September 2011

Standards Column -- Social Reading and the Problems of eBook Annotation and Referencing

Todd Carpenter

NISO, tcarpenter@niso.org

Standards Column — Social Reading and the Problems of eBook Annotation and Referencing

by **Todd Carpenter** (Managing Director, NISO, One North Charles Street, Suite 1905, Baltimore, MD 21201; Phone: 301-654-2512; Fax: 410-685-5278) <tcarpenter@niso.org> *www.niso.org*

Contrary to the stereotype of the introverted, quiet, bookish scholar, reading is a very social activity. The physical act of reading, as we all know is a seemingly self-focused activity. But while we may read to ourselves in a quiet environment, most of us do not read in an anti-social vacuum. Most of the time after reading, we share the book, article, or publication either directly or indirectly. This could be the physical item itself, the theme or contents of the book, or by simply retelling the stories in the publication to friends, family, or colleagues. We might also share in more formal contexts, such as book clubs, classrooms, or electronic forums by, for example, posting reviews on *Amazon.com*. Such sharing and discussion is especially prevalent in the scholarly community.

Social reading is also a crucial element of the publication process. Reading begets more reading. It informs writing, proposals, and future research; and hopefully, this reading is incorporated in citations and future publications. Thus the cycle begins anew. Although this process isn't as real-time as a phone conversation with a colleague, it is certainly social communication — simply proceeding at a different pace.

The new environment of electronic books poses some significant challenges to these social aspects of reading, in spite of the increasingly social nature of our online interactions. While we can read and share links to texts, new stories, and most online journals, the eBook market is decreasingly social. For a variety of business model and technological reasons, it is difficult to share eBooks. You generally cannot simply hand someone a digital file in the same way that one might give away a print copy of a book.

Similarly, referencing information within an eBook is a challenge. One cannot reference a quote from page 217 of a particular book, because in an environment where text is reflowable and reformattable based on screen size, device orientation, or user preference, the concept of "page 217" is meaningless. Page numbers are a useful relic of the print era, when they were critical to know the production order of pages when binding a physical copy of a book. Certainly, they have developed over time to have other uses, such as discovery through indices or tables of contents and the citation and referencing of texts.

Other tools can replace some elements of page number functionality. For example, the discovery process can be replaced with full-text search of an eBook. While some type of linking structure or reference system could involve identifiers or links to digital documents, precise location within a document is not currently an element of persistent linking systems like DOIs or PURLs.

A robust linking or annotation system needs a permanent and fixed location syntax and functionality. This is a critical requirement of the scholarly community, especially in the humanities, which more frequently rely on monographic citations. In our present information environment, there are generally canonical print versions that can be referenced. Also, some services, notably **Amazon**, are computing page numbers for selected texts when the publisher does not supply them in the electronic file. However, it won't be long before there are no longer print versions produced that can be referenced, at least of some books. In all likelihood, this transition will take place first among the already short-run scholarly monographs, where the University Press community has been struggling with the economic issues of short-print-run scholarship. In the near future, it won't be economical to produce print versions for many of these texts, apart from the occasional print-on-demand copy. Before we reach this tipping point where distribution in eBook format becomes the norm, it will be vitally important that annotation and reference structures be in place along with several other new standards on distribution, formats, and preservation of electronic texts.

It is for this reason, that **NISO** has partnered with the **Internet Archive (IA)** with the support of **The Andrew W. Mellon Foundation** to begin standards development work in this area. In May, **NISO** and **IA** hosted a meeting of key players in eBook reading systems and software manufacturers and distributors of content to discuss potential work in this space. Several producers and distributors of eBook content, as well as service providers, have already implemented some type of annotation capability, which requires location syntax. However, it will be important for disparate systems to have a standard basis for their reference location so that, for example, a Web-based eBook user and a **Nook** user can reference the same material and share their annotations. Such a standard is critical to allow the social interaction that is the basis for creating an annotation or reference — namely to share that annotation with someone else. Developing a system where these references or annotations can be shared across reading system platforms is the second critical component to creating an environment that supports social reading.

Among the leading actors in this space has been the **Open Annotation Collaboration** (*www.openannotation.org/*), a project also supported by the **Mellon Foundation** whose goal is to "facilitate the emergence of a Web and resource-centric interoperable annotation environment that allows leveraging annotations across the boundaries [of] annotation clients, annotation servers, and content collections; to demonstrate the utility of this environment; and to see widespread adoption of this environment." To date they have developed an ontology and an annotation data model. They also initiated a Request for Proposals, again with **Mellon Foundation** funding, earlier this year of prototypes and demonstrations of annotation sharing efforts (*http://www.openannotation. org/documents/openAnnotationRFP.pdf*) to generate interest in and facilitate the emergence of interoperable annotation environments.

Another significant player in this space has been the **International Digital Publishers Forum (IDPF)** (*www.idpf.org*), who developed the EPUB standard. During their Digital Book event (*http://idpf.org/news/digital-book-2011-presentation-slides-now-available*) at **Book Expo** in May, when the EPUB specification was officially finalized, there was discussion of annotation interoperability as an element of the standard. The working group involved in developing the standard is exploring a syntax for a text location within a file, based on some work initially done by Adobe.

NISO's grant from the **Mellon Foundation** will support two meetings that will be held in October 2011 to further explore these issues and focus on building from the work underway and from the planning meeting held in New York in May. The first meeting will take place on October 10th in Frankfurt, Germany, before the launch of the **Frankfurt Book Fair**. The second meeting will take place in San Francisco, California, on October 25th prior to the start of the **Books in Browsers Conference**. Each of these workshops will bring together players from the various eBook stakeholders to explore the issues, challenges, proposed syntax systems, and prototypes that exist or are being developed. The group will also contribute to a new work proposal under consideration within **NISO** to develop these systems as a standard in the information distribution community. Both of the meetings will be open to the community, however space will be limited. Contact the **NISO** office, if you're interested in participating to see if there is still space. Recordings of the meetings will be made available after the event for those unable to join us in person.

Hopefully, the community will be able to develop a common system for text location and sharing of annotations before all the players in the community develop their own systems that lock them into a proprietary solution. While such a solution is currently envisioned as replacing what readers can do now with print books, these readers will no doubt find creative new uses for these social reading capabilities. 🌰