Migrate, Publish, Repeat: TEI Journals in the Open Journal Systems Platform

Nicholas Homenda and Shayna Pekala

Electronic version
URL: http://journals.openedition.org/jtei/1746
DOI: 10.4000/jtei.1746
ISSN: 2162-5603

Publisher
TEI Consortium

Electronic reference

For this publication a Creative Commons Attribution 4.0 International license has been granted by the author(s) who retain full copyright.
1. Introduction

In recent years, many scholarly journals have sought to make their published content freely available. Publishing electronically helps by reducing overhead costs inherent in the print publication model. Academic libraries have become increasingly involved in facilitating the production and publication of open access electronic journals. Such efforts at the Indiana University (IU) Libraries were undertaken in parallel by two separate units: the former Digital Library Program (which now includes the Digital Collections Services department) and the Scholarly Communication department. In 2006, the Digital Library Program began a collaborative project with the Department of History to digitize and TEI-encode 102 years of the Indiana Magazine of History, a peer-reviewed scholarly journal that publishes quarterly issues focused on the history of Indiana and the Midwest. In 2007, the Libraries launched IUScholarWorks Repository, IU’s DSpace-based institutional repository supported by the Scholarly Communication department. A handful of journals leveraged this service to disseminate and preserve their articles, including The Medieval Review, an electronic book reviews journal that has published primarily via email since 2007.
1993 and also encodes its articles in TEI. In 2008, IUScholarWorks Journals emerged as a sister service dedicated to the publication of open access journals built on the Open Journal Systems (OJS) platform.

Developed by the Public Knowledge Project (PKP), OJS is a widely used, comprehensive open-source software platform that supports electronic journal publication, requiring only local server space and staff to maintain it. OJS supports two primary needs of the journal publication process: (1) the editorial workflow, including organizing submissions, managing peer review, and various stages of editing; and (2) providing access to journal content once it is published, whether it is open access or available for purchase. It allows articles to be published in a variety of formats, including PDF, HTML, and more recently XML. Using the XML Galley Plugin, a journal can upload an XSLT 1.0 file to transform XML article content in any markup schema into HTML on the fly for display on the site. In this manner, it is possible to use an XSLT file to convert TEI in OJS.

This paper aims to define potentially reusable workflows for publishing TEI-encoded electronic journals on the OJS software platform through two case studies. Existing research details considerations for using TEI for encoding electronic journal content, migrating journals to the OJS platform, and migrating TEI projects, but fails to document the use of OJS for publishing TEI-encoded journals. Besides documenting the migration process and publication workflow for two electronic open access journals at Indiana University, this paper will offer practical considerations for using OJS for TEI-encoded electronic journal publications.

2. Literature Review

Recent studies have examined the migration of digital projects, and themes from the literature run parallel to open access journal migration efforts at Indiana University. Pertaining to large-scale migration efforts, Stein and Thompson (2015) investigated the motivations behind institutions migrating digital asset management (DAM) systems, using survey findings to uncover common reasons for migration and desired functionalities in these systems (Stein and Thompson 2015). The Text Encoding Initiative Task Force on SGML to XML Migration issued guidance for tools and techniques to use when migrating resources encoded in SGML to XML (Bauman et al. 2002).
Concerning TEI projects at Indiana University, Dalmau, Hardesty, and Homenda presented about the challenges faced when performing large-scale TEI migrations while upgrading the underlying platform (Homenda, Dalmau, and Hardesty 2014).

Holmes and Romary examined the use of TEI to represent electronic journals. They identified the spectrum of file types used within electronic journal editorial and publication workflows, including DOC, PDF, HTML, and XML, and advocated for the development of a customization of the TEI schema specifically for encoding journal articles. They acknowledged the OJS editorial workflow but noted the platform’s limitations, including the need to produce and edit documents for publication on software outside of the system (Holmes and Romary 2011). At that time, OJS was a relatively new platform that had minimal XML support and no means of handling TEI files directly. The Journal of the Text Encoding Initiative has produced a highly constrained customization of TEI P5 (80 out of 600 elements) in order to be able to publish its own journal articles (Van den Branden and Holmes 2014). From the submitted TEI XML, the journal generates different outputs such as ODT for the editorial workflow, OpenEdition XML for the journals.openedition.org publication platform, and a PDF, which requires conformity of submissions to the standard. There is a jTEI template for the Oxygen XML editor that makes encoding within the schema easier.

A study by Dias, Delfino, and Silva investigated the feasibility of migrating electronic journals to OJS, and noted some of the advantages and shortcomings early in its life as an open source publishing platform (Dias, Delfino, and Silva 2007). Dalmau and Schlosser recognized that Indiana University’s IUScholarWorks Journals service began using OJS in a pilot project in 2008 after significant development of an alternate platform occurred. They outlined the history of the “Indiana Magazine of History,” the journal’s collaborative relationship with the IU Libraries, and the resulting accessible online index and the first iteration of its online open access platform (Dalmau and Schlosser 2010). Their work focused on developing a TEI model for the Indiana Magazine of History that featured issue-level encoding using the P4 version of the guidelines and independent header files for encoding article-level metadata.
3. Case Study A: Indiana Magazine of History

3.1 Project History

Beginning in 2006, the Indiana University Digital Library Program (DLP) began a project funded by a Library Services and Technology Act grant to digitize the back issues of the Indiana Magazine of History (IMH), encode the text using TEI, and deliver the page images and encoded text through an open access web application. Digitizing the journal entailed contracting an overseas vendor to scan the journal archives and deliver TIFF master files. Next, the DLP made JPEG, PDF, and OCR text derivative files from these TIFFs according to the content model for paged media in IU’s Fedora digital repository, where the derivative images were ingested. To prepare the journal for TEI encoding, optical character recognition (OCR) was used to produce plain-text files that were automatically placed into TEI files with bibliographic metadata in the header and page breaks inserted at the appropriate places. Back issues of IMH were encoded during the project years 2006 to 2008, and newly released issues continue to be encoded by IU Libraries interns and student employees seeking to gain experience with TEI. Issues are made open access with a two-year embargo.

The encoding workflow for this project was unique in that entire issues of the journal were encoded as single TEI documents with additional independent header files containing article-level metadata. Once issues and independent headers were initially encoded, they were subjected to quality control procedures involving markup inspection, additional validation, and visual inspection on a test server. After passing quality control, page image derivatives and TEI files were ingested into Fedora, and METS files were automatically generated for use by the DLP’s page-turning web application, METS Navigator. The TEI-encoded texts were then put into a custom eXtensible Text Framework (XTF) web application that provides indexing, searching, and browsing, and presents the encoded text rendered as HTML alongside page images (figure 1). Once the tasks of scanning and TEI encoding were completed, publishing an issue of IMH took an additional couple of weeks because of the need for software developers to interact with the various components that made up the web application.
Figure 1. The complex workflow steps for encoding an issue of *Indiana Magazine of History* in TEI and publishing it in XTF.

**Indiana Magazine of History XTF Workflow**

1. Entire issue provided as TIFF images
2. Image JPEG derivatives created along with METS file
3. Images and METS ingested into Fedora
4. METS Navigator within XTF application delivers page images
5. TIFF files' OCR is encoded at issue level in TEI
6. Article metadata encoded in independent TEI headers and all TEI ingested into Fedora
7. XSLT on XTF platform delivers article text rendered as HTML

3.2 System Limitations

Although the various systems that comprised the *IMH* online portal functioned stably and reliably, the amount of developer intervention required made it difficult to make new issues available quickly and even more difficult to make minor post-publication corrections. The underlying XTF platform natively supports XML collections and contains default stylesheets for TEI transformations; however, the level of customization needed to support the *IMH* impeded quick publication and indexing by search engines. A single incorrect page number in a page break element, bibliographic metadata field, or METS file would result in the entire article’s text and page images becoming inaccessible. Since these files were stored in the Fedora repository with no web-based user interface, changes to any of the numerous file types associated with each article would require a software developer to manually replace files. In addition to the complexity associated with several interlocking systems, the online portal of *IMH* also suffered from poor Google indexing: no individual articles appeared in Google search results, a problem that would require completely re-architecting the site to fix.
3.3 Journal Migration

Beginning in 2013, the Indiana University Libraries’ Digital Collections Services (DCS) department began investigating the feasibility of migrating IMH to OJS. DCS interns analyzed the existing XTF publishing platform, noted gaps and weaknesses, and investigated the functionality available using the OJS platform. DCS broached the idea of a migration with IMH staff and they indicated interest in transitioning to OJS to disseminate their open access issues, and possibly in exploring the use of OJS’s editorial workflow tools in the future.

DCS and Scholarly Communication staff worked together to execute the migration of 7,000 IMH articles and to transition the journal’s operational support from DCS to the Scholarly Communication department. During this process, the P4 TEI encoding was left unchanged; however, an article-level version of the encoding is required by OJS. Fortunately, the previous IMH platform contained a TEI article export function, which was used to obtain all IMH articles for the migration. This function took issue-level TEI files and automatically converted them to the article-level TEI files required by OJS, using encoded text from the issue TEI and article metadata from the independent header files developed by Dalmau and Schlosser during the project’s initial iteration. This function had been implemented within the XTF publishing platform to provide readers the option to download single TEI files for research or analysis purposes.

3.4 Workflow Solutions

In August 2014, a new workflow for publishing the IMH online was implemented using OJS. The IMH editors decided to continue conducting their peer review, editing, and typesetting processes outside of OJS. Once they release a new issue in print, they transfer PDF copies of the final, formatted files for each article to the library via IU Box. A Libraries intern or student employee copies and pastes the text of each article from the PDF into an article-level TEI template. They encode the article according to the local implementation guidelines established by the project in 2009 that were recently updated to reflect changes in encoding practices due to the journal’s migration. Most significantly, they no longer require the encoding of geographic places and personal names. Although this level of encoding was a key feature of the IMH project, the originally
proposed software features intended to leverage this encoding depth were never developed. Moreover, encoding these two components took the majority of the time spent on each issue, which further slowed down the publication timeline.

Once the initial TEI encoding is complete, Libraries staff check each encoded file for quality and return it to the intern or student employee with any necessary changes. This process is repeated until the file contains no encoding errors. Libraries staff run a prepared XQuery file on the set of encoded files to read metadata expediently from all of the XML files and to compose the single XML import file with the issue and article bibliographic metadata required by OJS. The TEI article files are stored on a local server in a location that is referenced by the import file. Next, Libraries staff upload the import file into a development instance of OJS using the “Articles & Issues” XML plugin. An XSLT file that is stored in the XML Galley plugin renders the TEI files automatically on the fly as HTML (figure 2). Once the file is successfully imported into the development instance of OJS, it is saved locally for two years after the print publication date (in accordance with the journal’s two-year embargo period) when it is imported into the production instance of OJS and the content becomes open access.

Figure 2. The workflow steps for encoding an “Indiana Magazine of History” issue in TEI and publishing it directly in OJS.
4. Case Study B. The Medieval Review

4.1 Project History

The Medieval Review (TMR), formerly The Bryn Mawr Medieval Review, is one of the oldest electronic journals in existence. It was published at the University of Washington from 1993 to 1995 and at Western Michigan University from 1995 to 2007. In 2007, the journal moved to Indiana University, where it has been published ever since. The journal has had a long history of encoding articles in TEI, using P3 encoding from 1993 to 2009. The encoding standard was updated to P4 in 2009 using an article-level model originally developed by the Digital Library program for the Indiana Magazine of History’s TEI article export function. At that time, however, the existing P3-encoded articles were not migrated to P4, which resulted in a corpus of TEI documents in both P3 and P4, making error correction and corpus-wide transformations problematic.

The publishing workflow for TMR relied on the use of a FileMaker Pro database built by the journal. The journal’s editorial assistant entered review articles into the database, which was also used to handle various aspects of the editorial workflow, including tracking books available for review, accepting incoming reviews, assigning editors, managing the revision process, publishing individual reviews via email, and exporting review articles as TEI XML. The editorial assistant placed exported files from the database into a drop box on a server managed by the Libraries. A script was then run each night to transform the files via XSLT and deposit them as HTML files into the IUScholarWorks institutional repository, which uses the DSpace open-source repository software. The script also extracted metadata from the files to complete each deposit (figure 3).
4.2 System Limitations

DSpace was developed to serve as an institutional repository, which Clifford Lynch (2003, 2) describes as “… a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.” Although some institutions have adapted DSpace to deliver digital collections, the software is not natively meant to support this: files stored in the repository are not viewable within the system and must be downloaded by users, and the visual styling of the repository is relatively inflexible across collections stored within.

The process for publishing TMR TEI was error-prone since incorrect encoding of metadata and text could result in skewed deposits in DSpace. Moreover, making updates to the XML files required developer intervention and an uncomfortably elastic approach to deleting and re-depositing that is normally not used in a secure institutional repository. Furthermore, the journal had little to no control over the appearance and functionality of the online portal or their articles, the latter of which were presented as downloadable HTML files with no styling and essentially appeared as plain text.
4.3 Journal Migration

In 2014, discussions with TMR editors about the viability of migrating to OJS were sparked by a slew of problems the editors experienced in trying to correct errors in previously published articles. Although OJS provides an “Objects for Review” plugin (formerly the “Books for Review” plugin), it is not comprehensive enough to manage TMR’s complex editorial workflow. Therefore, TMR editors were mainly interested in the features of OJS that provided a mechanism for publishing XML, enhanced indexability, and the ability to customize the appearance of the website. Before the journal accepted the terms of a migration to OJS, Libraries staff made prototype versions of TMR in OJS and imported articles and issues for journal staff to examine (as was done with IMH).

For TMR, which had nearly 4,000 articles at the time, DCS staff decided to simplify the journal’s setup in OJS by migrating both P3 and P4 articles to the current P5 markup in order to achieve consistency across the corpus. Had the encoding migration not been performed, the XSLT within the OJS XML Galley Plugin would have had to account for two different markup versions, and the P3 portion of the corpus would have had to be converted from SGML to XML. In addition, because the quality of the encoding was quite low in certain places, an item-by-item inspection was already necessary in order to correct significant metadata and encoding errors. These corrections were ultimately incorporated into the migration process to P5.

4.4 Workflow Solutions

TMR began publishing in OJS using a new workflow in June 2015. Libraries staff worked with TMR’s editorial staff to update their FileMaker Pro database export templates that conform and validate according to the TEI P5 Guidelines. Like the IMH editors, TMR editors decided to continue conducting their review, editing, and email distribution processes outside of OJS. Now, when a new review is ready for online publication, a TMR editorial assistant generates the TEI file from the FileMaker Pro database and enters the metadata for the article in the “Quick Submit” plugin in OJS (figure 4). This plugin allows journal staff to upload and quickly publish a single article within the system without needing to utilize the system’s editorial workflow steps. TMR staff use this plugin since they review and prepare their review articles within their own local database and need OJS to immediately publish the final versions of these articles without further editorial steps. Because the “Quick Submit” plugin is not currently compatible with XML files, it is not possible to upload...
the article file at this stage; the editorial assistant must edit the article after it has been created in OJS and attach the corresponding TEI file. The TEI file is then automatically rendered as HTML by an XSLT file (based on the one created for IMH) that is stored in the “XML Galley” plugin.

Figure 4. The workflow steps for encoding a single Medieval Review article in TEI and publishing it directly in OJS.

5. Discussion

Common concerns about reliability, discoverability, and lack of control prompted the migration of these two journals to OJS. The custom platforms and workflows developed for IMH and TMR required developer time to perform routine tasks and were prone to breaking. Moreover, having a custom-coded solution makes future maintenance more difficult when staffing changes occur, a problem slightly mitigated but not solved by extensive project documentation. These platforms also employed less-than-adequate indexing architecture resulting in low discoverability by users outside of IU. Finally, while many scholarly journals lack the staff necessary to design and manage a complex online publishing system, they still hope for a certain level of control over their publishing platform, and they desire the ability to change the appearance of the site, correct errors in markup and metadata, and perform basic publishing actions with minimal overhead. The custom platforms for IMH and TMR did not allow this control.
Both journals have benefited from using OJS as a publication platform. For IMH, the time to publication has been significantly reduced; after the TEI files are reviewed for quality, it takes just five minutes to publish a new issue online, a process which used to take weeks. Moving IMH to OJS also improved the visibility of its content; the site is now indexed by Google and individual articles are indexed by Google Scholar. From the Libraries’ perspective, moving IMH under the umbrella of IUScholarWorks Journals has helped to consolidate library journal publishing activities under one roof. Similarly, moving TMR from DSpace to OJS helped to consolidate all library-published open access journals within a single platform, making it easier to manage. This benefits journal staff, who now can receive a consistent level of publishing and technical support. Finally, OJS enables both IMH and TMR editors to have a greater amount of control over the appearance of their respective platforms. In the past, editors have had to ask library staff to make even the smallest of changes, which they can now accomplish themselves by logging in directly to OJS.

6. Recommendations and Future Directions

These two projects illustrate that while using XML to publish electronic journals is hardly a new idea, the technical complexity involved in designing highly customized publishing platforms makes them difficult and expensive to maintain. TEI’s flexibility makes it a good option for encoding journal content, but efficiently delivering this content requires an easy-to-use publication platform with minimal need for advanced software development skills and developer time.

OJS may be a solution for active TEI-encoded serial publications in need of a new publication platform. As an existing tool with a large community of support, OJS eliminates the need to build and maintain a boutique publication platform. While it is specifically designed to accommodate journals, its use can easily be extended to other types of serial publications, such as monographic series and conference proceedings. Furthermore, it provides a user interface that enables non-technically savvy editorial staff to easily update the content of their journal website. OJS, however, has its limitations for publishing TEI projects: it cannot interact with the TEI in any other way than simply transforming it to a different output format. For example, it is not possible to create
browse or search facets based on the encoding without making significant alterations to the core OJS code. These kinds of changes are not desirable, as they would make the process of upgrading the software much more challenging.

While OJS has demonstrated itself to be a flexible, easy-to-use publication platform for TEI-based journals, there are several considerations to be made before adopting it locally. First, journal staff must decide on a workflow for generating the TEI and secure the staffing resources to implement it. Second, server space and the ability to install and maintain the software is needed. If batch uploading of article files is desired, additional, albeit minimal, server space is needed to serve as a temporary holding place to facilitate this process.

As demonstrated through the case studies of IMH and TMR, XML publishing in OJS is by no means seamless. Fortunately, the PKP team is actively developing improvements in this area. The organization is currently focused on converting Word documents to Journal Article Tag Suite (JATS) XML, which can subsequently be converted to a variety of other formats including HTML and PDF. The PKP team is also working to perfect the Open Typesetting Stack (previously the XML Parsing Service), a collection of different software libraries that perform various steps in the conversion process. An OJS plugin that submits documents to the Parsing Service is currently available on Github, but it has not yet been integrated into the main software code. PKP’s efforts towards parsing minimally-structured text into XML could enable additional encoding standards and schemas to be used more easily within the platform, especially because the project is open source. Whether journals prefer to use JATS, TEI, or some other standard to encode their content, OJS as a platform is unlikely to influence the adoption of one XML standard over another, since their plugins and workflows have been built with a high degree of flexibility and the potential for customization to support individual journal needs.

7. Conclusion

While TEI remains a viable option among XML journal publishing schemas, the technical challenges associated with managing the publication process are daunting to many journal publishers. At Indiana University, the successful migration of two TEI-based journals to the OJS platform has provided a model for supporting journals that have developed or are interested in developing an XML publishing workflow. These projects have also helped the Libraries to forge
closer working relationships among their DCS and Scholarly Communication departments, and to strengthen their collective expertise in the realm of digital publishing. The cases of IMH and TMR demonstrate OJS’s ability to facilitate the publication of TEI-encoded journals, both of which previously relied upon complex platforms and workflows that required a high level of technical skill to manage and maintain. More importantly, these examples open the door for current or future journals to consider using this platform with an XML- or TEI-encoding workflow.

BIBLIOGRAPHY


NOTES

7 Indiana University, Box @ IU, accessed January 8, 2018, https://uits.iu.edu/box.

ABSTRACT

The Indiana University (IU) Libraries have a long history of using the TEI markup standard to encode and publish electronic texts, but choosing the best publishing platform has been challenging for certain projects. Before formally launching an open access journal publishing program in 2008, the Libraries collaborated with
two scholarly journals to provide open access publishing using P3 SGML and P4 XML TEI encoding delivered through the DSpace and XTF platforms. Both journals used complex encoding, transformation, and delivery workflows that required copious amounts of custom software development to function properly. As these systems aged, the time and effort required to maintain them steadily increased. In 2013, the Libraries began planning to migrate these journals into the Open Journal Systems (OJS) platform while preserving the TEI markup.

Both journals are now publishing using the OJS platform. The Indiana Magazine of History was successfully launched in OJS in August 2014, and The Medieval Review was launched in June 2015. Publishing in this manner leverages the IU Libraries’ strengths in electronic text projects and XML workflows within an easy-to-use, flexible platform that journal editors appreciate. The success of these migrations presents a new framework for future XML publishing of open access journals at Indiana University.

INDEX

Keywords: Open Journal Systems, open access publishing, electronic journals, digital project migration, XML publishing

AUTHORS

NICHOLAS HOMENDA
Nicholas Homenda is the digital initiatives librarian at Indiana University Bloomington Libraries, where he manages digital projects, services, and initiatives in the Digital Collections Services department. Nick holds a Master of Science in Information Studies from the University of Texas at Austin and previously worked as a music librarian and an orchestral clarinetist.

SHAYNA PEKALA
Shayna Pekala is the discovery services librarian in the Library Information Technology Department at Georgetown University Library. Previously, she led the open access publishing program at Indiana University Bloomington Libraries as the scholarly communication librarian. Shayna earned an MLS with a specialization in digital libraries from Indiana University Bloomington and a BA in English from Duke University.