April 2013

Don's Notes-Special Meeting Reports: TOC and PDA Conferences

Donald T. Hawkins
dthawkins@verizon.net

Follow this and additional works at: https://docs.lib.purdue.edu/atg

Part of the Library and Information Science Commons

Recommended Citation

DOI: https://doi.org/10.7771/2380-176X.6495

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Tools of Change and Personal Digital Archiving Conferences

Don’s Notes — Special Meeting Reports

“Connect/Explore/Create”: O’Reilly’s Tools of Change Conference

About 1,500 publishing aficionados gathered in New York on February 13 and 14 for the 2013 Tools of Change for Publishing (TOC) conference. Organized by O’Reilly Media, publishers of an extensive line of computer-related books known for the drawings of animals on their covers, the TOC conference began seven years ago and has grown into a significant and popular event on the information industry calendar. This article summarizes the plenary sessions; videos and slides of many of the presentations are available on the TOC Website at http://www.toccon.com.

Tim O’Reilly presented an optimistic view of the publishing industry in his opening keynote. He said that fear of the future seems to be abating and noted that eBook revenues now exceed those from hardcover books. The market seems to have stabilized, and the industry has come into equilibrium. People still love reading. Authors write and publishers publish to help people solve problems and give them instructions on how to do things. O’Reilly’s exhortation to his audience was to work on things that matter.

Brian David Johnson, Intel Corporation’s Futurist, said that we will soon be able to turn anything into a computer, and much of what publishers do is relevant to that. He identified science fiction and its steampunk sub-genre as helpful in prototyping the future. Sci-Fi can be used to understand the cultural impact of technologies, and steampunk is about the rise of technology. (See Vintage Tomorrows, co-authored by Johnson and published by O’Reilly for further information.) Publishers tell stories, so narratives, words, and opinions are important and can change how we reach people.

Matt MacInnis, founder and CEO of Inkling, a startup company devoted to changing the way that people access and read books, said that the model of a book has become unbound, and today’s new dominant medium has changed everything. For example, the concept of pages has no meaning in an electronic system because pages are a physical constraint on the amount of content that can be presented in a given space. In the electronic arena, we have unlimited memory instead of pages. E-readers are good for reading text, but they are much less suitable for images, videos, etc.

MacInnis does not think print books are dying; on the contrary, he sees new opportunities for them. At TOC, MacInnis announced the launch of the free Inkling Habitat (habitat.inkling.com) service. Habitat claims to be “the only collaborative publishing environment designed for professionals” and is paired with the Inkling Content Discovery Platform, allowing publishers to extract content from the book and get it into places where people can use it. Thus, users searching for knowledge can directly access quality curated content and easily share it with others, which is an implementation of two hot information technologies of today: discovery and social media.

Continuing the theme of making content available as widely as possible, John Wheeler, SVP at SPI Global, observed that publishers are all looking to place their content in multiple distribution channels, and electronic consumption of content will only increase. Mobile and tablet usage is skyrocketing, and tablets have become users’ primary computing devices. Publishers must ensure that people buying content on one platform can see it on others, so it has become important to author the content only once but distribute it on multiple platforms. In Wheeler’s opinion, HTML5, the base language of the EPUB3 platform, is best suited to leverage the capabilities of multiple platforms and enable cost-effective production of enhanced content because of its structured and complete set of supported features. Currently, there is tremendous support for activities based on EPUB3.

A conversation between Henry Jenkins, Professor at the University of Southern California; Brian David Johnson, Intel Corporation’s Futurist, and Cory Doctorow, science fiction author, journalist, and blogger, produced some interesting points about various aspects of print and electronic media (the entire conversation is available at http://www.toccon.com/toc2013/publisherschedule/detail/27069):

- Facts used to be expensive to rework and publish. Now all facts are available on the network, so one can write a novel with unfamiliar concepts and assume readers that have an implied “just Google it” search box available. This is the first hint that a 21st-century novel will always involve a search engine.
- A book that was too long for the publisher turned into a second eBook—a new approach to expanding on narratives. Intel has a “Tomorrow Project” to figure out how we can have conversations about the future and get as many people as possible to have them, which leads to enhanced experiences for readers.
- We need to change our ideas of how content travels and reproduces. Unauthorized circulation may actually increase value for original rights holders because even though they lose control, they will gain cultural currency. Figuring out how to make media transition smoothly to readers is very important.
- We need to ask people to react and tell us what they think about our scenarios. The ability to talk back and be heard is very important to the concept of spreadable media. Science fiction was built as a genre for talking about ideas and encouraging debate, and viral media began with science fiction in the 1950s as authors wrote about germs that infect large numbers of people.

Author Douglas Rushkoff discussed some concepts from his newest book, Present Shock (Current, 2013), that apply to the publishing industry. We are now in a society where the future is always on. Text invented history, but we now become informed by several different forms of writing. In a digitally dominated universe, everything becomes à la carte. As we have moved from an economy where value was stored over time towards a peer-to-peer economy, the present shock has impacted books as a business. Books are a sustainable business, not a growth business, but many publishers have not realized this. We are moving towards a model where people will subscribe to books and will have choices all the time. Our challenge is to show people that making a choice is entertaining and valuable.

The Web has caused a seismic shift in publishing as well as in many other industries. As Jeff Jaffe, CEO of the World Wide Web Consortium (W3C), observed, the early Web was no substitute for print; therefore, publishers were not motivated to participate in the development of its core standards. But now, new technologies such as broadband, social networks, and mobile platforms have made the Web a richer and better publishing platform.
And we can expect continuing improvements as the use of blogs, tablets, eBooks, and similar technologies grows. We have gone from a first generation publishing mechanism to one with a substantial impact on the entire publishing ecosystem, and we can expect further innovations in business models. Rich content, social media, and time slicing are more natural with the Web than with printed matter. eBooks are also heavily using Web technology, and with the advent of HTML5 (the most interoperable technology in the industry), the Web and publishing communities are beginning to communicate.

Jaffe listed four major strategies that will lead to convergence:

- Match Web technologies with current publishing practices,
- Leverage the value of the Web (operations such as annotations and cataloging will especially benefit),
- Support diverse business and distribution models, and
- Satisfy diverse consumer behaviors.

Plenary sessions on the second day began with an introduction of three winners of a “Startup Showcase” that were chosen by a vote of attendees:

- **Paperight** ([www.paperight.com](http://www.paperight.com)) was developed in South Africa as a means to get books into the hands of people in remote areas who cannot get to a bookstore. Publishers license their products to copy shops that legally print books on demand for customers and sell them at a low cost.

- **CartoDB** ([cartodb.com](http://cartodb.com)) is a cloud-based tool for dynamic, data-driven storytelling using a database and visualization engine. Users do not need to know how to write code to create visualizations; data can be simply dropped into the system from a spreadsheet to create map visualizations.

- **Borne Digital** ([www.borne-digital.com](http://www.borne-digital.com)) is a publisher of books, games, and software for children and educators. Books created with its system are not just simply repurposed print versions: value is added to produce unique interactive learning materials. As they are read, the system can measure the user’s reading level and automatically present more advanced versions as appropriate.

Following the Showcase winner presentations, **John Tayman**, founder and CEO of **Byliner**, an aggregator of book content, introduced his product, based on the principle that “Content is Still King.” **Byliner** commissions, collects, and curates stories from the world’s greatest authors and makes stories available in any way that a reader might enjoy them, viewing itself as more an entertainment company than a publisher. **Byliner** specializes in both fiction and nonfiction works specifically written to be read in two hours or less. Readers can subscribe to the stories or purchase them from digital bookstores. Many would-be readers of such works are currently frustrated in trying to find something to read. **Byliner** seeks to remove their frustrations by gathering all types of stories, curating and cleaning the data, and enabling discovery by time, mood, or location. One can create a live digital feed of a writer’s entire output; thus, **Byliner** becomes a powerful reader acquisition tool and quickly converts a first-time reader into a devoted fan of an author. **Byliner** can be regarded as a “Netflix for authors”: everything is available to readers for a single subscription price.

Comics are a widely-read genre by both adults and children. According to his biography on the **TOC** Speakers page, “Mark Waid has written a widely variety of well-known characters than any other author of American comics, from Superman to the Justice League to Spider-Man to Archie and hundreds of others.” **Waid** has discovered that there are significant challenges in adapting comics to digital media. Most significantly, they are almost exclusively published in a portrait format, but laptops and many digital readers operate in landscape mode. Reading comics on an e-reader therefore involves lots of scrolling and does not work very well, even though the story is available. **Waid** likened the experience to viewing a movie through a cardboard tube. His **Thrillbent.com** site takes advantage of the storytelling tools digital publishing allows and enables reading of comics on a digital medium. For example, things the reader needs to dwell on can be programmed to stay on the screen longer. Although comics sell well in print, their greatest production expense is the printing, which is escalating in price. The **Thrillbent** site provides a platform to avoid some of these costs. **Waid** echoed many of the other speakers by concluding with this advice: “Don’t sell pictures of books; sell a whole different publishing experience.”

At last year’s **TOC**, one of the presentations was from an executive at **Library Journal** (LJ) who discussed some statistics on eBook usage in libraries. This year, **Meredith Schwartz**, LJ’s News Editor, presented some findings from the second edition of that study (“Patron Profiles: Understanding the Behavior and Preferences of U.S. Public Library Users”). She noted that there are many libraries in the U.S., and they buy lots of books, including eBooks. Many library users are active book buyers as well as borrowers, and the library is the main local place where consumers go to find books they want to buy. Libraries are therefore an excellent example of showrooms with no guilt attached. Books on home decorating, humor, and gardening are the most popular ones bought.

About 40% of a library’s patrons are aware that their library has eBooks, and 30% have borrowed one. Patrons love the convenience of eBook borrowing but do not like restricted lending periods. Nor do they want to be required to come to the library to be able to borrow an eBook. Borrowing print books is still the king of library activities. Libraries and bookstores will not make book buying obsolete; in fact, they are working together to create a larger market.

**Maria Popova**, founder and editor of Brain Picks ([www.brainpickings.org](http://www.brainpickings.org)), a blog where she writes about a variety of subjects that interest her and which she calls a “human-powered discovery engine for interestingness,” gave the closing keynote. She explored the question of providing alternatives for ad-supported journalism and media and noted that advertorial stories have become quite common. She suggested these non-traditional alternatives to ad-supported journalism:

- **Longreads.com**: a searchable collection of in-depth stories suitable for a long train commute, airplane flight, etc.

- **Spot.us**: an open source project to pioneer “community powered reporting.” Through **Spot.us** the public can commission stories and participate with journalists to report on important and perhaps overlooked topics. Stories are published by local affiliates of **spot.us**.

- **Systems to distribute revenue through affiliate links or from large numbers of small donations**: for example **Flattr** ([flattr.com](http://flattr.com)) and **The Wirecutter** ([thewirecutter.com](http://thewirecutter.com)).

Joe Wikert, **TOC** Co-Chair, closed the conference with a list of things he had learned:

- Although we are still in the early days, we are seeing many advances in eBook production and delivery mechanisms.

- The startup space is most critical. Innovation is coming from everywhere. The traditional publishing world needs to begin collaborating with startup ventures, and there are many opportunities to do so.

- We are beginning to see companies leveraging technology for simple elegance, not just for technology’s sake.

- Today’s economic climates is difficult. We need to focus on what we do best and outsource the rest.

---

Continued on page 71
termine your “secret sauce”, invest in that, and be willing to partner with others for everything else.

- Community comes first, then e-commerce follows.

---

4th Personal Digital Archiving (PDA) Conference

The 4th Personal Digital Archiving (PDA) Conference and first one held on the East Coast convened on the University of Maryland Campus in College Park on February 21–22, 2013. It had quite a different tone than the first three, which were held in San Francisco. Due to the many technology and entrepreneurial startup companies in California, the first three PDA conferences had a significantly more technology bent than the fourth, which featured more of an emphasis on applications and services.

The Keynote Address: A Writer’s View of Personal Archives

In her keynote address, Sally Bedell Smith, a prominent author of biographies of famous people who were living at the time she was writing about them, such as the Clintons, the Kennedys, Princess Diana, and Queen Elizabeth II, presented a fascinating look at how writing has been impacted by changes in technology. Even in today’s electronic world, she continues to rely on printouts of files and accumulates 30 to 40 linear feet of paper for each book. (She would like to trust the cloud and abandon the printer but concedes that she may be a hard case.)

Although Smith uses interviews extensively because she writes about living subjects, she is also an enthusiastic user of archives, and especially likes the Hans Tasiemka Archive (for a description, see http://www.guardian.co.uk/media/2005/mar/13/pressandpublishing.observerreview) in London. She noted that researching archives can be very gratifying because it can provide the thrill of discovery, but it requires time, patience, and stamina. Archives allow one to see how a subject was viewed at a past time. Digitizing them is a noble effort, but it eliminates the ability to view the author’s handwriting, feel the different types of paper, smell the ink, etc.

Smith’s advice to writers is to organize all the source material and derive the structure of each chapter before sitting down to write. A chronology is the only way to organize non-fiction. Writers are planners and evolvers. Smith regards writing as a marathon and aims to produce about 1,000 words a day. She prints out completed chapters because it is easier to get the full picture of the work on paper than on a screen and said that nothing can ever substitute for the look and feel of pages between hard covers.

Acquisition of Personal Documents

According to Jenny Shaw, who works on a branch of the human genome project at the Wellcome Trust, a proactive approach toward acquiring records is necessary. The Wellcome library is a collecting repository and continues to establish new collections of the digital papers of scientists. With born digital documents, it is difficult to see at a glance what is in the collection, so software tools are useful in making acquisition decisions. Viruses are a currently a significant problem. The main challenges in building a repository are acquiring the material because the library has no power to compel contribution and can only encourage people to contribute. Thus, it is necessary to build relationships and trust with potential contributors and reassure them that personal information will be managed in accordance with legal and intellectual property requirements.

Engaging Users Through Email Gamification

In the first of his two presentations, Sudheendra Hangal, one of the developers of Stanford University’s Project MUSE (Memories USing Email) noted that email will always be a tool for examining the textual records of a person because it occurs in standard formats. Plug-in applications for popular browsers can do such things as extracting names from emails and highlighting messages already viewed, then subsequent searches can be restricted to only that subset of messages. The user, in effect, can create a personal email search engine.

Gamification refers to asking people questions about their friends and highlighting interesting relevant content in a social archive. Hangal showed an extract of a message from a person who suffered a severe brain injury and “used MUSE to try and figure out who I was and who was who to me in my past...” This experience shows that new memory tests based on personal archives may be better than those used currently. There are exciting possibilities in gamification of email archives, and these new uses for them may make people more aware of their value.

Connecting Local and Family History

Noah Lenstra, a student at the Graduate School of Library and Information Science at the University of Illinois, noted that many folk research efforts are done by people unaffiliated with academic research. Many public libraries that have long supporting genealogical studies are now moving to supporting personal digital archiving. Many public librarians are interested in this process, but they need help and guidance; some of them are now offering digitization of personal photos as a service to their users. Lenstra has recently conducted seminars for local librarians and is developing an instructional manual for them. Further details are available at http://manual.eblackcu.net/wiki/Main_Page/.

“Lightning” Talks — Round 1

Several brief presentations highlighted recent PDA developments:

- The Library of Congress is preparing several brief video tutorials that will be available on its YouTube channel. The first one is on preservation, and the next will be about scanning. These will be useful in outreach to public libraries.
- Video games can be an effective archiving experience because they are historical. One accumulates points in a game through achievements, which can form the story of how the play has been associated with the game.
- Archives are dependent on data storage, and for a large social network like Facebook, the infrastructure to house and support the servers is significant. For example, Facebook accounts for 1 of every 7 minutes spent online, uploads 3,000 photos/second, and its servers consume the same amount of electricity as 30,000 U.S. homes. How does this network affect our lives? The data cannot be separated from the network, and the archive may know us better than we know ourselves. The location of the servers has significant legal implications; for example, Max Schram, an Austrian student, was able to demand that Facebook provide him with a printout of his entire history because its European headquarters are in Ireland which has different privacy laws than the U.S. and Canada, whose residents do not have such a right. (For further details, see Schram’s Europe vs. Facebook Website at http://europe-v-facebook.org/F_AQ_ENG.pdf.)
- The usefulness of artifacts is based on their physical characteristics. Objects are important in the social
construction of a family’s identity, so we cannot leave them out of a personal archive. We must rethink our frame of reference, go beyond the physical, and not just limit ourselves to digital archives.

Scholarly Workflow in Personal Digital Archiving

Smiljana Antonijevic, John Meier, and Ellysca Cahoy from the Pennsylvania State University library conducted a study that investigated how academic faculty members create, manage, share, and archive their personal information collections. One of the goals of the study was to determine if there was a natural place to integrate personal digital archiving into the scholarly workflow. Can faculty be taught to do their archiving throughout their careers rather than at the end? What are the necessary critical digital literacies? They found the following:

- Endnote, Mendeley, and Zotero were the most commonly used citation management programs.
- Most stored information resides on the faculty member’s hard drive. Some is located on Dropbox or in a citation management system.
- Records are stored in a wide variety of formats, including PDF, Word files, spreadsheets, and email.
- Backups are generally done at least monthly, but some faculty back up on a daily basis.
- External hard drives, flash drives, and printouts are common backup formats.
- Emailing and Dropbox are the most commonly used sharing mechanisms.
- Younger faculty (less than 40 years old) are most likely to use cloud-based services and are less likely to curate their information. Older faculty are more likely to save Websites, images, and email collections.

Many faculty members feel that preservation is a task for publishers, scholars, and professional organizations. They regard sharing as a form of preservation. Many of them learn how to archive their materials on their own but recognize that they need help. The current academic reward system has a major influence on what scholars archive; they do not necessarily preserve what they care about but what they get credit for.

Further information on the study is available at http://scholarlyworkflow.org/.

Providing Access to Email Archives for Historical Research

Sudheendra Hangal followed up his earlier presentation with a description of Stanford University’s Email Process Appraise Discover Deliver (ePADD) system, an open source program to visualize and browse email archives in special collections. The ePADD system is powered by the MUSE platform. (MUSE is designed for personal use on a local computer; ePADD is for institutional repositories.) The public version of ePADD contains portions of several large email collections in the Stanford library. Messages can be analyzed and viewed (except for attachments which are accessible only in the library).

Libraries are good at capturing and preserving documents, but little progress has been made in processing and delivery. Email is an important part of these collections because everybody has it, and it will be around for a long time. According to Hangal, there are over 2 billion email users today and nearly 3.5 billion accounts. One of the interesting characteristics of an email collection is that it preserves a user’s incoming and outgoing messages, thus providing a picture of the viewpoints of both senders and recipients. An ePADD demonstration site is available at http://epadd.stanford.edu/muse/archives/b.

Narrative Searching of a Scholar’s Email Archive

Jason Zalinger, Assistant Professor at the University of South Florida, described his experiences in searching nearly 45,000 emails of Ben Shneiderman. (Shneiderman is Director of the Maryland Institute for Technology in the Humanities (MITH), one of the hosts of the PDA Conference, and is a renowned scholar as well as an author of a number of highly regarded books.) Shneiderman’s email archive contained over 4,000 personal relationships and was very well curated; the collection he gave to Zalinger contained no junk email, meeting announcements, etc. It was saved in folders organized by name and year.

Faced with this large collection, Zalinger tried several analytical approaches to it, but Shneiderman’s suggestion to identify influential correspondents provided the best insights. Zalinger went on to search on key transitional phrases or terms to construct interesting narratives from the emails. In a narrative, tension makes a good story, so Zalinger prepared lists of appropriate search terms related to anger, sudden events, or feelings. To create a narrative based on a collection of documents, one must think like a storyteller, and create an index of places, dramatic transition phrases, or emotion phrases. Zalinger concluded that personal digital archives are stories waiting for a narrator, and the archivist’s task is to find them.

Second Day Keynote: George Sanger, “The Fat Man”

The second day keynote was by George Sanger, composer of music used in over 250 video games. His archive contains many digital items, but they are only a small part of the entire collection. Sanger organized his records with little regard to format and experimented with different combinations of digital techniques and formats. Some items even contained data in several formats: for example, some CDs contained both audio files and images.

Sanger advocates getting rid of things and suggested giving them to archivists to keep them happy! He said that everyone should have an archive and should back it up continually for two reasons:

1. Posternity—someone might learn something from it, and
2. Safety and survival, particularly for a business.

Opportunities and Challenges of Personally Revealing Information in Digital Archives

Four academic faculty members participated in this panel discussion. Cal Lee from the University of North Carolina (UNC) said that digital materials can be represented in many ways, and every representation is a potential future interaction with the data. We must recognize the varying levels of interest by users and the ethical considerations of each level. Archivists often receive information that they are not supposed to keep, such as financial data, medical histories, etc. This type of information may be difficult to identify and segregate from the other information in an archive, and donors may be sensitive about it as well as its metadata. Naomi Nelson from Duke University’s Rare Book and Manuscript Library listed some questions we must ask before adding data to an archive:

- Is the information publicly and easily available elsewhere?
- Does it have enduring value?
- Can it be efficiently located and segregated?
- Can the donor imagine a time when it would be permissible to share it?

Kam Woods from UNC reported on the BITCURATOR project (http://www.bitcurator.net), “a research project to build, test, and analyze systems and software for incorporating digital forensics methods into the workflows of a variety of collecting institutions.” The goal is to protect the donor’s information. We might ask these questions:

- How do we know that we are not changing anything?
- Have we found everything there is to find?
- What file system are we dealing with?
- What files have potentially private data in them?

It is also important to recognize that even if someone does all their work in the cloud, traces of their activities are still left behind on their PC (such as Facebook and Twitter activity).

Finally, Matt Kirschenbaum from MITH said that we live in a time when machines are also actors. Automated agents mine the Web and harvest data. Objects and our relations with them change over time, and digital objects accelerate this process. Originators, donors, and archivists must negotiate what they will do with the information being considered for the archive.
Collecting to the Core — Women’s Studies in African History

by Summer Durrant (Assistant Librarian, Indiana University-Purdue University Indianapolis; African History, Languages and Literatures Editor, Resources for College Libraries) <durrant@iupui.edu>

Column Editor: Anne Doherty (Resources for College Libraries Project Editor, CHOICE/ACRL) <adoherty@ala-choice.org>

Column Editor’s Note: The “Collecting to the Core” column highlights monographic works that are essential to the academic library within a particular discipline, inspired by the Resources for College Libraries bibliography (online at http://www.rcweb.net). In each essay, subject specialists introduce and explain the classic titles and topics that continue to remain relevant to the undergraduate curriculum and library collection. Disciplinary trends may shift, but some classics never go out of style. — AD

African women received scant attention as historical subjects before the 1970s. This was largely because early historical records, written primarily by male European travelers, missionaries, and colonial officials, marginalized women’s roles and activities. Beginning in the 1980s, a handful of pioneering feminist historians began to focus their research on African women in an effort to counter negative stereotypes and present a more realistic portrayal of their daily lives.

Over the past three decades, the importance of women’s studies in African history has steadily grown. The ten monographs discussed in this essay are considered seminal works on the topic. Several have won prestigious awards like the African Studies Association’s Herskovits Prize, or have been named Choice Outstanding Academic Titles. The titles here cover major areas of scholarship, including production and reproduction, economic livelihood, and issues of domesticity, marriage, and sexuality.

Two texts offer authoritative overviews of women in African history. Catherine Coquery-Vidy-Vidrovitch’s African Women: A Modern History offers an excellent general introduction to the topic.1 Commonly used as a classroom textbook, this work provides a broad examination of how women’s lives changed over the course of the twentieth century. Coquery-Vidy-Vidrovitch focuses on several key aspects — marriage, wage labor, urban migration, education, economic activity, and political activism. It is one of the few books that spans the entire subcontinent and highlights how women’s conditions both improved and declined as a result of colonization. “Wicked” Women and the Reconfiguration of Gender in Africa, edited by Dorothy Hodgson and Sheryl McCurdy, is also often used as a classroom textbook, but primarily for upper-division courses.2 The seventeen provocative essays in the anthology explore how women challenged notions of “good” and “proper” behavior in society. In response, these women were often stigmatized, especially by men seeking to regain control over wives, daughters, and sisters.

Colonialism brought about sweeping social, economic, and political changes to the African continent, including profound impacts on conjugal and kinship relations. Cash cropping and the monetization of the economy led to domestic struggle as men and women sought to secure access to strategic resources like land, labor, and capital. New opportunities for education and wage labor, as well as the allure of city life, led many women to leave their villages in search of better opportunities in urban areas. This prompt-

continued on page 74