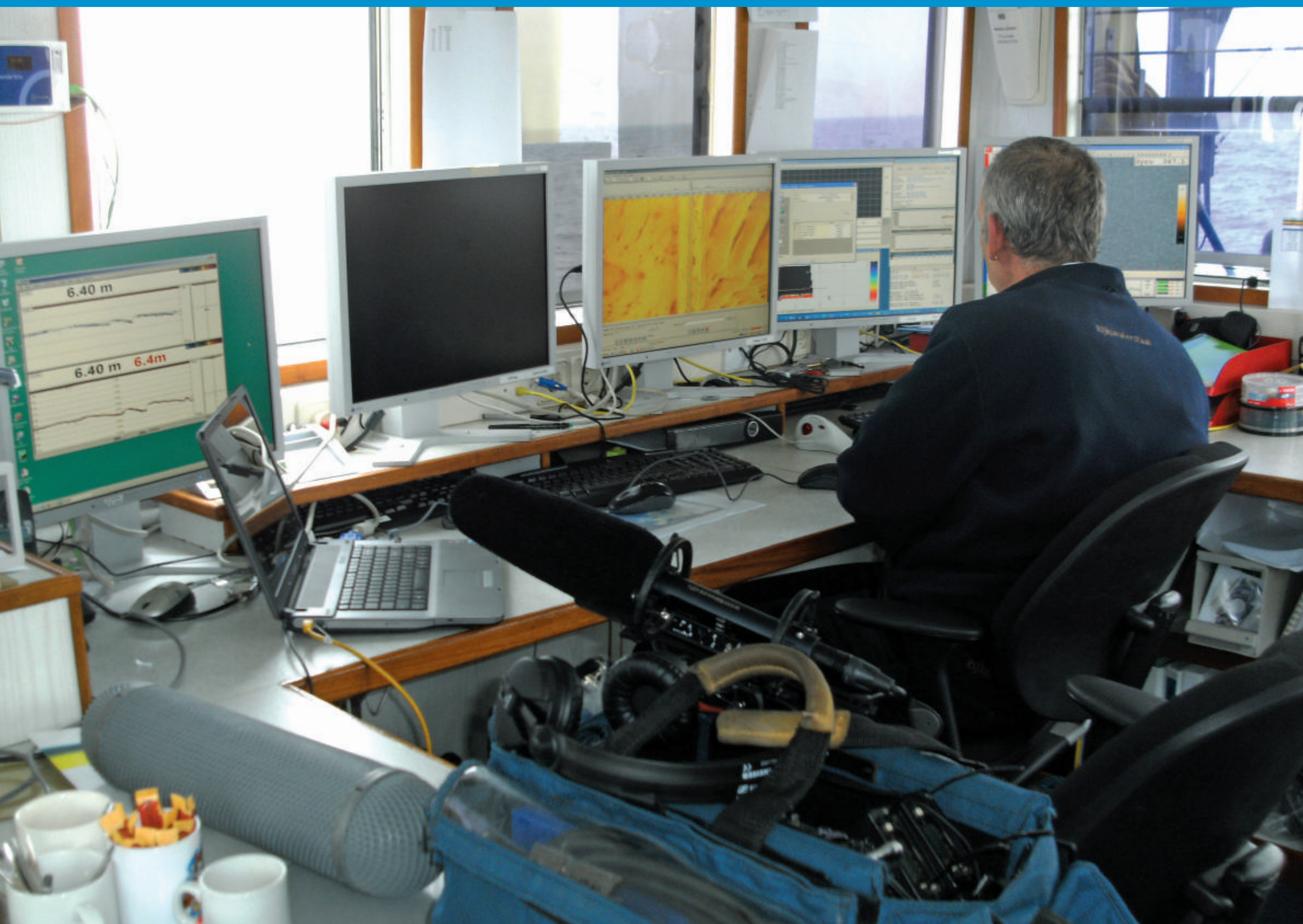
United Nations
Educational, Scientific and
Cultural Organization

The Protection of the
Underwater Cultural Heritage

UNIT 7

Author Peter R. Holt

Data Management in Maritime and Underwater Archaeology



Published by UNESCO Bangkok
Asia and Pacific Regional Bureau for Education
Mom Luang Pin Malakul Centenary Building
920 Sukhumvit Road, Prakanong, Klongtoey
Bangkok 10110, Thailand

ISBN: 978-92-9223-413-3 (Print version)
ISBN: 978-92-9223-414-0 (Electronic version)

Technical editing: Martijn R. Manders and Christopher J. Underwood
Copy-editing: Sara M. Mabelis
Design/Layout/Illustration: Warren Field
Cover photo: Archaeologists must carefully manage data from a variety of sources throughout each phase of a project. © Martijn R. Manders

Printed in Thailand

CLT/12/OS/015

UNIT 7

Contents

UNIT 7

Author Peter R. Holt

Data Management in Maritime and Underwater Archaeology

Core Knowledge of the Unit

This unit introduces students to data management in maritime and underwater archaeology and provides guidance on some of the techniques and methods that can be implemented during projects.

Upon completion of the Data Management in Maritime and Underwater Archaeology unit, students will:

- *Be able to identify and understand the main issues surrounding the management of archaeological data*
- *Understand the importance of managing data*
- *Be familiar with techniques and methods appropriate to each stage of a project*

Introduction to the Unit

This unit focuses on the issues relating to the management of digital data in maritime or underwater archaeology. It is important to take into account that although the problems and solutions often also relate to terrestrial and intertidal archaeology, the topics covered in this unit only mention these parallels where necessary.

1 Archaeological Archives

Archaeological fieldwork has the potential to generate a vast amount of data in the form of an archive. As a result, it is essential that the results of any fieldwork are properly documented, particularly as excavation is destructive, so that data and information can to be preserved by record.

The archive should contain all of the information available about the site, including:

- *Initial project proposals*
- *Project design documents*
- *Primary records*
- *Drawings, photographs and video*
- *Analysis results*
- *Research reports and interpretations*
- *Publications*
- *Computer generated models*

As the archive is used as the basis for future publications, project planning and museum displays, it is essential that the recorded information is both easy to recover and easy to use. A recording system used to capture site data can be as simple as a paper notebook or as sophisticated as a computerized digital information management system.

The data may be held on different media storage devices (such as CDs, DVDs, flash drives, etc.), which can often result in a variety of incompatible formats evolving as the project progresses. Any recording system must be capable of storing, managing and allowing access to this disparate set of data in an efficient way. It is, therefore, imperative that publication and archiving should be a central function of any recording system rather than an afterthought.

Suggested Reading

Andresen, J. and Madsen T. 1996. IDEA – The Integrated Database for Excavation Analysis, Interfacing the Past. CAA95. *Analecta Praehistorica Leidensia,* 28, pp 3-14.

Brown, D. 2007. Archaeological Archives: A Guide to Best Practice in Curation, Compilation, Transfer and Curation. *Archaeological Archives Forum*. London.

Drap, P. and Long, L. 2001. Towards a Digital Excavation Data Management System: the *Grand Ribaud F Etruscan* Deep Water Wreck. *Proceedings of the 2001 Conference on Virtual Reality, Archaeology and Cultural Heritage.*

Eiteljorg, H. 2007. *Archaeological Computing*. Centre for the Study of Architecture. http://archcomp.csanet.org (Accessed October 2011.)

Greene, K. 2002. *Archaeology: an Introduction*, Fourth Edition. London, Routledge.

Richards, J. 2004. *Internet Archeology 15.* http://intarch.ac.uk/journal/issue15/richards_index.html (Accessed 30 November 2011.)

2 Recording System Types

Recording systems used on maritime archaeological projects vary widely in their content and complexity. They are usually based on one of the methods outlined in this section.

2.1 Paper-based Systems

Paper based recording systems are appealing as the paper records are physical, can be created by anyone and they can be inspected or modified at any time.

An example of a paper-based system. © Peter R. Holt

2.1.1 Notebooks

Paper-based systems usually come in the form of notebooks. Some of the benefits and drawbacks of using notebooks are:

- *They are readily available and easy to use*
- *Flexible enough to be able to record any information, allowing for both text and illustrations to be made on the same page*
- *Unlike a digital system, a notebook can be read at any time without the need for electrical power or a suitable computer*
- *Unfortunately, it is difficult to be consistent in the information recorded in a notebook, especially if more than one person is compiling the data*
- *Training people in the use of the notebooks can be difficult if little consideration has gone into the structure and content of each book*
- *For large projects, the use of notebooks rapidly becomes unmanageable as the number of notebooks increases*
- *Notebooks are hard to copy and, thus, hard to archive, they are also both easy to lose and are easily damaged, so data security can be a problem*
- *The information in a notebook is hard to process and often hard to recover, especially if the handwriting is illegible*

2.1.2 Pre-Printed Forms

- *Pre-printed forms allow information to be structured under the designated framework of fields, with each field acting as a prompt. This makes the recording of the information more consistent and easier to locate and analyse. The information about a site can be sub-divided into separate forms, for example, for recording sketches, context and structural features of individual objects*
- *Unstructured text notes can also be recorded on areas of the form allocated for that use*

2.2 Digital Systems

Computer-based or digital recording systems are a modern alternative to paper-based systems and offer a range of advantages:

- *Digital systems often work with pre-printed forms as an interim step where direct entry of data is not possible, for example, for recording survey measurements underwater*
- *Information in a digital documentary archive is easily displayed, retrieved, shared and copied, leading to increased productivity and reduced data loss*
- *Digital recording places greater demands on the accuracy and availability of the data, which subsequently improves the quality of the information recorded*
- *Digital systems offer improvements in efficiency and can allow site records to be updated in the field.*
- *Offer the potential to publish both rapidly and widely using online digital archives*

Five types of digital systems have been developed:

2.2.1 Spreadsheet Systems

A simple, but effective recording system can be created using a spreadsheet programme, such as Microsoft Excel. A single spreadsheet contains rows and columns of cells that can be used to contain information. Each column can contain one particular property or item of information and each row can relate to one particular object. Separate sheets can be created within the same file for recording different categories, such as survey measurements, artefact information or details about features and contexts.

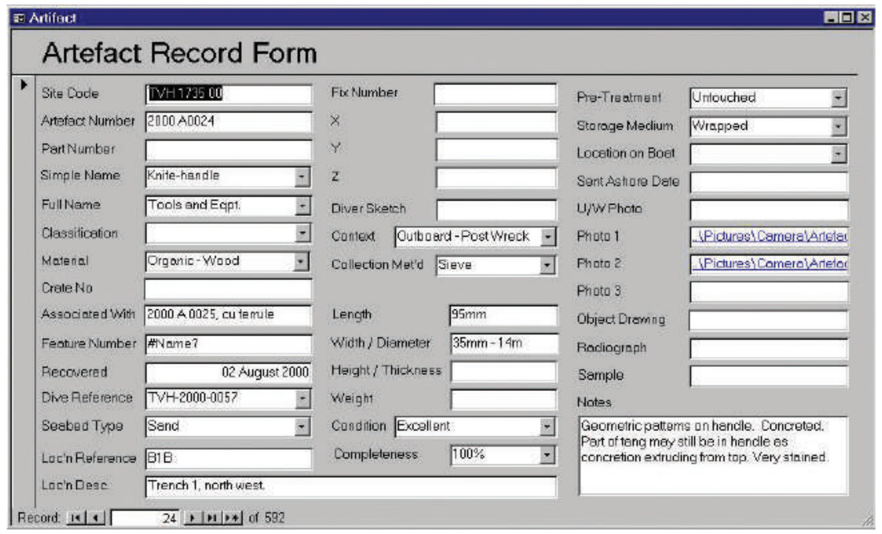
Each column can be configured to accept the correct type of information, such as text in a 'Description' column or numeric data in a 'Length' column. For cells that have a limited number of options it is possible to create a spreadsheet that only allows entry of a limited set of options, which can provide a solution to the problem of data validation. It is also possible to add links to cells, so that an image or text document can also be displayed. This can be useful for linking artefact images with artefact records.

The data contained within the spreadsheet can be searched and filtered simplifying the subsequent processing. The contents can be exported to files that can be read by other programs and the data can also be printed.

Spreadsheet systems are very easy to set up, use and extract information from, but are rather limited in capability compared with the other digital options.

2.2.2 Database Systems

Although similar to a spreadsheet, database systems offer more powerful data entry, searching, export and reporting options, which for larger projects can offer some advantages over a spreadsheet. Database systems are usually more difficult to set up and manage than a simple spreadsheet, requiring expert knowledge to do all, but basic tasks.



An example of a database artefact record.
© Peter R. Holt

2.2.3 Computer Aided Design (CAD) Systems

Computer Aided Design (CAD) systems are used to create drawings and 3-dimensional models. By providing a means of recording spatial information missing from a database, these systems can work in conjunction with a database to record the dataset for an entire archaeological project. Some modern CAD systems can be tightly coupled with a database, allowing properties to be recorded along with spatial information, blurring the edge between what is a CAD and what is a Geographical Information System (GIS).

2.2.4 Geographical Information System (GIS)

One major disadvantage of both spreadsheets and database systems is their inability to display maps. The position of objects is very important in archaeology, so the ability to display spatial information is a great benefit, as it helps us identify patterns not easily determined from non-spatial records. In its most simple form, Geographic Information System can be thought of as a database that can manage spatial data, so it can be used to display information on a site plan, therefore, increasing clarity and accessibility of the information. See Unit 8: *Geographical Information Systems (GIS) in Underwater Archaeology.*

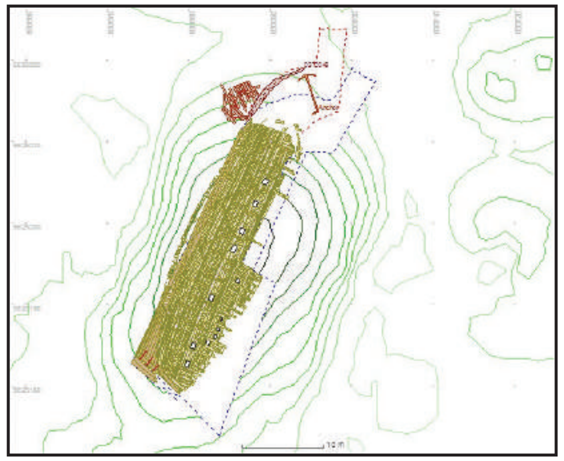
Most low cost commercial GIS are not designed for one particular task, but can be used to display and process any spatial data. These systems need to be configured for recording archaeological information, so they also require expert knowledge in their set up and usage.

2.2.5 Information Management Systems (IMS)

Information management systems that are designed for recording archaeological data have been developed. These systems incorporate the best features of database systems and GIS, but also include many other useful features not found in generic systems. As well as being more powerful than generic systems, they are also easier to use and have the advantage of being designed for the task, containing only tools that are useful.

Some of these systems are capable of managing data collection and processing in real time, an essential feature where timescales and funding are constrained or data is being collected at a very high rate.

Processing the data as it is collected allows problems to be identified and rectified before leaving the field. It also dramatically reduces the work that has to be done after the fieldwork is complete.



The Mary Rose *hull and bow positions shown in Site Recorder.*
© Peter R. Holt

An example of such a system is 3H's Site Recorder, which has been used on sites in over twenty countries including: the *Avondster* (Sri Lanka), *Mary Rose* (United Kingdom), Kizilburan Wreck (Turkey), *Le Forniche* (France) and *Pandora* (Australia).

2.3 Disadvantages with Digital Systems

As well as providing a significant number of benefits, there are some disadvantage to using a digital recording system. The most significant problem is the need for a computer to be available before the records can be viewed. This restricts the use of the recording system to sites where power is available and the working environment allows the use of a computer. Netbooks might provide an interim solution to the lack of mains supply, as they tend to have a longer battery life than most laptops and the battery can either be replaced with a spare or recharged overnight.

Digital records are sometimes seen as fragile or ephemeral and can be easily accidentally deleted or destroyed. Although this is true, problems can be avoided by a carefully designed recording system that is accompanied by a robust backup and archiving policy. It is important for measures to also take into account potential problems associated with the unknown lifetime of digital systems and changes in formats.

ADDITIONAL INFORMATION

Further information about the use of Site Recorder on sites around the world can be found on 3H's website: http://www.3hconsulting.com/SitesMain.htm (Accessed February 2012.)

2.4 Digital Recording System Requirements

Although systems may differ in what information they record, the basic requirements for any system are very similar. They include:

- *Information capture*
- *Ease of use*
- *Sharing*
- *Backup and archiving*
- *Searching, sorting and associating*
- *Data validation*
- *Collection management*
- *Publishing*

UNIT 7

Suggested Reading

Hildred, A. (ed.). 2001. *VOC Anniversary Shipwreck Project – Report on the Excavation of the Dutch East Indiaman Vliegent Hart*. Private Publication.

Holt, P. 2007a. *Development of an Object-Oriented GIS for Maritime Archaeology - Motivation, Implementation and Results*. CAA2007. Southampton.

Holt, P. 2007b. *The Site Recorder Database Schema*. http://www.3hConsulting.com/Research/research_schema.htm (Accessed October 2011.)

Lledo B. 2004. Field Example of a Database System Applied to Underwater Archaeology, the FileMaker Pro Approach. *TARAUDSTUA Conference*. Bodrum.

Nickerson, S. 1994. A Site Information System (SIS): CADD/Database Integration for Field Use. *APT Bulletin 24*, pp. 56-62.

3 Recording System Requirements

3.1 Information Capture

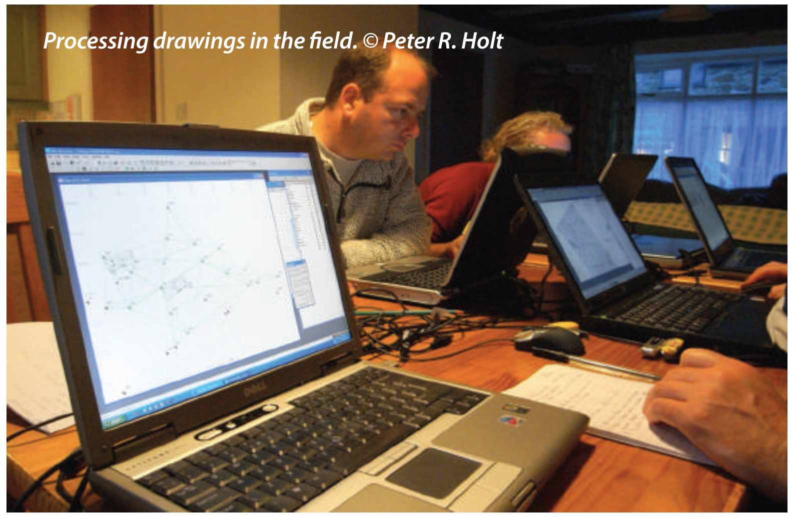
The most important requirement of any recording system is to be able to capture sufficient information about a site in a systematic way. The aim is to record a site accurately and completely, and to do the work efficiently. The recording system needs to be able to capture all of the diverse information about a site: the spatial information defining the positions of things, the descriptive information defining the properties of objects and the temporal information that records what happened and when. Information that is not recorded may be lost forever if the site is being excavated or if the site is damaged or destroyed at a later date.

The systems should primarily deal with information as hard facts and clearly separate these from interpretation, which being based on experience and background allows for different opinions. Interpretations of what things are or what they are used for should be kept separate from absolutes, such as position or dimensions.

3.2 Ease of Use

The second most important requirement is that the system is simple to set up and to use. The factors to consider are the same as those for any software application development and include:

- *Ensuring that the most common tasks and actions are simple and can be easily completed*
- *The way the system works should be consistent, so that the controls for similar tasks can be grouped together*
- *The information contained within the system should be well organized, conveniently accessible and easily retrieved*
- *The system must be scalable; it must work just as well for recording a simple survey, as it does for a full excavation*
- *The system must be efficient and must respond within a reasonable time. All common actions should be completed within 1 second for the system to feel responsive, rather than sluggish*
- *The system must be low cost and easily implemented. Any system that is too expensive to use will not be utilized and any system that is too difficult may be used incorrectly*


Processing drawings in the field. © Peter R. Holt

3.3 Sharing Information

Information about a site can be readily shared using a digital recording system. The complete set of records about a site may be held on several DVDs, external hard drives or flash drives, which enables the entire dataset from a project to be made available online.

3.4 Metadata

An integral part of the recording system is the additional information about the data that it contains, known as metadata. The metadata provides information about what has been recorded, how and when it was recorded, who did the work and who owns it. Metadata provides a useful summary and can be used by people and computers to determine whether the information is of interest, without having to examine a full set of data.

3.5 Backup and Archiving

It is easy to copy electronic information, make backups and create an archive. Having multiple copies of essential information mitigates loss or damage to the original and overcomes some of the problems associated with the fragility of digital data.

3.6 Searching, Sorting and Associating

Another powerful feature of a digital system is the ability to search rapidly and accurately through thousands of records, something that is impossible with paper records. Associated with this is the ability to filter records so that sub-sets of the data can be displayed in lists and on charts.

The ability to search and filter by object name, object type or any other property is essential. Keywords can be appended to each object to help with targeted search, so long as they are known in advance. Just as powerful is the ability to associate or relate information within the system, as this adds value to the data.

Associations form a direct cross-referenced connection between objects in the recording system, allowing a seamless jump between one and the other. An example of this would be to allow associations between information about dives recorded in dive logs with artefact records. From the artefact record, it is then possible to retrieve the dive record on which the artefact was discovered, along with the notes and sketches the diver made at that time.

3.7 Data Validation

A digital recording system limits the opportunities for making mistakes, either when adding data or when searching for information. This feature is less significant when data is added to the system by only a small number of key team members.

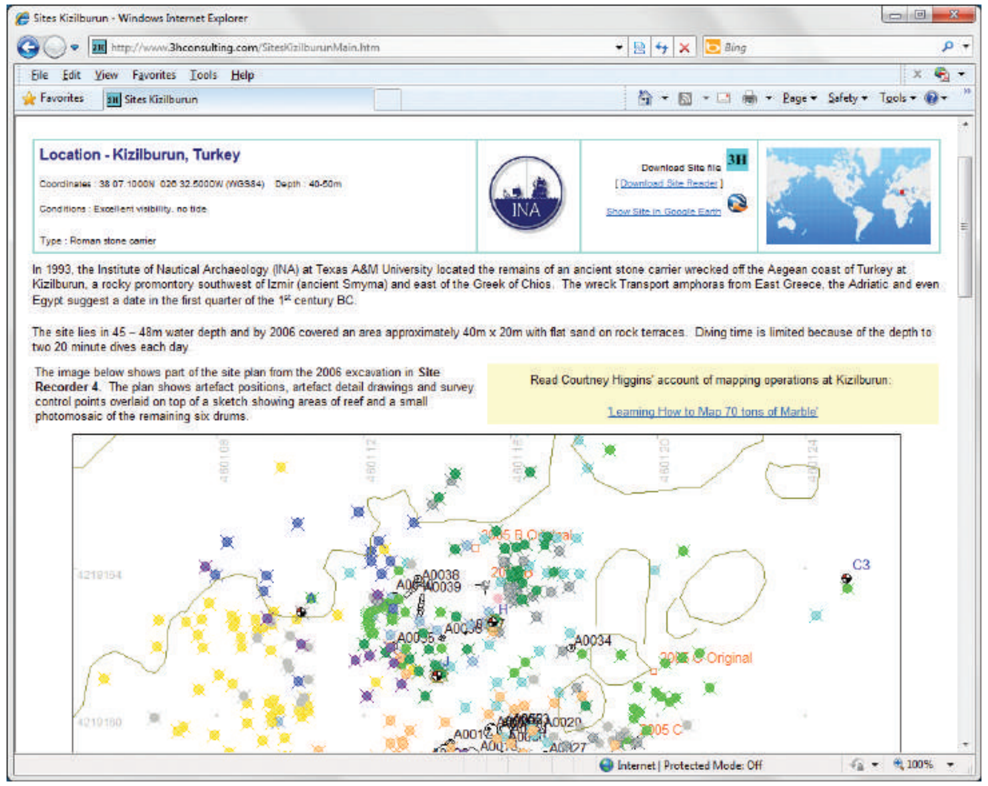
Some information entered into a digital system can be restricted to ensure that only valid information is provided. Examples of this include limiting input to only one of a set of valid values or ensuring that a numeric value entered is above or below a given number. This restriction of choice can be applied to descriptive terms, as well using a list of valid words defined in a word list or thesaurus, such as those that are part of the Information Management System, Site Recorder.

3.8 Collection Management

As well as being used as a repository for information, the recording system may be used as a tool for managing the collection of cultural material. The system contains information about each artefact in the material archive, so can be used to track its progress from discovery through recovery, registration, conservation and storage.

3.9 Publishing Information

Access to data is a major driving force in archaeology today and publication is a natural extension to the role fulfilled by a recording system. A digital recording system should allow for public access to primary data, perhaps even before preliminary and final interpretations have been made. Publishing information from a digital system is a simple process as the digital information can readily be converted to a format used by other computer programs. Information in a digital form can also readily be converted into a form that can be deployed on the Internet, either by using fixed pages of information or by using dynamic pages fed with data from an online database. Archaeological site information published online can then be directly searched by anyone or any program on the Internet.



Site information published on the internet.
© Peter R. Holt

Suggested Reading

- Madsen T. 1998. *Design Considerations for an Excavation Recording System, Design and use of Field Information Systems*, CAA-NL98.

4 Recording Systems: What Should be Recorded?

It is important to make sure that the recording system is capable of handling all of the different types of information that need to be recorded. The system may use a basic design and be limited to simple survey work or it may be more powerful and used for managing data from an excavation.

The first step is to make a list of all the information to be recorded about the site. The information can then be sorted into groups according to information type.

4.1 Information Types

Having decided on the scope of the system, the requirements can then be defined in detail. The aim is to record both the information contained in the archaeological site and that relating to the work undertaken to acquire it, as this helps with later analysis.

4.2 What Needs to be Recorded?

Primary data: original recordings taken at the site by archaeologists and site recorders.

Activities: what has been done at the site in terms of recording, excavation and monitoring activity.

Historical records: any documents or images relating to the site.

Reference documentation: information regarding other similar sites and historical events.

An archaeological site and the work done on that site can be recorded in a set of individual records that relate to each phase of work.

These records can include events that have already happened and any fieldwork planned for the future. Different records or different units of information in common records, will be required for each of the phases of work on a site: planning, searching, mapping, excavating, conserving, publishing and monitoring. By identifying what is to be recorded at each stage, the data can be defined and be managed by the recording system.

4.3 Records as 'Objects'

Computer records are most easily modelled as an individual 'object' rather than as something in a simple list. Recording 'objects' can be thought of as physical 'things' inside the computer, even though they do not physically exist. The advantage of using objects is that we can make our records more useful, as they can be associated with one another.

4.4 Object Properties

The framework of a database schema can be determined once what will be recorded has been decided. This involves determining which objects will be recorded and what properties they contain.

Object properties are the items of information that need to be recorded. Each recording system contains objects and each object contains a list of properties. For example, for an artefact it might be necessary to record properties such as its name, position, length, width and height. For a dive log, a record of the name of the diver, supervisor, time in, time out and maximum depth will be required.

4.5 Names and Object Identification

One of the fundamental requirements of a recording system is that every object must have a unique identifier so that it can separate one object from another. The name of the object is often used to do this. Using the name as the identifier means that two objects with the same name cannot be allowed to exist in the system, even by accident.

In practice, appending an object type identifier to the name can help interpretation of maps and charts. For example, all artefact objects use the prefix 'A', while all survey control points use the prefix 'CP'.

Adding a year code seems to be helpful in practice, even though it is not necessary, as the next free number can be used at the start of a new season and the date of recovery should identify the year. Two digit year codes are usually sufficient as the metadata associated with the site should identify the specific millennium. The year should be added at the front of the name, forming, for example, a name that reads '12A1564' for artefacts recovered in 2012.

Objects that are special and few in number, such as guns and anchors can have a specific numbering scheme for ease of recognition. 'Gun 12' is more readily identifiable than some generic artefact name such as '12A1284'.

The names used for dive logs should be chosen with care as often other things are associated with the logs, such as photographs or video. The computer recording system may have an upper limit to the number of characters allowed in a name, but it is sensible to limit them anyway as shorter names are easier to remember. There may also be a limit to the characters that can be used in a name. Often such characters as ' , ; / . : \ " ? * | ' are not allowed as they have special meanings.

All artefacts should be clearly marked with a unique identification. © Peter R. Holt

4.6 Notes

Another essential unit of information for almost every object is simple free form text to be used for adding notes. The recording system is unlikely to have sufficient scope to be able to record everything that you want to say about an object, so the notes are a useful place for the additional information.

4.7 Units

The units of length, weight, etc. used for recording will vary between projects. The most basic recording system may be fixed so that it uses only one set of units, an acceptable decision so long as the metadata associated with the project states which units are being used. A more flexible approach would be for the recording system to be able to display information in different units, but it will still need to work internally using only one particular set.

The majority of sites are recorded using the International System of units (SI); metres for survey measurements and millimetres for recording artefacts (although centimetres are used too). Weights should be measured in kilograms.

4.8 Time

An artefact on the seabed may be moved during the course of work on a site. The position of the artefact before moving is just as important as the position afterwards, so a recording system needs to be capable of recording both. This suggests that any artefact has a position at a given time, so it is important from the outset to consider this when recording positions.

4.9 Measurement Recording

A recording system should be capable of handling information about survey points, including both the measurements made at each survey point and between the survey points. Unfortunately, many of the measurements that are made will be mistakes, so it is important to make sure that these can be identified and removed. Later work may show up problems with measurements made earlier, so it is important that the raw measurements are recorded and not just the processed survey results.

4.10 Object Drawing

Ideally the recording system should also record shape information. For example, an artefact on the seabed can be represented on a site plan in a number of ways.

The simplest representation is as a point shape that identifies the position of the artefact. The size, colour and style of the point is used to identify some other property, such as the type or recovery date.

A more detailed representation would show a drawing of the artefact on the plan. Most site plans are drawn from above, so the artefact can be shown *in situ* as a 2-dimensional drawing on a plan view.

Drawing timbers at 1:1 scale. © Peter R. Holt

The most complex representation would show a 3-dimensional (3D) image or model of an artefact that can be viewed from any position of the site.

For each artefact there are two different sets of information. The first set of information is the drawing or 3D digital representation of the object. The other set is the recorded properties of the artefact, such as the material it is made from, its condition or weight.

4.11 Site Code

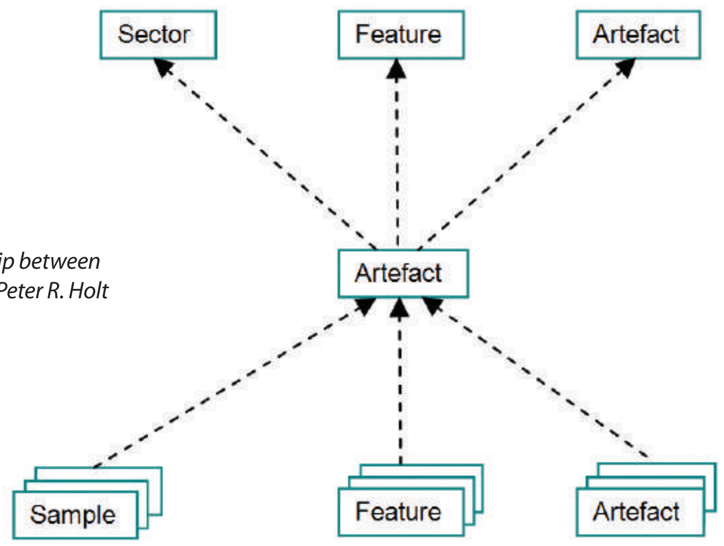
A unique code should be given to each site. Site codes need to be unique as they are used to form part of the name of every object, so that objects can be differentiated from one site to another.

With digital systems it is easy to make comparisons of datasets between sites. It is possible that the names of objects from one site are the same as those from another, so it is necessary to be able to make a distinction. To ensure that this distinction is made, the full name of each object should include the prefix of the site coded. For example, at POMR (Portsmouth - *Mary Rose*), an object can be labelled POMR/94/A1234 (location – site – year – artefact number), while at GAAV (Galle - *Avondster*) an object can be labelled GAAV/94/A1234.

4.12 Recording Associations

Archaeological objects can be associated with each other, providing extra information about the site and how it was recorded. This provides a way of relating groups of objects found in the same area that have some meaningful relationship, building up a hierarchy of associations within the site.

The relationship between these objects gives clues about their use, so the recording system should allow these relationships to be in a formal manner and provide tools to visualize them. The result is a hierarchical tree of relationships between artefacts, features and sectors.



Relationship between objects. © Peter R. Holt

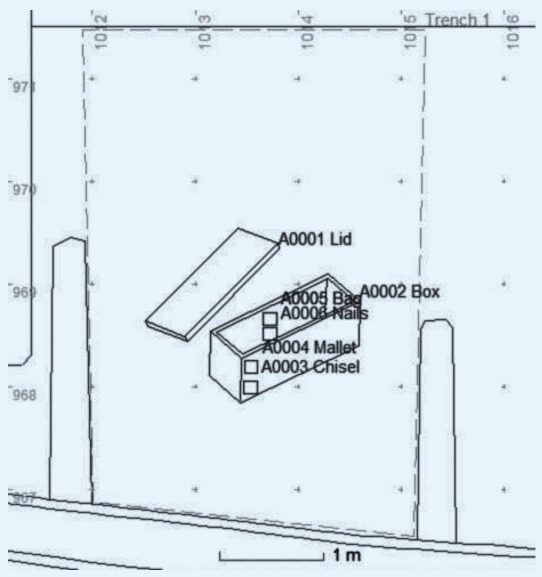
Associations between objects include:

Next to: *a loose spatial relationship.*
Above: *a loose spatial relationship more specific than 'next to'.*
Below: *a loose spatial relationship more specific than 'next to'.*
Contained by: *a tight spatial relationship.*
Part of: *a tight relationship.*

Objects can only be associated 'upwards' by being related to a 'parent' object. The 'parent' can have many 'child' objects below it, each linked once to the parent itself. An object does not have to be associated with a 'parent' and can remain unrelated to anything else.

The use of these associations can be represented by an example:

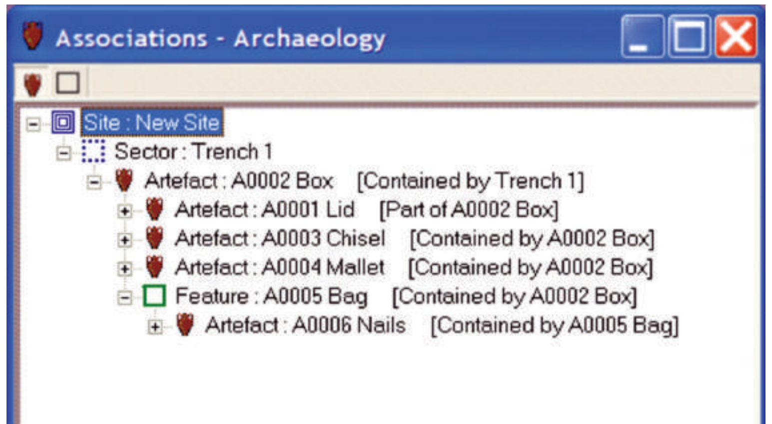
- *A wreck site had an excavation trench defined by the bulkheads of the carpenter's cabin*
- *In the trench was found the lid of a box*
- *Below the lid was found a box, the lid found earlier was the correct size to fit the box*
- *Found inside the box were a chisel and a mallet*
- *Also inside the box was the 'ghost' of a cloth bag, no cloth remained just a different colour sediment*
- *Inside the 'bag' were found 10 identical copper nails*



Representation of a relationahip between objects © Peter R. Holt

This image, taken from Site Recorder, shows the relationships between all archaeological objects in a hierarchy. The site is at the top, followed by sectors, features and artefacts. The type of relationship is shown in square brackets [ ]

At the top of the 'tree' is the site itself and below this is an object that represents Trench 1. Contained by Trench 1 is the Box, shown as Artefact A0002.



Object associations in Site Recorder. © Peter R. Holt

Below the Box is the Lid, Artefact A0001. The Lid is shown below the Box as it is related by being 'part of' the Box. Also below the Box are the Chisel and the Mallet represented by Artefacts A0003 and A0004.

Feature A0005 represents the remains of the Bag, also 'contained by' the Box. Contained by the Bag feature are the 10 Copper nails, represented by Artefact A0006.

Suggested Reading

Andresen, J. and Madsen, T. 1996. IDEA Integrated Database for Excavation Analysis. CAA95. *Analecta Praehistorica Leidensia*. No. 28, pp 3-14.

Quinn, R. 2001. The Assimilation of Marine Geophysical Data into the Maritime Sites and Monuments Record, Northern Ireland. *Historical Archaeology.*

5 Other Development Considerations

This section examines the more technical issues related to maintaining data quality, data description and data sharing.

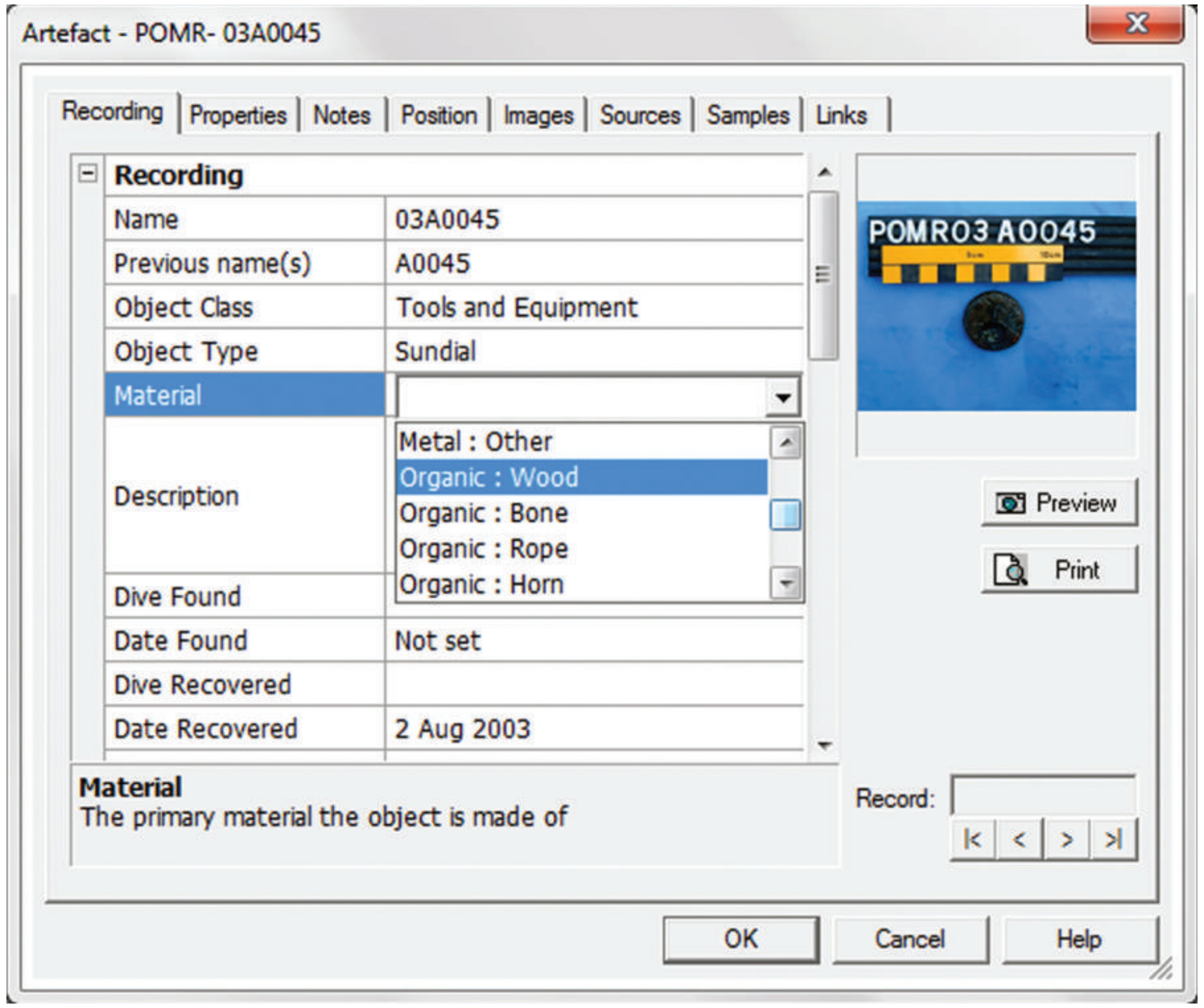
5.1 Controlled Vocabularies, Wordlists and Thesauri

One disadvantage of paper-based recording systems is the difficulty in maintaining consistency in the information that is recorded. Consistency is not only essential for ensuring the accuracy of the information, but also simplifies the process of searching and sorting. Digital systems can go further by limiting the choice of options available for some object properties using a controlled vocabulary, similar to the wordlists available in Site Recorder or the Inscription wordlists. For more information see http://www.fish-forum.info/i_lists.htm (Accessed Feb 2012.)

A controlled vocabulary is a collection of words that are allowed to be used and are useful in situations where clarity and consistency are important. Controlled vocabularies have many uses, but are particularly important where language translation occurs, as it is much easier to translate documents containing a limited selection of well defined words.

A digital recording system can readily enforce a controlled vocabulary by only offering the user the option of selecting one word or phrase from a list, known as a wordlist. Long wordlists can be laborious to work through, so the alternative is to offer a hierarchical 'tree' of options known as a hierarchical wordlist or thesaurus (Pl. thesauri). With a list, all the items are offered to the user at the same time, which may be difficult to manage if the list is long. With a thesaurus, the user can 'drill down' into the set of options starting with some general types at the top and more specific examples or component

UNIT 7



Artefact record from Site Recorder. © Peter R. Holt

parts at the bottom. If we use a sword as an example, the user is initially offered a simple 'sword', but further down the options may include specific types of sword, such as 'rapier' or 'cutlass'. This method can also be used to offer words to describe components of a sword, such as 'sword hilt' or 'sword blade'.

Using and sharing standard wordlists brings benefits to a project, particularly as time goes on and the same lists are used by more projects, providing uniformity within not only that project, but also between other projects.

The image above shows an artefact record in Site Recorder where the material type of an artefact can only be selected from a list of options and an incorrect or new option cannot be added directly. The list of available materials can be modified or extended, but this is usually done as a separate exercise.

Consistency within a dataset can be maintained using your own wordlists, but consistency between datasets can also be achieved using standard wordlists and thesauri.

5.2 Documenting the Archive

It is important to document the dataset so it will allow others to see what it contains and how it is structured, as well as defining the formats and conventions it uses.

The documentation includes:

- *Metadata*
- *Project title*
- *History of the project*
- *Purpose of the project*
- *Topics of research*
- *Geographic and temporal extents*
- *Information about methods*
- *Methods used to create the dataset*
- *Finds recording methods*
- *Survey and georeferencing methods*
- *Sampling strategy*
- *Details of source materials*
- *Archives used for initial assessment*
- *Maps and charts*
- *Descriptions of previous work on-site*
- *Known copyrights*
- *Content and structure*
- *List of file names and a description of contents*
- *Description of naming convention*
- *List of codes and what they mean (if used)*
- *Description of any known errors*
- *Description of any known areas of weakness*
- *Wordlists, thesauri*
- *Names of the primary project staff*
- *History of format changes to the dataset*
- *Archives and publications*
- *Bibliographic references to publications about the site*
- *Information about any museums or archives which hold related material*
- *Information about any non-public related material*

5.3 Exporting and Publishing Data

The recording system may only be suitable for collecting data rather than in depth analysis or public dissemination. As the data is to be used for analysis and in publications, it is necessary to ensure that the data in the system can be extracted easily and reliably.

Although any format for the data could be used, more widespread re-use would be possible if the data were in a standard format. The ideal mechanism for data interchange is now available as the extended Markup Language (XML), which is a non-proprietary standard available to all.

Coupled with related formats for graphics called Scalable Vector Graphics (SVG) and Geographic Markup Language (GML), both data and graphical elements can be exchanged. Use of these data formats brings with it the ability to easily exchange information between different types of computers (or computer systems), something which was a problem in the past. Another benefit is the ability to automatically validate the information before use and so avoid corruption of crucial data.

5.4 Archiving

The site data on a computer is vulnerable to deliberate or accidental destruction and steps should be taken during and after a project to ensure its survival. A computer used in the field may get stolen, dropped, water damaged or damaged by a poor quality power supply. The files containing the site data may get corrupted, deleted, overwritten or lost.

The first and most simple security measure is to make electronic copies of the dataset, in the hope that at least one of these copies survives if the original gets destroyed. To do this you only need to save a copy of the relevant files onto an external hard drive, CD or memory stick.

If multiple copies exist then it is essential to be able to tell which the master copy is. Each dataset that is copied should have a unique version number, so that older copies (usually with lower version numbers) can be easily identified.

Using the date stamp on a file for version control is not recommended as it is too easily altered. The date stamp on a file is the date and time shown in Windows Explorer or similar program. It is simply the date and time that the file was last saved, so you can get an older version of a file then save it, which would give it a new date stamp of a younger, more up to date version of the same data file.

5.5 Media Survival

The dataset can be copied to any media large enough to hold it. The favoured media include: external hard drives and DVDs for large projects or USB memory sticks for small projects. It is essential that the copies can survive on the media they are written on. As all digital storage media have a limited life, true long term survival of the data may require the use of redundant data servers and a robust data backup strategy.

The format of the data stored in the archive should also be considered, as curation problems can occur with proprietary data formats. Ideally, the digital data should be archived in a generic and open format, so the data can be recovered even if the viewing software is no longer available. However, it is essential that no data, links or associations are lost when converting the dataset from a proprietary format to the open format used for archiving.

The content of the documentary archive for a digital system is often self-documenting, especially if metadata has been included and digital copies of the final reports are linked to the archive.

For security reasons it is often desirable to produce a paper copy of the site records. This will be a 'snapshot' of the archive at the time of printing, so needs to be clearly identified with a date, time and version number.

5.6 Existing Standards

Wherever possible, it is recommended that recording systems are developed so that they comply with any existing international or local standards, such as Infrastructure for Spatial Information in Europe (INSPIRE), so that any information gathered can be used directly with no translation. Adhering to existing standards allows for a more ready exchange of information between one system and another that may wish to use or archive that data. In addition, you gain the benefit of the knowledge embedded in the standard, as the creators may have considered an issue that you have overlooked. Reviewing more than one standard may also provide ideas about additional information that you may wish to record.

6 Data Sources for Each Project Phase

Projects go through a number of phases during their lifetime, from initial inception through to the final end of project archiving. Each phase requires a different set of tasks to be undertaken using the recording system, to ensure that the information contained within it is accurate, comprehensive and up to date.

The phases of a typical maritime or underwater archaeology project are:

Planning and Assessment
- Search
  - Survey and Marketing
    - Excavation
      - Conservation
        - Analysis and Interpretation
          - Publication and Archiving

Phases of a project © Peter Holt

6.1 Planning and Assessment

During the planning phase, it is usual to gather together all of the readily available information about a site or its location. The information that exists can then be collated and reviewed to see what can be obtained from it or to determine what is missing.

6.1.1 Typical Sources of Information

Typical sources of information include:

- *Modern charts and maps*
- *Old charts and maps*
- *Reports from previous work*
- *Existing site plans*
- *Newspaper articles*
- *Web sites*
- *Environmental reports*
- *Photographs*

6.1.2 Typical Tasks

Typical tasks include:

- *Position site, geodesy, north*
- *Scan and import or digitise charts, chart copyright*
- *Import, geo reference and digitise site plans*
- *Import digital site plans*
- *Scan and link images*
- *Scan and link documents, copyright*
- *Create a timeline of events*

6.2 Search

If the project involves a search phase then geophysical survey techniques will be employed. Geophysical survey equipment produces georeferenced measurements of depth, magnetic field or sonar signals which can be incorporated into the recording system. The measurements themselves can be added or a more simple approach is to add just the list of targets detected by each survey. Often the raw survey data can be quite large in size; by processing survey data and extracting the position and size of any anomalies, a large raw data set can be reduced to a more manageable list of targets.

6.2.1 Typical Sources of Information

Typical sources of information include:

- *Magnetometer data*
- *Sub bottom profiler traces*
- *Raw multibeam echo sounder (MBES) data*
- *MBES data as a post-processed image*
- *Side scan sonar data as a post-processed mosaic*
- *Lists of targets*

6.2.2 Typical Tasks

Typical tasks include:

- *Import target lists*
- *Import raw survey data and reprocess to create target lists*
- *Import side scan sonar mosaics and multibeam images*
- *Check metadata*

6.3 Survey and Monitoring and *In Situ* Protection

Once a site has been located, the first task is to undertake a survey to record the site as it was found. This work may be a simple assessment survey which aims to create a simple site plan quickly and efficiently. A more accurate pre-disturbance survey is usually required in advance of any intrusive fieldwork; this has to be done carefully as the work cannot be repeated at a later date. High accuracy survey work usually continues during excavation as more of the site is uncovered.

In situ recording of artefacts and features can be undertaken at this stage. It is not necessary to excavate or recover finds to be able to record them sufficiently for identification or dating.

If a site is to be monitored, then this usually involves the recording of changes to the site over time. Typical measurements include the movement of tracer objects and relative movement of a structure as a way of monitoring gradual collapse, or the measurement of sediment depth in and around a site.

6.3.1 Typical Sources of Information

Typical sources of information include:

- *Primary, secondary and detail control points*
- *Distance, depth, height, offset, ties and radial measurements*
- *Surface position measurements from a Global Positioning System (GPS)*
- *Subsea position measurements from an Acoustic Positioning System (APS)*
- *Drawing frame drawings*
- *Photomosaics*
- *Point positions from 3-dimensional photogrammetry*
- *Tide measurements for correcting depths*
- *Survey measurements defined above*
- *Sediment depth measurements*

6.3.2 Typical Tasks

Typical tasks include:

- *Design the survey control point network*
- *Add survey points*
- *Add survey measurements*
- *Process measurements*
- *Scan, import and digitize drawing frame drawings*
- *Import and georeference photomosaics*
- *Import point positions from 3-dimensional photogrammetry*
- *Correct depth measurements for the effects of tide*
- *Changes on site after* in situ *protection*

6.4 Intrusive Fieldwork

If a site is excavated, then information becomes available about artefacts, features, samples and trenches, however, the process of recording finds may start earlier with the recording of finds *in situ* during the survey phase.

6.4.1 Typical Sources of Information

Typical sources of information include:

- *Artefact records*
- *Artefact* in situ *photographs and video*
- *Artefact recovery photographs and video*
- *Artefact registration photographs*
- *Artefact drawings*
- *Feature and context records*
- *Trenches and areas*
- *Sample records*
- *General site photographs and video*
- *Sections and stratigraphic records*
- *Dive logs*

6.4.2 Typical Tasks

Typical tasks include:

- *Add artefact, feature, trench and sample records*
- *Add linked images*
- *Add dive logs*
- *Scan, import and digitize artefact drawings or photographs and add to the site plan*

6.5 Conservation

During conservation work, further information is recorded about the artefacts being conserved. The artefacts will need to be cleaned and recorded before conservation, and then recorded after conservation to note any differences. The conservation treatments applied to each artefact will also need to be recorded as the artefact may need further treatment at a later date.

6.5.1 Typical Sources of Information

Typical sources of information include:

- *Pre and post-conservation artefact records*
- *Artefact conservation process records*
- *Artefact pre and post conservation photographs*

6.6 Analysis and Interpretation

During analysis and interpretation, the recording system will be used as a source of information about everything to do with the site and its environment. Information will be added in the form of results taken from the analysis of samples taken during fieldwork.

6.6.1 Typical Sources of Information

Typical sources of information include sample analysis reports.

6.7 Publication, Deposition and Curation

The publication of the site archive does not usually involve the addition of information to the recording system. The process of deposition may require the creation of indexes and summary documents, as specified by the archive repository. Curation involves the management of the archive to ensure its long term survival, but also includes management of access to the information; neither task should alter the archive itself.

Unit Summary

This unit highlights the importance of archaeological archives as the means for the long term preservation of archaeological information. The advantages of digital documentary archives and the problems associated with them have been illustrated. The significance of the quality recording of all aspects of the fieldwork has also been highlighted, as has the need for widespread re-use of the captured information.

Suggested Timetable

15 mins	Introduction to Data Management
75 mins	Data Management Part I - *Site Archive* - *Finds* - *Samples* - *Documentary archive initial project proposals* - *Project design documents* - *Primary records, finds records, dive logs* - *Drawings, photographs and video* - *Analysis results* - *Research reports and Interpretations* - *Publications* - *Computer generated models* - *Meta data*
	Break
30 mins	Data Management Part II - *Preservation by record* - *Recording system types* - *Paper notebooks and pre-printed forms*
30 mins	Data Management Part III - *Object drawing* - *Site codes* - *Recording associations* - *Wordlists and Thesauri* - *Archive version control* - *Documenting the archive* - *Data security* - *Existing standards* - *Sources of data for project phase* - *Data sources – planning and assessment* - *Data sources search, survey and monitoring, intrusive activities, conservation*
30 mins	Concluding Remarks and Closure

Teaching Suggestions

The topics covered during the unit provide students with the theoretical knowledge required to understand the importance of recording data in a systematic way and the advantages of using digital systems.

This unit is linked to Appendix D: *How to Use Site Recorder* and the survey and practical experience gained during Unit 12: *Practical Dive Session of the Foundation Course.* Aside from activities included in these units, there is also the opportunity for students to practice using the full licensed version of Site Recorder held by the centre, or the demo version, which allows most of the software's tools to be used.

To provide students with further insight and understanding of data management, it may be useful to include additional presentations prepared by 3H Consulting. These presentations can be downloaded from their website: www.3hconsulting.com.

PowerPoint lecture

Management of Digital Data in Maritime Archaeology.

www.3hconsulting.com/Downloads/SRPP03_ManagementOfDigitalDataV1_0.pdf (Accessed February 2012.)

Notes

Management of Digital Data in Maritime Archaeology.

www.3hconsulting.com/Downloads/SRDC03_ManagementOfDigitalData.pdf (Accessed February 2012.)

Site Recorder Database Schema

www.3hconsulting.com/Downloads/TheSiteRecorderDatabaseSchema.pdf

Suggested Reading: Full List

Andresen, J. and Madsen, T. 1996. IDEA Integrated Database for Excavation Analysis. CAA95. *Analecta Praehistorica Leidensia* 28, pp 3-14.

Brown, D. 2007. Archaeological Archives: A Guide to Best Practice in Curation, Compilation, Transfer and Curation. *Archaeological Archives Forum*. London.

Drap, P. and Long, L. 2001. Towards a Digital Excavation Data Management System: The *Grand Ribaud F Estruscan* Deep Water Wreck. *Proceedings of the 2001 Conference on Virtual Reality, Archaeology and Cultural Heritage.*

Eiteljorg, H. 2007. *Archaeological Computing*. Center for the Study of Architecture, http://archcomp.csanet.org/ (Accessed October 2011.)

Greene, K. 2002. *Archaeology: An Introduction,* Fourth Edition. London, Routledge.

Hildred, A. (ed.). 2001. *VOC Anniversary Shipwreck Project – Report on the Excavation of the Dutch East Indiaman Vliegent Hart*. Private Publication.

Holt, P. 2007a. *Development of an Object-Oriented GIS for Maritime Archaeology - Motivation, Implementation and Results*. CAA2007. Southampton.

Holt, P. 2007b. *The Site Recorder Database Schema* http://www.3hConsulting.com/Research/research_schema.htm (Accessed October 2011.)

Lledo, B. 2004. Field Example of a Database System Applied to Underwater Archaeology, the FileMaker Pro Approach. *TARAUDSTUA Conference*, Bodrum.

Madsen, T. 1998. *Design Considerations for an Excavation Recording System, Design and Use of Field Information Systems*. CAA-NL98.

Nickerson, S. 1994. A Site Information System (SIS): CADD/Database Integration for Field Use. *APT Bulletin 24,* pp. 56-62.

Quinn, R. 2001. The Assimilation of Marine Geophysical Data into the Maritime Sites and Monuments Record, Northern Ireland. *Historical Archaeology.*

Richards, J. 2004. Internet Archaeology 15. http://intarch.ac.uk/journal/issue15/richards_index.html (Accessed October 2011.)