The Challenge of Managing Archaeological Databases: Some Issues and Concerns

Maria Sagrario R. Simbulan [1,2]

Abstract

Archaeological databases contain information about archaeological sites – geographical information, type of archaeological operation, the site director, date(s) of the excavations, information about the structures or stratigraphic units in which the artefact, feature, human, animal or plant remains were found, their description, cultural, historical, and date-related information. The purpose of maintaining an archaeological database is threefold. Primarily, archaeological databases allow archaeological institutions to perform their roles as guardians and custodians by providing a repository for the long-term, archival storage of information about the country's archaeological heritage. Secondly, archaeological databases serve as research tools to help archaeologists and researchers in related disciplines such as anthropology, and history. Thirdly, archaeological databases provide the general public access to archaeological information and provide teaching support to teachers and students interested in their country's archaeological heritage.

This paper attempts to focus attention on concerns related to the management of archaeological databases as well as on issues relating to

[1]Assistant Professor 7, University of the Philippines Diliman, Extension Program in Pampanga

[2]Ph.D. Student, Archaeological Studies Program, University of the Philippines

Email: msrsimbulan@upd.edu.ph

archaeological research data residing in these databases. These concerns involve data quality, data access controls, data security, deployment channels, collaboration and continuity. Two important issues will be discussed: How can an archaeological institution balance the need to serve the public, honouring its right to know, with the need to protect archaeological sites and associated artefacts from individuals who seek information for the purpose of stealing our archaeological heritage? A second issue has to do with honouring the intellectual property rights of archaeologists and researchers vis-à-vis the rights of the public to access information generated as a result of publicly-funded research.

Introduction

What is heritage? According to Shanks (2005:166), "heritage can include just about anything—from landscapes to collections, buildings and institutions, living traditions, even impressions and orientations… The important thing about heritage is that it is about relationships with the past. Heritage is what the present values in the past, and the value of the past lies in its contribution to contemporary senses of worth and identity." The term 'heritage' has been used synonymously with 'cultural heritage' or 'national heritage' – a situation that does not sit well with some heritage experts who contend that national heritage as a concept is something totally different and has a meaning of its own. UNESCO (2008) defines heritage as:

> "the legacy of physical artefacts and intangible attributes of a group or society that are inherited from past generations, maintained in the present and bestowed for the benefit of future generations. Cultural heritage includes tangible culture (buildings, monuments, landscapes, books, works of art, and artefacts), intangible culture (folklore, traditions, language, and knowledge), and natural heritage (including culturally significant landscapes, and biodiversity)."

All remains and objects and other traces of humankind from past times are elements of the tangible archaeological heritage. The notion of archaeological heritage includes artefacts, remains, structures, constructions, groups of buildings, developed sites, moveable objects, monuments of other kinds as well as their context, whether situated on land or under water (ICOMOS 2010; UNESCO 2007). As such, this makes our archaeological heritage a finite non-renewable physical and material resource and every effort must be made to accurately document it for scientific study and to serve the needs of future generations.

Background

Archaeological databases contain information about archaeological heritage sites – geographical information, type of archaeological operation, date(s) of the excavations, information about the stratigraphic units in which the artefact, feature, human, animal or plant remains were found, their description, cultural, historical, and date-related information.

It is critical for archaeologists and researchers to manage data effectively as the quantity of data collected on surveys and excavations continues to grow. Dr. Nick Ryan (2004) succinctly described the nature of archaeological data thus:

> "Archaeologists …have to deal with a remarkable variety of different types of information. Archaeology is above all a multi-disciplinary subject drawing on a wide range of skills and specializations, from the arts and humanities through to the biological and physical sciences. From a computer scientist's perspective, archaeological applications provide some significant challenges, one of which is to develop information systems that can cope with this variety. In the processes of research, excavation, analysis and publication, each of the many specializations generates vast quantities of data, much of it of widely differing types, and the challenge is to provide ways in which this can be presented to and used by all who need it."

Given the ever-increasing amounts of information generated by archaeological projects, there is definitely a need to manage the data, to organise data for easy retrieval, and to allow information to be shared and used for publication. It is for these reasons that institutions engaged in archaeological research need to establish and maintain archaeological databases.

The purpose of maintaining archaeological databases is threefold. Primarily, archaeological databases allow archaeological institutions to perform their roles as guardians and custodians by providing a repository for the long-term, archival storage of information about the country's archaeological heritage. Secondly, these databases serve as research tools to help archaeologists and researchers in related disciplines such as anthropology and history. Thirdly, these databases provide the general public access to archaeological information, and provide teaching support to teachers and students interested in their country's archaeological heritage.

Brief overview of databases

A database is simply an organised, structured body of information. A database is not necessarily computerised though. A three-ring binder full of field forms is a database — but one that is cumbersome and difficult to query (Anderson et al. 2011). Banning (2000:62) defines an archaeological database as "a collection of interrelated data, stored with controlled redundancy to serve applications independent of programs that use the data". Users can query, add or modify data, and add or modify the database structure itself. The output consists of responses (onscreen or printed) to user queries, transaction logs (records of database changes), or updated data, and an updated database.

In a short introduction to database structure, Banning (2000:63) states that databases "can be as simple, flat-file databanks, similar to index card files and spreadsheets, or these can be complex, relational databases." A relational database consists of several tables, each of which has a number of fields and contain a number of records, and structured to recognise relationships between stored items of information. Banning also characterised a database field as containing "information on a particular attribute or characteristic of a particular item" while a record in an archaeological database "is analogous to a single file card in a card catalogue: it describes a single site, artefact or context by displaying several related fields".

It is crucial that archaeological research materials – artefacts, environmental and dating samples, field and laboratory documentation, records, notes, catalogues, photographs, drawings, historical documents, and reports – essentially all the data associated with an archaeological investigation, be stored appropriately and remain accessible to researchers and to future generations and be archived in a manner that ensures quick access.

Examples of archaeological databases

The following are four examples of archaeological databases that are accessible free of charge on the internet (all website information quoted below were obtained from the URLs cited):

A. Celtic Inscribed Stones Project

The Celtic Inscribed Stones Project (CISP) is based in the Institute of Archaeology, University College London. CISP's aim is to undertake a collaborative, interdisciplinary study of Early Medieval Celtic inscriptions, with the objective of compiling a comprehensive and authoritative database of all known inscriptions from Britain, Ireland, and Brittany. The goal is to turn what is a largely untapped resource into usable material. The database can be accessed at http://www.ucl.ac.uk/archaeology/cisp/database/.

The CISP database includes every non-Runic inscription raised on a stone monument within Celtic-speaking areas (Ireland, Scotland, Wales, Dumnonia, Brittany, and the Isle of Man) in the early middle ages (AD 400-1000). There are over 1,200 such inscriptions in the database. Information on the stones has been stored in three main databases:

- **Site**: Includes information on the physical character and/or history of the site.
- **Stone**: Includes information on discovery, location, condition, size, form, and decoration.
- **Inscription**: Includes information on legibility, position, script, linguistics. and readings.

Within each of these databases you can find bibliographic references while links to images of many of the stones can be found within the Inscription pages. CISP has given each site, stone, and inscription a 'unique identifier' to aid searching.

The inscriptions are faithfully recorded, whether these are written in the Celtic vernacular or in Latin, using the Roman alphabet or in Ogham script. Information for the database is drawn from published and unpublished works and internet resources. The databases incorporate the details of present and former locations, physical characteristics, readings, previous work, bibliographic info and, if available, images. The digital format has allowed researchers and scholars to periodically update the database as new researches on inscribed stones are published or more current field data becomes available. The CISP database has three subsystems: the Core, the Bibliography, and the Image subsystems. The Core subsystem is composed of five primary tables (Site, Stone, Inscrip, Reading, and Translat), secondary tables, and look-up tables. All in all, the CISP database consists of more than 40 interrelated tables.

B. Arachne

Arachne is the central object-database of the German Archaeological Institute (DAI) and the Research Archive for Ancient Sculpture at the University of Cologne. Arachne (http://arachne.uni-koeln.de) is intended to provide archaeologists and Classicists with a free internet research tool for searching hundreds of thousands of records on objects and their attributes. This combines an ongoing process of digitising traditional documentation (stored on media which are both threatened by decay and largely unexplored) with the production of new digital object and graphic data. In addition to extensive digital image archives integrated with detailed textual information and bibliographic databases, among the collections in the Arachne database are the following:

- The complete catalogue of sculptures in the **Antikensammlung der Staatlichen Museen zu Berlin** (Antiquities Collection of the Berlin State Museums). This database contains photos, descriptions and high quality scans of approximately 2600 Greek, Cypriot, Etruscan, and Roman sculptures in stone, as well as several large bronzes (including extant objects and those now lost).
- The **Foundation of Rhine-Westphalia Economy Archive of Cologne** (RWWA) has a collection of about 15,000 digitised glass plate negatives from the late 19th until the middle of the 20th century. These glass negatives show people at their workplace, production methods, machines, products as well as architectural or industrial plants – research materials important to industrial archaeology.
- The image database of the **Cast Collection of Ancient Sculpture** shows every plaster cast in Berlin representing an ancient object. In addition to data like provenance, state of preservation, and relevant bibliography, extant photographs were digitised.
- The **Hellespont Project** integrates Arachne with the Perseus digital collection at Tufts University, combining the digital collections of classical studies of both institutions into one of the most comprehensive and free online collections of Greek and Roman antiquity available for public and scientific use.

C. International Dunhuang Project: The Silk Road Online

The International Dunhuang Project (IDP) is a ground-breaking international collaboration to make information and images of manuscripts, paintings, textiles, and artefacts from Dunhuang and archaeological sites of the Eastern Silk Road freely available on the WWW (http://idp.bl.uk) and to encourage their use through educational and research programs.

While much of IDP's early work focused on conservation and cataloguing, these have been supplemented with digitisation, education, and research. IDP started digitising the manuscripts in 1997 with the aim of bringing together the collections in virtual space. As of September 23, 2013, the IDP holds 413,832 high-quality images of the manuscripts and other material, integrated with cataloguing and contextual information in its database.

The largest collections of material from the Silk Road town of Dunhuang and the surrounding region are held in libraries, museums, and research institutes in London, Beijing, Paris, St. Petersburg, and Berlin, with important holdings in Japan and smaller collections throughout the world. The geographical diversity of the collections is due to their having been removed from Central Asia by a succession of archaeological expeditions from different countries. These expeditions, which began in the late 19th century, uncovered and explored the ruined temples and settlements in the deserts of Central Asia.

The collections consist largely of items, dating from about 100 BC to AD 1200, including paintings, murals, artefacts, coins and manuscripts, the last in over twenty different languages and scripts. The size and scope of the collections, as well as their fragility and limited access, has meant that, while they constitute a primary research resource for the history and literature of the region, many of the manuscripts have yet to be studied in detail.

These first three examples of online archaeological databases offer datasets of a secondary nature (previously published research data) as opposed to the fourth example (see discussion on Lerna below) which offers primary datasets containing raw findings from actual excavations or field observations. CISP, Arachne, and Dunhuang feature easy-to-use, user-friendly interfaces, open access, and the presence of search tools that help users quickly locate or at least narrow their search for information.

The websites offer downloadable content such as database manuals (CISP), PDFs of journal articles and grey literature (Dunhuang, Arachne), and links to other sources. Multimedia content like maps, images, sound tracks, videos are available (Dunhuang) while user-selected database records and related content can be downloaded from Arachne. Only the Arachne webpage requires a user to login; new users need to apply for access to the site and wait for confirmation that access was granted. Both the Dunhuang and Arachne webpages offer a limited selection of content in other languages; Arachne allows the user to select different language versions of content shown using the various interfaces on its website. While all three webpages are expected to present updated content, the Dunhuang and Arachne webpages seem to offer newer, more contemporary content. Finally, all three online databases do not charge any fees for access to information. The fourth example discussed below, the Lerna pottery database, is also online but does not have a web-based user interface. Instead, the fourteen tables that make up the Lerna database can be downloaded free of charge in several file formats.

D. LERNA

The Lerna material is an example of a database comprised of a set of related tables that is downloadable from the archives of the Center for the Study of Architecture (CSA). The database can be downloaded in four formats: MS Access, Claris Filemaker Pro, ASCII, dBASE 5 from (http://www.csanet.org/archive/adap/greece/lernpot/lernameta.html#Download). While no online user interface is available, the database is freely downloadable and can be used by researchers without any charge.

Lerna, a type site for the pre-Mycenaean periods in southern Greece, was excavated from 1949 to 1959. The database contains information on 365 vessels belonging to the Early Helladic period that were found in Phase I, Level IV.

Fourteen data tables make up the Lerna IV pottery database. Each is a separate file in FileMaker, and each has its own ASCII file. In Access, however, all tables are included in the single Access file. Each table is also related to one or more other tables, and the relationships are crucial to the functioning of the whole. The fourteen data tables in the database are:

- The **Catalog1** table contains data such as the Catalogue number, Class, dimensions of the item (height, rim diameter, base foot), Form Shape, Form Type, descriptions of decorative aspects, Munsell numbers, and descriptions of the interior and exterior surfaces of each vessel.

- The **Classes** table contains data on decorative treatment, fabric, surface treatment and colour.
- The **Context** table contains basic information about site contexts - the phase(s) to which they belong and their relationships to other contexts and the site itself.
- The **FitchAnalysis** table contains results from neutron activation and atomic absorption spectrometry analysis performed by the Fitch Laboratory.
- The **LernaInvNo** table contains inventory numbers assigned to the vessels during the excavations.
- The **Morphology** table contains data describing the range of vessel forms and types current during the Early Helladic III period at Lerna, and illustrations of the vessel forms and types.
- The **Pattern** table contains descriptors of painted decorations along with some illustrations.
- The **PlanElevation** table contains all plans and elevations of Lerna IV Phase I.
- The **PotContext** table contains data on the find spots of the vessels.
- The **PrevPub** table contains references to publications which mention specific vessels.
- The **ProfileTerminology** table contains links to illustrations of Morphology profile descriptors.
- The **Sherds** table contains descriptions of existing sherd types from each catalogue entry.
- The **Syntax** table contains standard decorative schema used in Lerna pottery.
- The **ThinSection** table contains images of thin section petrography done at Temple University.

The preceding four examples of online archaeological databases serve to highlight the diversity of archaeological heritage material that can be stored in databases. This diversity and the significance of such heritage resources make managing these databases a significant challenge to archaeological institutions.

Managing Archaeological Databases

The focus of this paper, an eighteen-year old archaeological institution, is considering the use of digital databases across sub-disciplines within the organisation. Previously, spreadsheets, paper-based filing systems, and early versions of database software (Dbase and Access) have been used to store data generated during the different stages of its archaeological projects. However, no common file format or data storage system has ever been adopted by the institution. Essentially, this means that there are probably as many data storage methods in use as there are archaeological research teams.

The current effort to transition to the use of database software like FileMaker Pro offers the institution a chance to consolidate its research findings and field data into searchable archaeological databases that would include data from the sub-disciplines represented within the institution such as lithics, human osteology, botanical and faunal remains, pottery, and metal studies. However, with 18 years' worth of data, photo, illustrations, and reports to organise and digitise, the institution faces a huge challenge.

Taking steps to institutionalise the use of databases to store information relating to archaeological projects can contribute significantly toward protecting an archaeological institution's body of research. Since it is usual for sites to be excavated over several seasons in the very least, with especially rich archaeological sites exceeding 150 seasons (see Chersonesus, Ukraine), the ability to ensure the long-term accessibility and preservation of archaeological research data is a critical responsibility of archaeological institutions since the very same fieldwork that is the source of artefacts and associated field data also destroys the archaeological site and with it all traces of human activity and settlement (Niven and Thompson 2011). Simply put, you cannot dig a site twice.

Managing access to these databases, instituting procedural controls to protect these databases from malware and viruses, and taking steps to ensure data quality will serve to increase the value of this institutional asset over time; putting in place the technological infrastructure and support mechanisms will set the stage for the long-term curation of the research data, indefinitely extending the useful life of the archaeological data and allowing new generations of researchers and archaeologists to exploit the databases and to validate the data to further their own research agenda. Making these databases available as part of the institution's public archaeology initiatives also ensures that the nation will continue to benefit from the wealth of information found in these databases long into the future.

Dealing with the concerns

Managing archaeological databases is definitely not a walk in the park. There are many concerns that should be addressed by archaeological institutions that want to deploy these databases in a manner that ensures maximum exposure to the public and other academic research institutions, at the same time protecting the databases from unauthorised access or misuse.

Especially at risk are archaeological research databases that contain primary data sets and raw information derived from field observations and analysis. Discussed in this paper are concerns involving data quality, data access controls, data security, deployment channels, collaboration, and continuity. This paper will not cover archaeological database design and data conversion concerns; these topics deserve to be discussed more extensively in a future publication.

Maintaining Data Quality

Data quality has three facets: data accuracy, completeness, and consistency. These facets impact the data's usefulness which in turn determines the quality of the decisions based on that data. Below are some examples of basic questions that need to be answered to ensure that only high quality data are stored in the archaeological databases: Are the field measurements accurate? How precise should your data be? Have all measurements been recorded using the same system of measure? Have all stratigraphic data and data on artefacts been recorded?

Institutions should set up data entry protocols to protect the integrity of their databases. Field forms are the primary sources of inputs to archaeological databases. Great care should be taken to ensure that data points on the forms are completely and accurately filled up by the field investigators. This will facilitate the data entry process by reducing errors and eliminating the occurrence of incomplete database records.

Ensuring completeness is difficult, especially in large and long-running projects. Beyond data entry, post-field processing activities like collecting and organising digital photographs as well as digitising charts, diagrams, and illustrations generate digital files that will be part of a project's databases. Incomplete data will hamper the efforts of the researchers to make sense of the material gathered and will endanger the long-term success of the project. Involving team members who understand the importance of complete documentation will go a long way towards successful project completion and ensuring continued funding for future archaeological projects.

Selecting the Appropriate Information Channels

How should the institution's archaeological databases be deployed? The channels to be used to provide access to archaeological databases have implications on the technology infrastructure the institution needs to invest in and on the budgetary resources it needs (initially and on a continuing basis) to support the maintenance and updating of these databases. The institution can provide access to its archaeological databases either offline (within the premises of the institution or library) or online via the Worldwide Web (WWW). There are significant advantages and disadvantages to using either access channel. Offline access offers greater security and control over the institution's archaeological databases but is available to a limited audience only; online or web-based access offers access to a much wider audience but with much less control over the database and how it is used.

Opting to deploy the databases in-house via its research units offers the institution the best chance of protecting the data assets but severely limits public access to the information, significantly reducing the number of people who can benefit from the information. Permitting access to the archaeological databases via the institution's library facilities may offer a middle ground – using the library's built-in custodial procedures to monitor and control access while affording a bigger audience a better chance to view the contents of these databases.

A better appreciation of the work that the archaeological institution and its researchers have put in to populate these databases is a definite upside to this. Using the institution's library facilities to provide public access to these databases will also send a strong signal that the databases are primarily academic in nature, to serve academic research purposes (Lock 2003).

Interest in catering to an even broader audience brings up a related concern: Is the institution ready to use the WWW as an information channel and make the data in these databases available worldwide?

The question is deceptively simple; being 'ready' involves significant investments in networking technology and security software to safely deploy the databases in an online environment that will afford the institution complete control over the integrity of these archaeological databases and its content while presenting an easy-to-use interface that allow researchers and scholars to explore the databases.

While there are other national organisations tasked with protecting the national heritage and the repositories of information about this heritage, an online presence provides an important contact point between the research institutions maintaining archaeological databases and its public – this is particularly important when incidental finds or archaeological rescue situations need to be reported, or if unaccessioned, personal collections are to be repatriated, surrendered or deposited for safekeeping with the institution.

Persons reporting unauthorised excavations or looting might want to remain anonymous. This should be considered when developing the database's user interface. The user interface should provide the names of the institution's contact persons and their email addresses. Finally, feedback and ideas from viewers are important sources of inputs to improve the databases and access.

Implementing Data Access Controls

To help ensure data quality and to protect data integrity, two different sets of data access rights levels have to be established by the institution's database administrator. The first set involves access permissions for members of the institution (internal users) and the second set involves access permissions for the general public and to interested parties who are not associated with the institution (external users). A possible third set of access permissions might include research collaborators and funding institutions that have a stake in or are currently involved with research projects.

For internal users three levels of control would be ideal. Level 1 rights would grant a data entry operator the right to add new records and to input data into these new records but withhold the right to edit existing records. Level 2 rights would allow a supervisor to authorise staff with Level 1 rights to revise existing records. Level 2 rights holders can execute database backup and restore procedures to and from external data storage devices (see discussion of database protection concerns), extract subsets of databases, and delete database records. Ideally, a user with Level 3 rights can perform all the functions of Level 1 and Level 2 users. In addition, a Level 3 user defines the data access level to be granted to an authorised user.

Regardless of the access channels the institution chooses to use to disseminate the results of its archaeological research projects, being ready to cater to requests from internal and external users means setting up security protocols that establish the identity of the viewer (example: viewer can be Admin, Researcher, or Visitor) and on the basis of this identification, create a list of the viewer's rights vis-à-vis the databases, and enforcing these limitations to what the viewer can or cannot do. For example, as an Admin user, the viewer is permitted to update (Add, Revise, and Delete records) the database in addition to the Query and Browse rights that a Visitor is limited to.

A user identified as a Researcher will not have the right to update the database but is granted the right to view additional sensitive or proprietary information that is not normally shown on the standard browser interface shown to Visitors. Additionally, a running log of the database users/viewers and their activities (additions, revisions, deletions, and queries) will need to be maintained to track all changes to the database.

Protecting Archaeological Databases

The vulnerability of information systems increases as institutions move into a more networked world. Protecting the institution's archaeological databases and digital research materials from both unintentional (human error, environmental hazards, system failures) and intentional threats is essential. Applying Turban et al.'s (2008) description of intentional threats to the research setting, this type of threat can include theft of data (particularly researchers' primary data sets), inappropriate use of data, theft of data storage equipment, deliberate manipulation of data and programs, and destruction from viruses and malware .

Beyond controlling access to archaeological databases, it is important to establish procedures to physically protect the information assets of the institution. For institutions that choose to deploy the databases within their premises or in libraries, an important protective measure is to disable USB ports and CD/DVD drives. This will ensure that no unauthorised copying of databases occurs. This measure also prevents users from introducing viruses and malware into the system. Disabling internet access from the computers designated as data access kiosks will prevent unscrupulous users from sending copies of the databases to themselves or other recipients; this will also minimise the possibility that viruses and malware will get into the system. Choosing to use the Internet as an access channel may offer greater access to a wider audience but also increases the risk of exposure to internet-borne viruses, malware and unauthorised intrusions.

Additional ways that an archaeological institution can protect its information assets involve establishing policies that will govern data backup cycles and procedures, installing and regularly updating antivirus software on all devices connected to the institution's network and putting in place disaster recovery and continuity plans. Creating data backups should be a regularly scheduled activity to protect an institution's information assets. With the continuing drop in external hard disk prices and increasing drive capacities, the institution should not find it difficult to implement an institution-wide data backup policy to create copies of the databases and other related research materials stored in digital format. These backups should be stored in secure locations; ideally, copies of the data backups should be stored in multiple locations in different geographic regions or, if this is not possible, in different buildings in the same geographic region. Internet-based data repositories offer an alternative to the actual physical storage of backups in other locations. The Archaeology Data Service (ADS, http://ads.ahds.ac.uk/), the Digital Archaeological Record (tDAR, http://www.tdar.org/), and Open Context (http://www.opencontext.org/) are some examples of archaeological data repositories. A list of research data repositories is available from Databib (http://databib.org/), a searchable catalogue of online data repositories.

Insisting on the installation, use, and regular updating of antivirus software for all computers and devices that connect to the institution's network is the simplest way to avoid virus and malware infections that could render systems unusable and the storage devices connected to these compromised computers unreadable. A disaster is a situation where an institution is unable to access its information technology and systems resources as a result of fire, water, or other hazards or because of the catastrophic failure of its information systems (IS) due to equipment malfunctions, malicious 'exploits' or through virus attacks.

According to Turban et al. (2008:647), a disaster recovery and continuity plan (DRCP) "outlines the process by which an institution could recover from a major disaster". Developing a DRCP emphasises the willingness of the institution to be proactive about protecting its information and IT assets. A DRCP is essential to any data security system. While this paper will not discuss the DRCP in detail, answers to the sample questions below should give the reader a clear picture of what DRCP guidelines would show:

- What do you need to begin running in event of a computer outage?
- Who will do the data recovery work?
- How does the institution verify that data recovery worked?
- Who needs to be notified?
- What are priorities for recovery operations?

In the event that the institution's databases are destroyed or compromised by virus or malware attacks, the most recent database backup can be restored to provide a starting point for recovering the rest of the data affected by the attack. This is the best argument for instituting a regular backup schedule for all types of digital data, not just databases.

Establishing Ground Rules for Usage, Collaboration and Data Sharing

In an increasingly digital world where great value resides in information and information systems, it is important to be aware of the threats to these information resources. Institutions who choose to share data with the general public or with researchers in other institutions must establish ground rules for access to and usage of these information resources. One of the primary tools is the creation and adoption of an Acceptable Use Policy or AUP (also called Appropriate Use Policy) that any user who wishes to access the information resources of the institution must promise to abide by.

What is an AUP? The AUP is a formal or informal document that defines the intended use of the organisation's computing facilities and information resources, unacceptable uses, and the consequences for noncompliance. It is a document that outlines a set of rules to be followed by users of computing resources, which could be a computer network, a website, or information residing on a database. An AUP clearly states what the user is and is not allowed to do with these resources (Technopedia 2012). AUPs are created with three goals in mind: first, to educate users about activities which may harm the institution; second, to provide a legal notice of unacceptable behaviour and penalties for such behaviour; and third, to protect the institution from liabilities arising from the abuse or misuse of access facilities (Standler 2002). Putting in place these usage policies is an important facet of the institution's total information security program and should be a priority of the database administrator.

Security software should be deployed to periodically remind users of the terms of the AUP as well as to monitor their activities while accessing the information in the databases. A clear statement of access limitations as well as copyright notices should be clearly visible on the user interface. Collaboration and data sharing are also concerns that archaeological institutions must address.

What policies are in place to govern research partnerships and institutional linkages with regard to data sharing? What are the data access rights of researchers from other institutions who collaborate on archaeological research projects? A clear delineation of what is shareable and what is not must be communicated to all members of the institution. Beyond communicating data sharing policies, the institution should draw attention to the benefits of data sharing (Wallis et al. 2013).

Wiseman (2013) argues that 'data sharing' or making data open, is "building momentum to change the traditional approach to research publishing and unlock new research possibilities." While acknowledging that many researchers remain dubious about sharing their research data openly with the wider community, he maintains that doing so can actually bring them and their work wider recognition. He suggests that "when researchers make the data behind their work open, it enables others to use their datasets to enhance their own data, find new information in it or even use it for comparisons against their own work. This saves time, opens a world of opportunities and reduces inefficiencies when basic experiments are repeated unnecessarily. Allowing access in this way can also enable comparisons that have never been possible before, enhancing opportunities for cross-disciplinary research."

Ensuring Continuity

At the heart of this concern is the availability of long-term institutional support and stakeholder buy-in. Institutional support involves administrative support as well as long-term funding support for multi-year initiatives to hire and train technical staff, to setup the requisite technology infrastructure (facilities, hardware, software, and networks), and for the continuing maintenance of archaeological database systems.

Ensuring the continuity of programs to maximise the benefits derived from the institution's archaeological databases depends on factors like adequate numbers of trained staff and archaeologists to use and populate the databases, the availability of database software training for research staff and archaeologists, the continued support of the institution's administrators, the use of legally acquired software that will allow the institution to download updates and upgrades to keep the software current, and, possibly the most important factor, stakeholder buy-in.

Adequate funding will allow the recruitment and training for people who will operate and maintain the equipment on which the databases will reside, and for technical staff who will setup and maintain the databases and the information systems. Resources should be available to train archaeologists and research staff in the use of the databases. Ensuring that the institution's people know how to use databases is the best way to maximise the returns on the institution's investment. Funds should also be allocated for courses on data handling and database use for volunteers.

Stakeholder buy-in is critical because the best designed database systems are useless without people who will utilise the system for their research and who will populate the databases with their research data and findings. To help ensure researcher and staff buy-in, the institution must develop a data policy that reflects the institution's mission statement. The scope, according to Jones (2009), should be clearly defined as to the "type of research outputs covered by the policy and the context in which the policy is to be applied, i.e. across an entire institution or just a single department or research project".

Periodic reviews of the institution's data policy will encourage researchers and archaeologists to contribute ideas and share experiences with an eye towards improving the data policy. Their involvement in the implementation and review process exerts a subtle push towards acceptance and willingness to work within the bounds of the policy.

Resolving the issues

Shanks (2005:165) pointed out that "archaeologists have come to accept an obligation and professional responsibility to share their archaeological knowledge with the public as well as colleagues. And more—to carry out work for public as well as academic interests." However, there are issues that need to be resolved to ensure that the support for the use and maintenance of archaeological databases within the institution is unwavering.

Ownership of data residing in archaeological databases remains a thorny issue, whether an archaeological project is externally funded or supported with an academic institution's funds. Unresolved data ownership issues have resulted in the reluctance or even refusal of researchers to contribute research data to an institution's archaeological databases. This can result in islands of information that are accessible only to the research project team members, discouraging collaboration and data sharing within the institution and with other research institutions.

The primary issue involves finding a balance between the intellectual property rights of the archaeologists or researchers who participate in archaeological research projects and the public's expectation of access to the results of publicly funded research projects. How can the rights of researchers be protected while allowing them to fulfill their obligation to share the research results with the public?

Getting buy-in from researchers wanting to protect their primary datasets is a challenge. The problem lies in effecting a change in researcher attitudes towards data sharing and collaboration. This stems from the reluctance of researchers to share primary data without any assurance that their work (and their sharing of data) will be recognised in the same way that they do for publishing results in journals.

The following three points among several suggestions made by the Stellagroup (2012) seem to offer an equitable resolution to this challenge: First, "the creation of policies for reserving researchers' rights to first publication on primary datasets they have submitted for inclusion in the institution's databases." Second, "that dataset publication / dataset deposition / dataset disclosure be considered a contribution to be weighed in connection with tenure and promotion decisions." Third, "using the preceding suggestion as a carrot for data openness and data sharing to encourage collaboration with other researchers."

Changing this mindset carries with it the potential for creating 'new' knowledge from existing data. Some of the methods of new knowledge creation involve data re-use, data re-analysis, data mining, and combining datasets in innovative ways with research results obtained using new analytical tools and technologies.

An article by Costas, Meijer, Zahedi, and Wouters (2013) provides a perspective on "the current state of data sharing and explores how data publication can be encouraged, recognised, and simplified. The report recommends creating a reward system that will allow researchers to demonstrate the value of their work in an open/shared setting. Alongside this, it notes the need for data-citation standards so that usage of data can be tracked and recorded.

The report advises promoting the positive impacts of data sharing such as increased recognition, reduction in administrative costs, improved reputation, and ultimately the ability to attract the best new researchers."

Taking a longer term view, the institution can adopt a policy requiring new research proposals to include research data management plans. This way, at the outset of any project, the research team agrees to work within the data management framework laid out in the institution's data policy. This data management framework should clearly specify that the results of the research are to be made part of the institution's databases.

The second issue also involves another balancing act – finding ways to enable the institution to fulfill its mandate to provide public access to and to protect the archaeological heritage of the country by establishing and maintaining a facility for long-term data storage and curation of archaeological heritage materials while limiting the access to this resource by entities who seek to take advantage of the detailed information in the databases to locate and steal archaeological materials for monetary or personal gain.

The following quote from George Nicholas (2012: 108-109) on intellectual property issues in archaeology accurately describes the tension between archaeologists, researchers, heritage protection advocates, indigenous peoples, and the public:

> "The intersection of cultural knowledge, research practices, and the public domain produces many challenges for archaeologists and heritage stakeholders relating to intellectual property issues… How archaeologists and others respond to intellectual property issues has the potential to either positively transform research disciplines and their relations with stakeholders, or constrain the quest for equitable and productive research relationships and appropriate sharing of information. The challenges are enormous, and there is no single way to avoid or resolve intellectual property disputes that may arise within academia, the cultural heritage management, or descendant communities—or between any and all of these entities. However, a starting point is to recognize that intellectual property … is an inherent part of all human societies, and that archaeologists working at the intersection of tangible and intangible heritages, are well positioned to provide a fuller understanding of the nature of cultural knowledge and rights and thus facilitate more equitable sharing of information derived from the past."

Torsen and Andersen (2010) suggest that there are no clear cut 'recipes' for an archaeological institution to use to balance its mission to provide the broadest access to its archaeological databases through public archaeology initiatives, information dissemination and education while protecting the country's archaeological heritage from those who would misuse it. The level of interest (or disinterest) of indigenous groups, political will, enforcement concerns, and legal issues relating to data ownership are factors that make this balancing act unique for every situation and every institution. Combining information dissemination efforts and education through public archaeology activities offer a good starting point.

Conclusion

The four online archaeological database exemplars presented (CISP, Arachne, Dunhuang, and Lerna) drew attention to qualities that an institution wanting to setup access to its archaeological databases should emulate. These are ease of use, a user-friendly interface, the simplicity of data access procedures, the absence of fee-based access controls, the presence of website-based search tools enabling users to quickly locate information, the ability to construct user-defined queries that allow users to isolate data subsets for downloading in a range of file formats, and the availability of downloadable content such as scholarly papers and grey literature.

Beyond helping the institution fulfill its obligation to provide access to data in its archaeological databases (whether online or offline), developing a user interface that allows people to communicate with the institution directly provides a particularly important contact point, giving the concerned public a chance to provide feedback and suggestions, correct errors, and report illegal or unauthorised excavations and looting. This also opens the possibility of extending the functionality of the databases' user interface by providing a hyperlink to a registry system that will allow people to notify the institution of incidental finds as well as the location of unaccessioned, illegally obtained artefacts and heritage objects.

Making information in archaeological databases accessible despite the attendant costs and potential risks is offset by the resulting greater public awareness and concern for the country's archaeological heritage. This, in turn, will hopefully generate more interest and vigilance in heritage protection. Sustained interest from government agencies and pressure from indigenous groups and parties interested in protecting the country's archaeological and cultural heritage should provide impetus for stricter legal measures and enforcement protocols that support heritage protection initiatives at the local and national levels.

References

Anderson, D., N. Hermann, S. Miller, and S. Yerka. 2011. Archaeological Information System Standard Design Concepts. Digital Data Interest Group Electronic Symposium, SAA 2011. Accessed 01 November 2012. From http://visiblepast.net/see/archives/1066.

Banning, E. B. 2000. The archaeologist's laboratory: The analysis of archaeological data. New York: Kluwer Academic/Plenum.

Costas, R., Meijer, I., Zahedi, Z. and Wouters, P. 2013. The Value of Research Data - Metrics for datasets from a cultural and technical point of view. Accessed 15 November 2013. From http://www.knowledge-exchange.info/datametrics.

ICOMOS 2010. Heritage Definitions. Accessed 24 January 2013. From http://cif.icomos.org/pdf_docs/Documents on line/Heritage definitions.pdf.

Jones, S. 2009. A report on the range of policies required for and related to digital curation, v. 1.2. Glasgow: DCC pp. 29-31. Accessed 01 February 2010. From http://www.dcc.ac.uk/sites/default/files/documents/reports/DCC_Curation_Policies_Report.pdf.

Lock, G. R. 2003. Using computers in archaeology: Towards virtual pasts. London: Routledge.

Nicholas, G. P. 2012. "Intellectual Property Issues" in Companion to Archaeology 2nd ed. Edited by N. A. Silberman, pp. 106-109 Oxford: Oxford University Press.

Niven, K. and M.S. Thompson. 2011. Guides to Good Practice. Archaeology Data Service. Accessed 18 February 2013. From http://guides.archaeologydataservice.ac.uk/g2gp/ArchivalStrat_1-0.

Ryan, D. N. 2004. Managing complexity: Archaeological information systems past, present and future. Accessed 27 January 2012. From http://www.cs.kent.ac.uk/people/staff/nsr/arch/baas.html.

Shanks, M. 2005. "Public Archaeology/Museology/Conservation/Heritage", Archaeology: The key concepts. Edited by C. Renfrew and P. Bahn, pp. 164-167. London: Routledge.

Standler, R. B. 2002. Issues in computer acceptable use policy. Accessed 12 December 2002. From http://www.rbs2.com/policy.htm.

Stellagroup 2012. STELLA Unconference, Brooklyn College Library CUNY, November 9-10 2012. Accessed 9 January 2013. From http://stellagroup.wordpress.com/data-curation-notes/.

Technopedia. 2012. Acceptable use policy. Accessed 23 September 2013. From http://www.techopedia.com/definition/2471/acceptable-use-policy-aup.

Torsen, M. and Anderson, J.E. 2010. 'Intellectual Property and the Safeguarding of Traditional Cultures: Legal Issues and Practical Options for Museums, Libraries and Archives'. Heritage and Society, 3(1): 22-63. Accessed 23 September 2013. From http://works.bepress.com/jane_anderson/13/.

Turban, E., D. Leidner, E. McLean, and J. Wetherbe. 2008. Information

Technology for Management: Transforming organizations in the digital economy. Hoboken, NJ: John Wiley & Sons.

UNESCO 2007. Culture: Movable Heritage. Accessed 16 July 2013. From http://portal.unesco.org/culture/en/ev.php-URL_ID=35031&URL_DO=DO_TOPIC&URL_SECTION=201.html.

UNESCO 2008. What is meant by cultural heritage? Accessed 16 July 2013. From http://portal.unesco.org/culture/en/ev.php-URL_ID=34050&URL_DO=DO_PRINTPAGE&URL_SECTION=201.html.

Wallis, J.C., Rolando, E., and Borgman, C. L.. 2013. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. PLoS ONE 87: e67332. doi:10.1371/journal.pone.0067332. Accessed 18 September 2013. From http://www.plosone.org/article/info:doi/10.1371/journal.pone.0067332.

Wiseman, N. 2013. Encouraging scholars to embrace data sharing. Research Information. Accessed 22 September 2013. From http://researchinformation.info/news/news_story.php?news_id=1266.

Webpages Featured as Archaeological Database Examples

Celtic Inscribed Stones Project. Accessed 8 May 2013. From http://www.ucl.ac.uk/archaeology/cisp/database/.

Arachne. Accessed 25 September 2013. From http://arachne.uni-koeln.de/.

The International Dunhuang Project. Accessed 25 September 2013. From http://idp.bl.uk/.

Lerna. Accessed 24 August 2012. From http://csanet.org/archive/adap/greece/lernpot/lernameta.html#Download/.

Webpages Cited as Archaeological Data Repositories

Archaeology Data Service (ADS). Accessed 18 February 2013. From http://ads.ahds.ac.uk/.

the Digital Archaeological Record (tDAR). Accessed 2 May 2013. From http://www.tdar.org/.

Open Context. Accessed 13 March 2013. From http://

www.opencontext.org/.

Webpage Offering a Searchable List of Research Data Repositories

Databib. Accessed 13 March 2013. From http://databib.org/.