White Paper Report

Report ID: 104200

Application Number: HD5137811

Project Director: Jon Frey (freyjona@msu.edu)

Institution: Michigan State University

Reporting Period: 9/1/2011-12/31/2012

Report Due: 3/31/2013

Date Submitted: 4/2/2013

Final Report

Grant number HD-51378-11

Digging Digitally: Creating a More Dynamic Archaeological Field Journal Archive

Project Director:

Jon M Frey (MSU)

Project Participants:

Brian Adams (MSU)
Scott Schopieray (MSU)
Timothy E. Gregory (OSU)

March 31, 2013

Table of Contents:

I. Overview	**3**
II. Project Activities	**4**
III. Accomplishments	**8**
IV. Audiences	**9**
V. Evaluation	**10**
VI. Continuation of the Project	**10**
VII. Long Term Impact	**11**
VIII. Grant Products	**11**
IX. Appendix	**12**

I. Overview

The Archaeological Resource Cataloging System (ARCS) is an open source software solution designed by faculty, staff and student researchers at Michigan State University in order to provide low budget archaeological projects with a simple and inexpensive way to conduct research using digitized copies of documents contained in their traditional paper-based archives. While the initial stages of the development process began in early 2011 with the assistance of the Michigan State University College of Arts and Letters, it was the financial support of an NEH Digital Humanities Startup Grant that enabled ARCS to evolve into the innovative and intuitive web-based tool that it is today.

The experience of archaeologists and scholars working at the Ohio State University Excavations at Isthmia over the past few decades had shown that, while the development and adoption of digital tools for the collection and analysis of new archaeological data had become a widespread phenomenon, much less attention was given to the costly and difficult process of making pre-existing, non-digital archaeological information equally accessible and useful. What is more, for those archaeological projects that could afford to digitize their archival collections, the available services for the storage and dissemination of shared archaeological data did not provide a sufficient degree of flexibility to allow them to recreate the unique organizational structure of the recording system that documented the original process of archaeological discovery. As a result, projects were forced either to make their digitized data fit pre-determined systems or to engage in the costly process of developing their own electronic tools.

For those at Isthmia, the clearest example of these types of issues concerned the digitized copies of over fifty years worth of field journals. Because they represent the day-to-day record of an archaeologist's activities and ongoing interpretations in the field, these hand-written journals are irreplaceable and indispensible accounts of the archaeological process (fig. 1). What is more, at Isthmia and many other excavations, the field books are also the central unifying element for making sense of the various maps, drawings, reports, photographs, and object descriptions in the project archives (fig. 2). Yet it was difficult to enable these journals to continue to serve such a central role in a digital environment. On the one hand, a transcription failed to reflect the organic relationship of the words, emendations, photographs, drawings and even tracings of small artifacts that could be found on any page of a field book. On the other hand, a digital copy of a scanned notebook could not be searched by a computer nor linked to other forms of digitized archival records.

As such, the proposed project, entitled "Digging Digitally: Creating a More Dynamic Archaeological Field Journal Archive," aimed at achieving a small number of relatively simple goals. First, ARCS would be designed to display a digital copy of an archaeological field journal in a way that would allow users to

associated specific keywords with any single page, thereby allowing the document to be searched, yet retain its original appearance. Second, ARCS would allow users to form hyperlinks between those pages and any other digitized document in the archive or even other sites on the Internet, thus returning the notebook to its central place in the organizational structure of a digitized archive. ARCS would be made freely accessible over the Internet and would be compatible with a number of different hardware platforms and Digital Asset Management Systems, such as Resource Space (http://www.resourcespace.org/), Omeka (http://omeka.org/), and Kora (http://kora.matrix.msu.edu/). As the following record of development will show, the ARCS project has not only achieved nearly all of these limited goals, but also has grown into a program that holds great potential to serve as a powerful digitization and research collaboration tool for archaeologists who seek to employ the legacy data of an excavation in their interpretation of the past.

II. Project Activities

Initial development (August 2011 – January 2012)

Before the grant was awarded in August 2011, an initial version of ARCS had already been developed as a modification of the Resource Space Digital Asset Management System (http://www.resourcespace.org/) (fig. 3). Up to this point, this work had been carried out by a single programmer and it had become clear soon after submission of the grant proposal that a team of student programmers, designers and developers could more effectively address the different aspects of the project. The wisdom of this decision to change the makeup of the development team immediately become apparent when students, faculty and staff first met in September 2011 to discuss ARCS. The project team's efforts were first directed at improving the user interface and the functionality of the tools already offered in association with that system. In order to coordinate the efforts of a larger group, it was decided that changes to the source code would be tracked on github (https://github.com/calmsu/arcs) while conversations among the various members of the development team would take place on a collaborative developer blog (http://edblogs.cal.msu.edu/arcs/). A development version of the ARCS program would be maintained on a server at the College of Arts and Letters at Michigan State University, which could be accessed from any computer over the university's local area network or wireless network. The actual archival materials for any test of the ARCS system would come from the ongoing digitization project at the OSU Excavations at Isthmia (http://isthmia.osu.edu/). At the same time, developers would be careful to distinguish between the particular needs of the Isthmia project and the basic functions that could be utilized and modified by any archaeological project.

Due in large part to the input of several student members of the design team, the ARCS project soon began to evolve into a much more innovative and user-

friendly program (fig. 4). One of the first tasks involved the creation of a systematic vocabulary for the objects to be stored in ARCS (called "resources"), the information regarding those objects (called "info"), the search terms (called "keywords") and the objects to be linked to them (called "annotations"). Eventually, annotations and keywords came to be grouped under the general heading of "notations." Moreover, a transcription tool was created for cases in which users wished to make whole sections of text readable by a computer. Another improvement involved the association of a unique discussion board for each resource that allows individuals to hold an ongoing conversation regarding any aspect of the evidence it contains. Finally, a great deal of critical thought went into the best ways to search for and display resources in ARCS. The goal was to enable users to search resources not just by name, but also by type, name of their creators, dates of creation, etc. while at the same time encouraging creative play and serendipitous discovery. Wireframes for all of these utilities and displays were created using Basalmiq (http://www.balsamiq.com/) and became the foundation of the overall user experience that ARCS maintains today (fig. 5).

The most significant change made at this stage concerned the decision to employ a crowd source approach to the creation and augmentation of digital resources. The development team soon recognized that because archaeological archives contain thousands of documents that must be digitized, uploaded and properly augmented with keywords and metadata, the amount of effort that would be required to create a functional ARCS system would quickly overwhelm even a small group of dedicated archivists. Yet, if carefully supervised, a much larger group of interested individuals might achieve the same goal with less effort and in a shorter amount of time.

However, a crowd source approach required the development of a relatively complex system of access, tracking and control, in which users monitor one another's efforts, but only certain individuals have the privileges to make changes to the system and the information it contains. Thus, a new administrative architecture was created for ARCS that allowed both junior level users to upload new information, augment resources and report errors and senior level users to track individual use and resolve errors (fig. 6). Again, wireframes were developed for these features in Basalmiq.

Having completed many of these tasks by December 2011, graphic designers on the ARCS team then developed a more user-friendly interface based on the familiar look of current computer operating and file navigation systems. The overall look of ARCS was created starting with Twitter Bootstrap (http://twitter.github.com/bootstrap/) and further modified with custom components.

Database restructuring (January - February 2012)

Because it had developed into a much more robust suite of utilities and user controls than originally proposed, it became evident that ARCS was growing increasingly incompatible with the Resource Space system. Essentially, the programmers found themselves writing two different scripts—the first gave ARCS its functionality and interface while the second was responsible for translating the ARCS system over to the Resource Space code base. This double task not only increased the amount of time programmers spent on writing and debugging code, but also affected the speed and reliability of the ARCS system.

Thus, in January of 2012, the decision was made to develop an entirely new database system that would better meet the needs of the development team (fig. 7). This resulting code base is client-side JavaScript, written in CoffeeScript and built on open source libraries like Backbone.js, jQuery and jQueryUI. The programmers took advantage of modern browser technology to provide features like multi-file, asynchronous file uploads and in-page URL updates. On the server-side, a CakePHP application provides a RESTful JSON-based interface, renders HTML templates not rendered client-side, and handles image and document processing. Data is currently persisted in MySQL with plans for Apache SOLR search integration.

This significant change brought with it two important shortcomings that have yet to be fully addressed. First, in disassociating the ARCS program from Resource Space, the project team could no longer claim that ARCS was compatible with existing digital asset management systems. Secondly, because of the increasing complexity of the ARCS system and the limited resources for the project, the programmers eventually were forced to abandon the goal of making ARCS compatible with iOS technology as is found on Apple devices such as the iPhone and iPad. iOS devices are still able to display ARCS as the code base is PHP and HTML, but some functionality of the system is lost on a tablet computer.

Debugging and Beta Version Rollout (February – June 2012)

In February 2012, the first versions of ARCS using a limited subset of digitized records from the Isthmia Excavation archives were made available online. The system was presented informally to the MSU Cultural Heritage Informatics Graduate Group at Michigan State University (http://chi.anthropology.msu.edu/). In this meeting, it became apparent that ARCS required a much more carefully organized and standardized scheme for resource metadata. As a result, the ARCS development team explored a limited number of possibilities—CIDOC CRM (http://www.cidoc-crm.org/) and ArcheoML (https://github.com/neshmi/archaeoml)—before choosing to adapt the Dublin Core (http://dublincore.org/) core element metadata set as its schema.

By March 2012, a trial version of ARCS incorporating this metadata scheme was made available to a limited group of developers and archaeologists along with a request for feedback. Additional resources were also added to the ARCS system, including all excavation notebooks as well as a limited number of available excavation reports, digitized maps, scanned film-based photographs and digital photographs. Also, ARCS was also equipped with a utility that allowed individual users to gather any number of these unique resources into collections either for personal use or for sharing with other individuals be means of a stable URI locator. Finally, users were given the ability to edit the metadata of several similar resources at one time.

In April 2012, an additional undergraduate with experience in technical writing was brought onto the ARCS team in order to create the information and help pages that inform users about the ARCS system and its various utilities (http://arcs.cal.msu.edu/help and http://arcs.cal.msu.edu/about).

At last, in May 2012, the ARCS system was migrated out of the development site and into its current location (http://arcs.cal.msu.edu/) where it underwent its first full-scale test as an educational resource in conjunction with Dr. Timothy E. Gregory's History 306: Classical Archaeology class at Ohio State University. Over 100 students enrolled in this online course also participated in a research project in which they made use of a collection of resources associated with field investigations carried out in 2010. Instructional videos were created at this time in order to help students familiarize themselves with the ARCS system (http://www.youtube.com/user/caledtech/videos?query=ARCS). As a result of their experiences and feedback, additional changes were made to the ARCS system (e.g. a zoom feature for examining resources in greater detail, greater compatibility with alternate browser software (Chrome, Firefox, Safari, Internet Explorer), ability to turn off annotations in order to view a resource in an un-augmented state).

In June 2012, ARCS was used for the first time in the field by student interns from Michigan State University under the supervision of Dr. Jon M. Frey at the OSU Excavations at Isthmia. During the day at the excavations, students continued to digitize materials from the project archives, which were then processed in the afternoons and uploaded to the ARCS system in the evenings. In general, the ARCS system functioned as it was designed with the only shortcoming being the limited bandwidth of the high speed internet connections in Greece (figs. 8-10).

Reporting on ARCS (July 2012 – March 2013)

In July 2012, the project expended the last of its funds and began to depend on the efforts of student volunteers for further augmentation and development. In general, the ARCS code base has changed little since being migrated from the development site. Instead, the most significant change to the system concerns

the addition of nearly 5,000 digitized object inventory cards, which represent the sum total of all cataloged artifacts held by the OSU Excavations at Isthmia. As such these resources are, in addition to the excavation notebooks, the second data type to be fully transferred out of their paper format and onto the ARCS system. Students in Dr. Frey's HA 410: Greek Art and Archaeology course have been charged with the task of augmenting these resources with metadata transcribed from the information found on the inventory cards.

In January 2013 the ARCS system was presented to the scholarly community in a paper delivered at the AIA/APA Annual Meeting in Seattle Washington, as part of a session entitled "Managing Archaeological Data in the Digital Age: Best Practices and Realities." (http://www.youtube.com/watch?v=iasYBcpwdd4). In March 2013, the system was again presented at the 41st annual Computer Applications and Quantitative Methods in Archaeology (CAA) conference in Perth, WA Australia (http://www.caa2013.org/drupal/about). In July 2013, the ARCS system will again be presented in a session entitled "Current Research & Practice in Digital Archaeology" at the annual Digital Humanities conference (http://dh2013.unl.edu/).

The Isthmia version of ARCS can be evaluated at http://arcs.cal.msu.edu/ while the source code is freely available, along with instructions for installation, on Github (https://github.com/calmsu/arcs).

III. Accomplishments

Given the fact that the original goal of the proposed project was simply to test the ability of an open source software solution to allow a small group of users to tag an excavation notebook with search terms and connect it to other digitized documents in an online environment, it is difficult to exaggerate the degree to which "Digging Digitally" has been an overwhelming success. For this project has moved well beyond a "proof of concept" phase and has already been largely implemented at the OSU Excavations at Isthmia, Ohio State University online courses, and Michigan State University study abroad classes as a useful research, education and outreach tool. The clean, simple and intuitive user interface has succeeded in recapturing, and even improving upon, the benefits of being physically present at an archaeological research archive. The open-source code is already freely available to archaeological projects that want to create their own customized online research collection, but lack the funds or personnel to achieve this goal in the short term. In fact, it is no small boast to note the degree to which the ARCS system, which was independently developed in less than a full year on a modest budget through the efforts of a team of undergraduate students matches (or even outperforms) the system developed over the course of several years by the much more established and significantly better funded American School of Classical Studies at Athens (http://ascsa.net/research?v=default)

(http://www.ascsa.edu.gr/index.php/Spiffs/eea-project-feature).

In addition, the presentation of the ARCS system at conferences an in trial classroom experiences has sharpened our understanding of the place of this software within the larger world of digital archaeology. ARCS is not a digital publication service such as Open Context (http://opencontext.org/). Nor is it designed to serve as an archival storage solution for archaeological data. There are already services in existence that perform such a service much more dependably (http://archaeologydataservice.ac.uk/ and http://www.tdar.org/). Rather, ARCS is particularly well suited to provide small budget projects and individual researchers with the tools they need to prepare, to organize and to present digitized documents to the research/educational community, as well as the public at large, in a way that encourages exploration and collaborative research, but also maintains some small measure of control over the ways in which that evidence is interpreted. When used in conjunction with these other services, ARCS becomes an important part of the process of utilizing in the digital realm the primary documentation that represents our understanding of the archaeological past.

IV. Audiences

Although it was first proposed as a tool for scholarly collaboration, it is remarkable just how diverse the audience for the ARCS system has come to be. To be sure, among archaeologists and historians, ARCS is effective in providing immediate access to archival materials that had once required lengthy and costly visits to the excavation archives. Requests to the excavation director for specific records or forms of evidence are already being answered with a web link to the digitized resource, thus saving staff and researchers hours of effort and possibly even months of waiting. In this regard, ARCS also holds a great deal of potential as a means of satisfying the rules of and regulations for oversight in a host country. Through ARCS final reports to inspectors and regulators can now be accompanied by digital documentation of the end products of a season's excavation or survey.

In the classroom, ARCS is also proving to be a useful educational tool in its ability to expose students to the primary documentation and processes of investigation that eventually result in the published narratives they commonly regard as unquestionable fact. When given access to documents and materials that are too frequently considered the exclusive possession of professional archaeologists, students are empowered and challenged to think of themselves as scholars charged with the task of developing their own story of the past. Yet, even among the public, a crowd source approach to augmenting and interconnecting the information contained in ARCS encourages and benefits from the support of people who share our interest in the archaeological past. Through ARCS, it is possible for anyone to play an essential role in the generation of

archaeological knowledge regardless of whether they hold an academic or professional position.

In many ways, the audience for a project using ARCS is only limited by the time one is willing to spend to encourage its use among the professional community and the public in general.

V. Evaluation

While the original proposal called for specific periods of evaluation to occur at intervals throughout the development phase of this project, the growth in complexity of the ARCS system, combined with the later decision to abandon the use of Resource Space as the program's underlying architecture required a change to the original plan and schedule. Instead, the development team relied on feedback from the principle investigator in the early stages of the project, followed by student input in during the classroom implementation in May 2012. In order to speed and simplify the feedback process, programmers utilized the UserVoice service (https://www.uservoice.com/). In this way students and other users were able to immediately pinpoint problems with the software and make suggestions for improvements without disrupting their efforts to complete the classroom assignment. Typically, programmers were able to isolate and repair errors in the source code within hours of receiving a report. Occasionally, more involved improvements, such as a magnification tool to examine a resource in closer detail, took up to a week to create, to test for compatibility on the development site, and eventually to migrate to the active site.

Beyond this, the ARCS team has received valuable feedback from members of the academic community at the meetings and conferences where the software has been demonstrated. As the project timeline above clearly shows, this feedback has resulted in important improvements to the ARCS system.

In general, the response to ARCS has been very positive. Already, several archaeological projects have expressed interest in implementing ARCS. Of these, three projects in particular have agreed to cooperate with the ARCS team on an NEH Digital Humanities Implementation Grant which was submitted in January 2013.

VI. Continuation of the Project

In spite of the current lack of funding, ARCS continues to be developed, albeit at a slower pace. Student volunteers who had the opportunity to use the ARCS system in Greece in June 2012 remain committed to the project and continue to assist in all stages of preparing, uploading, and augmenting digitized records from the Isthmia excavations. Student programmers employed by the MSU

College of Arts and Letters also continue to offer limited services as a result of the college's continuing support of this important project. More significantly though, the positive evaluation of this software and its important place within the constellation of tools for the presentation, analysis and storage of digitized archaeological data clearly suggests that ARCS should be further developed and improved.

VII. Long Term Impact

Once these goals are achieved, it is clear that ARCS will serve as an easily accessible alternative for low budget projects that wish to make their archival documentation available to scholars, other archaeological projects and the public at large. Yet even in its present state of development, ARCS already is playing an important role in providing scholars and researchers with immediate and reliable access to documents that once were prohibitively difficult to obtain. What is more, ARCS will continue to assist the staff at the OSU Excavations at Isthmia by providing a digitally guided and supervised set of procedures for the digitization of archival documentation. Finally, and most importantly, in the classroom and on the Internet, the ARCS program will continue to play an important role in educating students and the general public about the archaeological process and the forms of primary evidence that form the foundation of our interpretations of the past.

VIII. Grant Products

As stated above, the source code, which is the primary product of this project, is currently available for free, along with instructions for installation at (https://github.com/calmsu/arcs). Those wishing to see a working version of ARCS can go to http://arcs.cal.msu.edu/ where they will find the OSU Excavations at Isthmia version of the system, as well as a description of the software and a number of informative pages regarding the technical aspects of the software. The Isthmia Excavations plans to continue to use ARCS as its primary software solution for online access to digitized research collections and scholarly collaboration. ARCS has now been presented at two international conferences (AIA/APA 2013 (http://www.youtube.com/watch?v=iasYBcpwdd4) and CAA 2013 (http://www.caa2013.org/drupal/about) with an additional conference paper scheduled for July 2013 (DH 2013).

At the present moment, an article is in preparation outlining recommendations for best practices in the digitization of archaeological legacy data, use of the ARCS system and the simultaneous maintenance of both digital and traditional analog archival information.

IX. Appendix:

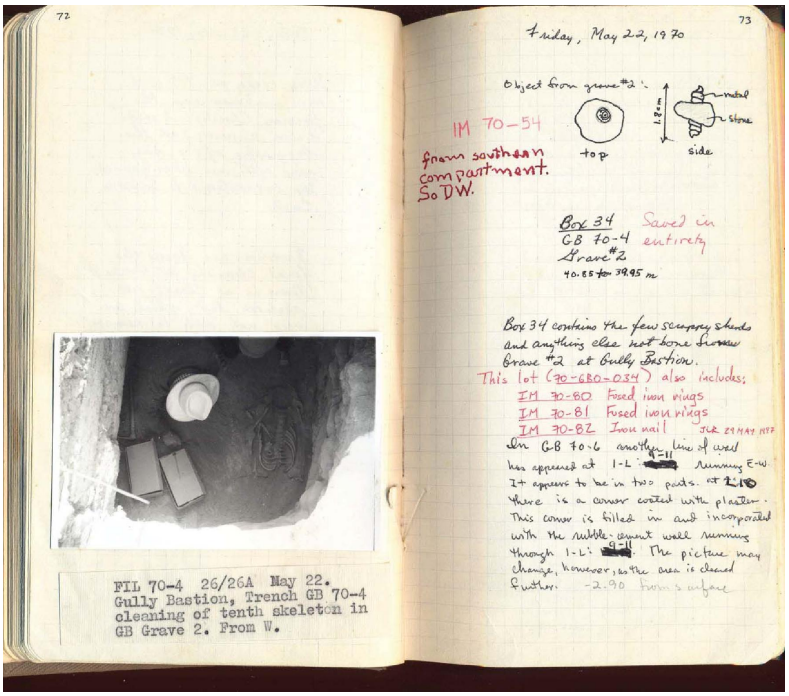

Fig. 1: Field book page from OSU Excavations at Isthmia Archive

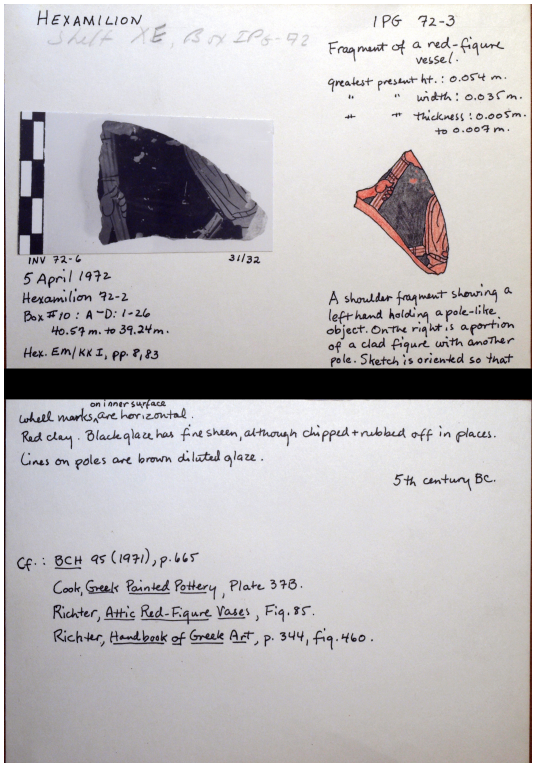

Fig.2: Inventoried artifact card from OSU Excavations at Isthmia Archive

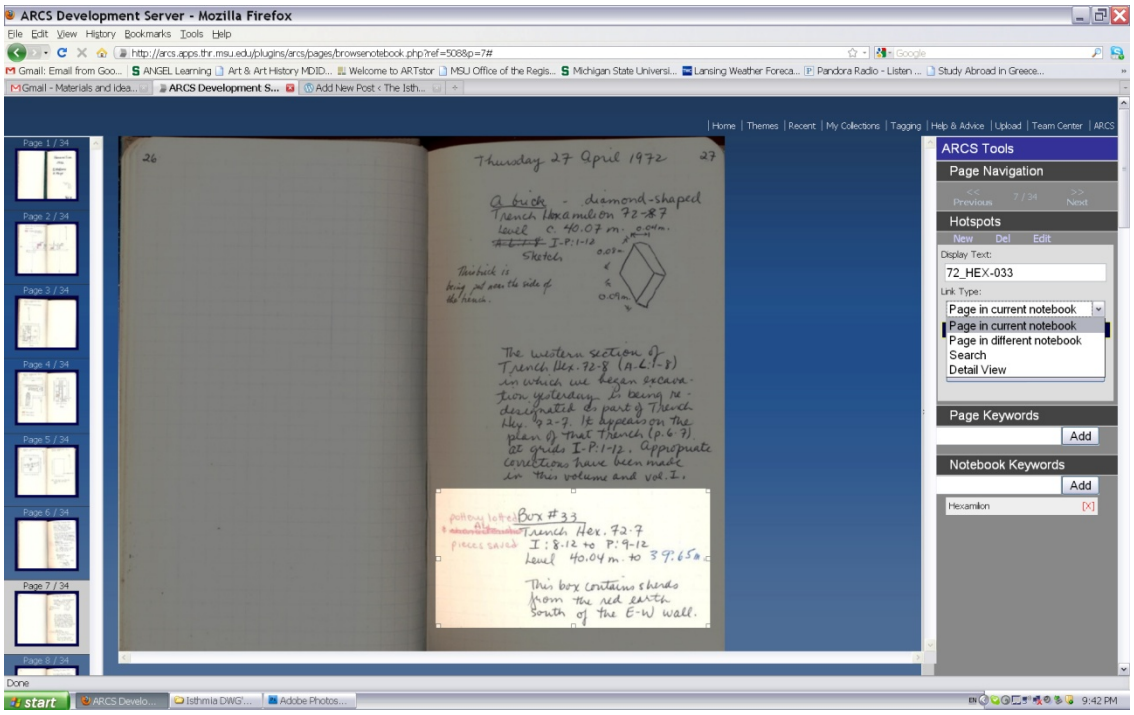

Fig. 3: Original ARCS interface based on Resource Space Digital Asset Management System

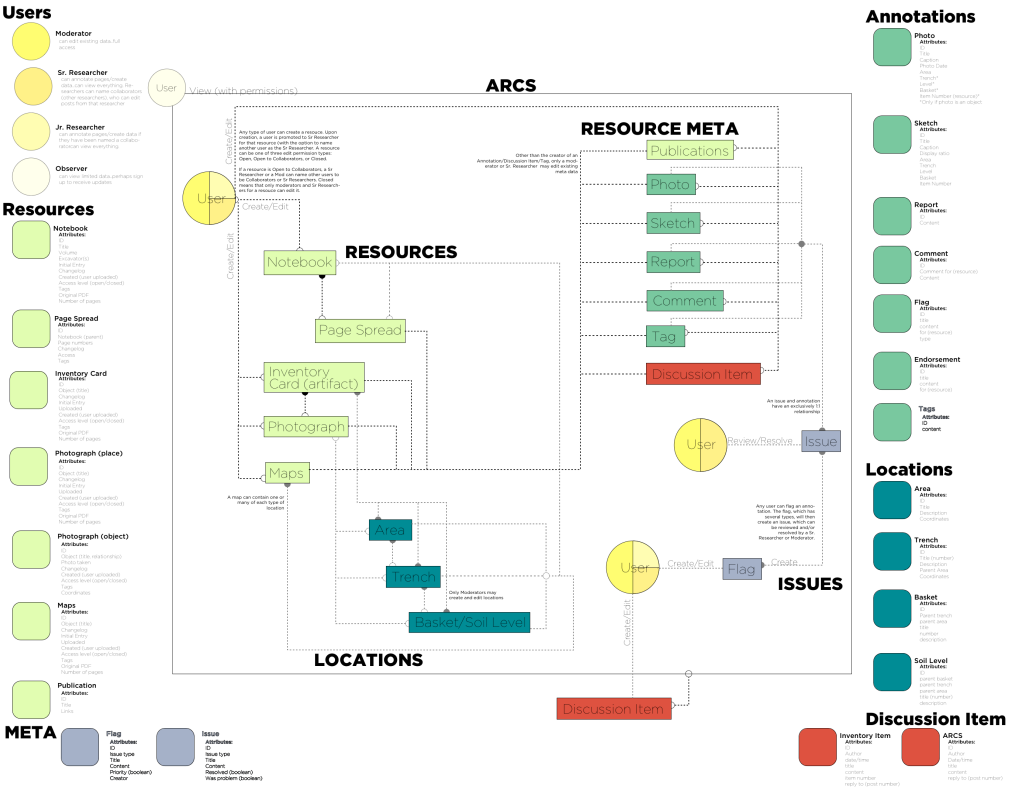

Fig. 4: Object Model showing organizational structure of ARCS

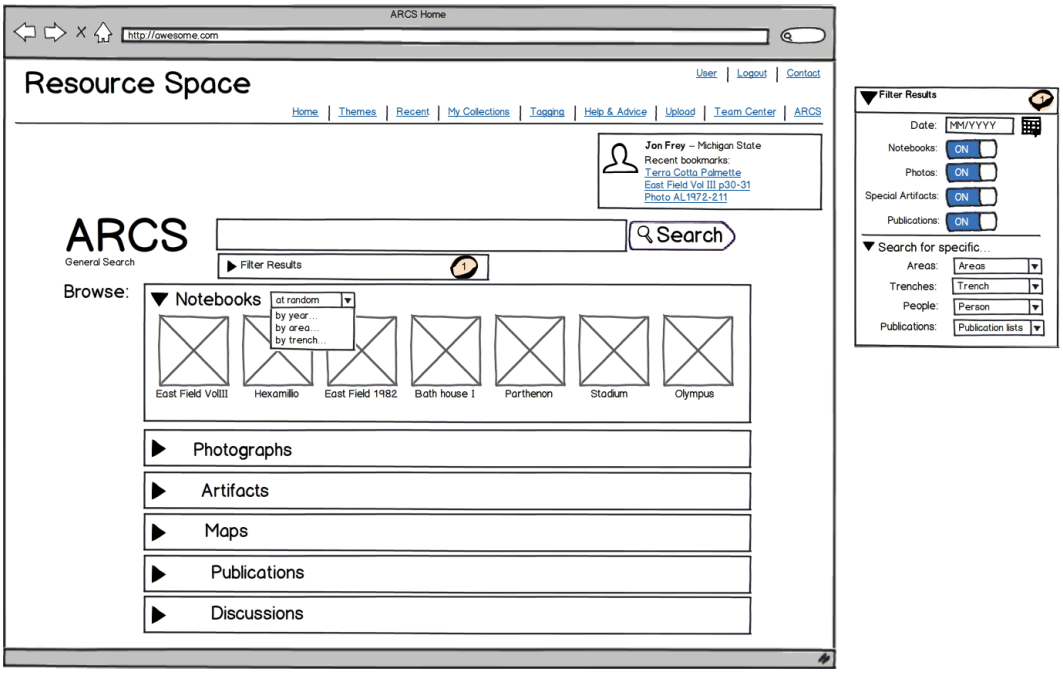

Fig. 5: Wireframe of opening page features created in Basalmiq

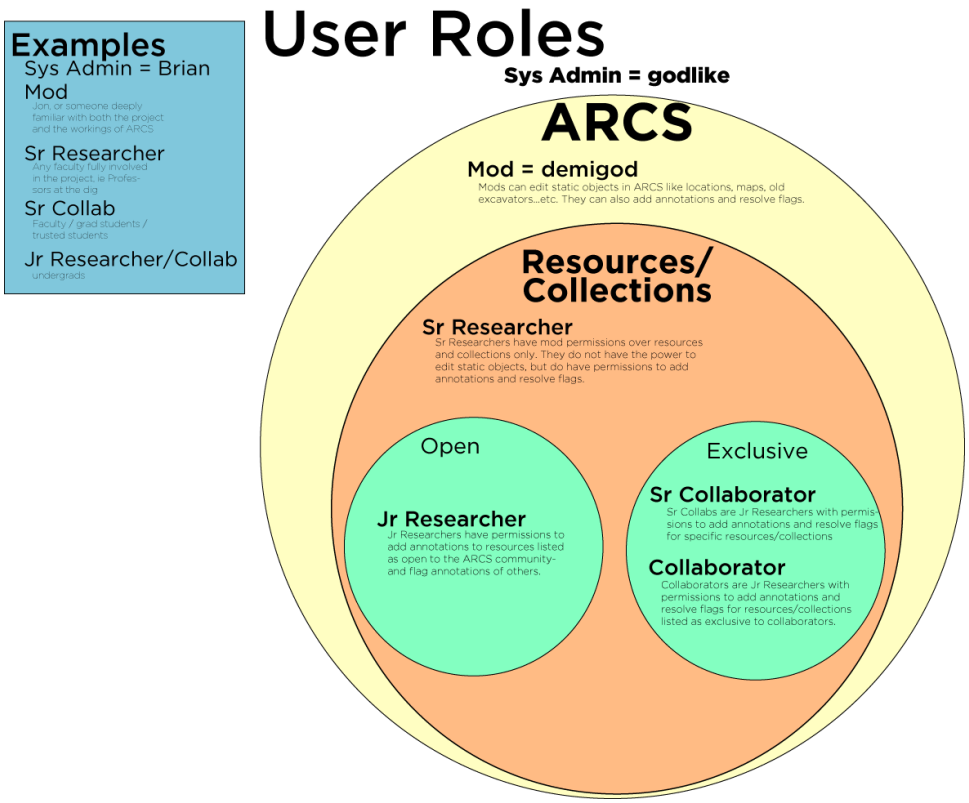

Fig. 6: Model showing user roles and privileges in ARCS

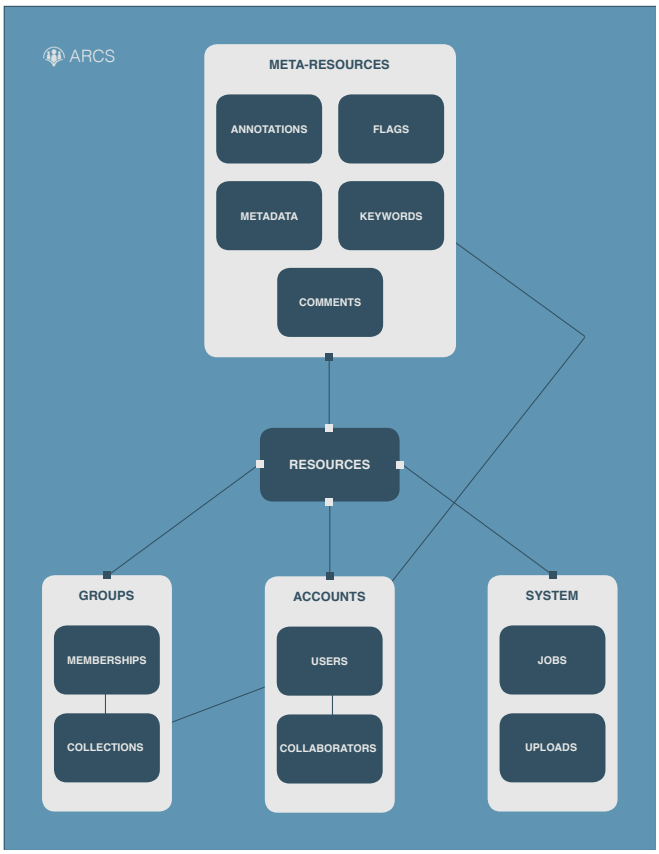

Fig. 7: Model showing database structure in ARCS

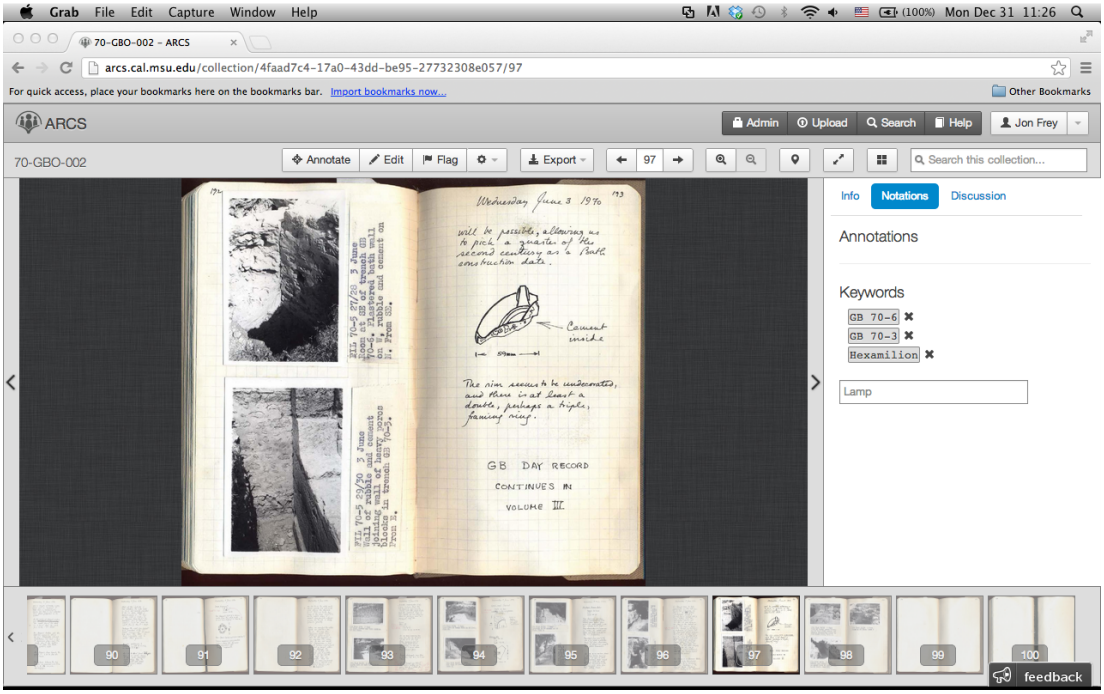

Fig. 8: Final ARCS interface showing keywording function

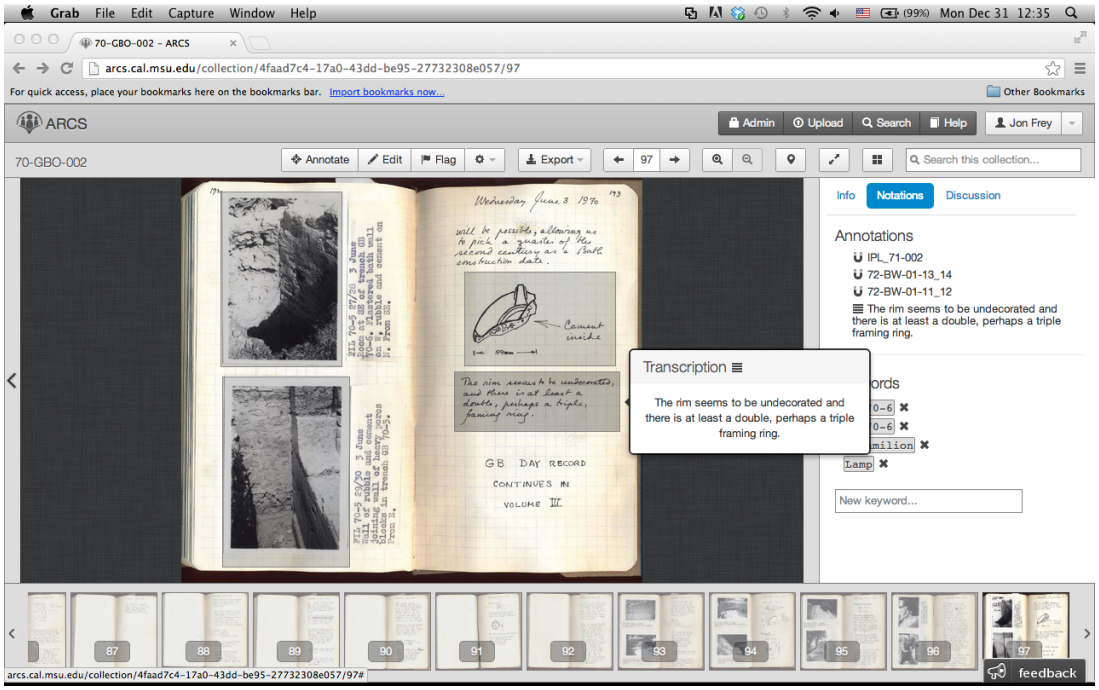

Fig. 9: Final ARCS interface showing annotation / transcription function

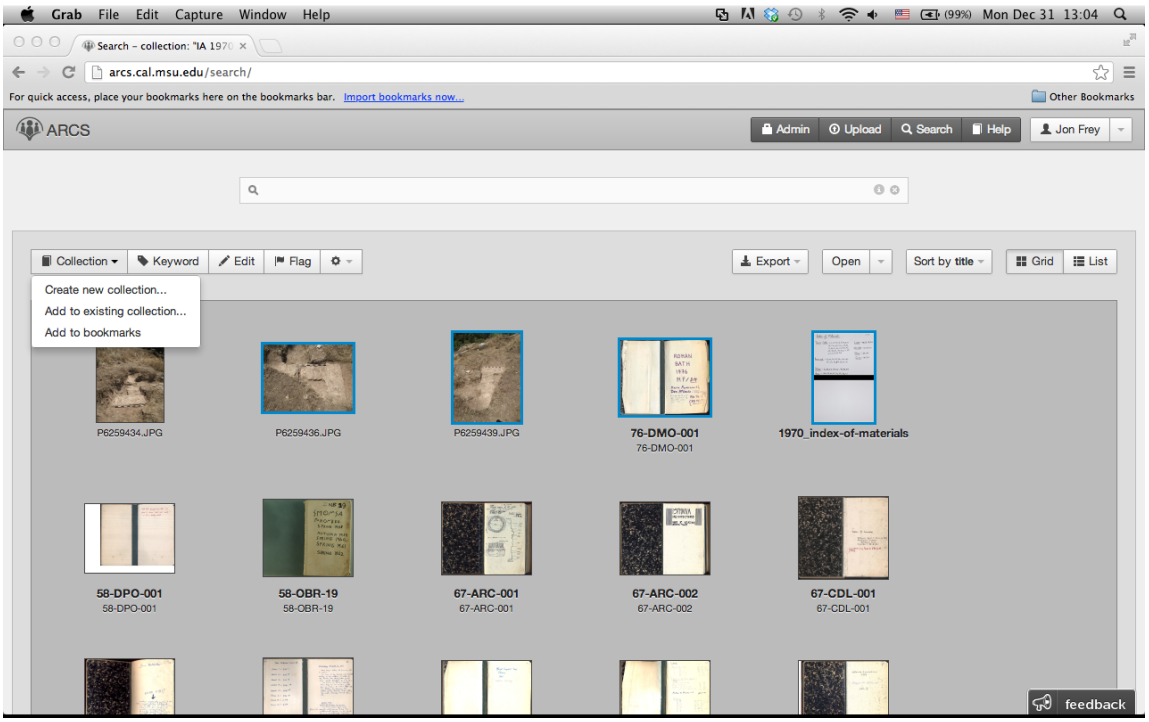

Fig. 10: Final ARCS interface showing creation of a research collection