September 2008

What About eBooks?

Toni Tracy

Portico, toni.tracy@portico.org

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

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

Recommended Citation

Tracy, Toni (2008) "What About eBooks?," *Against the Grain*: Vol. 20: Iss. 4, Article 10.

DOI: <https://doi.org/10.7771/2380-176X.5146>

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.

Was talking to **Janet Flowers** the other day. **Janet** has retired after 38 years at **UNC-Chapel Hill** (!) but not really if you ask me. **Janet** and her husband **Ed-die** are coming to **Charleston** in November AND **Janet** is doing a **Lively Lunch** at the **2008 Conference** called "**Web 2.0 Tools for Newbies**." This program (Thursday, from 12:30-1:40) will provide an orientation to the **23 Things approach to learning Web 2.0 tools**. Boy. You talk about taking a "**glass half full**" optimist's approach. Instead of being old, we are **newbies**. I like it!

www.katina.info/conference

And speaking of **UCSD**, the incredible **Brian E. C. Schottlaender**, the **Audrey Geisel University Librarian** at **UCSD**, is our keynote speaker on Saturday, November 8! Don't forget to be there! And, did you know that **Brian** once worked for **Harrasowitz**? www.ucsd.edu/

We're truly thrilled to have **Ned Kraft** back with his **Adventures** column. And the one in this issue is about **Heroes**, p 8.

From the *New York Times*, August 24, 2008 — A group of scholars are trying to make **George Orwell** more relevant to the younger generation by creating a blog of his extensive diaries. For example on September 7, 2008, **Orwell's** entry for September 7,

continued on page 44

against the grain people profile

Martha Anderson

Director of Program Management, National Digital Information Infrastructure and Preservation Program (NDIIPP)

101 Independence Avenue SE,

Phone: (202) 707-2598 • Fax: (202) 252-3173
<mande@loc.gov> • <http://digitalpreservation.gov>

PROFESSIONAL CAREER AND ACTIVITIES: Prior to my current position at the Library — I have served as project manager for the Office of Strategic Initiatives at the **Library of Congress** working with preservation architecture projects. I manage the capture of Web content in support of the **Library's Digital Strategic Initiatives Program** and I am the Library's program officer for the **Web-at-Risk, California Digital Library Project** and the **MetaArchive Project**.

During the first charter period, 2003-2006, of the **International Internet Preservation Consortium (IIPC)**, I chaired the Metrics and Testbed working group. The **IIPC** is an international organization of 37 national libraries and non-profit archives dedicated to collecting and archiving significant content from the Web. I now serve on the Steering Committee managing the communications.

During my tenure on the staff of the **National Digital Library Program** at the **Library of Congress**, I served in the roles of Manager of the Digital Conversion Group, Production Coordinator for American Memory historical Web collections, Contracting Officer's Technical Representative (COTR) for the multi-year paper scanning contract and as the **National Digital Library Program Liaison** to the Manuscript Division at the **Library of Congress**.

Prior to joining the **Library of Congress** in 1996, I worked in electronic publishing as editor and project manager for a variety of CD-ROM reference products including **Time Almanac** and the **Patrologia Latina Database**.

IN MY SPARE TIME: I am a Master Gardener.

PHILOSOPHY: There is much to learn about digital preservation from the natural world. Diversity survives in nature.

HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS: The explosion of user-generated content and the availability of personal publishing tools is changing the information environment for libraries and publishers alike. Within the next five years, more and more libraries will serve their online patrons in more dynamic and personalized ways. There will also be a greater demand for more robust searching and filtering mechanisms to help users manage information. 🌿

What About eBooks?

by **Toni Tracy** (Director, Publisher Relations, Portico) <toni.tracy@portico.org>

In response to a growing number of requests from the publisher and library communities to preserve non-journal scholarly literature, including eBooks, **Portico** developed a study in late 2007 to help offer insight to the community about the current eBook landscape, including business models, data formats, and preservation needs. The study was designed to understand what level of need and support for eBook preservation exists in the scholarly publishing and academic library communities and to analyze the structure and preservation requirements of eBooks to determine the operational and technological infrastructure required to offer eBook preservation.

Portico invited eleven publishers and eBook aggregators to participate in the study and nine accepted. These organizations all participated in an interview, and the publishers

supplied eBook content for technological and preservation assessment.

- American Mathematical Society
- Elsevier
- Morgan & Claypool
- Organization for Economic Cooperation and Development (OECD)
- Royal Pharmaceutical Society
- Taylor & Francis
- ebrary
- MyiLibrary
- NetLibrary

Portico also invited eleven libraries and one consortium to participate in the study and all accepted. These organizations also participated in an interview about eBooks within their own

institutional setting, or those of their members in the case of the consortium.

- California Digital Library
- Case Western Reserve University
- Cornell University
- McGill University
- Quinnipiac University
- SOLINET
- University College of London
- University of Montana
- University of New England (Australia)
- University of Texas
- Yale University
- York College (PA)

continued on page 42

Publisher Findings

Between September 2007 and February 2008, six scholarly publishers (three commercial and three not-for-profits) and three commercial eBook aggregators were interviewed and asked about various aspects of their eBook publishing and distribution activities. The aim was to discover how these organizations defined eBooks and the role they envision eBooks playing in their current and future programs. There was also an interest in understanding their business models for eBooks, what percentage of their sales were represented by eBooks, and what they estimated eBook sales growth would be over the next five years. The interview also explored their level of concern for the preservation of eBook content, any preservation strategies they currently had in place for this content, and what role they thought a trusted third party might play in eBook preservation.

- **eBook Definition:** There was general consensus that eBooks include e-only monographs, monographs appearing both in print and online, digitized print titles, and continuously updated reference databases. All of the participating publishers agreed that audio books are not currently included in the genre.
- **Role of eBooks:** 50% of the publishers interviewed agree that eBooks will eventually replace print in their publishing program while the other half view eBooks as experimental in nature as well as supplemental to a print publishing program that will endure.
- **eBook Sales/Business Models:** Participating publishers reported eBook sales of less than 1% of their total revenue to up to 100% of total sales for eBook-only publishers. All publishers expect double digit growth rates in eBook sales over the next five years.
- **Emerging Business Models:** Standard business models for eBooks have yet to emerge and the participating publishers currently employ a variety of channels and models including subscription, outright purchase with local loading rights, and exclusive as well as non-exclusive use of eBook aggregators. This is reminiscent of the early days of e-journals, and it is expected that as programs grow and revenues become more significant, standard eBook business models will emerge.
- **Preservation Concerns/Solutions:** All publishers are concerned about long-term preservation of eBooks and agree that the topic will gain importance as revenue from eBooks grows, but at the time of the study, no publisher had a preservation solution in place for their eBooks.

Toni Tracy

BORN AND LIVED: Born, New Brunswick, New Jersey. Have lived in New Jersey and Baltimore.

EARLY LIFE: Spent in New Jersey.

PROFESSIONAL CAREER AND ACTIVITIES: 25 years in medical publishing, first with **Williams & Wilkins** (now **Wolters Kluwer Health**) and then with **Churchill Livingstone** (now part of **Elsevier**) followed by early eBook development with **Franklin Electronic Publishers**.

IN MY SPARE TIME I LIKE TO: Travel, play golf.

HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS: I think that e-content formats will increasingly converge with less distinction made on the part of users whether the information they are accessing and using is a journal or a book or a database. This convergence will present new challenges for everyone in the scholarly publishing ecosystem. 🐼

- **Role of Portico in eBook Preservation:** All of the publishers interviewed expressed a desire for **Portico** to undertake eBook preservation in very much the same way as it has approached e-journal preservation, including “trigger event” access to content preserved in the archive.

Library Findings

During the same timeframe, eleven libraries and one consortium were interviewed and asked about trends in eBook adoption, usage and preservation. The aim here was to discover how these organizations defined eBooks and whether they saw eBooks moving into the mainstream of collections, much as e-journals did more than a decade ago. We were also interested in their definition of eBooks, trends in collection development and usage, concerns about preservation, and what role, if any, they felt **Portico** should play.

- **eBook Definition:** As with publishers there was general consensus that eBooks include e-only monographs, monographs appearing both in print and online, digitized print titles (including objects digitized by the library), and continuously updated reference databases. All those interviewed agreed that audio books were outside their collection scope although technically eBooks.

- **Trends in Collection Development:** With a few notable exceptions libraries regard eBooks as still

experimental content, but they express a strong expectation that over time (1-10 years) eBooks will become a core element of their collections much as e-journals have. One library reported an “aggressive” preference for eBooks and indicated that in selected disciplines only eBooks are purchased; print is no longer selected. Another library reported e-textbooks as an important component of ongoing eBook purchases.

“With a few notable exceptions libraries regard eBooks as still experimental content ...”

- **Discovery and Trends in Usage:** Most libraries reported minimal usage but expected usage to grow significantly over the next one- to two-year period. Two libraries reported very limited experimentation with handheld readers and did not expect to expand these activities. Several libraries reported that usage and requests for eBooks were subject to sudden spikes and that the source of these was difficult to identify. Usage across disciplines was highly varied and, as might be expected, was greatest in the sciences. Most libraries reported that users relied upon **Google** to discover eBooks; typically records for eBooks are loaded into the **OPAC**.
- **Business Models:** Typically libraries license large aggregated packages with ownership preferred, if possible. Post-cancellation access is typically available,

continued on page 44

but the reliability of implementation methods is regarded as unknown.

- **Preservation Concerns/Solutions:** All libraries reported that preservation arrangements needed to be more robust. Several noted that they do not (yet) raise this issue with publishers routinely but expect to do so as eBooks become a more permanent part of the collections.
- **Role of Portico in E-Book Preservation:** All of the libraries interviewed expressed a desire for **Portico** to undertake eBook preservation as a way of facilitating a more secure transition to reliance upon eBooks and permanent inclusion of this content into the collections, but the expressed urgency of this need varied considerably across institutions and disciplines. The greatest sense of urgency was expressed by librarians responsible for medical collections.

Technical and Preservation Summary

The publisher participants in the survey provided sample content for the technical portion of the study. One reason these particular publishers were selected to be invited into the survey was to cover as wide a variety of formats and platforms as possible, including eBook products in standard formats such as OEB, monographs created and/or hosted in journal systems, high-end reference publications, and output from digitization projects for older monographs.

“... libraries are still experimenting with eBooks and how to properly include them in their collection development strategies ...”

Technical Findings:

- **Content in Standard eBook Reader Formats:** There are a number of standard eBook readers and formats, including **Open eBook Publication Structure (OEB)** and its successor, the **.EPUB** family of standards which has been recently released. eBook content produced for eBook readers is inherently standards-driven and highly uniform across publishers; this is in contrast to e-journal content which is designed to work specifically in the publisher's own Website or delivery platform of choice and can be highly idiosyncratic.
- **Content Created or Hosted on e-journal Systems:** This content is typically PDF plus HTML with XML source files behind the HTML. It is similar to the vast majority of the content in **Portico's** current e-journal preservation program. It may present some metadata challenges as a number of publishers may be offering monographs via a journal delivery system, bending the metadata structures in the process.

against the grain people profile

Carol Richman

Director of Licensing, SAGE Publications
2455 Teller Road, Thousand Oaks, CA 91320
Phone: (410) 327-6808 • Fax: (410) 327-7045
<carol.richman@sagepub.com> • www.sagepub.com

BORN AND LIVED: Boston, MA, El Salvador, Los Angeles, Chicago, Baltimore.

PROFESSIONAL CAREER AND ACTIVITIES: English teacher and professor.

FAMILY: My husband and our pets!

IN MY SPARE TIME I LIKE: Bicycling, hiking, sitting on the beach reading.

FAVORITE BOOKS: Anything by **Eudora Welty** and great non-fiction.

PET PEEVES/WHAT MAKES ME MAD: People who stand at the top of the escalator and not move — looking bewildered...

PHILOSOPHY: Sighing helps clear the mind...

MOST MEMORABLE CAREER ACHIEVEMENT: ?

GOAL I HOPE TO ACHIEVE FIVE YEARS FROM NOW: Decide what I would like to do in retirement...not retire but do something else...

HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS: I think it will be greatly changed and much the same...technology will truly guide events and how we learn, do research, and conduct business. But libraries will still exist...I have been in this industry for a while and I remember the head of **Welch Library** at **Johns Hopkins** telling me that libraries were going to be a thing of the past, that everything would be online...that was over 20 years ago...and, here we are today. So, in five years, there will be a few changes, but libraries will be that much stronger. 🐘

- **Content on Custom Websites:** This was the most problematic data in the study as there were no other data formats beyond the Website HTML available. HTML Web pages are notorious for lack of standards compliance and none of the sites examined had fully valid HTML and many suffered from hard-coded relative file paths.

- **High-end Reference Works:** These are, in effect, custom-designed and semantically-rich databases and they do not lend themselves to generic preservation approaches, at least not without serious loss of meaning. Each preservation project of this type will require a custom preservation plan.

In comparison to the current e-journal content seen by **Portico**, the best eBook con-

tent is much more straightforward: the eBook standards have jumped ahead of the e-journal standards and provide a more robust basis for content management and long-term content preservation. The worst eBook content is no worse than the messiest e-journal content — the difficulties that may arise for preservation will be due to lack of scale and resulting costs. For journals, the significant variations in data tend to be between titles, not at any smaller levels; in contrast, some of the eBook data studied varied from monograph to monograph.

Conclusions

While libraries are still experimenting with eBooks and how to properly include them in their collection development strategies, the eBook publishers are immediately ready to implement a preservation strategy. The technological analysis of the study shows that even the more difficult eBook content will be no more difficult than e-journals. 🐘

Rumors from page 40

1938, will be posted from his political diary when **Orwell** was “consumed by the clouds gathering” for the second World War. **Jean Seton**, a professor at the **University of Westminster** in London, had the idea for the blog and administers the project. **Seton** says the site

will be publishing at least until 2010 and has received many visitors. The original material can be found in **University College London** where many of our **ATG** friends are located — **Anthony Watkinson**, **David Nicholas**, **Liz Chapman** are just a few of them. But. As we all know, putting the material on the Web will make it more visible and accessible. See — “What George Orwell Wrote, 70 years Later

continued on page 55