Nicholas Homenda, Visiting Digital Projects Librarian, nhomenda@indiana.edu

Michelle Dalmau, Interim Head of Digital Collections Services, mdalmau@indiana.edu

Juliet L. Hardesty, Metadata Analyst, jlhardes@iu.edu

All the Wright Moves: Migrating Wright American Fiction for Streamlined TEI Publishing

Collections of electronic texts in academic libraries and the platforms that deliver them do not often age gracefully. Eventually, the need to "update the website" becomes painfully obvious, and with electronic texts this problem is magnified by the seemingly simple requests to "make the search work better," "toggle between text and page images," and other features that the majority of current electronic text websites now support. Academic libraries do not want to make each migration project akin to starting from scratch, yet it often feels that way with the daunting amount of development time needed to retool or reimagine a legacy website. Hope comes from the possibility of building upon, extending, or better yet, seamlessly integrating with more recently published electronic text websites. At this point, updating the text encoding is often inevitable, and even desirable, usually using the most recent version of the Text Encoding Initiative Guidelines for Electronic Text Encoding and Interchange (TEI). In theory, by using predictably-encoded texts, one can design and utilize delivery platforms to handle storing, indexing, and delivering all encoded texts across collections without needing intensive customization from project to project. However, while we have seen great strides in TEI infrastructures to streamline the publication of TEI-encoded texts as evidenced in the June 2013 special issue of the *Journal of the Text Encoding Initiative,* namely in the TAPAS, PhiloLogic4, and TextGrid/TEXTvre/DARIAH initiatives (Flanders & Hamlin; Hedges et al.; Allen, Gladstone & Whaling), we also see breakdowns in abstracting even something as seemingly straightforward as bibliographies encoded in TEI (Dombrowski & Denbo).

As proposed by Glorieux and Jolivet (2013), it is advantageous for TEI encoded text projects to design schemas and markup structure with simultaneous goals: faithful representation of the text in digital form, and deep, consistent markup that will allow for reliable access and presentation via an intended platform, or "processing expectations". Recent migration work on the *Wright American Fiction* project (http://dlib.indiana.edu/collections/wright/) at the Indiana University Libraries allowed us the opportunity to transform previously encoded texts into TEI P5 documents following a common core of encoding guidelines and electronic text publishing strategies originally derived as part of rapid parallel web development for three other electronic text projects, *Victorian Women Writers Project*, *Brevier Legistlative Reports*, and *Indiana Authors and Theirs Books*, as detailed by Dalmau, Hardesty, and Floyd (2012). This allowed for an improved user experience while minimizing the time and resources needed to create a customized delivery website for the *Wright* texts.

The Wright American Fiction project, conceived in 2000 collaboratively among numerous Committee on Institutional Cooperation (CIC) Libraries, digitized microfilm page images of the nearly 3000 works of fiction by American authors identified by Lyle H. Wright in his bibliography *American Fiction 1851-1875: A Contribution Toward a Bibliography.* The resulting digital files were converted to text via optical character recognition (OCR), encoded in P3 TEI in SGML, and displayed and indexed via the Digital Library Extension Service (DLXS) software. The Indiana University Libraries hosted the files and the

underlying platform, and were finally able in 2012 to migrate the *Wright* project to California Digital Library's open source eXtensible Text Framework (XTF) platform. Following a long history of implementing one-off, customized electronic text projects under the Digital Library Program, the IU Libraries have more recently shifted approaches. We now embark on new text encoding projects within a services-based model, utilizing a common core of software platforms, workflows, TEI schemas, and XSLT transformations. New projects and migration of older content are ideally launched with minimal customization, provided that the encoding can conform to a core set of guidelines.

Like those behind the TAPAS initiative, we do not necessarily see the "variability [of encoded texts] as a disadvantage to be overcome," but instead value the "set of research questions about the nature and purpose of the actual variability exhibited by the TEI data" not only for their unique readings and contributions to research, but also as a way to enhance or extend baseline publications of text collections (Flanders & Hamlin, para. 4). As stated above, we are also working toward supporting a core set of schemas and transformations, much like how TAPAS will provide a "target schema" to support a "basic publication interface" with a building-block approach so that more advanced markup practices can eventually be codified across scholarly projects for the creations of timelines, map-based browsing, and other advanced functionality (Flanders & Hamlin, para. 11). While we have not yet attained the sophistication of PhiloLogic4's TEI-compliant abstract data model and established parsing procedures (we are mostly still performing full-text or bibliographic queries only), the *Wright American Fiction* project brings us a step closer in that direction (Allen, Gladstone & Whaling, paras. 2-6). Further, it is important to note that increasingly open-source content management and publishing systems like XTF and Islandora (through the IslandLives project) are becoming even more keenly aware of the TEI, and how the encoding scheme shapes the editorial and publishing workflows and functionality of these systems (Stapelfeldt & Mosses). In fact, in recent posts (April 2014) to the XTF Users List, we see an active community of distributed XTF developers rallying to move the XTF codebase to GitHub (https://github.com/cdlib/xtf). Many of these developers are advocating for and implementing much needed improvements within the XTF framework in response to TEI projects (https://groups.google.com/d/topic/xtf-user/W9G-wVTIyr0).

This paper will present the achievements of the Wright American Fiction migration project, centered on improving the user experience by: (1) implementing a nearly seamless integration of scanned page images and encoded text rendered in HTML via XSL transformations, (2) augmenting approximately 1200 fully encoded texts with an additional 1800 minimally-encoded texts automatically generated through OCR processes, and (3) creating collection-wide XSL transformations that enhance the richness of encoding without losing any previous markup through programmatic logic. Additionally, due to limited project staff time and resources, this migration project was completed on a shoestring budget, requiring the migration of these texts to a new platform to be as efficient as possible. We will share how we managed these accomplishments by: (4) using only existing electronic text services software platforms, (5) minimizing the amount of editorial time spent on any particular text, (6) updating the encoding based on a pre-existing TEI P5 model with a common schema, and (7) making minimal and only critically necessary customizations to the XTF platform that would benefit future TEI-based projects to come. Finally, we will take a closer look at the XTF platform, and how we at Indiana University Libraries can contribute to the codebase the abstractions we have developed and perfected thus far.

Works Cited

Allen, Timothy, Gladstone, Clovis, and Richard Whaling. 2013. “PhiloLogic4: An Abstract TEI Query System.” *Journal of the Text Encoding Initiative* 5. http://jtei.revues.org/817.

Dalmau, Michelle, Floyd, Randall and Julie Hardesty. 2012. “Electronic Text Services, from Project to Portfolio Management.” Paper presented at 2012 Annual Conference and Members’ Meeting of the TEI Consortium, College Station, Texas, November 7-10. http://idhmc.tamu.edu/teiconference/program/papers/.

Dombrowski, Quinn and Seth Denbo. 2013. “TEI and Project Bamboo.” *Journal of the Text Encoding Initiative* 5. http://jtei.revues.org/787.

Flanders, Julia and Scott Hamlin. 2013. “TAPAS: Building a TEI Publishing and Repository Service.” *Journal of the Text Encoding Initiative* 5. http://jtei.revues.org/788.

Glorieux, Frédéric and Vincent Jolivet. 2013. “Documenter des ‘attentes applicatives’ (processing expectations).” Paper presented at 2013 Annual Conference and Members’ Meeting of the TEI Consortium, Rome, Italy, October 2–5. http://digilab2.let.uniroma1.it/teiconf2013/program/papers/abstracts-paper#C167.

Hedges, Mark et al. 2013. “TextGrid, TEXTvre, and DARIAH: Sustainability of Infrastructures for Textual Scholarship.” *Journal of the Text Encoding Initiative* 5. http://jtei.revues.org/774.

Stapelfeldt, Kirsta and Donald Moses. 2013. “Islandora and TEI: Current and Emerging Applications/Approaches.” *Journal of the Text Encoding Initiative* 5. http://jtei.revues.org/790.