

D-Lib Magazine

July/August 2014

Volume 20, Number 7/8

[Table of Contents](#)

The SIMP Tool: Facilitating Digital Library, Metadata, and Preservation Workflow at the University of Utah's J. Willard Marriott Library

Anna Neatrour, Matt Brunsvik, Sean Buckner, Brian McBride and Jeremy Myntti
University of Utah J. Willard Marriott Library

Point of contact: Anna Neatrour (anna.neatrour@utah.edu)

doi:10.1045/july2014-neatrour

[Printer-friendly Version](#)

Abstract

This article presents a case study of the Submission Information Metadata Packaging (SIMP) tool developed at the University of Utah's J. Willard Marriott Library. The Library designed this platform-independent tool to facilitate the deposit of descriptive metadata and derivative formats into CONTENTdm, the Library's current digital asset management system. It also supports the submission of technical metadata and archival content into the Ex Libris Rosetta digital preservation system. The Marriott Library deliberately developed the SIMP tool to accommodate multiple workflows and ingestion processes in a modular fashion, which allows the Library to easily modify the tool to extend its functionality to other digital asset management and preservation systems or enterprise repositories.

Introduction

Academic libraries find themselves confronted with the challenge of adding digital preservation activities to their ongoing services¹ while continuing to provide access to digital collections through digital asset management systems (DAMS). The Marriott Library, which uses CONTENTdm as its DAMS, recently implemented Ex Libris Rosetta digital preservation system. In order to accommodate unrelated preservation and access systems, the Library was forced to revise its workflow for processing digital content. One major element of this revision was the Marriott Library's development of a platform-agnostic tool which would help relevant departments manage the digitization workflow, input and edit descriptive metadata, and package digital content for ingestion into disparate systems.

The objective of digital preservation at the Marriott Library is to "preserve and sustain long-term accessibility to all digital collections created or collected throughout the Library by maintaining a comprehensive digital preservation program."² In order to meet this goal, the Marriott Library began searching in 2012 for a scalable, flexible, and standards-based instrument with which to manage its sizable, ever-expanding digital archives. [Rosetta](#) was selected as the digital preservation system that would best meet the Marriott Library's needs. The multifaceted and complex Rosetta implementation process spurred a series of productive discussions among the Marriott Library departments that participate in the processes of digitization, metadata creation, and digital preservation. Representatives from each department collaboratively developed new workflow requirements and improved best practices, which led to a greater mutual understanding of interdependent tasks and an increase in standardization across departments.

Literature Review

The types, configurations, and implementations of digital library software, tools, and workflow management procedures vary in libraries. The University of North Texas Libraries use open source command line tools for digital curation tasks.³ In the first stage of metadata authoring, digital library staff may rely on .csv or .tsv files for batch upload into a DAMS. Protocols like Simple Web-service Offering Repository Deposit (SWORD) serve to facilitate workflows for depositing materials into repositories⁴, and many institutions use locally or consortially developed workflow management systems. At Rutgers University, the Workflow Management System supports metadata creation, item deposit, collection management, and more.⁵ The Capture, Ingest, and Checksum (CINCH) tool developed by the State Library of North Carolina is an example of a tool used to help libraries capture and preserve web-based files along with accompanying metadata.⁶ Harvard's [Batch Builder](#) tool supports the deposit of materials into their Digital Repository Service. The joint [Avalon Project](#) between Indiana University and Northwestern University provides a system for managing multimedia files, including descriptive metadata and deposit into preservation systems.

Managing both the display and preservation of digital content has created a number of different systems implementation models for digital libraries and digital archives. Some institutions may use [Fedora](#) for a preservation solution with Islandora available for access to public collections, as in the case of the University of Prince Edward Island's Island Archives Project. The LDS Church History Library utilizes two heavily customized and separate instances of Ex Libris Rosetta, one for preservation and one for display.⁷ The Kentucky State Archives uses DSpace for public access to digital collections and Preservica for their digital archive.⁸ The top priority for the Marriott Library was to develop a solution for depositing digital objects that would not necessarily be tied to any particular software platform, especially since the specific implementations of both its DAMS and digital preservation system could change over time.

Metadata plays an essential role in the discoverability of those digital assets that are managed in discovery and preservation systems. While it is recognized that the quality of metadata in digital collections is important, there are not always useful mechanisms for implementing quality control procedures within a DAMS.⁹ Catalogers in academic libraries have many of the skills needed to contribute quality metadata for digital library programs, including the knowledge of current standards and quality control methods to ensure high quality metadata.¹⁰ A current trend in many cataloging departments has been to refine workflows to better use the skills of catalogers in the design, implementation, and maintenance of digital library metadata. These workflows make it possible for catalogers to contribute in the initial planning stages of a digital project in order to provide guidance in creating the metadata templates to be used.¹¹ Common methods for ensuring high quality metadata include creating metadata guidelines, best practices, and application profiles that can be reused in multiple digital collections.¹² In a study of digital repositories, it was found that the time taken to invest in creating a metadata editing tool "would help to minimize the amount of customization each institution has to do in order to produce metadata that meets their requirements."¹³

Digital Library and Digital Preservation at the Marriott Library

The [Marriott Library](#) has used CONTENTdm as its DAMS for 14 years, and uses the software for multiple purposes to include the display of its digital collections as its [institutional repository](#), and for the [Utah Digital Newspapers](#) collection. Additionally, the Marriott Library collaborates in its use of CONTENTdm with the University of Utah's S. J. Quinney Law Library and Spencer S. Eccles Health Sciences Library. The Library's DAMS also serves as a regional hub for many digitization projects supporting external clients, including the Utah Division of State History, the Utah American Indian Digital Archive, Westminster College Library, Uintah County Library, and other cultural heritage institutions in the region.

Within the Marriott Library, digital library functions stretch across many departments but those most involved in the development and initial launch of the SIMP tool were Digital Operations (digitization and technical metadata), Cataloging & Metadata Services (descriptive metadata), and Digital Ventures/Digital Preservation (data management and preservation via Rosetta). The Application Development department developed the software and helped implement a user interface designed by the Discovery and Web Development department. Prior to the SIMP tool's development, metadata creation and data upload for the Library's digital content was usually done by submitting batches to the CONTENTdm Project Client with tabs-delimited text files generated through Excel and by using the CONTENTdm web interface to upload other small collections. Frequently, staff among departments would share metadata in Excel files via email, and while this informal system worked for many years, it was neither adequate nor efficient when faced with the growing complexity of the digital library environment.

The Marriott Library's [Digital Preservation](#) department is relatively new and was designed and developed by the Library's Digital Preservation Archivist in 2010. The department's first priority was to establish a digital preservation policy and mission, but in 2011 the department shifted its focus to developing and implementing a multi-stage plan for preserving the Library's robust digital collections. Successful in its initial stages of implementation, the Marriott Library is now advancing the program and its

new digital preservation system for expansion not only within the Library's divisions and departments, but also the University of Utah's various data-producing departments, colleges, schools, and institutes.

SIMP Tool Development and Testing Process

The workflow for processing digital collections in the Marriott Library is primarily shared among the three previously mentioned departments: Digital Operations, Cataloging & Metadata Services, and Digital Preservation. The principal administrators of these sections, in conjunction with a number of key contributors from other departments, formed a team charged with developing an interdepartmental workflow and accompanying tool for data packaging, enrichment, and delivery. Technical considerations and conceptual discussions regarding the functionality and integration of the SIMP tool focused on exactly which components of each digital object should be stored in either CONTENTdm or Rosetta.

After much deliberation, the SIMP development team decided that technical metadata would remain with the preservation master copies in Rosetta and that all descriptive metadata would be kept together with the access derivative copies in the DAMS. The team determined that storing the descriptive metadata exclusively in the DAMS was necessary because of the need to add and update metadata on a regular basis. Were the descriptive metadata to be housed in both systems simultaneously, the data could not be concurrently maintained and would inevitably become out of sync unless constantly updated. This scenario would have required the Library to develop a method of communication between the Rosetta and CONTENTdm systems; however, rather than attempt to develop a large-scale mechanism for pushing updates from one proprietary system to another, it became apparent that the most feasible solution to the dilemma would instead be to link the two records via a persistent identifier.

While CONTENTdm provides a reference URL for digital objects, in considering the Library's long-term commitment to its digital library and digital preservation program, [EZID](#) was chosen to provide Archival Resource Keys (ARKs) as persistent identifiers for the Library's digital collections. EZID offered a number of advantages such as a reasonable pricing structure, the ability to register both ARKs and DOIs, a clearly documented API, and the opportunity to store the ARK identifier information both with EZID and on a local backup database.

Throughout the SIMP development process, testing involved a variety of scenarios for processing and delivering content to both the digital library and digital preservation system. The development group repeatedly tested the application's capabilities and processing capacity by ingesting data packages containing single item and/or compound objects of varying sizes and formats, while constantly refining the functionality of the embedded metadata editor (online spreadsheet). System stress testing was also conducted on the SIMP tool by ingesting large multimedia files and collections of high resolution images. In conjunction with testing of the SIMP tool, the development group tested the system setup and configurations of the recently installed Rosetta system in a similar fashion.

Technical Details

The SIMP tool was developed with the goals of modularity, scalability, and cross-platform support utilizing the traditional LAMP (Linux, Apache, MySQL, PHP) stack in addition to other open source libraries and applications such as [ImageMagick](#), [Libav](#) tools, and [HandsonTable](#). The tool was also designed to integrate with the University of Utah's Central Administration Service (CAS) for user authentication. Modularity was a critical consideration in order to allow for expanding publishing functionality to platforms other than Rosetta and CONTENTdm, which will allow the Library the flexibility to continue to grow and innovate within the changing environment of its digital library and digital preservation systems.

Overview of the SIMP Tool

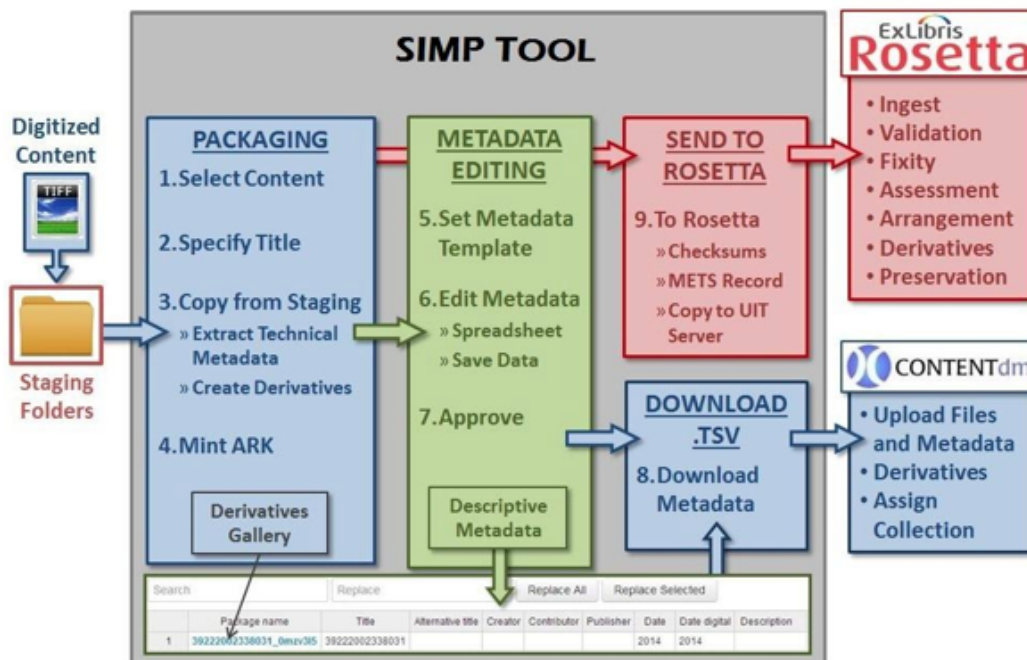


Figure 1: Diagram of SIMP tool workflow

An overview of the Library's current digital workflow utilizing the SIMP tool is depicted in the image above (Figure 1), with the blue-colored boxes representing the contributions of the Digital Operations department, green constituting the descriptive metadata work done primarily by Cataloging & Metadata Services, and red denoting Digital Preservation activities. General quality assurance and data management oversight throughout the workflow is also performed by the Digital Preservation department.

In the initial step, Digital Operations digitizes material and places the master copies of the content into one of a number of staging folders located on various library servers. Within the staging folders, directories are created and named to correspond with the collection titles of the digitized content. An additional sub-folder must subsequently be created, either manually or using a built in directory splitting application, for each [Intellectual Entity](#) (IE) therein as the SIMP tool was designed to package data exclusively at the IE level. An IE represents all content considered to be one complete accessioned unit such as an individual photograph, 10-page letter, three-part video, series of blueprints, or full book.

The screenshot shows the 'Browse Servers' tab in the SIMP tool. The breadcrumb path is '/mnt / POD1 / DigOps'. Below the path is a table with columns for 'Invert', 'Name', 'Size', and 'Type'. The table lists several folders, each with an unchecked checkbox in the 'Invert' column. At the bottom of the table area, there are two buttons: 'Create Packages' and 'Create Compound Packages'.

Invert	Name	Size	Type
<input type="checkbox"/>	American_Westward_Migration		
<input type="checkbox"/>	Rare_Books		
<input type="checkbox"/>	Univ_of_Utah_-_John_Mills_Whitaker_Papers		
<input type="checkbox"/>	Univ_of_Utah_-_Manuscripts		
<input type="checkbox"/>	Univ_of_Utah_-_Multimedia_Archives_Photos		
<input type="checkbox"/>	Univ_of_Utah_-_Undergrad_Research_Abstracts_Journal		
<input type="checkbox"/>	Univ_of_Utah_-_Utah_Humanities_Research_Foundation		
<input type="checkbox"/>	Univ_of_Utah_Richard_Wright_Young_Papers		
<input type="checkbox"/>	University_of_Utah_-_Abravanel_Studio_Collection		
<input type="checkbox"/>	Utah_State_Historical_Society_-_Classified_Photos		
<input type="checkbox"/>	Utah_State_Historical_Society_-_Fine_Arts		
<input type="checkbox"/>	Utah_State_Historical_Society_-_Shieler_Commercial_Photos		

Figure 2: Screenshot of SIMP tool: Browse Servers tab

Once the data has been placed into the staging folders, Digital Operations staff can browse the folders using the SIMP "Browse Servers" tab (Figure 2) and select all IEs that are ready for packaging.

The screenshot shows the 'Browse Servers' tab in the SIMP tool. The breadcrumb path is '/mnt / POD1 / DigOps'. Below the path is a table with columns for 'Invert', 'Name', 'Size', and 'Type'. The table lists several folders, each with a checkbox in the 'Invert' column. The checkboxes for folders 0044_14_01 through 0044_14_07 are checked. At the bottom of the table area, there are two buttons: 'Create Packages' and 'Create Compound Packages'.

Invert	Name	Size	Type
<input type="checkbox"/>	0044_13_06		
<input type="checkbox"/>	0044_13_07		
<input type="checkbox"/>	0044_13_08		
<input type="checkbox"/>	0044_13_09		
<input checked="" type="checkbox"/>	0044_14_01		
<input checked="" type="checkbox"/>	0044_14_02		
<input checked="" type="checkbox"/>	0044_14_03		
<input checked="" type="checkbox"/>	0044_14_04		
<input checked="" type="checkbox"/>	0044_14_05		
<input checked="" type="checkbox"/>	0044_14_06		
<input checked="" type="checkbox"/>	0044_14_07		

Figure 3: Screenshot of SIMP tool: Create Packages

With the content selected and still within the "Browse Servers" tab, each IE can then be packaged (Figure 3). The package title is assigned by default as the IE-level directory name unless it is manually altered, and then the package is copied to the SIMP tool server space. During data movement, the SIMP tool places the packages in a queue in order to extract technical metadata and create derivative copies, which are placed into a gallery later viewable from the metadata editor or "Assess Packages" tab.

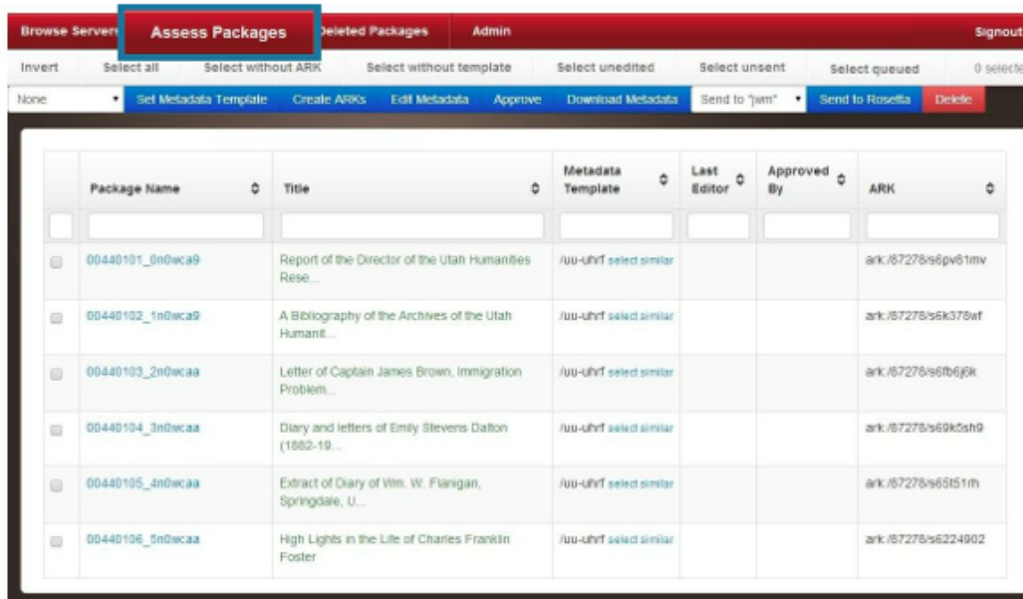


Figure 4: Screenshot of SIMP tool: Assess Packages tab

Once packaged and processed, each IE can be viewed and, if selected, can be worked on within the "Assess Packages" tab (Figure 4). Within this tab, the ARKs are minted and the corresponding metadata template is assigned from a drop down menu of preconfigured collection templates (via the [CONTENTdm API](#)). Additionally, the tab displays the current status of each IE, who uploaded and/or last edited it, who may have checked out or locked it for editing, and its approval status, among other information. Search boxes and options for selecting groups of IEs are also provided to facilitate the user's work with large numbers of items.

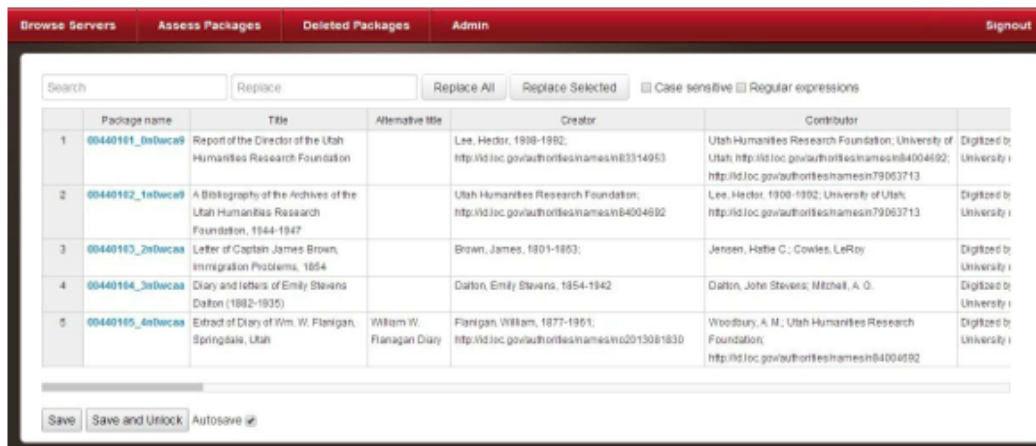


Figure 5: Screenshot of SIMP tool: Metadata Editor spreadsheet

At this point, the workflow passes from Digital Operations to Cataloging & Metadata Services staff who edits the data, filling in the inline spreadsheet with descriptive metadata for each IE. The SIMP tool's metadata editor possesses many of the typical spreadsheet capabilities such as sort, find and replace, fill up and fill down, and other basic functions. In addition, the initial column contains a link to the corresponding digital image or audio/visual track in the derivatives gallery, allowing the cataloguer easy access to the items, which helps to synchronize the metadata with the appropriate content and reduce human error. As the metadata entry process can be lengthy, taking days or even weeks for larger batches, all IEs being worked on by an individual metadata cataloguer can be indefinitely locked and their work continually saved. Once all of the descriptive metadata has been

input to the spreadsheet, the IEs can be unlocked, reviewed, and approved by an administrator. With that approval, Digital Operations staff is then able to download a .tsv file containing all of the metadata and file information needed for upload, along with the corresponding derivative access copies of the content, to the CONTENTdm Project Client from the "Assess Packages" page (Figure 4).

The final step involves the Digital Preservation department sending the master copies of the fully processed IEs to Rosetta. Upon executing this process, the SIMP tool queues the selected IEs and generates a Metadata Encoding and Transmission Standard (METS) record that includes checksums, file titles, and some structural and technical metadata. The descriptive metadata is excluded for reasons previously mentioned. The master copies and METS record of each IE are packaged in accordance with the OAIS model to create a SIP (Submission Information Package), which is then copied over to the Library's server space, where the Rosetta program resides, for subsequent ingest and preservation. This final packaging process was designed to support future change as the .xml generated by the SIMP tool can be formatted to any platform and easily adapted to different systems as the digital library and digital preservation programs continue to evolve.

Reviewing SIMP Tool Implementation and Future Directions

The SIMP tool has been actively used for more than four months by those Marriott Library departments that consulted on its development, namely Digital Operations, Cataloging & Metadata Services, and Digital Preservation. A committee has recently been formed to oversee training and possible enhancement requests for the SIMP tool as it is gradually rolled out to additional departments within the Library that are engaged in collection digitization, descriptive metadata assignment, and/or digital preservation activities.

For Digital Operations, the implementation of the SIMP tool inserted another step to the workflow process when adding to the digital library in CONTENTdm. This added step slowed the progress of submitting content to the digital library, but support activities for Digital Operations were also expanded to accommodate both digital library ingestion and digital preservation activities. Efficiency and consistency increased with regard to the Library's archived files and descriptive metadata. Digital collection items gained truly unique identifiers, which facilitated searching for and retrieving copies of the high resolution and high definition files from the digital repository in Rosetta. Also, the descriptive metadata can now be compiled in a single spreadsheet location, along with the ARKs, before being uploaded to CONTENTdm, which was seldom the case with past projects utilizing the previous process. Furthermore, the SIMP tool currently allows departments to ingest into CONTENTdm as well as Rosetta via a single workflow.

The SIMP tool increased efficiency for descriptive metadata services. The workflow resulting from use of the SIMP tool makes it possible for digital objects to be uploaded to CONTENTdm with descriptive metadata already in place. Workflows used before the SIMP tool often required that the descriptive metadata be entered after the content had been uploaded to CONTENTdm, which meant that many digital objects were available to public display without any useful metadata. Since the step of metadata creation now precedes the ingestion process, all items are now searchable and complete from the moment they are uploaded. Another key advantage to creating the metadata within the SIMP tool is that all of the data is maintained and stored in a centralized database in real time rather than shared in multiple versions and copies of Excel spreadsheets, which could easily be altered, lost, or corrupted. In the online editor, if data becomes corrupted or is inadvertently lost or altered, it is relatively easy to retrieve previous versions of the data set.

With the SIMP tool's metadata editor, there are also easier and more reliable ways of reviewing and auditing the metadata before it is ingested into CONTENTdm. This makes it possible to more easily standardize data and produce higher-quality metadata. The process can often be automated more easily through the use of the fill up/down options in addition to tailored macros utilized within the tool. Projects are currently being explored to automate more of the metadata creation for certain types of objects. This includes re-using the metadata for electronic theses and dissertations that is provided by ProQuest and mapping certain pieces of metadata for collections where an EAD register exists.

Conclusion

The recently implemented SIMP tool has streamlined digitization, metadata creation, and digital preservation activities in the Marriott Library. Three different departments are now able to more efficiently coordinate their efforts through a single workflow utilizing the SIMP tool. The Library is better able to support the digital preservation program in addition to the ongoing digital library and descriptive metadata work. As the Library investigates potential systems and services for the future, the SIMP tool will remain a core element of its technical infrastructure. By developing a tool that supports descriptive metadata editing as well as the packaging of digital objects for long-term preservation, the Library will be less dependent on vendor-based solutions for digital library and digital preservation services. The Library will be able to more easily respond to the changing environment for digital library systems since the output for the workflow supported by the SIMP tool can be configured for a new DAMS.

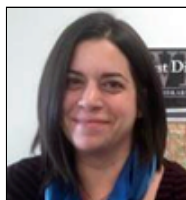
Acknowledgements

The SIMP tool was developed by the J. Willard Marriott Library's Application Development department. Alan Witkowski was the lead developer on the SIMP tool project with Curtis Mirci working on the EZID API integration. Leah Martin developed the user interface. A cross-departmental collaborative team from the Library contributed to the planning, development, and testing of the SIMP tool, including the authors of this paper, John Herbert from Digital Ventures, Tawnya Keller from Digital Preservation, and Kinza Masood from Digital Operations. The authors wish to thank Rebekah Cummings, Sarah LeMire, and Ken Rockwell for providing feedback on a draft of this paper.

Notes

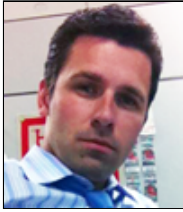
- ¹ Banach, M., & Yuan, L. (2011). Institutional repositories and digital preservation: Assessing current practices at research libraries. *D-Lib Magazine* 17, (5/6). <http://doi.org/10.1045/may2011-yuanli>
 - ² Keller, T. (2012). [University of Utah J. Willard Marriott Library Digital Preservation Program: Digital Preservation Policy](#).
 - ³ Weidner, A. J. & Alemneh, D. A. (2013). [Workflow tools for digital curation](#). *Code4Lib Journal*, 20.
 - ⁴ Lewis, S., de Castro, P., & Jones, R. (2012). SWORD: Facilitating deposit scenarios. *D-Lib Magazine*, 18, (1/2). <http://doi.org/10.1045/january2012-lewis>
 - ⁵ Agnew, G. & Yu, Y. (2007). [The Rutgers workflow management system: Migrating a digital object management utility to Open Source](#). *Code4Lib Journal*, 1.
 - ⁶ Rudersdorf, A. (2012). Digital preservation ingest can be a "CINCH". *Library Hi Tech* 30 (3), 449-456. <http://doi.org/10.1108/07378831211266591>
 - ⁷ Laxman, Rick. (2013). [Selecting, migrating, and preserving digital records](#), Best Practices Exchange 2013.
 - ⁸ Evans, M. (2013). [Its all about the metadata](#), Best Practices Exchange 2013.
 - ⁹ Park, J. R. & Tosaka, Y. (2010). Metadata quality control in digital repositories and collections: Criteria, semantics, and mechanisms. *Cataloging & Classification Quarterly*, 48(8). <http://doi.org/10.1080/01639374.2010.508711>
 - ¹⁰ Boydston, J. M. K. & Leysen, J.M. (2006). Observations on the catalogers' role in descriptive metadata creation in academic libraries. *Cataloging & Classification Quarterly*, 43(2). http://doi.org/10.1300/J104v43n02_02
 - ¹¹ Valentino, M. L., (2010). Integrating metadata creation into catalog workflow. *Cataloging & Classification Quarterly*, 48(6-7). <http://doi.org/10.1080/01639374.2010.496304>
 - ¹² Park, J. R. (2009). Metadata quality in digital repositories: A survey of the current state of the art. *Cataloging & Classification Quarterly*, 47(3-4). <http://doi.org/10.1080/01639370902737240>
 - ¹³ Chapman, J., Reynolds, D. & Shreeves, S. A., (2009) Repository metadata: Approaches and challenges. *Cataloging & Classification Quarterly* 47(3-4). <http://doi.org/10.1080/01639370902735020>
-

About the Authors



Anna Neatrou is the Digital Metadata Librarian for the Mountain West Digital Library. She also works with digital projects at the J. Willard Marriott Library and the Western Waters Digital Library. She received her MLIS from the University of Illinois Urbana-Champaign.

Matt Brunsvik is the Digital Operations Coordinator for the University of Utah J. Willard Marriott Library. He oversees the creation and management of the Digital Collections for the J. Willard Marriott Library's



Digital Library. He received his Bachelors in History from the University of Utah.



Sean Buckner is the Digital Preservation System Coordinator for the University of Utah's J. Willard Marriott Library. He helps to manage the Library's digital archives and administers the Rosetta digital preservation system. He received his MS in Information with specializations in Archives and Preservation from the University of Michigan-Ann Arbor.



Brian McBride is the Head of Application Development for the University of Utah's J. Willard Marriott Library. He actively leads a talented team of developers who create and maintain applications that enrich the lives of students, staff, and faculty.



Jeremy Myntti is the Head of Cataloging & Metadata Services for the University of Utah J. Willard Marriott Library. He is responsible for optimizing metadata creation for the library's physical and digital collections by leading a team of dedicated librarians, staff members, and student workers. He received his MLIS from the University of Alabama.

Copyright © 2014 Anna Neatrou, Matt Brunsvik, Sean Buckner, Brian McBride and Jeremy Myntti