Archaeologies: Journal of the World Archaeological Congress (© 2010)
DOI 10.1007/s11759-010-9146-4

RESEARCH

Googling the Grey: Open Data, Web Services, and Semantics

Eric C. Kansa, School of Information, UC Berkeley, Berkeley, CA, USA
E-mail: ekansa@ischool.berkeley.edu

Sarah Whitcher Kansa, Alexandria Archive Institute, San Francisco, CA, USA

Margie M. Burton and Cindy Stankowski, San Diego Archaeological Center, Escondido, CA, USA

ABSTRACT

Primary data, though an essential resource for supporting authoritative archaeological narratives, rarely enters the public record. Lack of primary data publication is also a major obstacle to cultural heritage preservation and the goals of cultural resource management (CRM). Moreover, access to primary data is key to contesting claims about the past and to the formulation of credible alternative interpretations. In response to these concerns, experimental systems have implemented a variety of strategies to support online publication of primary data. Online data dissemination can be a powerful tool to meet the needs of CRM professionals, establish better communication and collaborative ties with colleagues in academic settings, and encourage public engagement with the documented record of the past. This paper introduces the ArchaeoML standard and its implementation in the Open Context system. As will be discussed, the integration and online dissemination of primary data offer great opportunities for making archaeological knowledge creation more participatory and transparent. However, different strategies in this area involve important trade-offs, and all face complex conceptual, ethical, legal, and professional challenges.

Since Open Context is operated in the United States, the European Union's database protection laws are less applicable. Similar systems in the European Union would need appropriate policies to insure the legal reusability of database content.

Résumé: Les données primaires, bien que sources d'information essentielles pour étayer les récits archéologiques faisant autorité, deviennent rarement de notoriété publique. L'absence de publication des données primaires est également un obstacle majeur pour la préservation du patrimoine culturel et pour les objectifs de gestion des informations culturelles. En outre, l'accès aux données primaires est un élément crucial pour contester les affirmations sur le passé et formuler d'autres interprétations plausibles. En réponse à ces préoccupations, les systèmes expérimentaux ont mis en œuvre un certain nombre de stratégies pour soutenir la publication en ligne des données primaires. La diffusion de données en ligne peut être un outil puissant pour répondre aux besoins des professionnels de gestion des informations culturelles, établir une meilleure communication et développer des liens de collaboration avec des collègues dans les milieux universitaires, ainsi que pour favoriser la participation du public envers la documentation du passé reposant sur des recherches sérieuses. Cette étude présente la technologie de référence ArchaeoML et sa mise en œuvre dans le système Open Context. Comme on le verra, l'intégration et la diffusion en ligne de données primaires offrent de grandes possibilités pour rendre la création des connaissances archéologiques plus participative et transparente. Toutefois, les différentes stratégies dans ce domaine impliquent d'importants compromis, et tous font face à des défis complexes d'ordre conceptuel, éthique, juridique et professionnel.

Resumen: Aunque los datos básicos constituyen un recurso de apoyo esencial para las narrativas arqueológicas acreditadas, rara vez entran en el registro público. La no publicación de los datos básicos es un obstáculo importante para preservar el patrimonio cultural y los objetivos de la gestión de recursos culturales. Es más, el acceso a los datos básicos es clave para responder a reclamaciones sobre el pasado y para formular interpretaciones alternativas y creíbles. En respuesta a estas preocupaciones, sistemas experimentales han puesto en marcha una serie de estrategias para fomentar la publicación en línea de los datos básicos. La difusión de los datos en línea puede ser una poderosa herramienta para cubrir las necesidades de los profesionales de la gestión de recursos culturales, establecer una comunicación mejor y nexos de colaboración con los compañeros en los entornos académicos y fomentar el compromiso público con el registro documentado del pasado. Este trabajo presenta el estándar ArchaeoML y su implementación en el sistema Open Context. Según se argumenta, la integración y la difusión en línea de los datos básicos ofrecen grandes oportunidades para lograr que la creación de conocimientos arqueológicos sea más participativa y transparente. No obstante, las distintas estrategias en este campo implican importantes pros y contras, y todas encaran complejos retos conceptuales, éticos, legales y profesionales.

KEYWORDS

Archaeology, Primary data, Cultural resource management, Intellectual property, Open Data, Grey literature

Primary archaeological data has received little theoretical attention partially because such datasets typically see minimal exposure, especially in academic settings. Primary datasets are arguably of the greatest public importance for cultural resource management (CRM) because they are evidence of regulatory compliance and proper curation practice. Yet in this case, too, primary datasets are largely relegated to the appendices of project reports that remain as 'grey literature' because they are rarely published or otherwise disseminated. The creation and maintenance of the often voluminous tables and cumbersome spreadsheets that are primary datasets are generally regarded as background processes. Cost and time-constraints work against their formal publication, reproduction and distribution. Nevertheless, primary data remains a vital and sometimes contested aspect of archaeological knowledge production. The organization, dissemination, and ownership of primary data all help shape interpretive possibilities in the discipline. The role and purpose of primary data also touches on important ethical and conservation issues. Lack of access to primary data negatively impacts the archaeological community's capacity for research, cultural heritage preservation and public education. Economic and efficient methods of primary data dissemination, capable of capturing and distributing complex sets of non-standardized documentation, are urgently needed to support collective advances in archaeological evidence and interpretation. At the same time, these methods must respect the legal and ethical concerns of a variety of stakeholders in the archaeological data.

This paper explores approaches to data sharing and data integration based on the implementation of different "meta models" that can be used to organize archaeological information. Such models have varying levels of semantic content, and as such, have different practical, usability, and theoretical implications. To explore these issues more thoroughly, this paper will focus discussion on the Archaeological Markup Language (ArchaeoML) (Schloen 2001; http://ochre.lib.uchicago.edu/index_files/ArchaeoML_Schema.htm). ArchaeoML describes a very expressive and generalized data model. It sees implementation in a number of different systems, particularly the

University of Chicago OCHRE project and the *Open Context* (www.opencontext.org) project. While ArchaeoML is not a "universal" solution for archaeological data sharing, it is sufficiently generalized to support wide application. As such, ArchaeoML-based systems offer valuable test-beds to explore the various conceptual and practical challenges associated with data sharing.

Examples from *Open Context* are used here to highlight how some of the theoretical, practical, and incentive issues involved in CRM data dissemination. *Open Context*'s main purpose is to help make primary archaeological datasets available on the Web for the long term, along with textual narratives and media (images, maps, drawings, videos). A second related goal is to make the publication of primary field documentation more attractive to researchers by facilitating data discovery and use and by situating data dissemination within familiar scholarly publication norms and practices. However, as discussion of *Open Context*'s development efforts will show, data sharing is a complex and theoretically challenging goal. Information models and standards, user interfaces, professional incentives, intellectual property, and the larger ecosystem of worldwide information systems all impact interpretive possibilities in archaeology. Technology by itself does not necessarily expand interpretive possibilities. Some information systems may severely constrain how the past can be represented and how the past can be viewed. Thus, data models, standards, access and intellectual property issues, all have critical theoretical importance to the discipline.

Finally, it should be noted that the bulk of this discussion centers on primary data dissemination, not preservation. Data preservation is a closely related issue, but is largely beyond the scope of this paper. The pioneering activities of the Archaeology Data Service (ADS) (Richards 2004) and, more recently, the US-based Digital Antiquity (Kintigh 2006; Snow et al. 2006; McManamon and Kintigh 2010) initiative have made great strides in archaeological data preservation. The main thrust of this paper, with its focus on data access and use, touches upon data preservation only in that datasets that are available for duplication and reuse are more likely to stand the test of time (Reich and Rosenthal 2001). Though data longevity requires sustained institutional commitments, access and reuse help justify commitments toward digital preservation.

Documentation and Cultural Resource Management

The activities of CRM archaeology produce vast quantities of rich documentation about the past. In contrast to archaeology that takes place in university contexts, CRM archaeology tends to be more highly regulated and operates under more formal reporting requirements Briefly, federal [such as the National Historic Preservation Act (NHPA) and National Environmental Policy Act (NEPA)], and state [such as the California Environmental Quality Act (CEQA)] laws require federal and state agencies that either conduct land development projects or otherwise fund, permit, or approve projects to consider the effects of their undertakings on historic resources. In order to satisfy this requirement, archaeological studies can be commissioned to identify and evaluate sites within a project area. CRM archaeology typically proceeds as a three-stage investigation: the identification of historic properties and the evaluation of properties by applying the National Register criteria; and, if a property is determined eligible and the adverse effects cannot be avoided or minimized, mitigation of adverse effects is often accomplished through data recovery which may involve full or partial excavation of a site. Each stage is designed and required to answer specific, pre-determined questions and therefore the corresponding reports tend to be standardized in terms of format and content (though methods, theoretical perspectives, and recording practices may differ widely from project to project, see below). Project reports usually include maps, a description of methods used and results, and artifact catalogues, among other elements.

Despite the important and accumulating body of CRM-generated archaeological data, a number of factors limit the comprehensiveness and detail of CRM reporting as well as its dissemination and further synthesis. First, time and cost are often critical parameters for archaeologists working within a competitive bidding environment. Projects must be completed on time and on budget and this may constrain how much detail will be sampled, recorded and reported. Secondly, the CRM archaeologist's primary function is to provide data and advice to the agency or client that is obligated to perform the work (with very different consequences, depending on the stakeholders involved). The report is thus subject to comment and revision by supervising state and federal authorities. Further restrictions on data dissemination are imposed due to the need to protect sites from looting or vandalism. Finally, there are relatively few financial incentives for CRM archaeologists to publish or otherwise disseminate their data beyond the required agency reporting, resulting in a large body of "grey literature." Differing ideas among practitioners about what is required and appropriate for "compliance" further inhibit dissemination. Moreover, as Seymour notes in this volume, there are differences in content and format of contract reports versus journal articles that discourage attempts to publish and often lead to rejection of such efforts by reviewers unfamiliar with this difference. All of these factors work to limit the amount of information that reaches the public and, ultimately, its impact on archaeological knowledge.

The Shift Toward Open Access

Declining costs of Internet connectivity, digital storage, processing and software mean that global dissemination of Web-based content is nearly free. These economic realities coupled with new social movements, in particular, the "Free Culture" movement and related "Open Source" communities, have made possible large-scale social and collaborative information production. The Wikipedia (http://www.wikipedia.org), a vast information resource (of sometimes uneven quality, see Duguid 2006) created through volunteer collaboration, today stands as one of the top ten most accessed of all websites on the Internet (Benkler 2006:70–72). The Wikipedia is important in how it helps illustrate the impact advantages of free and open access (OA). However, other aspects of the Wikipedia, especially its fluid nature and absence of "write protections" make it problematic as a publishing platform for researchers. Nevertheless, in demonstrating the reach and impact of free information, the Wikipedia represents an excellent model for the research community to consider.

Researcher communications, too, are rapidly changing as expectations, economics, and dissemination channels evolve. While some scholarly communication models have elaborate permission systems that restrict content to subscribers, open access models dispense with permission systems and offer free Web-based access to scholarly content. Traditional peer review vetting systems see continued use in *both* closed access and OA models. Open access practices are more widely adopted in some areas of the natural sciences, particularly physics. Many policy makers and funding bodies are supporting a move toward OA by increasingly requiring some form of OA to publicly supported research. Efforts to maximize public engagement with research, as well as the effectiveness of funding partially motivate these policy changes. Open access publications routinely see greater citation impact rates than restricted access papers (Harnad and Brody 2004; Hajjem et al. 2005; Brody et al. 2006). Through legislative mandate, the US National Institutes of Health now requires OA to preprint versions of peer review papers that result from government-supported biomedical research.

The economics of CRM archaeology derive from the compliance requirements of federal, state, and sometimes local heritage preservation laws that are met through project funding from both private and public sources. As in the case of other areas of publicly mandated or supported research, similar public policy arguments can be made in favor of OA and transparency to CRM archaeology. In the US, the national government sponsors the creation of many significant datasets as diverse as weather, health statistics, census data, geological mapping, forestry statistics, economic data, and transportation data. All of these government-produced datasets are released into the public domain, though often at a reduced level of resolution or specificity so that privacy can be protected. In particular, for CRM archaeology, site protection (from looting and destruction) is a vital concern, meaning that access to certain kinds or levels of information, especially with respect to location, must remain restricted and are not subject to the Freedom of Information Act.

Challenges in Documenting, Preserving, and Sharing the Past

Open access models are proliferating, not only for sharing traditional forms of scholarly production (peer-reviewed papers), but also among new forms of content, especially databases and media archives. New analysis and recording tools, such as electronic distance measurement devices (EDMs), global positioning systems (GPS), digital cameras and video recording, as well as the growing popularity of handheld data entry devices, mean that the practice of archaeology increasingly results in ''born digital'' documentation. The proliferation of such tools as well as the continued decline in storage costs help fuel this drive for more comprehensive field recording and documentation. Besides making distribution highly cost-effective, the Internet is a powerful means to share large collections of rich media and complex data. These types of content are important components of both museum collections and excavation documentation. Many museums now display portions of their collections online and some research projects have online databases documenting their excavation and survey results. Çatalhöyük (http://www.catalhoyuk.com/database/catal/), the Digital Archaeological Archive of Comparative Slavery (http://www.daacs.org/), and many other projects have a rich online presence. The CyArk 3-D Heritage Archive Network (http://archive.cyark.org/) provides a searchable archive of free 3-D scans and maps of World Heritage sites. The pioneering Perseus Digital Library (http://www.perseus.tufts.edu/) has a rich and ever growing collection of texts, images, and other media for Classical studies and other areas, while the Cuneiform Digital Library (http://cdli.ucla.edu/) makes an impressive collection of early Near Eastern texts openly accessible. The public is getting involved as well. For instance, the commercial photo-sharing site Flickr (http://www.flickr.com) currently has over 50,000 photos of items in the British Museum, contributed by public enthusiasts fascinated by the historical and aesthetic achievements of the past.

In spite of these recent developments, archaeologists still lack the means to easily share their field research. The impediments toward more open and comprehensive dissemination include a variety of professional, conceptual, and technological challenges that are common to many ''small science'' domains. Small science typically works with very case-specific research questions, often using customized methodologies and recording systems and individually maintained data resources (Borgman et al. 2007). Similarly, archaeologists typically adhere to few specific methodological or recording standards, and often make their own customized databases to suit the needs of their individual research agendas (Baines and Brophy 2005). If anything, the particular nature of archaeology, a discipline that straddles the humanities, social sciences, and natural sciences, further encourages a diversity of documentation needs and methods. Data, evidence, interpretations, and syntheses all have very different roles across this diverse community (for discussion of social science material, see Paterson 2003). Also, practical and budgetary factors external to scientific aims are very important in shaping documentation strategies. In the case of CRM archaeology, excavation sampling strategies and the types of laboratory analyses conducted, may all be shaped by construction timelines and imperatives, permitting requirements, property owners, and community interest groups. Archaeology's (and museum studies') early development, often in colonial contexts, further colors and shapes conflicts over some archaeological classification systems (Barringer and Flynn 1998; Bowker and Star 2000). As a consequence, archaeological excavation results, specialist analyses and museum collection databases are highly variable (Kintigh 2006).

In spite of the highly situated and case-specific nature of small science research, field data often promise rich and under-realized interpretive potential. An example from the biological sciences helps to illustrate this point. In 1898, Hermon Bumpus published a landmark study on the evolutionary process of stabilizing selection by investigating mortality among house sparrows. Unlike most of his contemporaries, he comprehensively published his primary observations along with his theoretical interpretations. His set of raw data has proven to be tremendously valuable to later researchers, and has helped inspire the publication of many (some highly influential) peer-reviewed papers. If one measures the value of raw data by the number of publications it spawns, then sharing this set of raw data made it at least ten times more valuable than it would have been without dissemination. Such reuse is likely to increase dramatically if the raw data are made available over general public networks such as the Internet (Kansa et al. 2005). The Bumpus dataset has even more value if we consider how useful it has proven for student instruction and exploration of ''real world'' evidence.

Because of the variability of small science data, databases need extensive documentation for others to decipher their contents. This type of documentation is often called ''metadata,'' a term that is commonly defined as ''information about information.'' Metadata, such as titles, keywords, author and catalogue numbers, enable library users to find relevant publications. The United Kingdom-based ADS has made impressive strides in defining the metadata requirements for managing archaeological data. Metadata documentation associated with archaeological datasets can help others find and decode those data. However, many large archaeological databases have complex structures and include hundreds of thousands of individual records created by multidisciplinary teams. If a dataset needs to be downloaded and deployed on appropriate software, it will still be very difficult to use even with adequate metadata documentation. Once deployed, users will have to familiarize themselves with a project's database organization and interface. The steps involved in downloading and deploying such databases require far too much effort for casual browsing and searching. Thus, making datasets available for download (even with adequate metadata) is not an ideal solution for archaeological communication if the data are not easily "digestible" by others.

In an effort to avoid cumbersome data downloads, many data dissemination initiatives have turned toward providing access to databases dynamically via the Web. Web-based access helps to make content easier to browse and explore because they require no special software or downloads of large complex files. Unfortunately, this typically requires complex and expensive custom Web development. Only a handful of well-funded projects offer access to databases of primary results via the Internet. Thus, observations on thousands of bones, seeds, potsherds, lithics, and other artifacts and ecofacts, including maps, photos, and log entries associated with a typical project, almost never see publication beyond summarized forms. Dissemination support for these small, under-resourced projects is an important goal. Ideally, even small project datasets should be available for casual inspection and analysis without requiring the user to download large data files or launch special software. This kind of access requires some level of data integration and a Web-based infrastructure that can enable dynamic interaction with pooled datasets.

Accommodating Data Diversity: The ArchaeoML Data Model

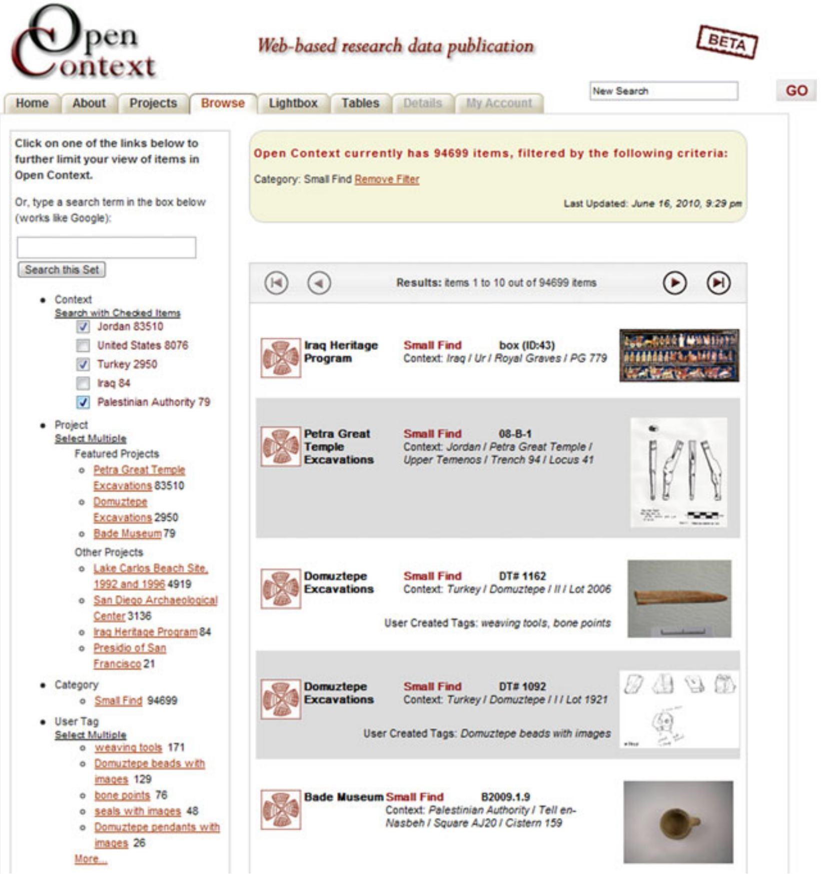
To be cost-effective, dissemination systems need to work for more than one project. However, in doing so, dissemination systems must accommodate the wide diversity of archaeological recording standards. A database is a model or representation of observed and interpreted reality. Such models help to organize and guide interpretations (Kansa 2005).

One strategy to accommodate such diverse archaeological data is to build data dissemination systems around generalized and abstracted data models. ArchaeoML provides a common and highly abstracted framework for expressing archaeological observations, their descriptive properties and their contextual relationships (Schloen 2001). To achieve this flexibility, ArchaeoML has a very generalized, item-based information model. Individual atomic units of observation are related to each other and their descriptive attributes. Each item does not belong to a predetermined observational class (pottery, bone, deposit, grave good, etc.), but is, instead, an abstract entity that has descriptive properties and different forms of linking relations with other items. ArchaeoML's key features include:

- *Flexibility in scale* An ArchaeoML item can be any type of archaeological observation at any scale, ranging from a region, to a site, to a specific deposit, to an artifact, ecofact, or even microscopic observation. Each item has its own unique label (site name, context ID, bone ID, etc.) created at the discretion of the researcher.
- *Flexibility in description* Similarly, the names, terminologies, and values of the descriptive properties of each item are also created at the discretion of the individual researcher. For instance, one is free to describe the composition of pottery with a property like fabric, ware type, or any other set of variables. In other words, descriptive variables and terminologies are left to the researcher's discretion, and are not hard-coded into the data structure. Multiple media, including video, images and GIS can be used in addition to alphanumeric text to describe specific items.
- *Accommodation of heterogeneity* ArchaeoML allows new descriptive variables to be tailor-made for a specific unit without changing the descriptive framework for a whole class of artifacts. Researchers can create new observational criteria and descriptive properties very easily if they encounter unexpected or unique items.
- *Multiple observations and observers* ArchaeoML easily represents multiple observations (even contradictory ones) made on a single item. Each observation is individually authored, thus making the process of knowledge construction transparent. This feature also enables ArchaeoML to represent multiple descriptions of items created for multiple purposes. Museum catalogue data and archaeological contextual observations and descriptions can coexist on the same system.
- *Representation of contextual relationships* Extrinsic contextual relationships in ArchaeoML organize the mass of individual items into archaeologically meaningful structures. These relationships include spatial hierarchies (some items contain smaller items, which contain even smaller items), stratigraphic relationships of sequences of deposition (shown graphically), and relationships of spatial adjacency. These archaeologically meaningful structures (many of which are recursive)

provide the framework that guides searches and analytically powerful queries (see Figure 1). Users also have the option to define their own customized types of relationships.

The primary intent behind the ArchaeoML data model is to provide a useful and widely applicable data structure on which to build archaeological data management tools. As discussed, ArchaeoML models archaeological data in small "atomic" units. Because of this feature, it is easy to publish archaeological collections on the Web in a way that facilitates reuse. *Open Context*, an editorially supervised OA data publication system implements ArchaeoML. Each item in *Open Context* has its own URL (Web address). This "one URL per potsherd" approach makes it easy to



Click on one of the links below to further limit your view of items in Open Context.

Or, type a search term in the box below (works like Google):

- Context
  Search with Checked Items
  ☑ Jordan 83510
  □ United States 8076
  ☑ Turkey 2950
  □ Iraq 84
  ☑ Palestinian Authority 79
- Project
  Select Multiple
  Featured Projects
  - Petra Great Temple Excavations 83510
  - Domuztepe Excavations 2950
  - Bade Museum 79

  Other Projects
  - Lake Carlos Beach Site, 1992 and 1996 4919
  - San Diego Archaeological Center 3136
  - Iraq Heritage Program 84
  - Presidio of San Francisco 21
- Category
  - Small Find 94699
- User Tag
  Select Multiple
  - weaving tools 171
  - Domuztepe beads with images 129
  - bone points 76
  - seals with images 48
  - Domuztepe pendants with images 26

  More...

Open Context currently has 94699 items, filtered by the following criteria:

Category: Small Find Remove Filter

Last Updated: *June 16, 2010, 9:29 pm*

Results: items 1 to 10 out of 94699 items

Iraq Heritage Program | Small Find box (ID:43)
Context: *Iraq / Ur / Royal Graves / PG 779*

Petra Great Temple Excavations | Small Find 08-B-1
Context: *Jordan / Petra Great Temple / Upper Temenos / Trench 94 / Locus 41*

Domuztepe Excavations | Small Find DT# 1162
Context: *Turkey / Domuztepe / II / Lot 2006*
User Created Tags: *weaving tools, bone points*

Domuztepe Excavations | Small Find DT# 1092
Context: *Turkey / Domuztepe / I / Lot 1921*
User Created Tags: *Domuztepe beads with images*

Bade Museum Small Find B2009.1.9
Context: *Palestinian Authority / Tell en-Nasbeh / Square AJ20 / Cistern 159*

Figure 1. A view of Open Context, showing how the faceted-browse tool at the left can be used to filter for items of interest

reference archaeological data at a very granular and specific level. This enables other systems to annotate and reference *Open Context* data, perhaps using some domain ontology (a formal conceptual and classification system, such as the CIDOC-CRM, see Doerr 2003; Doerr and Iorizzo 2008; Lampe et al. 2008) or with Web 2.0 user generated tags (Bearman and Trant 2005; Trant 2006; Boast et al. 2007; Kansa et al. 2010). As discussed below, this granularity is also important for bibliographic citation of data. Thus, ArchaeoML-based systems can support useful data dissemination functions that are flexible enough to meet the needs of diverse user communities in archaeology.

ArchaeoML Schema Mapping

Open Context, is an open source, free, web-based system that serves as a data-publishing tool for individual researchers and small institutions. The generally limited technological expertise of its users requires that standards for data archiving not be overly complex or prescriptive. Because ArchaeoML data structures are highly abstracted and generalized, mapping a given project's database schema into the ArchaeoML global schema implemented by *Open Context* is relatively simple and fast. *Open Context* provides a data-mapping tool, called "*Penelope*," which guides data contributors through the process of uploading their content into the system. The *Penelope* import tool guides users through a mapping of their data schema to ArchaeoML and also gathers high-level descriptive metadata (Dublin Core elements) about the content (Kansa 2007).

To facilitate mapping into ArchaeoML in *Open Context*, *Penelope* guides individual data contributors through a step-by-step process to classify each field in their legacy data table according to the above schema. Each step has a manageable level of complexity. *Penelope* provides users with immediate dynamic feedback that illustrates the effect of selected mapping parameters. The immediate feedback helps users learn how the mapping works and correct mis-mappings as they occur. For example, in one step, the user is asked to describe spatial containment relationships in their imported dataset. After each relationship is defined, *Penelope* generates an example spatial hierarchy tree that illustrates containment relationships that the user defined. To help troubleshoot poor mappings, the user has the option to undo mapping and data imports that were unsuccessful.

Finally, once a contributor has finished importing datasets and has corrected errors, *Penelope* provides a form for the user to provide some standard metadata about their project. This metadata includes Dublin Core elements, as well as some more discipline-specific metadata promoted by DigitalAntiquity.org.

Penelope stores data that has been mapped and imported in a MySQL relational database. Currently, the import process allows users to map and import one table at a time, but complex excavation datasets can also be accommodated by joining multiple tables. For example, some 15 years of field data collected by the Brown University excavations at Petra were prepared and successfully mapped into *Open Context* over the course of 2 weeks. This dataset included some three Filemaker relational databases, and nine other single table spreadsheets. Several days were spent in data clean up and in associating images with specific locations and objects, a time consuming process because these images were described only by their filenames according to no clear conventions. This large-scale effort stands in contrast to a small, single-spreadsheet project, which can take only an hour to describe, map and upload via *Penelope*.

Open Context works under a "data sharing as publication" model, and all data sees editorial vetting and review. Data mapping and publication requires some time and some moderate training in the ArchaeoML data model and in the *Penelope* software application. Data clean up and editing also requires effort. Funding streams are needed to support this efforts. Currently, these editorial processes are funded through grant support. However, *Open Context* will charge publication fees to support editorial and data preservation services (offered through the California Digital Library (CDL), see below). New National Science Foundation mandates require grant-seekers to include data access plans in proposals. This new requirement incentivizes use of *Open Context* as a data publication service, even if this service requires publication fees. In fact, several pending NSF grant applications include *Open Context* publication fees in their budgets. In the case of CRM, government agencies can make similar regulatory decisions to require use of fee-for-service data access and curation services. Without such regulatory pressure, CRM archaeologists will admittedly have weaker incentives to pay for data publication. Though "sticks" such as regulatory mandates will provide powerful incentives, the discussion about citation (below) describes some "carrots" that can also motivate openness with data in the CRM community.

Interestingly, over time, the schema mapping process via *Penelope* has great potential for becoming an automated or at least partially automated process. *Penelope* records information about how each contributed dataset maps into the ArchaeoML global schema. As the number of mapped datasets grows, it is likely that statistical commonalities in schema mappings will emerge. In this case, future versions of *Penelope* may be able to recommend schema mapping parameters based on previously imported datasets. Automating or partially automating the process of schema mapping will lower labor costs associated with publishing datasets in ArchaeoML-based systems.

The Open Context Experience: Working with Pooled Data

Once users have uploaded project data into the system via *Penelope*, *Open Context* pools the data with that from other projects and collections in the system. Because each project has its own idiosyncratic recording system, making a system that is intelligible to users is a major concern. To work with diverse datasets, *Open Context* currently offers several navigation, search, and query tools and features through a standard interface. However, user experience evaluation may find that the standard interface may be inappropriate or confusing to some users. Some user communities may have more need for data analysis tools, while others may have more need for easy browsing and retrieval of images or other media. Detailed analysis of server statistics, together with extensive qualitative user experience evaluation, will guide future revisions of *Open Context*'s interface to best meet the diverse user needs.

Faceted Navigation

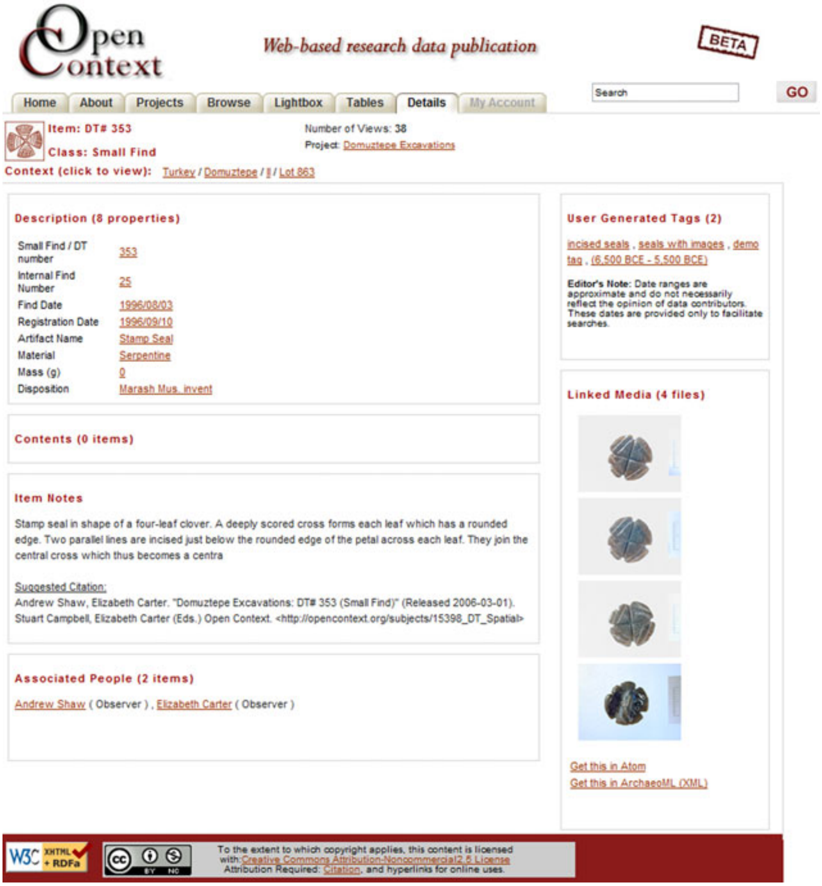
Searching and retrieving relevant information from large bodies of complex data is a challenge for many digital libraries and information services. Keyword searches are common solutions to this problem. However, keyword searches often yield incomplete and ambiguous results. This kind of uncertainty is particularly problematic for professional research applications, since the "hit or miss" nature of keyword searches adds a critical element of uncertainty to information retrieval.

To avoid some of the difficulties associated with keyword search, *Open Context* offers a "faceted search" system. In faceted browsing applications, users navigate through hierarchically structured metadata to progressively select more specific information from a larger collection. Because such filters are applied across an entire collection, users have greater certainty in the comprehensiveness of their results than they do with keyword searches. Navigation involves simple and intuitive "point and click" selection through increasingly narrow filters, allowing users to hone in on their desired results. Feedback, in the form of subtotals for the numbers of items that fall under each available facet, helps guide users in the selection of additional filters. This feature helps give users a good overall understanding of important characteristics of the particular filtered collection they are viewing. Thus, faceted navigation offers users important information cues about the size and composition of the collection they are searching. The feedback offered with faceted navigation is important for navigating and understanding complex archaeological data resources (Ross et al. 2007).

Faceted navigation is based on the organization of collections according to a common data structure. The types of facets available for navigation will depend on the nature of the data structure. The ArchaeoML data structure is very well suited to enable a great deal of fine-grain control and flexibility in information retrieval. As is the case with the recent "Archeotools" faceted navigation developed by the ADS, ArchaeoML enables *Open Context* to offer faceted filtering of content based on project or collections level metadata. In addition, because ArchaeoML represents each project and collection dataset in the same way, *Open Context* offers facets based on the contextual and descriptive properties of items *within* each project and collection (see Figure 1). In other words, *Open Context*'s facet navigation enables users to selectively discover and filter records of individual items (contexts, artifacts, ecofacts, etc.) that are contained within diverse projects. Thus, using *Open Context*'s faceted browser tool, users can seamlessly and simultaneously navigate *between* and *within* individual projects and collections.

Mashups and Alternate Visualization

Faceted navigation offers a potentially useful strategy for enhancing the interoperability of distributed systems (e.g., systems from many sources can work together). User-selected views of the faceted navigation tool do not have to be expressed in a human readable webpage. They can also be expressed in other formats (particularly Atom and KML), that better lend themselves to machine processing. Expressing data in these formats can facilitate use of *Open Context* content in other applications. The use of Atom can also make aggregation of content from multiple sources besides *Open Context* easier. Such portability can encourage user created "mashups" or ad hoc juxtapositions and recombinations of content from various sources. Such capabilities give users far greater flexibility to explore and create new meanings with archaeological data (Kansa et al. 2010; Kansa and Bissell 2010). For example, although *Open Context* itself has some limited visualization capabilities (see Figure 2), other online applications have more sophisticated visualization capabilities. Data portability can enable *Open Context* users to use these other powerful visualization tools. Thus, *Open Context* makes data available in KML to enable visualization in *GoogleEarth* (see Figure 3). Viewing *Open Context* content in *GoogleEarth* can be particularly useful for instructional purposes because *GoogleEarth* aggregates spatially referenced content from many other sources on the Web. For example, archaeological data can be shown along with geo-referenced photos shared by tourists. This can offer a potentially valuable approach toward exploring relationships between popular and professional representations of the past.



Item: DT# 353
Class: Small Find
Number of Views: 38
Project: Domuztepe Excavations
Context (click to view): Turkey / Domuztepe / II / Lot 863

Description (8 properties)

Small Find / DT number	353
Internal Find Number	25
Find Date	1996/08/03
Registration Date	1996/09/10
Artifact Name	Stamp Seal
Material	Serpentine
Mass (g)	0
Disposition	Marash Mus. invent

Contents (0 items)

Item Notes

Stamp seal in shape of a four-leaf clover. A deeply scored cross forms each leaf which has a rounded edge. Two parallel lines are incised just below the rounded edge of the petal across each leaf. They join the central cross which thus becomes a centra

Suggested Citation:
Andrew Shaw, Elizabeth Carter. "Domuztepe Excavations: DT# 353 (Small Find)" (Released 2006-03-01). Stuart Campbell, Elizabeth Carter (Eds.) Open Context. <http://opencontext.org/subjects/15398_DT_Spatial>

Associated People (2 items)

Andrew Shaw (Observer) , Elizabeth Carter (Observer)

User Generated Tags (2)

incised seals , seals with images , demo tag , (6,500 BCE - 5,500 BCE)

Editor's Note: Date ranges are approximate and do not necessarily reflect the opinion of data contributors. These dates are provided only to facilitate searches.

Linked Media (4 files)

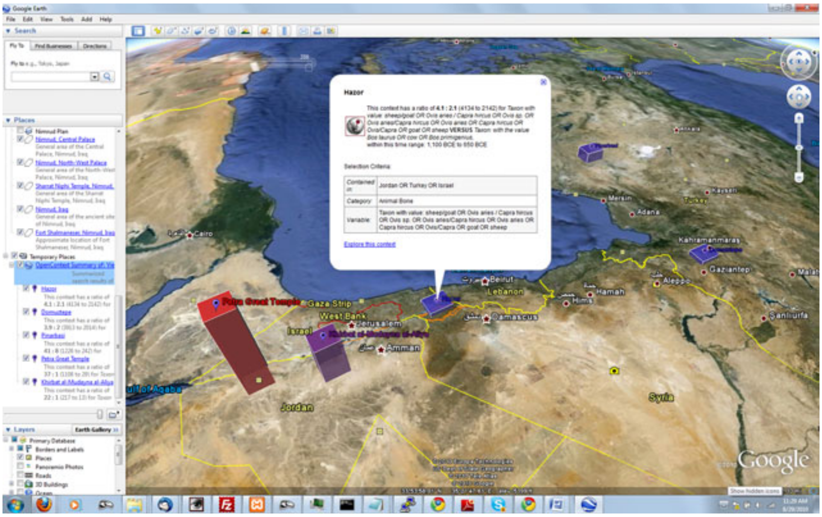
Get this in Atom
Get this in ArchaeoML (XML)

To the extent to which copyright applies, this content is licensed with: Creative Commons Attribution-Noncommercial2.5 License
Attribution Required: Citation, and hyperlinks for online uses.

Figure 2. A detailed view of one item in Open Context, with images and all related descriptive information, including people and user-generated tags associated with the item

Citation and Incentives

Professional incentives help shape archaeological data sharing practice. This is true for both university and CRM professionals. The Society for American Archaeology's ethical code suggests a great deal of interest in promoting openness and the sharing of information (also see Seymour this volume). However, the actual practice of archaeology suggests that the great mass of primary excavation observations and interpretations are seen as proprietary knowledge that is set aside until appropriate opportunities for dissemination present themselves. In particular, many researchers have



Hazor

This context has a ratio of **4.1 : 2.1** (4134 to 2142) for Taxon with value: *sheep/goat* OR *Ovis aries / Capra hircus* OR *Ovis sp.* OR *Ovis aries/Capra hircus* OR *Ovis aries* OR *Capra hircus* OR *Ovis/Capra* OR *goat* OR *sheep* **VERSUS** Taxon with the value *Bos taurus* OR *cow* OR *Bos primigenius*.
within this time range: 1,100 BCE to 650 BCE

Selection Criteria

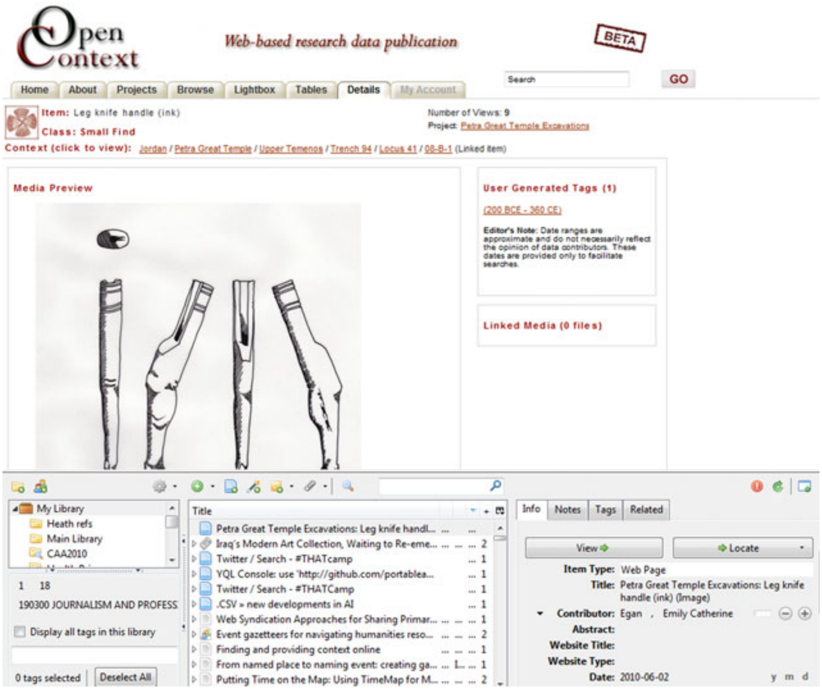
Contained in:	Jordan OR Turkey OR Israel
Category:	Animal Bone
Variable:	Taxon with value: sheep/goat OR Ovis aries / Capra hircus OR Ovis sp. OR Ovis aries/Capra hircus OR Ovis aries OR Capra hircus OR Ovis/Capra OR goat OR sheep

Explore this context

Figure 3. A view of data from Open Context's faceted search tool visualized in GoogleEarth

deep concerns that they will not be properly acknowledged for their contributions unless their research is disseminated in print publication (see Seymour this volume). In recognition of these incentive issues, *Open Context* has tools to generate clear citation information in a form that closely mimics bibliographic information for printed material. The goal is to help fit data publication within frameworks established for other forms of professional contributions. The system automatically generates citation information and a stable hyperlink for each item in the database (Figure 4). Finally, the bibliographic metadata stored in *Open Context* is also expressed using the COinS (ContextObjects in Spans; see http://ocoins.info/) standard. COinS is a micro-format for expressing Dublin Core metadata and is readable by the new *Zotero* (http://www.zotero.org) citation tool. Using *Zotero*, investigators can automatically capture bibliographic information associated with *Open Context* materials. Compatibility with *Zotero* makes using and citing *Open Context* easy and convenient. Finally, Open Context content is archived by the University of California's CDL. The CDL is a key participant in National Science Foundation funded efforts in building "cyberinfrastructure" for scientific data sharing and preservation. CDL services include:

- *Minting and binding of ARKs ("Archival Resource Keys")* ARKs are special identifiers managed by an institutional repository. The CDL will help insure the objects associated with these identifiers can be retrieved in the future, even if access protocols such as "HTTP" change.



Item: Leg knife handle (ink)

Class: Small Find

Number of Views: 9

Project: Petra Great Temple Excavations

Context (click to view): Jordan / Petra Great Temple / Upper Temenos / Trench 94 / Locus 41 / 08-B-1 (Linked item)

User Generated Tags (1)

(200 BCE - 360 CE)

Editor's Note: Date ranges are approximate and do not necessarily reflect the opinion of data contributors. These dates are provided only to facilitate searches.

Linked Media (0 files)

1 18

190300 JOURNALISM AND PROFESS

Title	
Petra Great Temple Excavations: Leg knife handl... ...	...
Iraq's Modern Art Collection, Waiting to Re-eme... ... — ...	2
Twitter / Search - #THATcamp ...	1
YQL Console: use 'http://github.com/portablea... ...	1
Twitter / Search - #THATCamp ...	1
.CSV » new developments in AI ...	1
Web Syndication Approaches for Sharing Primar... ... — ...	1
Event gazetteers for navigating humanities reso... ... — ...	2
Finding and providing context online ... — ...	1
From named place to naming event: creating ga... ... I... ...	1
Putting Time on the Map: Using TimeMap for M... ... — ...	2

Item Type: Web Page
Title: Petra Great Temple Excavations: Leg knife handle (ink) (Image)
Contributor: Egan , Emily Catherine
Abstract:
Website Title:
Website Type:
Date: 2010-06-02 y m d

Figure 4. A view of the use of the Zotero citation management tool to capture reference information for a context from the Petra Great Temple Excavations

- *Data archiving* The CDL also provides data curation and stewardship to maintain integrity of digital data and to migrate data into new computing environments as required.

The University of California provides *Open Context* with a strong institutional foundation for citation and data archiving. The CDL also participates in the DataCite (http://datacite.org), an international consortium of libraries and data publishers establishing standards for the citation of published datasets. Such standards can further enhance the prestige and professional returns of data publication in both CRM and academic contexts.

Citation can be a powerful motivation for organizations and individuals to openly publish data. In some domains, publication of data can enhance the impact of associated papers (Piwowar et al. 2007). Each record in Open Context is associated with the names and even organizational affiliations of responsible analysts. Since Google crawls and indexes all of *Open Context*, these citations are very easy to discover on the Web. As more people reference *Open Context*, the more search-engine "Page rankings" (see Brin and

Page 1998) will improve, further elevating the visibility of research in the system. *Open Context* therefore represents a good solution for people and organizations, now struggling in some obscurity, to gain more exposure for their professional outputs. Thus, though *Open Context* both academic and CRM researchers can publish in a way that builds their public visibility. These "carrots" of positive incentives helped to motivate several organizations and researchers to publish with *Open Context.* Even before the NSF requirement for data access plans takes effect in October 2010, *Open Context* reached a publication rate of about one new project and collection per month.

Copyright and Licensing

Open Context is an OA publication system. All content is freely available on the Web. However, in order to encourage the legal use and reuse of this content, intellectual property issues need to be addressed. The multidisciplinary nature of archaeology complicates intellectual property concerns. Archaeology has one foot in the humanities and social sciences, and another in the natural sciences. As a result, intellectual property tools and conventions that are emerging in some scholarly domains may map poorly onto some archaeological datasets.

Open Context has adopted a policy that allows contributors to retain copyright to their content. This policy is intended to encourage dissemination through *Open Context* by not precluding publication in other more established venues (especially journals and books). *Open Context* currently requires all data contributors to license content with a Creative Commons license. *Creative Commons* licenses give explicit permissions for users to freely and legally use the material so long as they properly attribute the original creator (Brown 2003a). Creative Commons licenses include machine-readable metadata that is captured by commercial search engines such as Yahoo and Google (Kansa et al. 2005). This metadata facilitates discovery of openly licensed content, including *Open Context* resources through commercial search engines.

While search engines are an increasingly important feature that helps shape scholarly communication, intellectual property concerns are another. *Creative Commons* licenses are applicable to copyright protected material. Because a great deal of field documentation relies heavily on written narrative, drawing and photography, much of the content in *Open Context* has a high degree of originality in expression (in the sense of intellectual property law). Such "expressive" forms of field documentation have a high degree of authorial voice. In other words, archaeological data is often a "cultural expression" in its own right, and not simply a set of objective physical measurements. In some ways, the expressive nature of field narratives relates to archaeology's ties with the humanities and also with theoretical trends in archaeology that emphasize reflexivity and view objective truth claims with skepticism. Thus, copyright protections legally apply to much of *Open Context*'s content.

The practice and funding structure of CRM archaeology also impacts intellectual property considerations. In the United States, there is considerable ambiguity about the intellectual property status of archaeological data developed under contract for a commercial client. Many CRM archaeologists work in contract conditions that give clients ownership of archaeological documentation created in the Section 106 process. In these circumstances, dissemination can only occur if clients give explicit permissions. It is only after the Section 106 process is complete can data enter the public domain, but this process may take many years. Delays in dissemination and archiving can have serious implications for digital data that are very vulnerable to loss without active efforts in curation and archiving.

The intellectual property rules for archaeological documentation created under government contract have more clarity. In the United States, data collected under the auspices of government mandates or contracts typically fall into the public domain. Access restrictions, if present, are typically imposed if there is a security or privacy concern. For example, location information that may compromise site protection goals will be restricted. But when data fall within the public domain Creative Commons licenses should not be applied, and the data content should be clearly tagged as belonging to the public domain. Though legally public domain, social norms appropriate for scholarship (see below) should require that intellectual products, such as reports and published datasets must still be cited and credited to the original researchers. In practice, government contracted archeologists are generally assumed to own the copyrights to reports and images in reports and other textual material. Object catalogues, however, are seen as "factual" public domain resources. As for objects themselves, the San Diego Archaeological Center owns the objects themselves for accessioned collections, though some objects are owned by the Federal Government and are curated under contract.

Science Commons, a branch of *Creative Commons* focused on the natural sciences, recently announced an "Open Data Protocol" to deal with this issue. The Open Data Protocol recommends that scientific data repositories use legal instruments to remove this ambiguity by declaring their content to be part of the public domain. In other words, data repositories adopting the Open Data Protocol would renounce copyright (even if applicable) and other protections (such as the European Union's database protections). This new "CC-Zero" declaration, will legally waive copyright on content while requesting, but not legally compelling, attribution for content providers. Instead of using legal means to force citation of data, Science Commons (2006) instead calls on research communities to rely on community social norms. Social norms are an important force in science (and the humanities), and many researchers probably mistake the social norms of their fields with copyright or other legal protections. For example, archaeologists (and other researchers) often publish non-copyrightable facts in traditional journals. These "facts" include counts of species, dimensions of artifacts, etc. Citation is still expected in the use of these published (public domain) facts, even through that expectation has nothing to do with copyright law.

Finally, the intellectual property issues take on added importance and complexity when one looks beyond professional research circles. Various segments of the public, especially indigenous and descendant communities may have strong claims about the past and documentation about the past (Brown 2003b; Hayden 2003; Nicholas and Bannister 2004; Brown 2005). Here the application and perceived benefit of *Creative Commons* licenses are more problematic (Christen 2005; Kansa et al. 2005; Kansa 2009). Standards, including standard licensing tools such as those offered by *Creative Commons*, are not politically and culturally neutral. Standards express and help reinforce particular world-views and agendas (Bowker and Star 2000; Boast et al. 2007). In the case of *Creative Commons*, diverse ideas and concerns over knowledge privacy and custodianship shared among some indigenous peoples may map poorly to these standard licenses.

Conclusions

We are just beginning to demonstrate the feasibility of low cost approaches toward greater accessibility of CRM data. In spite of the technical, ethical, legal, and professional issues that complicate web-based dissemination, the benefits of improved data sharing are compelling. These include:

- An expanded information base to aid public and private agencies and institutions in their efforts to manage and preserve archaeological sites and materials
- Enhanced collaboration among CRM, museum, and university-based archaeologists as well as increased cross-disciplinary collaboration
- Better adherence to best practices as a result of higher visibility and hence accountability

- Greater public engagement and appreciation of local historical and cultural landscapes
- Greater potential for use of cultural heritage management data for other applications, particularly for instructional purposes.

Taken together, these benefits will lead to broader participation in archaeological knowledge creation while at the same time making its processes more transparent.

As discussed in this article, *Open Context* represents an attempt to address the divergent needs of the varying communities in archaeology, as well as the challenge of sharing and integrating diverse archaeological data. To demonstrate its potential to improve collaboration and communication of archaeology, *Open Context* must build a critical mass of data. To do this, users must see an advantage in sharing and that advantage must outweigh the costs, time commitments, and professional fears that currently inhibit greater data transparency. Thus, current developments on *Open Context* aim to build content and tools with the needs of specific research communities in mind (CRM, small museums, specialist groups, and research excavations). Having demonstrated a viable technological platform for data sharing, we must now understand user needs and experience requirements in greater detail so that we can optimize *Open Context* to better meet the day-to-day goals of individual practitioners. In other words, any successful technology for data transparency must be carefully designed to work in the social and professional context of its community of users.

Acknowledgments

The authors would like to thank those involved in the creation of *Open Context*, in particular David Schloen and the University of Chicago OCHRE project for their continued support and partnership. Their efforts have provided the essential conceptual groundwork needed for developing the *Open Context* system. Needless to say, any omissions or errors in this work are solely the fault and responsibility of the authors. Finally, *Open Context* and many other open education initiatives directly result from the enabling financial support of the William and Flora Hewlett Foundation and the National Endowment for the Humanities and Institute for Museum and Library Services. Their support, together with the generous financial contributions of Doris and Donald Fisher and the Joukowsky Family Foundation, help ensure that *Open Context* can serve as a free and open access resource for the community.

Open Access

References Cited

Baines, Andrew, and Kenneth Brophy
2005. What's Another Word for Thesaurus? Data Standards and Classifying the Past. In *Digital Archaeology: Bridging Method and Theory*, edited by Thomas Evans, pp. 236–250. Routledge, London.

Barringer, Timothy J., and Tom Flynn
1998. Introduction. In *Colonialism and the Object: Empire, Material Culture, and the Museum*, edited by Timothy Barringer and Tom Flynn. Routledge, London.

Bearman, David, and Jennifer Trant
2005. Social Terminology Enhancement Through Vernacular Engagement: Exploring Collaborative Annotation to Encourage Interaction with Museum Collections. D-Lib Magazine 11(9). http://www.dlib.org/dlib/september05/bearman/09bearman.html.

Benkler, Yochai
2006. The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press. New Haven, CT. Full text at http://www.benkler.org/wealth_of_networks/index.php?title=Download_PDFs_of_the_book.

Boast, Robin, Michael Bravo, and Ramesh Srinivasan
2007. Return to Babel: Emergent Diversity, Digital Resources, and Local Knowledge. *The Information Society* 23:395.

Borgman, Christine, Jillian Wallis, and Noel Enyedy
2007. Little Science Confronts the Data Deluge: Habitat Ecology, Embedded Sensor Networks, and Digital Libraries. *International Journal on Digital Libraries* 7:17–30.

Bowker, Geoffrey C., and Susan Leigh Star
2000. *Sorting Things Out: Classification and Its Consequences* (2nd ed.). MIT Press, Cambridge, MA.

Brin, Sergey, and Lawrence Page
1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems* 30:107–117.

Brody, Tim, Stevan Harnad, and Les Carr
2006. Earlier Web Usage Statistics as Predictors of Later Citation Impact. *Journal of the American Society for Information Science and Technology* 57(8):1060–1072.

Brown, Glenn Otis
2003a. Out of the Way: How the Next Copyright Revolution can Help the Next Scientific Revolution. PLoS Biology 1(1):e9. http://biology.plosjournals.org/perlserv?request=get-document&doi=10.1371/journal.pbio.0000009.

Brown, Michael
2003b. Who Owns Native Culture? Harvard University Press, Cambridge, MA.

2005. Heritage Trouble: Recent Work on the Protection of Intangible Cultural Property. *International Journal of Cultural Property* 12(1):40–61.

Christen, Kimberly
2005. Gone Digital: Aboriginal Remix and the Cultural Commons. *International Journal of Cultural Property* 12:315–345.

Doerr, Martin
2003. The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine* 24(3):75–92.

Doerr, Martin, and Dolores Iorizzo
2008. The Dream of a Global Knowledge Network – A New Approach. *Journal on Computing and Cultural Heritage* 1:1–23.

Duguid, Paul
2006. Limits of self-Organization: Peer Production and "Laws of Quality". First Monday 11. http://firstmonday.org/issues/issue11_10/duguid/index.html. Accessed 25 June 2008.

Hajjem, Chawki, Stevan Harnad, and Yves Gingras
2005. Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. IEEE Data Engineering Bulletin 28(4):39–47. http://eprints.ecs.soton.ac.uk/11688/.

Harnad, Stevan, and Tim Brody
2004. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. D-Lib Magazine 10(6). Full-text: http://dlib.org/dlib/june04/harnad/06harnad.html.

Hayden, Cori
2003. *When Nature Goes Public: The Making and Unmaking of Bioprospecting in Mexico*. Princeton University Press, Princeton.

Kansa, Eric
2005. A Community Approach to Data Integration: Authorship and Building Meaningful Links Across Diverse Archaeological Data Sets. *Geosphere* 1(2):97–109.

2007. An Open Context for Small-Scale Field Science Data. Proceedings of the International Association for Technical University Libraries Meeting (IATUL07), Stockholm, Sweden, July 2007. Full-text: http://www.lib.kth.se/iatul2007/assets/files/fulltext/Kansa_E_full.pdf. Accessed 1 Sept 2008.

2009. Indigenous Heritage and the Digital Commons. In *Traditional Knowledge, Traditional Cultural Expressions and Intellectual Property Law in the Asia Pacific Region*, edited by Christoph Antons, pp. 219–244. Kluwer Law International, Alphen Aan Den Rijn.

Kansa, Eric C., and Ahrash Bissell

2010. Web Syndication Approaches for Sharing Primary Data in "Small Science" Domains. *Data Science Journal* 9(2010):42–53.

Kansa, Eric C., Jason Schultz, and Ahrash Bissell

2005. Protecting Traditional Knowledge and Expanding Access to Scientific Data. *International Journal of Cultural Property* 12(3):97–109.

Kansa, Eric C., Tom Elliott, Sebastian Heath, and Sean Gillies

2010. Atom Feeds and Incremental Semantic Annotation of Archaeological Collections. In Computer Applications and Quantitative Methods in Archeology – CAA 2010, edited by J. Melero and P. Cano. CAA, Fargo, ND.

Kintigh, Keith

2006. The Promise and Challenge of Archaeological Data Integration. *American Antiquity* 71(3):567–578.

Lampe, Karl-Heinz, Klaus Riede, and Martin Doerr

2008. Research Between Natural and Cultural History Information: Benefits and IT-Requirements for Transdisciplinarity. *Journal on Computing and Cultural Heritage* 1:1–22.

McManamon, Francis P., and Keith W. Kintigh

2010. Digital Antiquity: Transforming Archaeological Data into Knowledge. *The SAA Archaeological Record* 10(2):37–40.

Nicholas, George P., and Kelly P. Bannister

2004. Copyrighting the Past? Emerging Intellectual Property Rights Issues in Archaeology. *Current Anthropology* 45(3):327–350.

Paterson, A

2003. The Design and Development of a social Science Data Warehouse: A Case Study of the Human Resources Development Data Warehouse Project of the Human Sciences Research Council, South Africa. Data Science Journal 2:12–24.

Piwowar, Heather A., Roger S. Day, and Douglas B. Fridsma

2007. Sharing Detailed Research Data is Associated with Increased Citation Rate, PLoS One 2(3): e308. http://www.plosone.org/article/fetchArticle.action?articleURI=info:doi/10.1371/journal.pone.0000308. Accessed 20 April 2007.

Reich, Vicky, and David S. Rosenthal

2001. LOCKSS: A Permanent Web Publishing and Access System. D-Lib Magazine 7(6). http://www.dlib.org/dlib/june01/reich/06reich.html.

Richards, Julian

2004. Online Archives. Internet Archaeology 15. http://intarch.ac.uk/journal/issue15/richards_toc.html.

Ross, Kenneth, Angel Janevski, and Julia Stoyanovich

2007. A Faceted Query Engine Applied to Archaeology. Internet Archaeology. http://intarch.ac.uk/journal/issue21/3/intro.html. Accessed 18 March 2008.

Schloen, David

2001. Archaeological Data Models and Web Publication Using XML. *Computers and the Humanities* 35:123–152.

Science Commons

2006. Scholar's Copyright Project: Background. Sciencecommons.org. Creative Commons and Science Commons. http://sciencecommons.org/projects/publishing/background.html.

Snow, Dean R., Mark Gahegan, C. Lee Giles, Kenneth G. Hirth, George R. Milner, Prasenjit Mitra, and James Z. Wang

2006. Cybertools and Archaeology. *Science* 311(5763):958–959.

Trant, Jennifer

2006. Exploring the Potential for Social Tagging and Folksonomy in Art Museums: Proof of Concept. *New Review of Hypermedia and Multimedia* 12(1):83–105.