

Against the Grain

Manuscript 8385

Biz of Digital – Why Reinvent the Wheel? The Benefits of Updating Preexisting Digital Collections Metadata

Michelle Flinchbaugh

Leigh A. Martin

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



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

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.

Biz of Digital — Why Reinvent the Wheel? The Benefits of Updating Preexisting Digital Collections Metadata

by **Leigh A. Martin** (Resource Description Associate, Boatwright Memorial Library, University of Richmond, 261 Richmond Way, University of Richmond, VA 23173; Phone: 804-289-8808) <lbrosnih@richmond.edu>

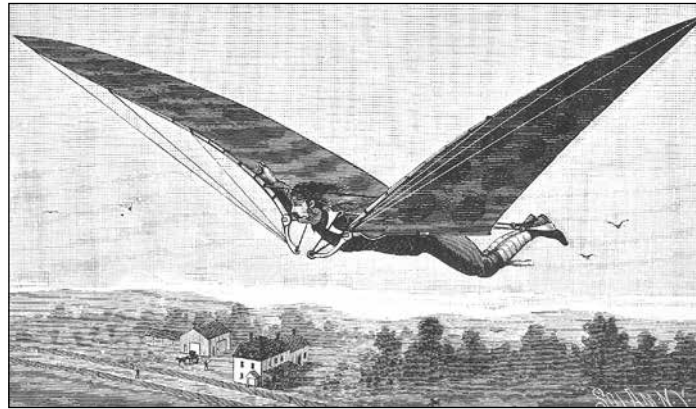
Column Editor: **Michelle Flinchbaugh** (Acquisitions and Digital Scholarship Services Librarian, Albin O. Kuhn Library & Gallery, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250; Phone: 410-455-6754; Fax: 410-455-1598) <flinchba@umbc.edu>

Boatwright Memorial Library is the main campus library for the **University of Richmond**, a private liberal arts institution serving just over 4,000 students, primarily undergraduates. Our digital collections, available at <https://richmond.access.preservica.com/>, are created and maintained by four members of our library's Digital Engagement Department and their seven student employees. Item-level metadata is produced in collaboration with the three-person Resource Description Department as items are digitized. Current ongoing joint digitization efforts include campus publications such as yearbooks, alumni magazines, student literary magazines, and honors theses. Accompanying this steady workload are various new requests, which range in complexity from the single-page scan for an academic course to the digitization of groupings of materials from our Rare Books and Special Collections Department for exhibition purposes.

Over the last year, we have added to this list of projects a comprehensive revision of all metadata for items digitized prior to our interdepartmental partnership. Each of these collections were developed as separate projects, and as such were inconsistent in their choice of metadata standards and descriptive depth. Though the process of updating the descriptive data for all these documents is an ongoing one, we have found it to have been extremely instructive. We hope that this account of our experiences might communicate the value in this task to other digital collections teams facing similar challenges.

Boatwright Memorial Library began its digitization efforts in 2003 with two simultaneous projects. One, the America at War collection, was an interdepartmental collaboration that digitized a body of World War II-era pamphlets and newsletters. The other was an Institute of Museum and Library Sciences grant funded project which digitized archival copies of the *Richmond Daily Dispatch*, a Civil War-era newspaper. As **Boatwright** did not yet have a department dedicated specifically to the creation and management of digital collections, these initial endeavors were developed by working groups with each collection's metadata standards determined by its respective project lead. Individual websites were created on a collection-by-collection ba-

sis; the first collections were made available using the DLXS software suite (<http://www.dlxs.org/>), and later instances made available using the open-source web publishing platform Omeka (<https://omeka.org/>). Digital Initiatives (now Engagement) was formed in 2006, but the project-based model remained for several subsequent self-contained collections as the new department focused on codifying the workflows for continuous campus resources.



In 2016, we adopted a digital preservation solution, Preservica (<https://preservica.com/>), for our ongoing storage and access needs. For our older collections, **Preservica's** WordPress-based front end, **Universal Access**, solved multiple problems, but in the process, illuminated others that would be beneficial to address in the future. Though the scope of each collection differed, many overlapped in their relation to our institutional or state history. By publishing this content through **Universal Access**, the collections would be for the first time available in a single location, rather than spread out over multiple unconnected websites, allowing users to easily browse related content. However, after nearly a decade of independently conceived projects, the resultant body of metadata was anything but unified, and this had consequences for our user experience.

None of the older collections used metadata formats natively supported by **Preservica**. The earliest collections had used Text Encoding Initiatives (TEI) standards, which were not amongst the formats supported out-of-box by the system, and two others, including our largest, 3,000 item correspondence collection, had employed custom, collection-specific XML schemas. Fortunately, we were able to ingest these projects in their original format thanks to Preservica's custom schema support, but this process required the creation and maintenance

of several Extensible Stylesheet Language Transformation (XSLT) documents, which was less than desirable.

In the end, though, the decision to align our older collections metadata with the standards we had newly established was primarily driven by a desire to better facilitate browsing and discovery. The use of multiple schemas in our repository hampered the functionality of **Universal Access's** faceted searching capabilities.

The facets were divided up by schema, such that selecting a facet checkbox or clicking on a subject from an item's record would only return results from works using the same metadata standard. With so many different schemas in play, this rendered our faceting functions too limited to be useful, thus spurring the decision to undertake a large scale metadata revision project alongside our ongoing workflows. At present, we continue to add new collections using the **Library of Congress's** MODS standard while simultaneously

updating the metadata from our previous projects to match. Updating this data, from its initial test phase to its current streamlined process, has assisted us greatly in refining a list of best practices for current and future collections.

As a first step in the metadata revision process, we evaluated the existing collections to identify a candidate for our first conversion project, separating out those records which would require extensive to full recataloging from those which simply utilized a nonstandard format. After reviewing all of our digitized materials, we decided upon our Civil War Sheet Music Collection, an exhibit comprised of 32 digitized objects with thorough descriptive data. The collection's detailed metadata included subject fields, which would allow us to confirm the functionality of the faceted browsing after conversion. This, along with its manageable size, made it an ideal candidate for a small-scale experiment.

The collection had been described using a custom schema, which meant that no data crosswalk would exist to easily transform it into MODS, so we converted the data manually using Oxygen XML Editor (<https://www.oxygenxml.com/>). This method was comparatively time consuming, but allowed the team to become more comfortable with the MODS standard through practical application. Though

continued on page 51

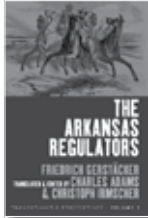
Essential reading from berghahn



MANAGEMENT BY SECLUSION

A Critique of World Bank Promises to End Global Poverty
Glynn Cochrane

"[This book] offers many significant insights regarding the World Bank, its institutional outlook, and [its] practices. The author, given his early involvement at the Bank, as well as his subsequent experience with NGOs, the private sector, and other organizations, is an ideal candidate to provide such an account." • A. Peter Castro, Syracuse University
May 2019



THE ARKANSAS REGULATORS

Friedrich Gerstäcker

Translated and edited by Charles Adams and Christoph Irmischer
"An accurate, full-length translation of this text has been needed for many years, and it is exciting that Friedrich Gerstäcker's work is being made available to an English readership." • Kathleen Condray, University of Arkansas
January 2019 • Transatlantic Perspectives

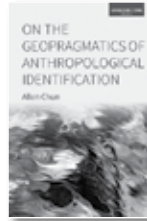


POLITICS OF SCALE

New Directions in Critical Heritage Studies

Edited by Tuuli Lähdesmäki, Suzie Thomas, and Yujie Zhu

Politics of Scale offers a global, multi- and interdisciplinary point of view to the scaled nature of heritage, and provides a theoretical discussion on scale as a social construct and a method in Critical Heritage Studies.
January 2019 • NEW SERIES: Explorations in Heritage Studies



ON THE GEOPRAGMATICS OF ANTHROPOLOGICAL IDENTIFICATION

Allen Chun

"Allen Chun's book is a wide-ranging, intelligent, critical and required manifesto for a reconfigured anthropology-cultural studies-social sciences." • John Hutnyk, Ton Duc Thang University
April 2019 • Loose Can(n)ons



REFUGEES WELCOME?

Difference and Diversity in a Changing Germany

Edited by Jan-Jonathan Bock and Sharon Macdonald

Combining in-depth anthropological studies with more long-term analyses, this volume examines the responses to and implications of the arrival in 2015 and 2016 of over one million asylum seekers and refugees in Germany – widely seen as the most major and contested social change in the country since reunification.
January 2019



PUBLIC ENGAGEMENT AND EDUCATION

Developing and Fostering Stewardship for an Archaeological Future

Edited by Katherine M. Erdman

Public Engagement and Education shares effective approaches for engaging and educating learners of all ages about archaeology and how one can encourage them to become stewards of the past.
February 2019



Follow us on Twitter: @BerghahnBooks
www.berghahnbooks.com

Biz of Digital from page 50

all were familiar with the schema through previous continuing education coursework, applying the standard to an active collection assisted us to further refine and codify our best practices into fully-realized documentation. We also determined during this process that working with small batches of files at a time helped to keep our workload manageable and trackable, and have utilized this approach with subsequent collections.

Once we had completed our test collection, we felt more comfortable taking a batch approach with the next group of records to be converted. For this, we used the TEI-encoded America at War collection. TEI and MODS are both widely-used encoding standards, so we were able to employ a modified version of an XSLT crosswalk available on the TEI Wiki (<https://wiki.tei-c.org/index.php/Crosswalks>), with some minor edits to better suit our data input preferences. Though this approach does not completely eradicate the need for human intervention (for example, the transformation cannot discern the difference between a personal or corporate name entry), it does drastically reduce the amount of data entry required to transform records between formats.

XSLT crosswalks have also been useful in enabling members of our team less comfortable

with processing raw XML to participate fully in the digital collections metadata process. For our catalogers, who are familiar with XML but find the interface of our Voyager integrated library system to be more comfortable and user-friendly, we use the built-in transformation capabilities of popular data cleanup tool MarcEdit (<https://marcedit.reeset.net/>). Our catalogers can then create the record in Voyager, export it into MarcEdit, and there transform that data into a fully-functional MODS record. Another popular option for transforming raw XML into a more user-friendly format is **OpenRefine** (<http://openrefine.org/>), which presents data in a format visually reminiscent of Microsoft Excel. Leveraging transformative tools such as these creates less of an initial barrier for participation, allowing a team to delegate tasks to the appropriate subject matter experts.

As we have worked through each collection, we have simultaneously assessed how they display to our users through **Universal Access**, improving our end-user experience in the process. Our digital collections are very diverse — we have collections of oral histories, pamphlets, letters, architectural diagrams, and sheet music — and each group of materials has been unique in its descriptive needs. In order to keep our digital metadata display consistent and appealing to our users, we have customized the XSLT transformations governing display in **Universal Access**. This has allowed us to

choose which elements are visible, hide unused fields, choose the order in which information is arranged, and modify the labels to match those used in our library's OPAC, giving our users a more visually consistent experience between the two platforms.

When this project was initially proposed, some of our team wondered if it was worth our effort to revise materials that had already been completed, regardless of quality. New digitization requests showed no sign of slowing down, and there was a perception that we would only be able to keep up with one or the other. A year later, we are more practiced, our procedures for both new and existing materials have been refined, and metadata transformation has become an established part of our workflow, one that is applicable to any body of metadata that requires batch alteration. We have completed the conversion process for the majority of our older collections, creating a repository of consistently applied metadata that better facilitates user research and browsing. With these results, none amongst us feel that the time spent was at all wasted. Instead, the process has helped us to develop the tools to manage metadata generation for our digital collections with greater speed, thoroughness, and understanding. 🌱