

Archive Digitization Application (ADA)

Dedicated to Steven Puglia (1961-2013), Manager of Digital Conversion Services at the Library of Congress, a Digital Preservation Pioneer and friend, who supported, cheered and guided our development of ADA from 2009.



In this project¹ we partnered with local, state, and federal actors to digitize and exhibit special collections from the Macon Middle Georgia archives. As a result we have:

- Consolidated, converted and imported all indexes of the Macon Archives holdings
- Established the Macon Archives in-house scanning lab for a continuing process of scanning
- Built a Macon Archives website with access to publicly available holdings
- Assembled two online exhibitions of the scanned collections
- Launched an open source digitization workflow management software solution (ADA) which can be used by any Archive

A significant challenge to a community archive digitization effort is the diverse array of actors necessary to assure a sustainable solution. One institution will have the archival holdings, another will have infrastructure for imaging and storage, yet another will have access to low cost or volunteer labor. But, involving this diverse set of actors can threaten the carefully shielded order and authenticity of an archive. To assure central control and authenticity along with dispersed production roles, we delivered an open source solution for archive digitization workflow management.

We want to empower citizens with the knowledge that public history is a public good, and that interpretation of history is and should be a participatory act of the community itself.

¹ This project was made possible the generous support of the John S. and James L. Knight Foundation

The Middle Georgia Archives

The Middle Georgia Archives is a part of the Middle Georgia Regional Library System with its headquarters in Macon. The public library is at the same time a vibrant community center and in the morning when its doors open there are already people waiting in line to get in.

The Library offers free broadband internet access; space to explore and read everything from daily newspapers to old and rare books; numerous free courses for community members – generally focused on job search and computer skills development; a children's floor with a separate program designed for community youth; and also access to a wide collection of online books at the regional repository hub.

In the genealogical and archival department, the situation is a bit different. The majority of the holdings are in hard copy. These Archives preserve the records of the life of Middle Georgia: institutional, governmental, organizational, and business, but also the history of ordinary people.

The Macon Historical Society is very active, but their interaction with the Archive is a series of ad hoc one directional inquiries without a long term strategy on how to contribute in a more sustainable way. The head of the genealogical department, Muriel Jackson, is the memory keeper of the archives holdings with the honor and huge responsibility of being the only person who knows the paths within the holdings.

A traditional stop on the way to sunny Florida and home of Georgia's famous peach cobbler, Macon has a rich and textured history. It is a Civil War trail; it is stories of the painful path of the slavery tradition; it is the birth place to many famous people like the poet Sidney Lanier and many others;

One of the nice local surprises is the recently opened Macon Mall, which features overblown Macon Archive photos from the XIX and beginning of the XX century. The Coca Cola plant, famous baseball games, the YKK zipper manufacture, and early Macon Cherry Blossom Festivals are easier to relate to in connection with a modern mall structure and contemporary stores than they might be in the Archive itself.

At the Macon Archives, local history is buried in big bound and unbound newspaper collections from the Middle Georgia region; hidden in the boxes of donated private collections of photos and documents, and numerous collections of business, churches, associations, cemeteries, and architectural drawings.

Most of these collections are indexed in a very limited way, meaning the metadata are not unified and not sufficient, which is an obstacle for researchers and renders connections to other archives and repositories nearly impossible.

Moreover, the Archives continues to collect and to accept donated collections, so many recently added collections are not processed at all.

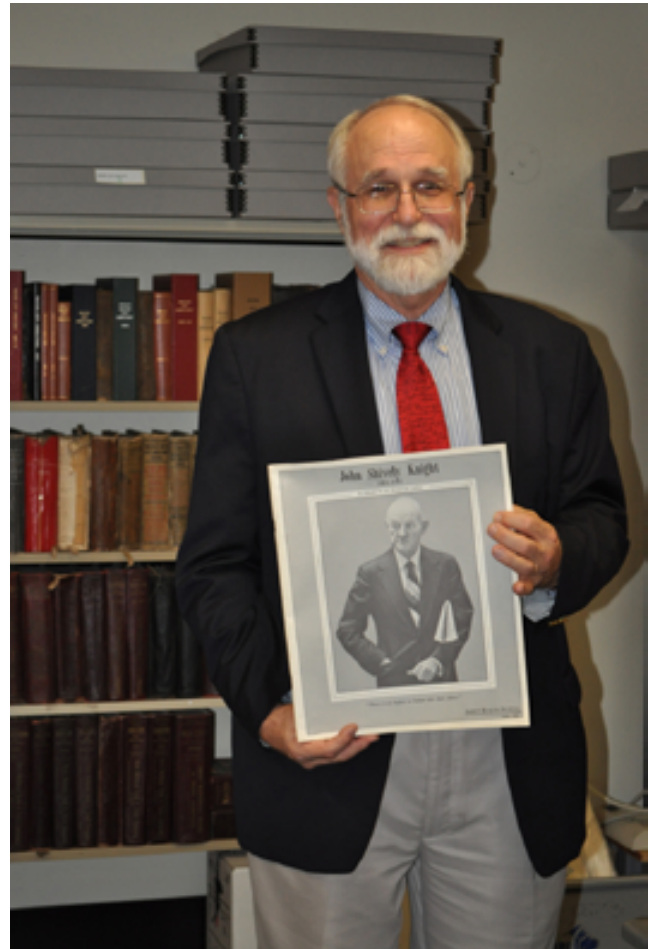


Figure 1. Dr Edward Papenfuss, MD State Archivist with the photo of John S. Knight at the Macon Archives October 12, 2012 Macon, GA

Identifying Digitization Objectives

The first objective was to apply interoperable indexing to metadata across all collections. The next biggest obstacle was a lack of in-house capabilities for scanning and very limited capabilities for metadata entry. Thus the second objective was to connect the already active community outreach programs of the library with the Archive digitization efforts and to engage active community members as a part of this process.

Finally, since there were not sufficient repository capabilities in house for digital content, our goal was to create a sustainable solution that can rely on existing external repository hubs for digital archives, such as the Digital Public Library of America.

ARCHIVE DIGITIZATION APPLICATION SOLUTION

Our solution was to start with consolidating all existing metadata files and combining them into one database, while at the same time applying commonly adopted Dublin core initiative standards.

This metadata task was a challenge because the files were in various and often out of date hardware and software technologies.

Capturing the unique Macon Archives needs made it necessary to give flexibility to Dublin core fields. We overcome this by using Dublin core fields as the basic fields for every archival unit, but with the possibility to add optional single line or descriptive fields where needed.

These are not challenges unique to Macon. Before we started working on the Macon digitization process and the public version of ADA, we led several large archive and library digitization processes where we used earlier versions of ADA. These included projects in the Military Archive of Serbia resulting in 3.5 million documents digitized; digitization of the photo archive of the Ministry of Defense of Serbia; digitization of the National Archive of Bosnia and Herzegovina; and the library digitization of the Military Academy of Serbia.

As a founding member organization of the US Library of Congress National Digital Stewardship Alliance, we also learned from other members of the archival community and connected the lessons of existing efforts with our own. Our goal was to establish a vivid solution that can be applied to small and medium sized archives, one that would be supported by a large pool of contributors, and integrate smoothly with already existing open source management tools for digital archives and repositories.

Simultaneous to our efforts with Macon, we worked closely with the Baltimore City Archive and Maryland State Archive to be sure our public ADA solution was able to capture needs of various archives structures and digitization the capabilities.

The ADA process is adaptable by design, to meet the inherently unique needs and conditions of individual archives, and is fully ready to incorporate digital-first materials produced by archives and other content owners. The ADA platform provides an advanced set of tools for management of a digitization process: to scan, store and back up archival documents in digital format and protect them from further physical handling, building a database containing fully cross-referenced metadata, and a highly-efficient search interface for data access.

Its workflows cover most typical use cases, including handling newly scanned documents, integrating previously scanned digital collections, with or without metadata, incorporating unprocessed archival collections, supervising crowd sourced metadata input, and more. A powerful search engine provides archive users with keyword, category and advanced search, and gives archive managers a multilayer overview of the archive structure, as well as user and archival statistics.

ADA can also serve as a temporary document repository, offering controls for metadata mapping into Dublin Core meta tags.

ADA is built as a Drupal 7 distribution package, containing the Drupal core, a number of contributed community modules, several modules specifically built for ADA, the ADA theme, and a pre-defined configuration. The standard ADA configuration can be quickly installed on any pHp 5.3x / Mysql 5.x enabled web server, following the common Drupal installation procedure. Additional advanced features, requiring the server level installation of third party open source software are supported by detailed documentation.

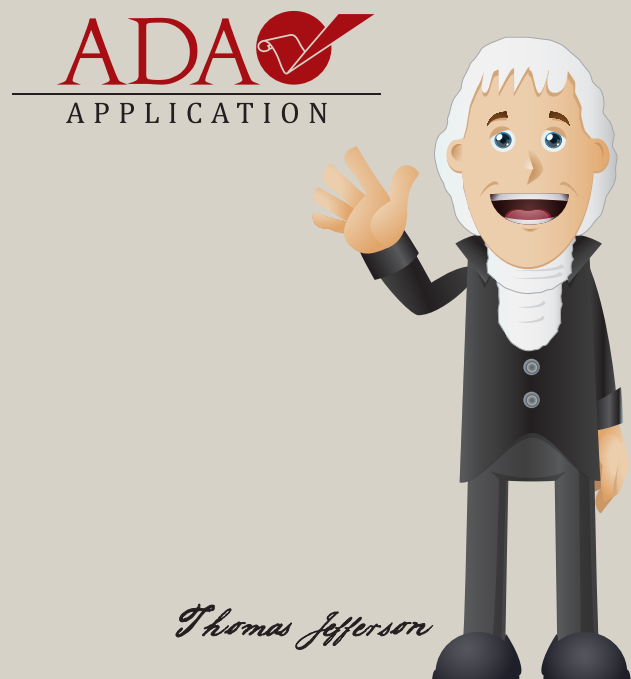


Figure 2. Archive Digitization Application logo

KEY FEATURES:

- separate workflows for scanning, metadata input and quality control
- correction loops for all segments of the process
- strict division of tasks between predefined roles
- detailed overviews and statistics for all segments of the process
- bulk metadata import
- flexible mapping into Dublin Core
- support for major digital document types and formats
- automated thumbnail extraction and watermarks
- supervised metadata crowd sourcing interface
- advanced faceted search
- enhanced user experience through personalized research interface

ADA predefines specialized interfaces for several platform specific roles:

- archive admin - management of the archive structure and users
- supervisor - digitization tasks distribution and quality control
- scan operator - scanning tasks
- metadata operator - metadata input tasks
- archivist - management of collection access permissions, researchers, researchers' requests and supervision of crowd sourced metadata input
- researcher - bookmarks, requests, comments and advanced search interface

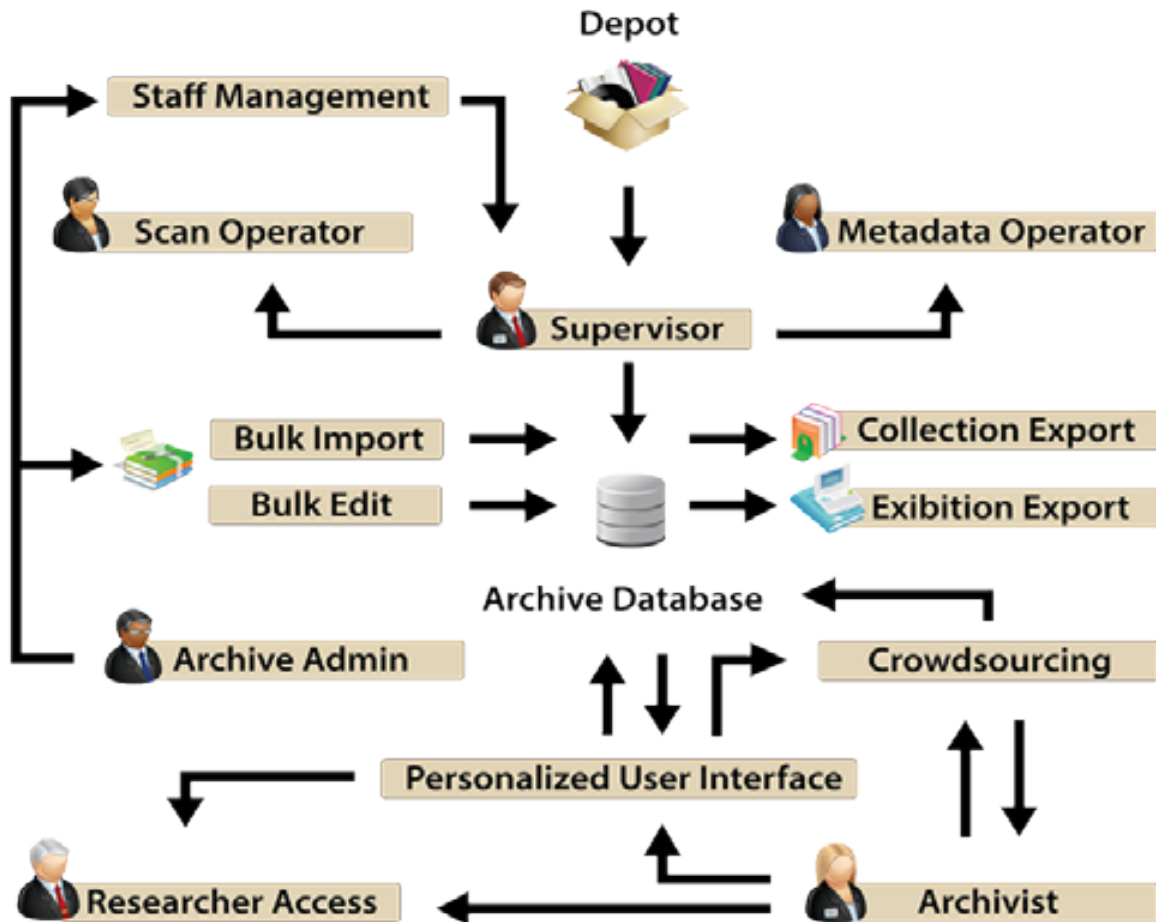


Figure 3. ADA Workflow

We also established a small scanning lab at the Archives enabling the personnel to invite library volunteers to join the digitization process with scanning and/or metadata entry with supervision of the Archive staff.

We designed and built a new website for the Macon Middle Georgia Genealogical Department of the Macon Middle Georgia Public Library, with ADA at its core. The researchers can access ADA and request a password, which they receive through email. They can view all collections and documents that the Macon Archivist assigns as public and the descriptions of some collections assigned as public overview collections.

Among many fascinating collections, two were selected for special online exhibitions. The Middle Georgia Archives hold an expansive collection of postcards from throughout Georgia. In the times without email and cell phones, a postcard to your family or friends, sent from a visited place opened a whole new world of communication. Once upon a

time, visitors to Macon sent greetings to their loved ones by post. The built exhibition layered these postcards, along with descriptive metadata, as pins across a map of Macon. Seeing how these greetings from everyday life long ago look today on the Google street view map serve as an engaging and playful online invitation to visit the Middle Georgia Archive and explore more.

The Middle Georgia Archives' stereographs collection was meant to be viewed using stereoscopes to entertain the public of the 19 century with the illusion of 3d scenes.

For a second on-line special exhibition, we created digital replicas of the stereographs to bring this magic alive. You can choose viewing options with or without viewing glasses to reanimate these vivid photos of old times. We presented only a glimpse of the huge set of stereograph holdings, but offered an interactive invitation for the public to visit, search and value the treasures of the community's past hidden at the Middle Georgia Archives.



Figure 4. 3d anaglyph from *The Middle Georgia Archives' stereographs collection*: 1. Central City Park - Editors Home; 2. Central Railroad Yard

TESTING ADA

Our cooperation with the Baltimore City Archive and the support of the Maryland State Archivist, Dr Edward Papenfuse, during the development of ADA and during its implementation in the Macon Middle Georgia Archives, lead us to testing ADA systems in another case of another archive type.

The Baltimore city archives has already established both a metadata system that corresponds seamlessly with commonly established archive standards and an expansive in-house scanning operation with a large number of overhead scanners and other equipment.

ADA served in this test case as an external tool for managing current digital exhibitions while integrating into the existing system of archive management.

Given that the Baltimore City Archive already has a number of digital documents ready for exhibitions, we used the opportunity to place the existing sets of Baltimore postcards and a selection of their holdings of stereographs and are preparing two similar exhibitions: Greetings from Baltimore and a Modern digital display of stereographs from Baltimore.

At the www.archivedigitization.org website, we have open for free download and use a distribution pack of the ADA application under a creative commons license. You are free to: Share — copy and redistribute the material in any medium or format, and Adapt — remix, transform, and build upon the material for Noncommercial purposes.

It comes with predefined roles for all key users in the system and is ready to be used for any archive solution.

Its basic search capabilities can be upgraded and tuned through the server administration preferences.

Ideally, this high profile pilot project for ADA will inspire and provide a foundation of tools for replication across city archives in the United States, and directly engage neighborhood communities in better understanding how they came to be and how they interrelate with one another in confronting the municipal challenges of the present.



Figure 5. ADA Platform ScreenShots

References

- [1] www.archivedigitization.org
[2] Presnall B. and Hrcic Z. "Digitization of the Military Archive of Serbia", Jefferson Institute, 2008
[3] Presnall B, " Digitization of the Archive of Bosnia and Herzegovina" , Jefferson Institute, 2012

Author Biography

Biljana Presnall is an online educator, author, future thinker and Vice President of Jefferson Institute. She and her team of developers and designers apply emerging technologies for use in democratization processes. She has over 15 years of development and management experience across the education, business and technical domains. She led several breakthrough ICT projects including: Digitization of the Military Archive of Serbia, Advanced Distributed Learning in MOD of Serbia, Development of the VIDI Data visualization Tool Set, and is author of numerous online courses, tutorials and books. She began her career working at the East West Institute in Prague, one of the leading think tanks in the world. Her educational background is Slavic languages with a MA in Czech Language and Literature. She is currently working on her doctorate on digital communication. She is married and has two children, and lives in Washington, DC.

