbalances that exist” around us, and (b) our conscious choices. We choose “how our technologies handle metadata and catalog records,” and these choices “have consequences for how existing biases and exclusions get perpetuated” in our digital libraries. Furthermore, basing collection development on popularity ensures that our collections reflect existing biases and inequalities. Similarly, search engines that base relevance ranking on popularity ensure the same biases persist online.

The Feral Librarian is Chris Bourg, Director of Libraries at MIT.

### Read *In the Library with the Lead Pipe*

The subtitle of this [open access journal](#) is: “The murder victim? Your library assumptions. Suspects? It could have been any of us.”

### Read *Now I am become DOI, destroyer of gatekeeping worlds*

Intelligent, insightful, hilarious, this [article](#) by Ral Yarkoni – initially published as a blog post without a Digital Object Identifier (DOI) – explores the publisher injunction against citing documents that don’t have a DOI, what Yarkoni calls the “no-DOI-no-shoes-no-service policy.”

Yarkoni believes people deserve credit for interesting things they say, whether or not what they said is associated with a DOI. He’s sure that if Mark Twain were alive today “he’d write the best tweets EVER” and wouldn’t publish at all because “he’s too impatient to rinse-and-repeat his way through the revise-and-resubmit process.” He wonders why Elsevier hates 21st Century Mark Twain.

#### Asides

- When researchers complained to ORCID that they wanted to import citations that did not have a DOI, ORCID added the functionality to import BibTeX files. Users can also manually add citations to their ORCID profile.
- Having been told that relying on DOIs disenfranchises new niche journals, conference proceedings, blogs and other research and scholarly communication products, Kudos is “working towards a more flexible ingestion system that will enable these other kinds of outputs to be recognized and enriched within Kudos.” (email to Denise Troll Covey, May 19, 2015)

### Get intellectually engaged!

#### Become part of the solution

Libraries are, perhaps unknowingly and unwillingly, instruments of institutional suppression, complicit in the subjugation of knowledge in the academy. We provide and support practices and technologies that have inherent biases. What might we do to correct or disclose the issues raised on this page?