DCPAPERS
Proc. Int'l Conf. on Dublin Core and Metadata Applications 2014

Metadata Integration for an Archaeology Collection Architecture

Sivakumar Kulasekaran
Texas Advanced Computing Center
The University of Texas at Austin
siva@tacc.utexas.edu

Jessica Trelogan
Institute of Classical Archaeology
The University of Texas at Austin
j.trelogan@austin.utexas.edu

Maria Esteva
Texas Advanced Computing Center
The University of Texas at Austin
maria@tacc.utexas.edu

Michael Johnson
L - P : Archaeology, United Kingdom
m.johnson@lparchaeology.com

Abstract

During the lifecycle of a research project, from the collection of raw data through study to publication, researchers remain active curators and decide how to present their data for future access and reuse. Thus, current trends in data collections are moving toward infrastructure services that are centralized, flexible, and involve diverse technologies across which multiple researchers work simultaneously and in parallel. In this context, metadata is key to ensuring that data and results remain organized and that their authenticity and integrity are preserved. Building and maintaining it can be cumbersome, however, especially in the case of large and complex datasets. This paper presents our work to develop a collection architecture, with metadata at its core, for a large and varied archaeological collection. We use metadata, mapped to Dublin Core, to tie the pieces of this architecture together and to manage data objects as they move through the research lifecycle over time and across technologies and changing methods. This metadata, extracted automatically where possible, also fulfills a fundamental preservation role in case any part of the architecture should fail.

Keywords: archeology; collection architecture; metadata integration; automated metadata extraction; ARK; iRODS rules; Corral; Rodeo; Ranch

1. Introduction

Data collections are the focal point through which study and publishing are currently accomplished by large research projects. Increasingly they are developed across what we refer to as *collection architectures*, in which data and metadata are curated across multi-component infrastructures and in which tasks such as data analysis and publication can be accomplished by multiple users seamlessly and simultaneously across a collection's lifecycle. It is well known that metadata is indispensable in furthering a collection's preservation, interpretation, and potential for reuse, and that the process of documenting data in transition to an archival collection is essential to those goals. In the collection architecture we present here, we use metadata in a novel way: to integrate data across recordkeeping and archival lifecycle phases as well as to manage relationships between data objects, research stages, and technologies. In this paper, we introduce and illustrate these concepts through the formation of an archaeological collection spanning many years. We show how metadata, formatted in Dublin Core (DC), is used to bridge data and semantics developed as teams and research methods have changed over the decades.

The model we propose differs from traditional data management practices that have been described as the "long tail of research" (Wallis et al., 2014), in which researchers may store data in scattered places like home computers, hard-drives and institutional servers, with data integrity

https://doi.org/10.23106/dcmi.952136410

potentially compromised. Without a clear metadata strategy, data provenance becomes blurry and integration impossible. In the traditional model, archiving in an institutional repository or in a data publication platform comes at the end of the research lifecycle, when projects are finalized, often decades after they started, and sometimes too late to retain their original intended meaning (Eiteljorg, 2011). At that final stage, reassembling datasets into collections that can be archived and shared becomes arduous and daunting, preventing many from depositing data at all. Instead, a collection architecture such as the one presented here, which is actively curated by the research team throughout a project, helps to keep ongoing research organized, aggregates metadata on the go, facilitates data sharing as research progresses, and enables the curator-researcher to control how the public interacts with the data. Moreover, data that are already organized and described can be promptly transferred to a canonical repository.

For research projects midway between the "long tail" and the new data model, the challenge is to merge old and new practices, to shape legacy data into new systems without losing meaning and without overwriting the processes through which data were conceived. We present one such case: a collection created by the Institute of Classical Archaeology (ICA, 2014) representing several archaeological investigations (excavations, field surveys, conservation, and study projects) in Italy and Ukraine going back as far as the mid-1970s. As such, it includes data produced by many generations of research teams, each with their own idiosyncratic recording methods, research aims, and documentation standards. Integrating it into a collection architecture that is accessible for ongoing study while thinking ahead about data publishing and long-term archiving has been the subject of ongoing collaboration between ICA and the Texas Advanced Computing Center (TACC, 2014) for the last five years (Trelogan et al., 2010; Walling et al., 2011; Rabinowitz et al., 2013).

In this project metadata is at the center of a transition from a disorganized aggregation of data—belonging to both the long tail of research, and new data that is being actively created during study and publication—into a collection architecture. The work has involved re-engineering research workflows and the definition of two instances of the collection with different functions and structures: one is a stable collection which we call the *archival instance* and the other, *a study and presentation instance*. Both are actively evolving as research continues, but the methods we have developed allow researchers to archive data on the fly, enter metadata only once, and to move documented data from the archive into the presentation instance and vice versa, ensuring data integrity and avoiding the duplication of effort. The DC standard integrates the data objects within the collection and binds the collection instances together.

2. Archaeology as the Conceptual Framework for a Collection Architecture

Archaeology is an especially relevant domain for exploring issues of data curation and management because of the sheer volume and complexity of documentation produced during the course of fieldwork and study (Kansa et al., 2011). Likewise, because a typical archaeological investigation requires teams of specialists from a large number of disciplines (such as physical anthropology, paleobotany, geophysics, and archaeozoology) a great deal of work is involved in coordinating the datasets produced (Faniel et al., 2013). Making such coordination even more challenging is the tendency for large archaeological research projects, like those in the ICA collection, to carry on for multiple seasons, sometimes lasting for decades. Projects with such long histories and large teams can contain layer upon layer of documentation that reflect changes in technologies, standard practices, methodologies, teams, and the varied ways in which they record the objects of their particular study.

As in an archaeological excavation, understanding these sediments is key to unlocking the collection's meaning and to developing strategies for its preservation. Due to the inevitable lack of consistency in records that span years and specialties, these layers can easily become closed-off information silos that make it impossible to understand their purpose or usefulness. The work we are doing focuses on revealing and documenting those layers through metadata, without

https://doi.org/10.23106/dcmi.952136410

erasing the semantics of past documentation, and without a huge investment of labor at the end. To address these challenges within the ICA collection, we needed a highly flexible, lightweight solution (in terms of cost, time, and skills required to maintain) for file management, ongoing curation, publishing, and archiving.

3. Functional and Resource Components of the Collection Architecture

Currently the ICA collection is in transition from disorganized data silos to an organized collection architecture, illustrated in Figure 2. The disorganized data, recently centralized in a networked server managed by the College of Liberal Arts Instructional Technology Service (LAITS, 2014), represents an aggregation of legacy data that had been previously dispersed across servers, hard-drives and personal computers. The data were centralized there to round up and preserve disconnected portions of the collection so that active users could work collaboratively within a single, shared collection. Meanwhile, new data are continuously produced as paper records are digitized and as born-digital data are sent in from specialists studying abroad. To manage new data and consolidate the legacy collection, we created a recordkeeping system consisting of a hierarchical file structure implemented within the file share, with descriptive labels and a set of naming conventions for key data types, allowing users to promptly classify the general contents and relationships between data objects while performing routine data management tasks (see Figs. 1 and 5). The recordkeeping system is used as a staging area where researchers simultaneously quality check files, describe and organize them (by naming and classifying into labeled directories) and purge redundant copies, all without resorting to time-consuming data entry. Once organized, data are ingested into the collection's archival instance (See Fig. 2) where they are preserved for the long term and can be further studied, described, and exposed for data sharing.

3.1. Staging and recordkeeping system: gathering basic collection metadata

Basic metadata for the collection is generated from the recordkeeping system mentioned above. Using the records management big bucket theory (Cisco, 2008) as a framework, we developed a file structure that would be useful and intuitive for active and future research and extensible to all of the past, present, and future data that will be part of the ICA collection (Fig. 1). This file structure was implemented within the fileshare and is mirrored in the archival instance of the collection for a seamless transition to the stable archive. The core organizing principle for the data is its provenance as the archaeological "site" or "project" for which it was generated. Within each of these larger "buckets", we group data according to three basic research phases appropriate to any investigation encountered in the collection, be it surface survey, geophysical prospection, or excavation[1]: 1) field, 2) study, 3) publication. These top two tiers of the hierarchy allow us to semantically represent, per project, what we consider primary or raw versus processed, interpreted data, and the final polished data that are tied to specific print or online publications. The third tier includes classes of data recorded during fieldwork and study (e.g. field notes, site photos, object drawings) and the subjects of special investigations (e.g. black-gloss pottery, physical anthropology, or paleobotany). The list was generated inductively from the materials produced during specific investigations and is applicable to most ICA projects. As projects continue through the research lifecycle this list may expand to add other materials that were not initially accounted for. Curators can pick the appropriate classes and file data accordingly. Files are named according to a convention (Fig. 5), which encodes provenance, relationships between objects found together, the subject represented (e.g. a bone from a specific context), as well as the process history of the data object (e.g. a scanned photograph).

This recordkeeping system is invaluable for the small team at ICA managing large numbers of documentation objects (>50,000 per each of over two dozen field projects). Because many

[1] This is, in fact, an appropriate way to describe the lifecycle of any kind of investigation – archaeological or otherwise – that involves a fieldwork or data-collection stage.

https://doi.org/10.23106/dcmi.952136410

projects in ICA's collection are still in the study phase and do not yet have a fully developed documentation system, the filenames and directories are often the sole place to record metadata. As the data are moved to the new collection architecture, the metadata is automatically mapped as a DC document with specific qualifiers that preserve provenance and contextual relationships between objects. Metadata is thus entered only once, and is carried along through the archival to the study and presentation instances where specialists may expand and further describe them as they study and prepare their publications.

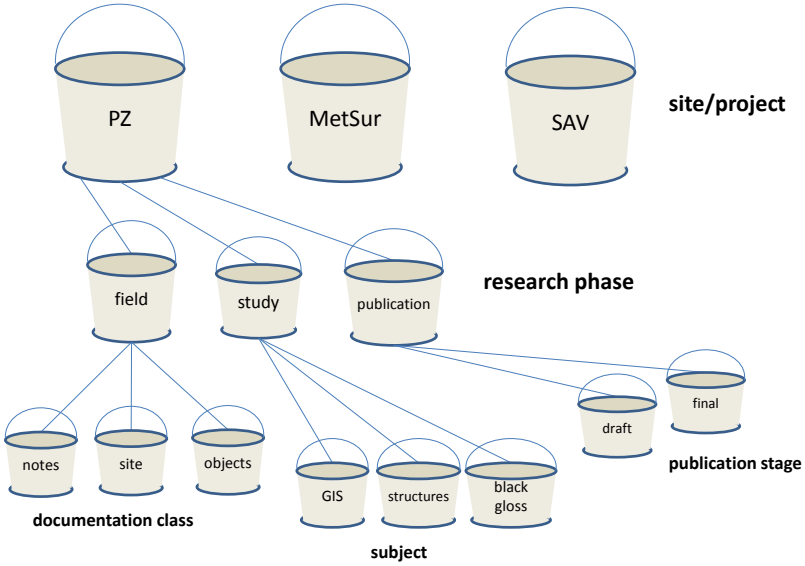

FIG. 1. The highest levels of the file structure, represented here as "big buckets" whose labels embed metadata about the project, stages of research, classes of documentation, and subjects of specialist study.

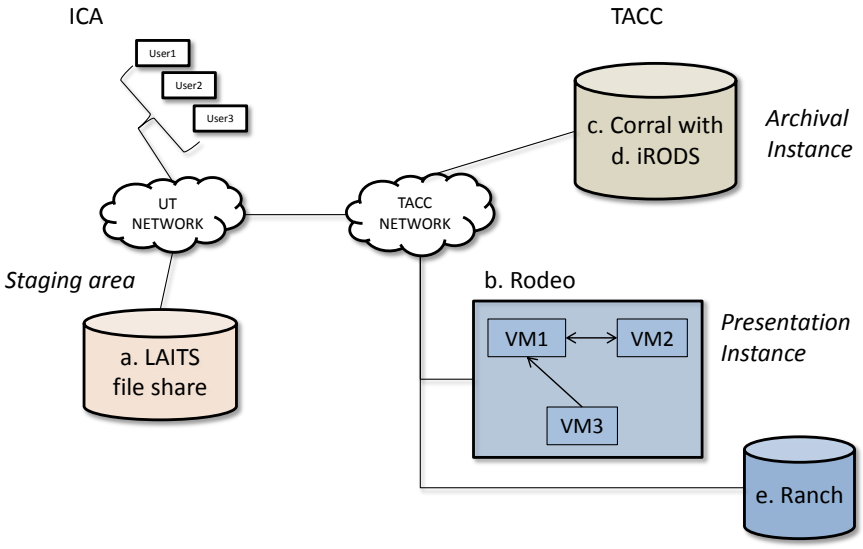

FIG. 2. Resource components of ICA's collection architecture: a. LAITS file share (staging area); b. Rodeo, cloud computing resource that hosts Virtual Machines (VMs); c. Corral, storage resource that contains active collections; d. iRODS, data management system; e. Ranch, tape archive for backups and long-term storage.

https://doi.org/10.23106/dcmi.952136410

3.2. Archival instance: Corral/iRODS

Corral is a high performance resource maintained by TACC to service UT System researchers (TACC, 2014; Corral, 2014). This system includes 6 petabytes of on- and off-site storage for data replication, as well as data management services through iRODS (integrated Rule-Oriented Data System) (iRODS, 2014). iRODS is an open-source software system that abstracts data from storage in order to present a uniform view of data within a distributed storage system. In iRODS a central metadata database called iCAT holds both user defined and system metadata, and a rule engine is available to create and enforce data policies. We implemented custom iRODS rules to automate the metadata extraction process. To access the data on Corral/iRODS, users can use GUI-based interfaces like iDROP and WebDAV or a command-line utility. Data on Corral/iRODS are secured through geographical replication to another site at UT Arlington.

3.3. Presentation instance

3.3.1. ARK

To provide a central platform for collaborative study of all material from each project, to record richer descriptions and interpretations, and to define complex contextual relationships, we adopted ARK, the Archaeological Recording Kit (ARK, 2014). ARK is a web-based, modular "toolkit" with GIS support, a highly flexible and customizable database and user interface, and a prefabricated data schema to which any kind of data structure can be mapped (Eve et al., 2008). This has allowed ICA staff to create—relatively quickly and easily—a separate ARK for each site or project, and to pick and choose the main units of observation within that (e.g. the "site" in the case of a survey project, or the "context" and "finds" for an excavation project). At ARK's core are user-configured "modules", in which the data structure is defined for each project. In terms of the "big buckets" shown in Fig. 1, each of the top tier (site/project) buckets can have an implementation of ARK, with custom modules that may correspond to the documentation classes and/or study subjects represented in the third tier of buckets, depending on the methodological approach.[2] Metadata mappings are defined within the modules in each ARK (e.g., Fig. 6). This presentation instance allows the user to interact with data objects that reside in the archival instance on Corral/iRODS, describe them more fully in context of the whole collection (creating more metadata), and then push that metadata back to the archival instance.

3.3.2. Rodeo

Rodeo is TACC's cloud and storage platform for open science research (RODEO, 2014). It provides web services, virtual machine (VM) hosting, science gateways, and storage facilities. Virtual machines can be defined as a "software based emulation of a computer" (VM, 2014). Rodeo allows users to create their own VM instance and customize it to perform scientific activities for their research needs. All of the ARK services, including the front-end web services, databases, and GIS, are hosted in Rodeo's cloud environment. We use three VM instances to host each of these services. To comply with best security practices we separate out the web services from the GIS and the databases. If the web service is compromised or any security issues arise, none of the other services are affected and only the VM that hosts the affected web service needs to be recreated. During the study and publication stages, data on iRODS are called from ARK, and metadata from ARK is integrated into the iCAT database.

[2] We currently have three live implementations of ARK hosted at TACC, one housing legacy data from excavations carried out from the 1970s to the 1990s, recorded with pen and paper and film photography with finds as the main unit of observation; a contemporary excavation, from 2001 to 2007, which was mostly born digital (digital photos, total station, in-the-field GIS, etc.) and focused on the stratigraphic context; and one survey project, from the 1980s to 2007, consisting of a combination of born digital and digitized data and centered on the "site" and surface scatters of finds.

https://doi.org/10.23106/dcmi.952136410

3.3.3. Ranch

Ranch is TACC's long-term mass storage solution with a high-performance tape-based system. We are using it here as a high-reliability backup system for the publication instance of the collection and its metadata hosted in Rodeo on the VMs. We also routinely back up the ARK code base and custom configurations. Across Corral and Ranch, the entire collection architecture is replicated for high data availability and fault tolerance.

4. Workflow and DC Metadata

4.1. Automated metadata extraction from the recordkeeping system

To keep manual data entry to a minimum, we developed a method for automatically extracting metadata embedded in filenames and folders of our recordkeeping system. We used a modularized approach using Python (Python, 2014) and customized iRODS rules so that individual modules can be easily plugged in or reused for other collections. One module extracts technical metadata using FITS (FITS, 2014) and maps the extracted information to DC and to PREMIS (PREMIS, 2014) using an XSLT stylesheet. Another module creates a METS document (METS, 2014) also using a XSLT stylesheet transformation from the FITS document. The module focusing on descriptive metadata extracts information from the recordkeeping system and maps it to DC following the instructions from the data dictionary. Metadata is integrated into a METS/DC document. Finally, metadata from the METS document is parsed and registered in the iCAT database (Walling et al., 2011). Some files do not conform to the recordkeeping system because they could not be properly identified and thus named and classified. For those, the descriptive metadata will be missing and only a METS document with technical metadata is created, with the technical information added into iCAT. This metadata extraction happens on ingest to iRODS, so it occurs only as frequently as the users upload data that are understood and organized by the researchers. The accuracy of the extracted metadata depends upon the accuracy of the filenames (e.g., adherence to naming convention or correctness of object identification). These are then further quality checked within the ARK interface during detailed collaborative study, and corrections are pushed back to the iRODS database as needed by the user.

4.2. Syncing data between ARK and iRODS

The next phase was to sync metadata between the two databases: ARK and iCAT/iRODS. A new function was created within ARK to pull in metadata from iRODS and display it alongside the metadata from ARK for each item in a module (e.g. object photographs).

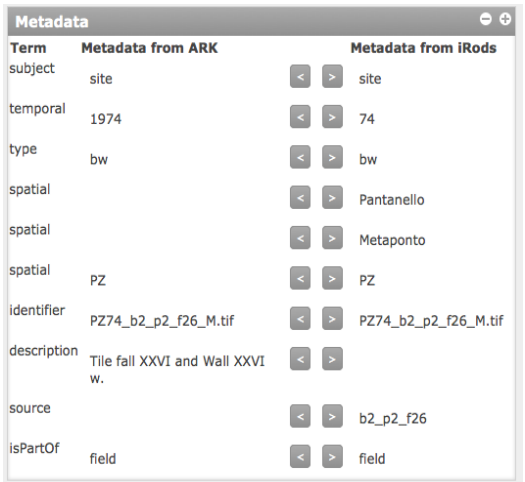

FIG. 3 Metadata subform from ARK, allowing user to compare the information from the two collection instances.

https://doi.org/10.23106/dcmi.952136410

Fields in ARK are used to define what data are stored where in the back-end ARK database, the way that they should be displayed on the front-end website, and the way that they should be added or edited by a researcher. The data classes used in ARK are specific to that environment and have been customized and defined according to user needs within each implementation. The mapping between the DC term and the corresponding field within ARK is defined in the module configuration files.

While research progresses, data and metadata are added and edited via the ARK interface. The user can update the metadata in iRODS from ARK or vice versa, using arrow buttons showing the direction that the data will move. The system automatically recognizes if the user is performing an add or edit operation. PHP is used to read and edit the information from ARK and iRODS, and Javascript is used to give the user feedback and confirm the modifications (Fig. 3). The metadata linked to either the DC term or the ARK field are then presented and updated through the ARK web interface.

The workflow represented in Fig. 4 allows us to transition data into the collection architecture and to perform ongoing data curation tasks throughout the research lifecycle. Note that in this workflow, data are ingested first to the archival instance of the collection. This allows archiving as soon as data are generated, assuring integrity at the beginning of the research lifecycle.

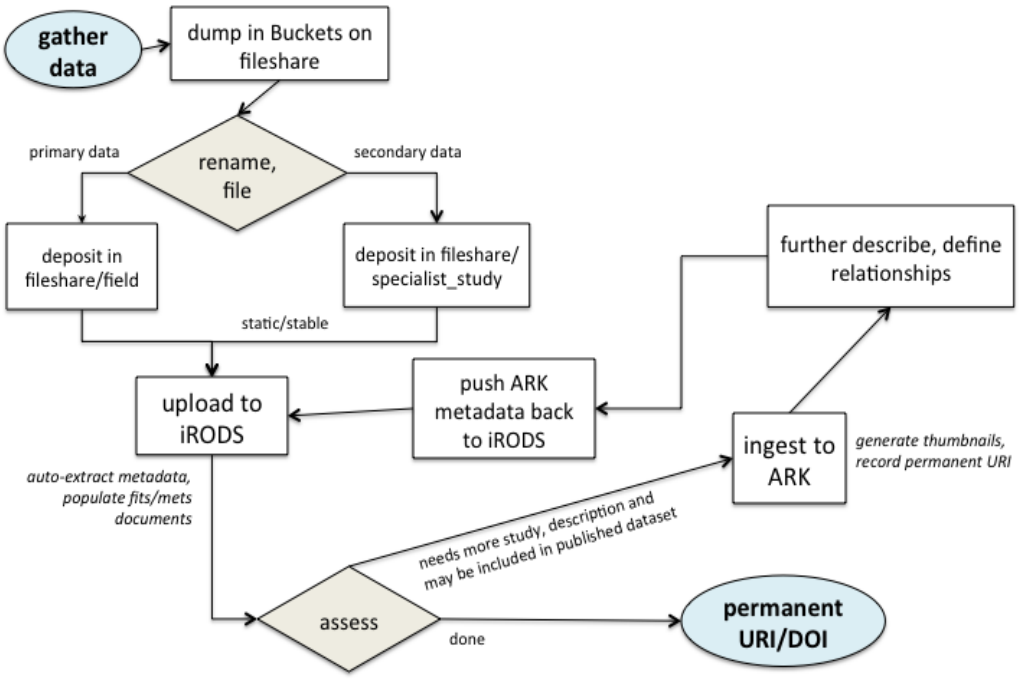

FIG. 4. Curation workflow.

4.3. Dublin Core metadata: the glue that binds it all together

Metadata schemas are typically used to describe data for ease of access, to provide normalization, and to establish relationships between objects. They can be highly specialized to include elements that embed domain-specific constructs. A general schema like DC, on the other hand, can be used in most disciplines, if fine-grained description is not a priority. In choosing a schema for this project we considered its ability to relate objects to one another, its generalizability in representing the wide range of recording systems represented in the collection, and its ease of use. With this in mind, we chose to use DC, which is widely used for

https://doi.org/10.23106/dcmi.952136410

archaeological applications, including major data repositories like the UK-based Archaeology Data Service (ADS, 2014) and, in the US, the Digital Archaeological Record (tDAR, 2014).

In this project the DC standard is a bridge over which data are exchanged between collection instances and across active research workflows, turning non-curated into curated data, while providing a general, widely understood method for describing the collection and the relationships between the objects. Given the need for automated metadata extraction and organization processes, we required higher levels of abstraction to map between the different organizational and recording systems, data structures, and concepts used over time. Furthermore, DC is the building block for future mapping to a semantically rich ontology like CIDOC-CRM (CRM, 2014), a growing standard that is used for describing cultural heritage objects that is particularly relevant for representing archaeology data in online publishing platforms (OpenContext, 2014). CIDOC-CRM provides the scope to fully expose the richness of exhaustive analysis, and allows the precise expression of contextual relationships between objects of study, as well as the research process and events (historical and within an excavation or study), provenance (of cultural artifacts as well as of data objects), and people. Such semantic richness, however, only fully emerges at the final stages of a project, and we are here concerned with ongoing work resulting in a collection that is still in formation and evolving rapidly.

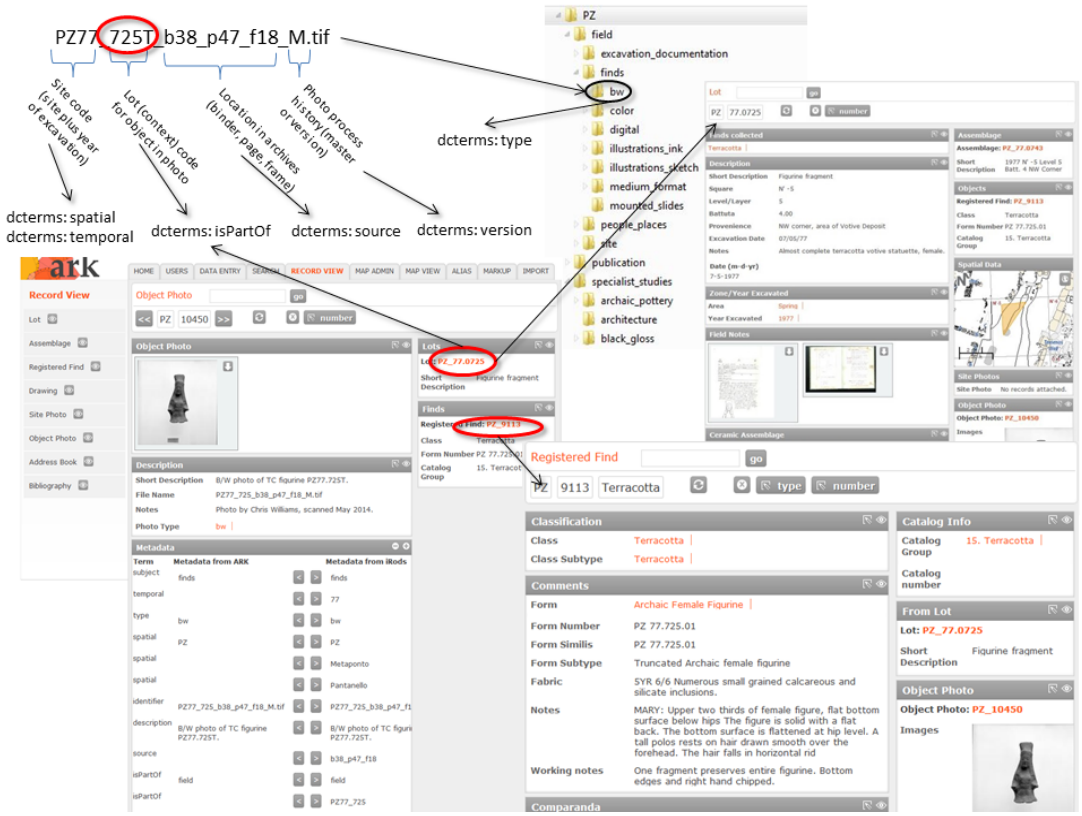

FIG. 5. Metadata extracted from filename and folder labels are mapped to DC terms. Once in ARK further descriptive metadata can be added and pushed back to iRODS.

4.4. Metadata mapping and its semantics

The mapping to DC for this project was considered in two stages. For the archival instance of the collection, we focused on expressing relationships between individual data objects (represented by unique identifiers) through the DC elements "spatial," "temporal," and "isPartof." This allows grouping, for example, of all the documentation from a given excavation, or all

https://doi.org/10.23106/dcmi.952136410

artifacts found within the same context. We also categorized documentation types and versions to help us relate data objects to the physical objects they represent (e.g., a drawing or photo of an artifact). For the publication instance presented in ARK, mapping focused on verbal descriptions, interpretations, and the definition of relationships produced during study. These then populate the "description" and "isPartOf" elements in the DC document. As a data object enters the collection to be further analyzed and documented in ARK, all the key documentation related to that object is exchanged over time throughout all pieces of the collection architecture and remain in the archival instance once complete. For example, when a photo is scanned, named, and stored in the appropriate folder, this embeds provenance information for the object in the photo (e.g., context code, site and year of excavation), the provenance of the photo itself (e.g., location of negative in physical archive), the process history of the data object (e.g., raw scan or an edited version), its relations to other objects in the collection, and the description created by specialists in ARK (see Fig. 5). For the data curator, the effort is minimal, and information is extracted automatically and mapped to terms that are clearly understood. The information is carried along as the data object moves from the primary data archive to the interpretation platform, and is enhanced through study and further description every time the metadata is updated. By mapping key metadata elements to DC (Table 1) we reduce data entry and provide a base for future users of the collection.

TABLE 1. Extract of a data dictionary that maps the fields in an ARK object photo module to the recordkeeping system and DC elements.

ARK term	ARK field	Record Keeping Example	DC Term
Short Description	$conf_field_short_desc	Terracotta Figurine	description
File Name	$conf_field_filename	PZ77_725T_b38_p47_f18_M.tif	identifier
Photo Type	$conf_field_phototype	PZ/field/finds/bw	format
Date Excavated	$conf_field_excavyear	1977	date
Date Photographed	$conf_field_takenon	1978	created
Photographed by	$conf_field_takenby	Chris Williams	creator
Area	$conf_field_area	Pantanello	spatial
Zone	$conf_field_zone	Sanctuary	spatial

4.5. Metadata for integrity

In addition to the technical metadata extraction, descriptive metadata added throughout the research lifecycle assures the collection's integrity in an archaeological sense by reflecting relationships between data objects. Moreover, because we have the same metadata stored in both the archival and presentation instances, if one or more parts of the complex architecture should fail, the collection can be restored. Once the publication instance is completed and accessible to the public, users will be able to download selected images and their correspondent DC metadata, containing all the information related to those images.

5. Conclusion

This work was developed for an evolving archaeological dataset, but can act as a model to inform any kind of similarly complex academic research collection. The model illustrates that DC metadata can act as an integrative platform for a non-traditional (but increasingly common) researcher-curated, distributed repository environment. With DC as a bridge between collection instances we ensure that the relationships between objects and their metadata are preserved and that original meaning is not lost. Integration also reduces overhead in entering repetitive information and provides a means for preservation. In the event that a database fails or becomes obsolete, or if ICA can no longer support the presentation instance, the archival instance can be sent to a canonical repository with all its metadata intact.

https://doi.org/10.23106/dcmi.952136410

Finally, we can also attest that the model enables an organized and documented research process in which curators can conduct a variety of tasks including archiving, study, and publication, while simultaneously integrating legacy data. Our whole team, including specialists working remotely, can now access our entire collection as a whole, view everything in context, and work collaboratively in a single place. Because this work was developed with and by the users actively testing it during ongoing study, we can also speak to the real benefits that have been gained. In the course of this work, ICA lost over 2/3 of its research and publication staff due to budget cuts. While this was a serious blow, the collection architecture we have described here has allowed us to radically streamline our study and publication process enough that, despite losing valuable staff, we are actually producing our publications much more efficiently than we ever have before and have helped ensure a future for the data behind them.

References

ADS, Archaeology Data Service. (2014). Retrieved May 9, 2014 from http://archaeologydataservice.ac.uk/.

ARK, the Archaeological Recording Kit. (2014). Retrieved May 9, 2014 from http://ark.lparchaeology.com/.

Cisco, Susan. (2008). Trimming your bucket list. ARMA International's hot topic. Retrieved May 9, 2014 from http://www.emmettleahyaward.org/uploads/Big_Bucket_Theory.pdf.

Corral. (2014). Retrieved August 14, 2014 from https://www.tacc.utexas.edu/resources/corral.

CRM. (2014). CIDOC Conceptual Reference Model. Retrieved May 9, 2014 from

http://www.cidoc-crm.org/.

Eiteljorg, Harrison. (2011). What are our critical data-preservation needs? In: Eric C. Kansa, Sarah Whitcher Kansa, & Ethan Wattrall (eds). Archaeology 2.0: New Approaches to Communication and Collaboration. Cotsen Digital Archaeology series 1, 251–264. Los Angeles: Cotsen Institute of Archaeology Press.

Eve, Stuart, and Guy Hunt. (2008). ARK: A Developmental Framework for Archaeological Recording. In: A. Posluschnya, K. Lambers, & I. Herzong. (eds). Layers of Perception: Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Berlin, Germany, April 2–6, 2007. Kolloquien zur Vor- und Frühgeschichte 10. Bonn: Rudolf Habelt GmbH. Retrieved from: http://proceedings.caaconference.org/files/2007/09_Eve_Hunt_CAA2007.pdf.

Faniel, Ixchel, Eric Kansa, Sarah Whitcher Kansa, Julianna Barrera-Gomez, and Elizabeth Yakel. (2013). The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse. JCDL 2013 Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, 295–304. New York: Association for Computing Machinery. doi:10.1145/2467696.2467712

FITS, File Information Tool Set. (2014). Retrieved May 9, 2014 from https://code.google.com/p/fits/.

Harris, Edward C. (1979). Laws of Archaeological Stratigraphy. World Archaeology Vol. 11, No. 1: 111–117.

ICA, Institute of Classical Archaeology. (2014). Retrieved May 9, 2014 from http://www.utexas.edu/research/ica/.

iRODS, A data management software. (2014). Retrieved May 9, 2014 from http://irods.org/.

Kansa, Eric C., Sarah Whitcher Kansa, & Ethan Watrall (eds). (2011). Archaeology 2.0: New Approaches to Communication and Collaboration. Cotsen Digital Archaeology series 1. Los Angeles: Cotsen Institute of Archaeology Press.

LAITS, College of Liberal Arts. (2014). Retrieved May 9, 2014 from http://www.utexas.edu/cola/laits/.

METS, Metadata Encoding & Transmission Standard. (2014). Retrieved May 9, 2014 from http://www.loc.gov/standards/mets/.

OpenContext. (2014). Retrieved May 9, 2014 from http://opencontext.org/.

PREMIS, Preservation Metadata Maintenance Activity. (2014). Retrieved May 9, 2014 from http://www.loc.gov/standards/premis/.

Python, a programming Language. (2014). Retrieved May 9, 2014 from https://www.python.org/.

PHP, A hypertext preprocessor. (2014). Retrieved May 9, 2014 from http://www.php.net.

Rabinowitz, Adam, Jessica Trelogan, and Maria Esteva. (2012). Ensuring a future for the past: long term preservation strategies for digital archaeology data. Presented at Memory of the Worlds in the Digital Age Conference: Digitization and Preservation, UNESCO, September 26–28, 2012, Vancouver, British Columbia, Canada.

https://doi.org/10.23106/dcmi.952136410

Rodeo. (2014). Retrieved August 14, 2014 from https://www.tacc.utexas.edu/resources/data-storage/#rodeo.

TACC, The Texas Advanced Computing Center. (2014). Retrieved May 9, 2014 from https://www.tacc.utexas.edu/.

tDAR, Digital Archaeological Record. (2014). Retrieved May 9, 2014 from http://www.tdar.org/.

VM, Virtual Machine. (2014) Retrieved May 9, 2014 from http://en.wikipedia.org/wiki/Virtual_machine.

Walling, David, and Maria Esteva. (2011). Automating the Extraction of Metadata from Archaeological Data Using iRods Rules. International Journal of Digital Curation Vol. 6, No. 2: 253–264.

Wallis, Jillian C., Elizabeth Rolando, and Christine L. Borgman. 2013. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. PLoS ONE 8(7): e67332. doi:10.1371/journal.pone.0067332

https://doi.org/10.23106/dcmi.952136410