Management of Digital Data in Maritime Archaeology

Version 1.0

Management of Digital Data in Maritime Archaeology

Prepared by:

Peter Holt
3H Consulting Ltd.
6 Honcray, Plymouth, PL9 7RP UK
pete@3HConsulting.com

September 2008

Acknowledgements

This document was prepared with contributions from:

- Alex Hildred, Mary Rose Trust
- Sarah Ward, Nautical Archaeology Society

Pictures are courtesy of:

- The Mary Rose Trust
- Cornwall and Isles of Scilly Maritime Archaeology Society (CISMAS)
- University of Bristol
- VOC Anniversary Shipwreck Project

Contents

1 Introduction

1.1 Scope

This document discusses the issues relating to the management of digital data in maritime archaeology.

The problems and solutions discussed often relate to terrestrial and intertidal archaeology as the products from both are the same; however there are numerous differences in the methods used and difficulty of collecting information. Although briefly mentioned, this document deliberately avoids long discussions about paper based recording systems as the reasons to still use them are so few in the 21st century.

Part 1 of the text concentrates on describing the problems associated with data collection, management, publication and archiving. Various solutions are offered to each of the problems, each are digital in nature and each has its own advantages.

Part 2 focuses on the Site Recorder software application, an Information Management System (IMS) designed specifically for maritime archaeology projects.

1.2 Target Audience

The text is aimed primarily at those responsible for the management of project archives but the explanations of the issues surrounding data management will be relevant to all.

1.3 Training Course

This document forms the basis of a training course that provides an understanding of the principles and practice of archaeological data management for maritime sites. This will be achieved by a combination of lectures and practical sessions involving data from real underwater archaeological sites. The course is aimed at NAS Part 3 level, as well as professional archaeologists wanting to increase their knowledge of data management techniques.

It is intended that participants will complete the course with an understanding of the principles and practice of archaeological data management. Participants will:

- be able to identify the issues surrounding archaeological data management,
- be familiar with techniques and methods appropriate to the particular stage of the project;
- have the ability to plan and execute recording fieldwork (although to some extent this is assumed knowledge)
- be able to use Site Recorder to complete an NAS Part 2 survey project
- be able to provide advice and guidance to NAS members on issues of data management

Part 1 – Digital Data Management

2 Archaeological Archives

Excavation, survey and other archaeological fieldwork has the potential to generate a vast amount of data in the form of an archive. No one person can remember all of the details of even a small survey task so it is essential that the results of any fieldwork are properly documented.

The archive should contain all of the information about a site, both the documentary archive in the form of records as well as the material archive comprising all of the recovered finds and samples. The documentary record should contain all of the information available about the site:

- 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

Figure 1: Recording timbers

As sites contain an infinite amount of detail only part of the data can ever be recorded and captured. Not all of the recorded data is then converted into useful information and even less is made available outside the project itself. Information buried in the archives of a project is often as inaccessible as it was when under many metres of sediment. As the archive is used as the basis for future publications, museum displays and future project planning, it is essential that the recorded information is easy to recover and easy to use.

The archive may be the only record of a site that remains if the site has been excavated or has subsequently been destroyed so the documents become the only means of reconstructing the site itself. 'Preservation by record' as it is known therefore relies heavily on the recording system to ensure that the information about the site has been recorded in sufficient detail and with sufficient accuracy. The intention is for long term preservation and use of the hard-won and often expensive dataset.

A recording system used to capture site data can be as simple as a paper notebook or as complicated as a digital Information Management System on a computer: both methods are a way of storing information. Both systems can describe physical objects found on site such as artefacts and contexts, but can also record other things such as activities undertaken on the site and relevant information sources.

When recording an object such as a find or artefact a simple set of data is often noted for each, examples include the overall length of an artefact or the start date of an excavation. More complicated data can include free-form notes, sketches, photographs and video clips. The data is captured on a myriad of different media types often with incompatible formats that may have evolved 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.

Recording systems are often not designed with dissemination in mind so the only published product of extensive fieldwork may be a highly distilled paper report. Where the destructive process of excavation has effectively turned the material of site itself into recorded information the importance of the primary and processed data cannot be overstated. Difficulty in sharing data is usually coupled with difficulty in archiving so the data may be at risk if there is only one copy. Damage or loss can be catastrophic, particularly in the case of hand-drawn charts where accurate or up-to-date copies may not exist. Publication and archiving should be a central function of any recording system rather than an afterthought.

Recent papers on the subject have predicted that access to information would improve and that we would build information management systems to support the full range of archaeological activity from planning, through fieldwork and excavation, to archiving and publication of results. This suggests that there is work to be done in the analysis and design of future recording systems. With these issues in mind we look at the need for recording this information, what information should be recorded, how it should be captured and what to do with it once it has been collected.

Technical Note: *Archaeological Computing*

For a comprehensive guide to the use of computers in archaeology I can recommend *Archaeological Computing* by Harrison Eiteljorg II. (2007, Center for the Study of Architecture) available from http://archcomp.csanet.org/

Technical Note: *Archaeological Archives*

A guide to best practice in the creation, compilation, transfer and curation of archaeological archives can be found in *Archaeological Archives* (Brown D., 2007, Archaeological Archives Forum). Much that could have been added to this document has already been published in *Archaeological Archives* so this should be seen as essential reading and a companion text.

3 Recording System Types

3.1 Introduction

Recording systems used on maritime archaeological projects vary widely in their content and complexity. Recording systems are based on one of the six methods below:

- Paper recording systems
  - Notebooks
  - Pre-printed Forms
- Digital recording systems
  - Spreadsheet systems
  - Database systems
  - CAD Systems
  - GIS Systems
  - IMS Systems

3.2 Paper Based Systems

3.2.1 Introduction

Paper-based recording systems are appealing as the paper records can be held in the hand, anyone can create them and they can be inspected or modified at any time. A set of files containing reams of paper readily demonstrates that work has been done and results have been achieved.

3.2.2 Notebooks

Notebooks used to be the standard way for the archaeological director to record the details of a site. Notebooks are readily available, easy to use and flexible enough to be able to record any information and the ability to record text and sketches on the same page is useful. 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 adding to it. Training people in the use of the notebooks can be difficult if little thought has gone in to the structure and content of the books. 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 easy to lose and are easily damaged so data security is a problem. The information in a notebook is hard to process and often hard to recover, especially if the handwriting is poor.

Figure 2: A system with poor search capability

3.2.3 Pre-printed Forms

The next step up from using unstructured notebooks is the use of pre-printed forms. Here the information is recorded within separate fields on a form so the form acts as a prompt for the information, so the information recorded is more consistent, easier to find and easier to analyse.

The information about a site can be sub-divided into separate forms so there may be a set of forms to record an artefact and another to record a context. A separate form is needed for each category to be recorded. Forms can be created for recording sketches or a sketch may be an integral part of a specific form. Unstructured text notes can also be recorded on areas of the form allocated for that use.

Digital systems often work with pre-printed forms as an interim step where direct entry of data is not possible. Forms are most often used for recording survey measurements underwater or during initial finds recording where the wet and dirty work area is not suitable for computers.

3.3 Digital Systems

3.3.1 Introduction

Computer based or digital recording systems are a more modern alternative to paper based systems. Sometimes these systems are simply a digital replica of an existing paper based system but the main benefits of going digital appear when the advantages of digital systems are employed.

Information in a digital documentary archive is easily displayed, retrieved, shared and copied leading to increased productivity and less data loss. Digital recording places greater demands on the accuracy and availability of the data so inherently assisting with improving the quality of the information recorded. Digital systems offer improvements in efficiency and can allow site records to be updated in the field so when fieldwork ends the results can be published almost immediately. The potential to publish rapidly and widely using online digital archives is another benefit of digital system.

Five types of digital systems have been developed:

- Spreadsheet
- Database
- Computer Aided Design (CAD) drawing systems
- Geographic Information Systems (GIS)
- Information Management Systems (IMS)

Digital systems also have their limitations and these are discussed at the end of this section.

3.3.2 Spreadsheet Systems

A simple but effective recording system can be created using a spreadsheet program such as Microsoft Excel. A single spreadsheet contains rows and columns of cells that can be used to contain information, so 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 so set up the spreadsheet to only allow entry of a limited set of options, this helps with the problem of data validation discussed later. It is also possible to add links to cells so that clicking on a link displays an image or a text document, this can be useful for linking artefact images with artefact records.

The data contained within the spreadsheet can be searched and filtered making subsequent processing straightforward. 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, easy to use and easy to extract information from but are rather limited in capability compared with the other digital options.

3.3.3 Database Systems

One alternative to a simple spreadsheet system is the database. Although similar to a spreadsheet, the database offers more powerful data entry, searching, export and reporting options and for larger projects a database can offer some advantages over a spreadsheet. Database systems are usually more difficult to set up and manage than a simple spreadsheet and some expert knowledge is required to do all but the simplest tasks.

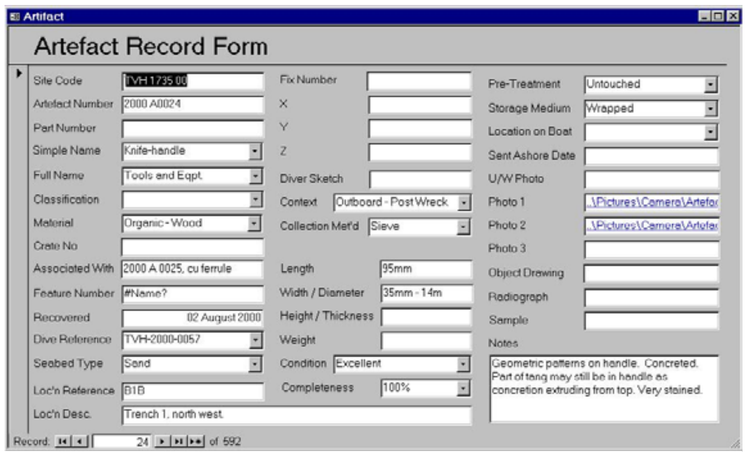


Figure 3: A database artefact record form

Database programs that have been used for creating recording systems include Microsoft Access (Hildred 2001) and FileMaker Pro (Lledo 2004).

3.3.4 Computer Aided Design (CAD) Systems

Computer Aided Design (or Drafting) (CAD) systems are used to create drawings and 3D models. By providing a means of recording spatial information missing from a database, these systems have been loosely coupled with a database to record the dataset for an entire archaeological project (Hildred 2001).

Some modern CAD systems can be tightly coupled with a database allowing properties to be recorded along with spatial information, blurring the edge between CAD and GIS systems.

3.3.5 Geographic Information Systems (GIS)

One major drawback with spreadsheet 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 to any recording system as it helps identify patterns not easily seen from non-spatial records. In its most simple form a Geographic Information System (GIS) can be thought of as a database that can manage and display spatial data so it can be used to display information on a site plan, increasing clarity and making the information more accessible.

Most low-cost commercial GIS are not designed for one particular task but can be used to display and process any spatial data. As such, these systems need to be configured for recording archaeological information so they require more expert knowledge in their setting up and use. Generic GIS systems also contain many features that are not useful for archaeological recording which make them more complex with no added benefit. In fact much of the capability of generic

GIS systems can only be used for analysis once the information has been captured and are of no use during data collection.

3.3.6 Information Management Systems (IMS)

Computer systems that are designed for recording archaeological data have been developed. As well as basic recording, these systems can be used for pre-fieldwork planning, aiding real-time excavation decision support and electronic publication.

These systems incorporate the best features of database and GIS systems 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, they have the advantage of being designed for the task so are already set up for recording and only contain features and tools that are useful.

For terrestrial work a number of systems have been developed:

ARK	http://ark.lparchaeology.com/
Nabonidus	http://www.nabonidus.org/
RAMSES	http://www.disi.unige.it/person/DoderoG/ramses/main.html

For maritime and intertidal archaeology projects there are:

VENUS project	http://piccard.esil.univmed.fr/venus/index.html (Drap & Long 2001)
Site Recorder 4	http://www.3HConsulting.com

Some of these systems are capable of managing data collection and processing in real time, an essential feature where timescales and funding are in short supply 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 but also dramatically reduces the work that has to be done after the fieldwork has finished.

Examples of this in practice are given below:

- During the excavation of the Mary Rose in 2003 the Site Recorder program was used to record information from two divers working at the same time, each fitted with a voice communications system, hat camera and acoustic positioning system. The archaeologists on the surface could plot the position of in-situ finds in real time directly into the recording system, annotating comments from the diver.
- A recent survey on the Colossus site in the Isles of Scilly had a team of divers recording the remains using planning frames onto plastic paper. At the end of each day the drawings from that day were scanned, digitised and added to the main site plan so problems or omissions could be remedied during the next day's dives. By the end of the fieldwork the site plan already included all of the drawings that had been made during the previous week.

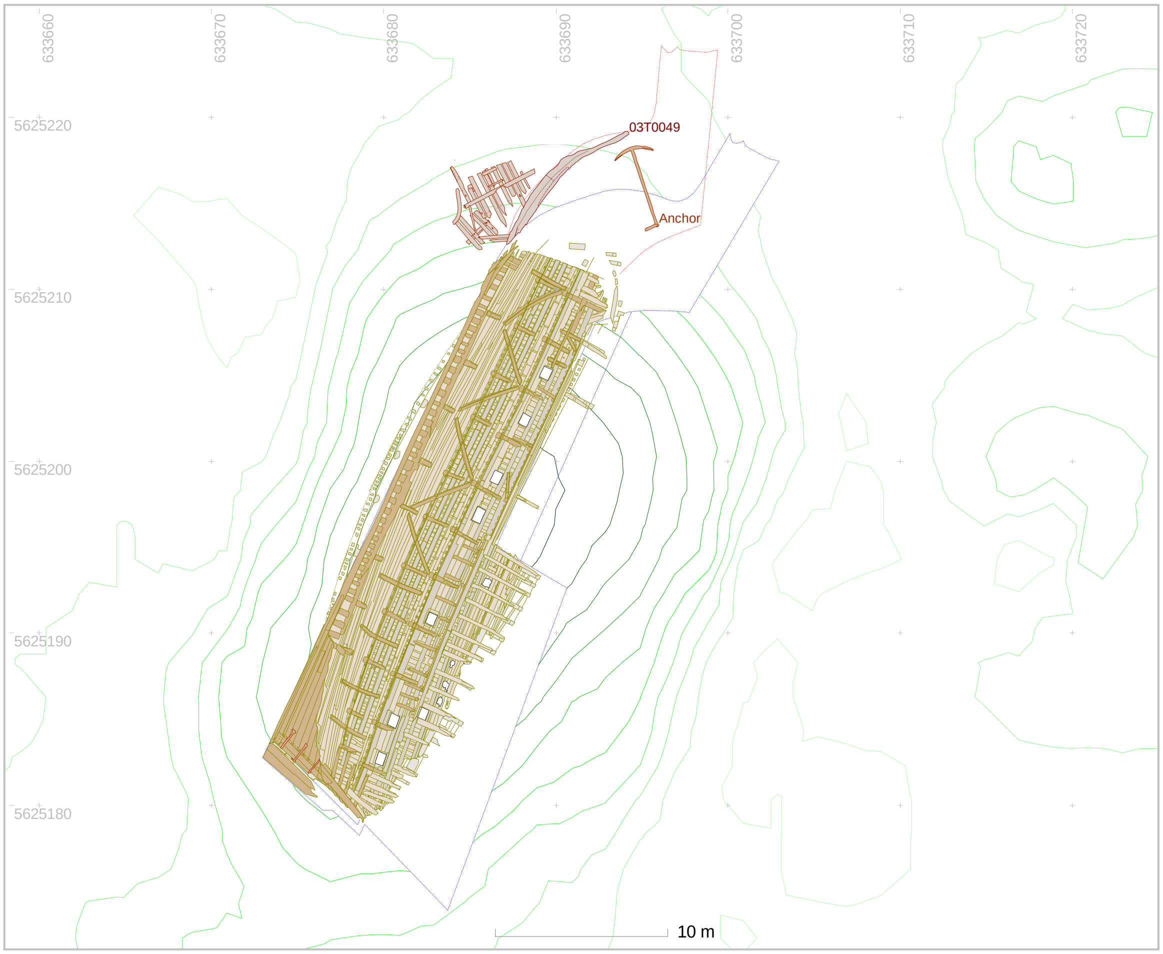


Figure 4: The Mary Rose *hull and bow positions shown in Site Recorder 4*

3.3.7 Drawbacks with Digital Systems

As well as providing a significant number of benefits there are some drawbacks to using a digital recording system. The most significant problem is the need for a computer with a reliable source of power 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.

Lack of familiarity with computers within the project team is also a limiting factor, but this is becoming less relevant now that computers are so common. Problems in this area can be overcome with training, care when designing the recording system and the delegation of more complex tasks to those with the aptitude for that kind of work.

Digital records are sometimes seen as being rather fragile or ephemeral and that they can be deleted or destroyed without effort. Although this is true the problems can be surmounted. The ability to delete or damage the site records can be mitigated by a carefully designed recording system and with the assistance of a robust backup and archiving policy. The ability to lose or destroy digital records is easily overcome by making many and frequent digital copies of the data or by the periodic creation of paper records where a small project size allows it.

4 Recording System Requirements

4.1 Introduction

Although systems may differ in what information they record, the basic requirements for any system are very similar. The list of factors to consider for any recording system is shown below:

- Information Capture
- Ease of use
- Sharing
- Backup and Archiving
- Searching, Sorting and Associating
- Data Validation
- Collection Management
- Publishing

Although paper based systems have some benefits we will put them to one side from now on and only consider digital systems. The benefits of digital systems easily outweigh the drawbacks and they allow us to fulfil more of the requirements outlined above.

4.2 Information Capture

Most important requirement of any recording system is to be able to capture sufficient information about a site in an efficient and systematic way. The aim is to record a site accurately, 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 for or what they are should be kept separate from hard facts such as position or dimensions.

The question of how much to record is not easy to answer as archaeological sites are inherently ‘fractal’ in nature - the closer you look the more detail you see. If we record too little information we may miss some crucial aspect necessary for reconstruction but if we attempt to record too much then we run the risk of running out of time or money. The scope of the operation is also significant as it is less important to capture all of the information if the work on site is non-intrusive, because you can always go back if you miss something. For assessment or survey operations it is acceptable to simply record sufficient information to answer the current questions you have about a site. For an excavation this is not the case, recording only the information you think you need at that time may miss something crucial that can never be recovered.

Preservation by record demands high quality excavation and recording of all archaeological features and some of the problems associated with it are discussed in detail in the technical note below.

Technical Note: Problems with Information Recovery during Archaeological Recording

In principle, everything about any archaeological site should be recorded. In practice this cannot happen as it is not physically possible to record everything and both available time and budgets limit the scope of work. Thus any recording work is a compromise and it is important to understand the issues before making judgements about what to record.

Archaeological sites can provide information at different levels of detail and it is important to consider these levels in any recording process. If we consider a shipwreck site, on the large scale we can record simple facts about the site such as the length and breadth of the ship, the materials used in construction or the method of propulsion. At the next level of detail we can record the position and nature of structural elements and other remains on the seabed. Looking at the site in even greater detail shows us fastenings and tool marks on the structure or patterns on ceramic finds. At each increasing level of detail we have the potential to recover more information and the work involved in its recovery increases too.

If we consider our shipwreck site as a pool of information we can see that not only can we sort this information by level of detail but we can also assign a level of importance. For example, knowing the length of an unknown ship is generally more important than knowing the number of peppercorns found in a box on board as the length will give more clues about the type or name of the ship itself. However, recording the patterns of ceramic finds may be very useful in helping to identify the ceramic and thus provide a date range for the ship. Importance of information is relative and changes as work on site progresses so what was considered less important at the start of a project may become crucial later on. As we can only guess at what would be important at any time it suggests that we either allow more work to be done at a later date by not excavating a site or by ensuring a rigorous recording strategy from the start.

Partial excavation of a shipwreck site may not help as unlike terrestrial sites the site itself is most often a single entity formed in a single event. The excavation of only one part of the ship removes all of the information that exists about that part – for example there is only one captain's cabin that can be recorded.

Allowing an incomplete recording strategy for non-intrusive work may not work either. All underwater sites that are visible (and many that are buried) are degrading at varying rates and often the rate of change can vary over a very short time. The degradation of the site is information loss – erosion loses structural material, collapse loses shape, pilfering of finds loses a wealth of information. So the record of any site is just a capture of that site at that time and if you go back at a later date then the same information is unlikely to be available.

Available time and resources usually has the largest effect on the amount and quality of information recorded. It is unlikely that there has ever been any fieldwork anywhere that had sufficient manpower, skills or tools to do the work to the highest possible standard so each project makes economies where it can. A problem occurs when work is undertaken with insufficient resources and it is that which determines the quality and quantity of information. In some cases this is justifiable as the site is under threat and the recovery of some information is better than none at all but for relatively stable sites this is less acceptable.

The problems associated with information capture in archaeological recording underwater are not well known. Lack of internationally agreed standards in recording and the distillation process that is publication effectively hide the achieved quality of work on any site and thus hide the problem. Work on site is often done using inappropriate methods or is incomplete largely because there is no standard to comply with. This begs the question. How much information is lost due to poor recording and publication compared with the loss caused by treasure hunters, and how much time and effort are being spent on each issue?

4.3 Ease of Use

The second most important requirement is that the system is easy to set up and easy to use, as if it is not then it will not be used or it will be used incorrectly. This factor is one that has often been forgotten in the development of previous systems and has often been the main cause of failure; in fact this is the prime cause of failure of software applications as a whole. The factors to consider are the same as those considered for any software application development so we can use the recommendations from this field when defining the requirements.

- We need to ensure that the most common tasks and actions are simple can be completed easily. For example, it would be awkward if an action like adding an Artefact record were to take 15 keystrokes rather than 2 mouse clicks.
- The way the system works should be consistent so the controls for similar tasks can be found grouped together and they work in a similar way.
- The information contained within the system should be well organised, conveniently accessible and easily retrieved.
- The system must be scalable; it must work just as well for recording a simple survey as 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 used and any system that is too difficult to set up will make users

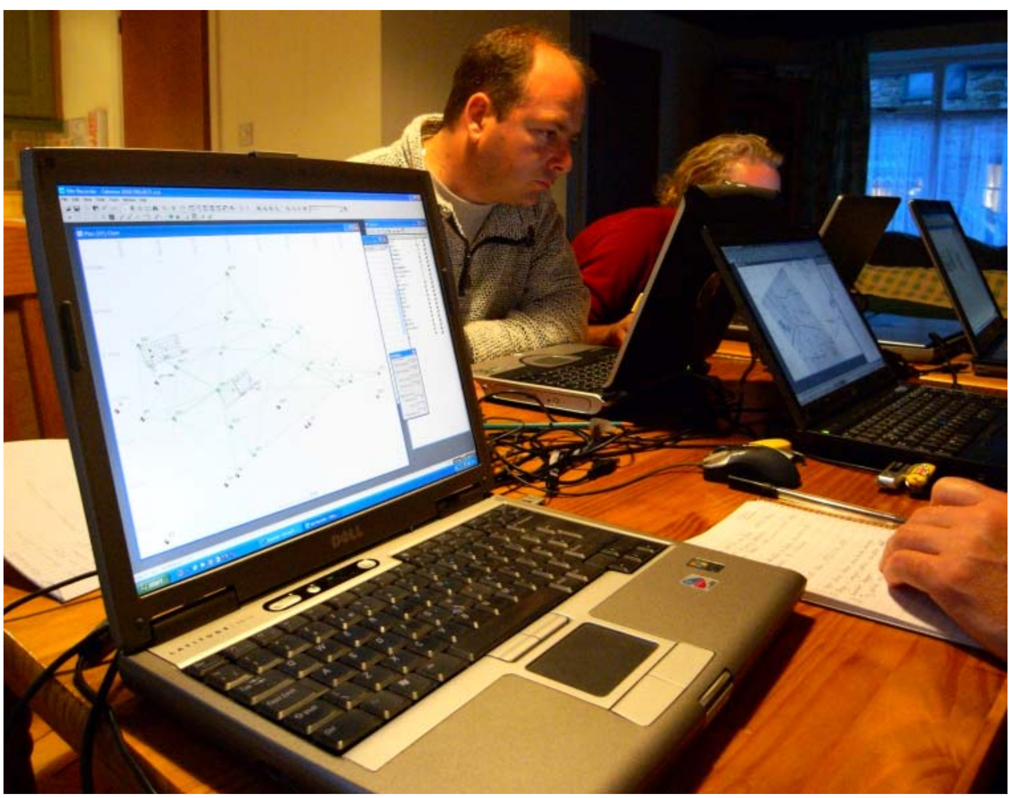


Figure 5: Processing drawings in the field

4.4 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 a few CDs or DVDs, small enough to allow widespread sharing of all of the primary and processed information. The entire dataset from a project can now be made available to the public by publication online. Ease of sharing data not only supports dissemination within the archaeological community but makes information more easily accessible to the general public, often the reason for doing the work in the first place.

If the first step is to be able to publish all of the data then the next is to allow its use, particularly for intra-site analysis. Interoperability or the ability to seamlessly and simply exchange information between computer systems is currently a subject of much research. Information extracted from the recording system can be imported into other database and spreadsheet programs for further analysis, to graphics programs that can show the site in 3D or show changes in the site over time. The options for reuse are only limited by the tools available to process the data and the ease which the data can be accessed. The goal would be for different projects to be able to use information about each others sites directly without any tedious and expensive data entry or data translation. Researchers could collate information from a number of sites without having to understand the details of the recording scheme used by each site. This idea is developed further when we look at the advantages of a standard recording method or schema.

As an integral part of the recording system we need to provide information about the information that it contains, this is known as metadata. This additional information describes such information as what is recorded, how it is recorded and when it was recorded, who did the work and who owns it. The metadata provides a useful summary and can be used by people and computers to see if the information is of interest without having to inspect all of the information first.

Once captured in a digital form, all of the data available can be published rather than just a filtered subset; this raises interesting questions about ownership of data.

4.5 Backup and Archiving

Archaeologists have an ethical obligation to preserve the information they collect during archaeological projects for future generations. Because it is easy to copy electronic information, is also easy to backup and archive that same information so it can more easily be kept forever. Having multiple copies of essential information mitigates against loss or damage to the original and overcomes some of the problems with the fragility of digital data. In addition, storing information in a digital form takes up considerably less space and is more portable than paper. There is much written material available on the subject of archiving and a particularly good source of this is the Archaeology Data Service (ADS) at the University of York, UK.

Some recording systems will contain a mix of original paper records along with their digital counterparts as capturing data straight to computer may be difficult. Survey records are often written underwater so are recorded on plastic paper while finds records may also be recorded on plastic paper as the dirty finds handling environment is not suitable for computers. Here the recording sheets can be scanned and transcribed for use in the digital system but they also need to be kept as part of the archive.

4.6 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 subsets 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 searching so long as the typical search queries are known in advance. The ability to search through free-form text such as notes and descriptions is also of benefit so long as the search tool is sufficiently capable of providing useful answers.

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 open the record for dive on which the artefact was discovered, along with the notes and sketches the diver made at that time.

4.7 Data Validation

Using a digital recording system it is easier to limit the opportunities for the user to make mistakes when adding data to the system or when searching for information. This feature is less significant when data is added to the system by one or two key team members working full time but crucial when the team is large or includes volunteers or part-time workers.

Some of the 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 wordlist or thesaurus. Simple wordlists are simply a list of words that are allowed to be used, but with a list it is sometimes difficult to manage generic and specific terms for similar objects. One solution is to use a hierarchical ‘tree’ of words with more generic terms as the ‘branches’ and more detailed terms at the ‘leaves’. The most detailed form is a thesaurus where terms have a hierarchy and have additional information describing use and suggesting alternative terms. Limiting the choice of options for objects’ properties using wordlists is straightforward on a computer but rather laborious using a paper-based system. This helps ensure that data within a site is consistent, as everyone has used the same words to describe the same kinds of objects and has spelled the words correctly. You also improve efficiency in data entry as selection from a list is usually quicker than typing.

4.8 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 finds. 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.

4.9 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. With the Internet allowing access to information from other areas of our lives at previously unimaginable quantities it was a natural progression to try and extend this to archaeological data. The currently accepted but rather limited method of publication of archaeological data widely used today is at odds with our newly data-rich world. Huge amounts of data are gathered in a site archive yet the evidence presented in a summary or full publication can be highly filtered and does not do justice to the archaeological record. Richards suggests that ‘There is a tension in the publication of archaeological fieldwork results between a synthetic readable account, accessible to the intelligent layperson which 'tells a story', and the scientific presentation of interpretation backed up by supporting data’ (Richards 2004). A digital recording system should allow immediate public access to primary data, even before preliminary and final interpretations have been made.

A distinction has to be made between a well published archive that is complete, has supporting documentation, metadata and allows ready access and one that has simply been dumped on a web site. There is more to publication than the presentation of a documentary archive.

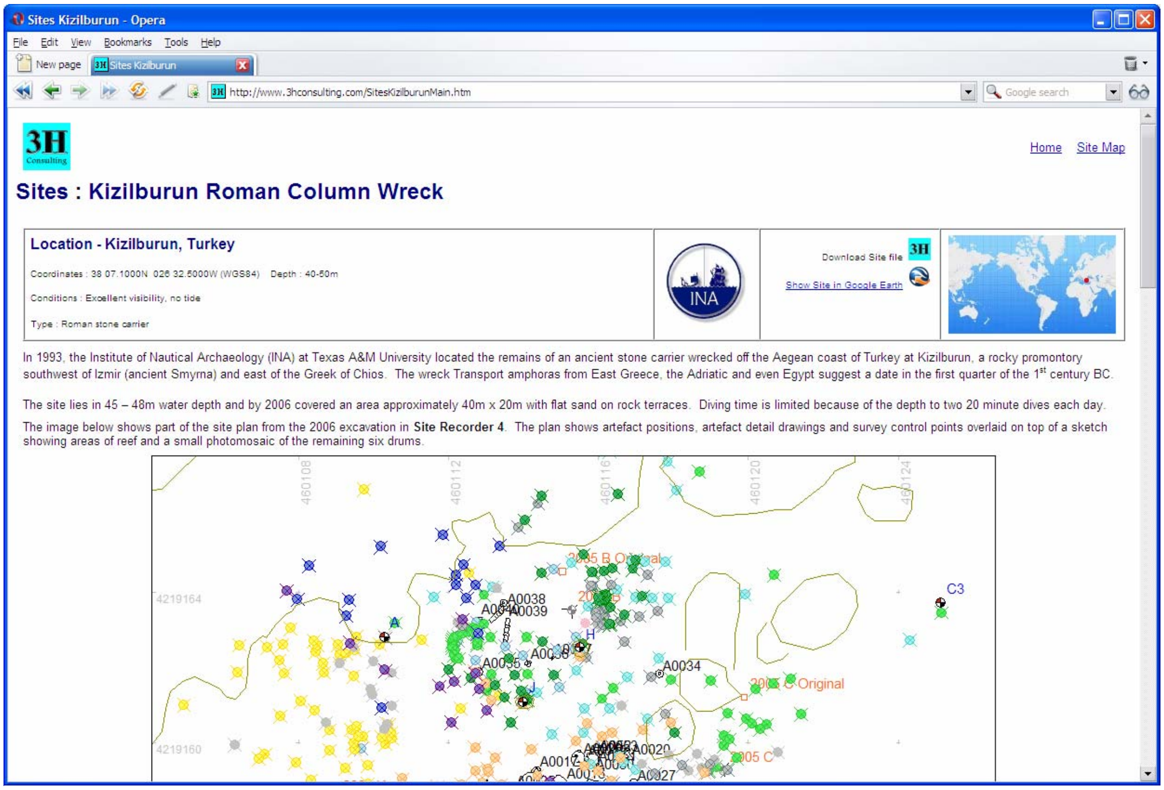


Figure 6: Site information published on the Web

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. With a degree of co-operation and some existing technology we could make huge gains in access to research material. At the most basic level, textual data from the system can be used within published papers created on a word processor as can digital site plans and sections. The next level of access would be to publish complete data sets or selected subsets of them. The World Wide Web is a globally available medium ideally suited to publishing this information to the widest possible audience. Information in a digital form can readily be converted into a form that can be deployed on the Web, either by using fixed pages of information or by using dynamic pages fed with data from an online database. Quite often the amount of information available is too large to be put on the Web in its entirety so as an alternative complete web sites can be published on CDROM or DVD using similar web pages.

The technology is now available to publish archaeological site information in such a way that it can be directly searched by anyone or any program on the Internet. This puts demands on the dataset and the method used for publication to allow others to easily find out what a dataset contains. A computer connected to the Internet running a suitable web service program can be set up to publish information so that another computer can access it directly. The metadata defined earlier gives a clue to the other computer that the data contained may be of interest and the web service on our computer provides the mechanism for the other computer to search the records for what it is looking for.

5 Inside a Recording System

5.1 Introduction

We have looked at the different kinds of recording system that can be developed and we have defined the requirements for a fully-featured digital system. Now we need to investigate the inside of a recording system, the details about what information to record and how best to record it.

5.2 What Should Be Recorded?

How much information is contained in the recording system depends on what it is to be used for and how it is to be used. A system developed specifically for recording simple pre-disturbance survey work on a site would be simpler than one developed for recording information generated by an excavation. A system designed for the management of sites and monuments would deal with sites and finds at a higher or more abstract level than would be required for excavation recording. Creating systems that are scalable and can manage small projects as well as larger ones is possible but requires careful analysis and design.

The types of archaeological sites to be recorded will also affect what is to be captured, so a system specifically designed for recording crannogs may be different from one used to record submerged shipwrecks. Ideally we would want to create a system that could be used for both, so that the system itself could be shared and reused. To make a reusable system that could be applied to both types of sites we need to decide which core information is common to both and which is specific to crannogs or wrecks. Allowing too few units of core information makes the recording system useless as the extra information will still need to be recorded somewhere else. Adding too much core information will make the system unwieldy as only a small proportion of the information possible to be recorded is actually applicable in any one case.

A more simple system may be found to be sufficient for the survey and excavation phases of the project and that something more sophisticated is only really required for post-excavation analysis. There are advantages in using the simple system at the stage where there may be more people associated with the project who are not experts. Burdening a volunteer with a system only understood by an expert will be counter-productive. So long as the systems are well-designed the information in a simple system can be transferred to a more powerful one at any stage in the project then be updated periodically as more data becomes available.

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

Technical Note: System Reuse

As well as enabling the easy sharing of information from a project, using a digital recording system can also allow sharing of the recording system itself. So long as the system has been designed so that it is not specific to one particular project it can be reused on similar projects. The recording system will can then be used to create further data sets that conform to whatever standards have been embodied in the system, further increasing the likelihood that tools are developed to reuse those data sets.

5.3 Information Types

Having decided on the scope of the system (what it will do) the requirements can be defined in detail. The aim would be to record the information contained in our archaeological site along with recording the work undertaken to acquire that information as this helps with later analysis. The process of recording is as important as the data itself as often it gives clues useful in when interpreting results or when dealing with mistakes.

We need to record:

- The primary data: original recordings taken at the site by archaeologists and site recorders
- Activities, such as what has been done at the site in terms of recording, excavation, monitoring activity and so on
- Historical records: any documents or images relating to the site
- Reference documentation: information regarding other similar sites and historical events

Figure 7: Finds registration

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 contain information about something that physically exists, such as an artefact, something that happened, such as a geophysical survey or they can describe other objects such as the site itself.

These records can relate to any phase of work on a site, so can include events that have already happened and any fieldwork planned for the future. A survey may be done to search for a known shipwreck and this may generate records containing information about geophysical target points. Subsequent dives to investigate these targets may need artefact records to capture information about what was found on the seabed and each dive itself will be recorded in a dive log so a dive log record will also be needed. A pre-disturbance survey will generate survey points, measurements and dive logs, while an excavation will add features, drawings, photographs and so on.

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 the information to be recorded at any stage we can define the data to be managed by the recording system.

5.4 Records as 'Objects'

Working on a computer, each of these records is most easily modelled as an individual 'object' rather than as something in a simple list. Record '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. In technical jargon our recording system uses an 'object oriented' design, but this simply means that it manages objects rather than simple lists.

We could keep all of our information in a simple list on the computer and a spreadsheet is ideal for this task, unfortunately problems occur where lots of identical data is to be recorded. For example, if one artefact needs information about the dive on which the find was found then this can easily be typed in to the relevant column in the list. However, if on that dive there were fifty artefacts tagged then each artefact record would then need a copy of the same dive information making the spreadsheet rather large. A more efficient way would be to associate the dive log record with each artefact record so only the link between them takes up space in the list. Although this is possible to do in a spreadsheet, it is considerably easier to do using a relational database or GIS.

5.5 Recording an Object

5.5.1 Object Properties

The details of a database schema can be considered having decided what to record, this involves determining which 'objects' will be recorded and what properties they contain.

Object properties are the items of information that we want to record, so our recording system contains objects and each object contains a list of properties. For example, for an Artefact we may want to record properties such as its name, its position, its length, width and height. For a dive log we could record the name of the diver, supervisor, time in, time out and maximum depth.

These properties themselves are one of a few simple types: numbers, text, link and association.

- A numeric or number unit would be used to record such things as the length or weight of an object. We often differentiate between whole number integers (such as 12 and 42) and real numbers (such as 2.451 and 30.3) but this is largely for convenience within the computer recording system. We may also put defined limits on what the number is allowed to be, such as bigger than zero or less than 3000.
- Text units can be used to store the name of an object, something longer such as a description, or something very long such as the transcription of a relevant document. We may also want to limit which words can be used; this has been mentioned above where the text is limited to a set of preferred words in a wordlist.
- Links are used to connect an object such as an image or a video clip with a digital bitmap file held somewhere on the computer. This allows us to show the image or video from within the computer recording system.
- Associations are similar to links but connect objects within the recording system to one another. For example, Dive log records can be associated with an artefact record. Anywhere in the system where one object is related to another we can add an association.

5.5.2 Names and Object Identification

One of the fundamental requirements of the recording system is that every object must have a unique identifier so that we can tell one object from another, the name of the object is often used in this role. 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. Furthermore, the name for that object can never be changed as this would break any links to it. It is better to create a unique identifier in the form of a number as well as the name so that even if the name is changed the identifier still remains the same. On the 'inside' of the system the objects are referred using the unique number identifier, while the user sees only the friendlier object name.

Almost every project has its own naming scheme for artefacts, samples, survey points and other similar objects. The naming schemes were often created when paper records were the only option, so complex schemes were devised in an attempt to manage the paper more readily. Often the schemes were used to separate object types or objects made of a particular material in order to allow more rapid searching of the data.

When using a computer recording system these schemes are an over-complication best avoided. Generating a list of all of the timbers recovered in one particular season's excavation can be created quickly and easily by filtering, so long as the computer knows the object type and the year it was recovered. The simplest scheme would be to give each object a sequential number, whatever its type. Giving objects elaborate names that involve the object type, the material and the year will not make the search any faster on a computer but will slow down the initial recording process and be prone to errors.

In practice appending an object type identifier to the name can help interpretation of maps and charts. For example, all artefact objects have the prefix 'A' and all survey control points use the prefix 'CP'. A list of suitable prefixes is given below:

Artefact	A
Feature	F
Sector	Usually given a name such as 'Trench 1' or 'East Site'
Sample	S
Control point	CP
Detail point	DP
Dive log	DIV
Wreck	WK
Target	T
Image	IMG
Event	EVT
Contact	Use the name of the person or organisation
Source	SRC

The numbering scheme is important too. It is better to use leading zeroes for numbers creating sequences such as 'A00001, A00010, A00100, A01000' rather than 'A1, A10, A100, A1000'. Five digits are usually enough for most sites as it allows up to 99999 objects of any given type (A00001 to A99999). There are not usually more than 99 survey control points on a site so numbering them CP01 to CP99 seems sensible.

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 found or recovered 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 giving names in the form '08A01564' for artefacts recovered in 2008.

Objects that are special and few in number such as guns and anchors can have a specific numbering scheme for ease of recognition as ‘Gun 12’ is more readily identifiable than some generic artefact name such as ‘08A01284’.

The names used for dive logs should be chosen with care as often other things are associated with the logs such as photographs taken or finds recovered on the dive. Problems with dive log names can occur when many divers are put in the water at the same time in small teams but are given different tasks. The ‘dive’ may involve lots of divers but each team doing a different task should create a separate dive log as the dives may be in different locations and we want to accurately record the tasks allocated. One solution is to allocate the whole ‘dive’ a single number but then add a suffix to the dive for each team so dive 34 undertaken by three teams of divers would create dive logs called DIV-034a, DIV-034b and DIV-034c.

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 limits to the characters that can be used in a name, often such characters as ‘ , ; / . : \ " ? * | ’ are not allowed as they have special meanings.

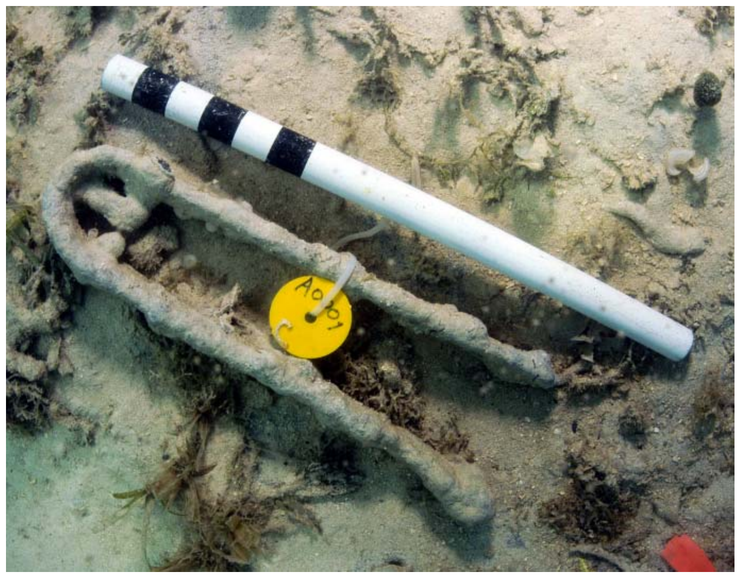


Figure 8: All artefacts should be clearly marked

5.5.3 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 useful as a place for the extra information. The notes often become the place for discussion about the interpretation of an object so contributions can be added by many people and by appending each contributor’s initials you can capture a complete trace of ideas. If you find that the notes are regularly being used to store one particular piece of information then this suggests that the recording system needs to be extended by adding that piece to the core set of properties for that object.

5.5.4 Units

The units of length, weight and so on 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 one particular set.

The idea that a shipwreck site should be recorded using the same units as those used to build the ship causes problems in practice and is best avoided. The first problem is that the older units may be unfamiliar to some members of the team leading to measurement errors. There is also a tendency to round up measurements to the nearest whole unit, recording timbers as they were assumed to be before erosion rather than as they are now.

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

5.5.5 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 at the outset to consider time when recording positions.

5.5.6 Measurement Recording

A system capable of recording underwater or inter-tidal excavations should be capable of handling information about survey points and measurements. The main reason for this is that a significant proportion of measurements are mistakes, so raw measurements need to be recorded in the archive so that they can be reprocessed at a later date. These errors are an inherent feature of the types of measurement found on these sites. Later work can often suggest that there may be errors in earlier work, so it is essential that earlier results can be reconstructed from the original measurements.

5.5.7 Object Drawing

As well as being able to record text and numeric information about objects on the site, the recording system could also record shape information as well. If we consider an artefact on the seabed then this can be represented on a site plan in a number of ways:

- The simplest representation is as a point shape where the position of the point shape identifies the position of the artefact and the size, colour and style of the point is used to identify some other property such as the artefact 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 could be shown in-situ as a 2D drawing on a plan view.
- In the most complex example the artefact could be drawn as a 3D or solid model that could be viewed from any position in a 3D representation of the site.

Figure 9: Drawing timbers at 1:1 scale

In each of the cases above the information needed to show the artefact is additional to its archaeological properties that exist within the recording system. There is a clear distinction between the properties that describe what the object is from those used to show what it looks like.

The amount of information that is required to display the artefact on the 2D or 3D site plan gets larger as the complexity of the model increases. For the case where the artefact is shown as a point then only the position, point style, colour and size are required. For a 2D plan view we would need to record how the artefact appears from above as a collection of polylines, perhaps with different colours and fill styles. A 3D model would require a 3D representation of the artefact to be generated which requires a huge amount of information if the representation is to be realistic rather than stylised.

In practice, most plans show artefacts as simple points or plan view outlines of the artefact shape as the information needed to be recorded is reasonably small. The information needed to draw an artefact in plan view can be obtained from the drawings created from the recording the site using planning frames, scanning and digitising these sheets provides the outlines and edges required.

Some research is being done into 3D representation of archaeological sites underwater but to date they are limited to representing complex and rather 'organic' objects as ideal models. These models are also limited to representing simple, common shapes such as amphorae, cannons and anchors as only one has to be drawn and the others are copied from it. The problem is not in rendering the information but in collecting sufficiently detailed information; you only have to look at a modern 3D computer game to see what could be achieved if the 3D data were available. The seabed can often be approximated from high-resolution single or multibeam sonar data so you can create approximate models of a complete site that are useful for interpretation. However we still do not have the ability to collect sufficient information to be able to accurately render a complete shipwreck containing eroded structure and concretion.

5.5.8 Site Code

A unique site code should be given to the each site, a code that is used to uniquely identify the site you are working on. Site codes need to be different as they are used as part of the name of every object on the site so you can differentiate objects from one site from one of the same name on another.

With digital systems it is easy to make comparisons of datasets from different sites. On those sites it is possible that the names of some of the objects are the same so we need to be able to tell artefact '06A0453' from one site from the one of the same name on the other. In fact, the full name of each object should have the site code added to it so each name then becomes unique.

The site code may be allocated based on a standard naming practice, if not then you will have to make up your own code. It is common to create a site codes from the name of the location of the site and the name of the site itself, sometimes the year of finding is added too. For example, the site code for the Mary Rose is 'POMR' created from 'Portsmouth' and 'Mary Rose', the Cattewater wreck is 'PLYCAT' from 'Plymouth' and 'Cattewater' while the USS Monitor has a site code of 'MNMS' from 'Monitor National Marine Sanctuary'.

Technical Note: Data Types

Another way of looking at the different kinds of data to be recorded is how the information is displayed on the computer. This includes:

- Point data
- Line or vector data
- Area data
- Complex data
- Scaled image, raster or coverage data
- Non-spatial data
- Temporal data

Point data is represented by a single point object with a 3D position, artefact objects are often represented this way on the chart. Line data is comprised of a number of line segments joined together in what is known as a polyline, this is in effect a series of points joined together by lines and can be used to represent a line object such as a depth contour. Closing the ends of a polyline forms a closed polygon shape which defines an area, for example the extents of a trench.

By combining lines and polygons into a single object we can form a complex shape; these can represent the plan drawing of an artefact on the seabed.

We can also show a scaled image on the chart, such as a scanned site plan or an aerial photograph. Each pixel of the image then covers a known area of the site plan and the pixel can be used to represent a measurement or value at that point, as with the depth from a rendered image from a multibeam sonar.

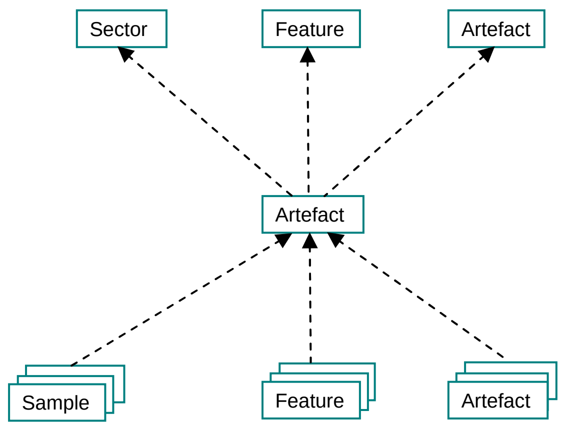
We also need to be able to record non-spatial information to record things that do not appear on the chart such as source documents and people. Typically these kinds of objects are shown in lists. We can also record information about events such as a survey or a dive, these temporal objects are also shown in lists rather than on the chart but they also have a defined start and end time associated with them.

Images can be represented in a number of ways: a publication photograph of an artefact can be shown in a list but will not appear on the chart, a general photograph of the seabed can be represented as a point object on the chart that marks where the photo was taken. A photograph from the site taken vertically downwards can also be shown on the chart as a scaled image.

5.6 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.

Objects can be used to define areas of the site or seabed such as trenches, groups of scattered finds or areas that have been investigated; these are collectively known as Sectors. Feature objects can be used to represent such things as large concretions, areas containing particular deposits or the 'ghost' remains of a container that has long since disappeared leaving just the contents. A large concretion recorded as a feature may contain a number of artefacts and other, smaller features. Artefacts themselves can be associated with other Artefacts, either by containing them or simply by being close to them.



The relationship between these objects gives clues about their use so the recording system should allows these relationships to be recorded in a formal manner and provide tools to visualise them. The result is a hierarchical tree of relationships between Artefacts, Features and Sectors.

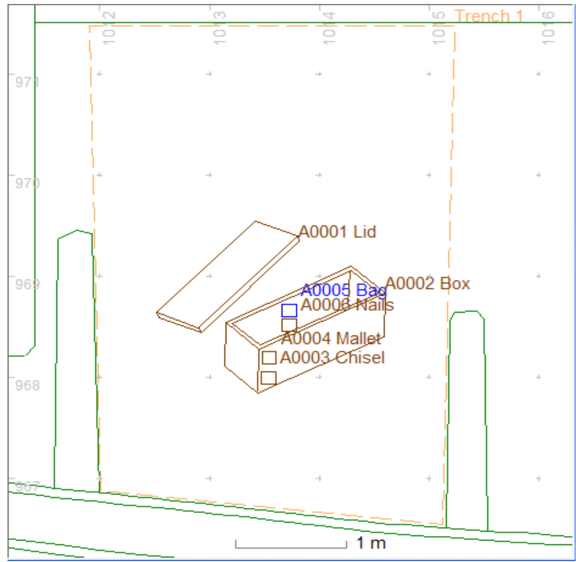
Associations between objects include:

- 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



This shows the relationships between all archaeological objects in a hierarchy with the site at the top followed by Sectors, Features and Artefacts. Objects are then shown related to other objects with the type of relationship 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.

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.

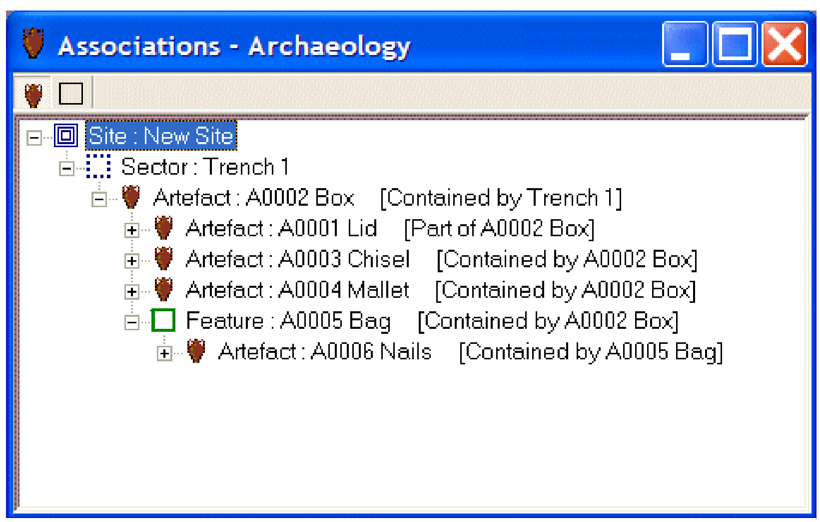


Figure 10: Object associations in Site Recorder 4

These relationships can be shown in ***Figure 10*** below taken from the Site Recorder 4 program.

6 Other Development Considerations

6.1 Introduction

In the previous section we looked at some of the fundamental issues of what to record and how to record it. In this section we consider more technical issues to do with maintaining data quality, data description and data sharing.

- Controlled Vocabularies, Wordlists and Thesauri
- Describing the Data with Metadata
- Exporting and Publishing Data
- Archiving
- Existing Standards
- Documenting the Recording System

6.2 Controlled Vocabularies, Wordlists and Thesauri

One of the drawbacks of paper based recording systems is the difficulty in maintaining consistency in the information that is recorded. Consistency is essential for ensuring accuracy of the information but also simplifies the process of searching and sorting. A search for objects of a particular type or style is far more likely to be successful if the objects are described in the same way and the words used to describe them are known.

The records for two identical objects recorded by two different people should be the same however this rarely is so in practice. With paper based systems a degree of consistency can be achieved using pre-printed forms as they ensure that the same information is recorded for objects of the same type. With digital systems we can go further by limiting the choice of options available for some object properties using a controlled vocabulary.

A controlled vocabulary is simply a collection of words that are allowed to be used and they are useful in situations where clarity and consistency are important. The English language, like many others, has many words that mean nearly the same thing but with subtle differences. If a controlled vocabulary is used then the large set of similar words that could be used are limited to one or two. Controlled vocabularies have many used but particularly for where language translation occurs as it is much easier to translate documents containing a limited selection of well defined words. For an information recording system a controlled vocabulary helps ensure that descriptions are consistent.

In a digital recording system we can readily enforce a controlled vocabulary by only offering the user the option of selecting one word or phrase from a list. As well as maintaining consistency this increases efficiency as no typing is involved, simply a few mouse clicks.

Where the number of options is short then they can be offered to the user as a simple list, known as a wordlist. A wordlist can be enhanced by adding short descriptive text to sit alongside each option, providing the user with more information about the choices available. Long wordlists can be laborious to work through so the alternative is to offer a 'tree' of options rather than a list, often implemented as a hierarchical wordlist or thesaurus (plural - thesauri). The user starts at the top of the tree by being offered a few options then clicks on branches of the tree to 'dig down' into further choices. This top-down approach ensures that only a limited set of options are offered to the user at any time, but the drawback is that multiple choices need to be made to make just one selection.

Figure 11 shows an artefact record in Site Recorder 4 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 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. Using and sharing standard wordlists will bring benefits to your project, particularly as time goes on and the same lists are used by more projects.

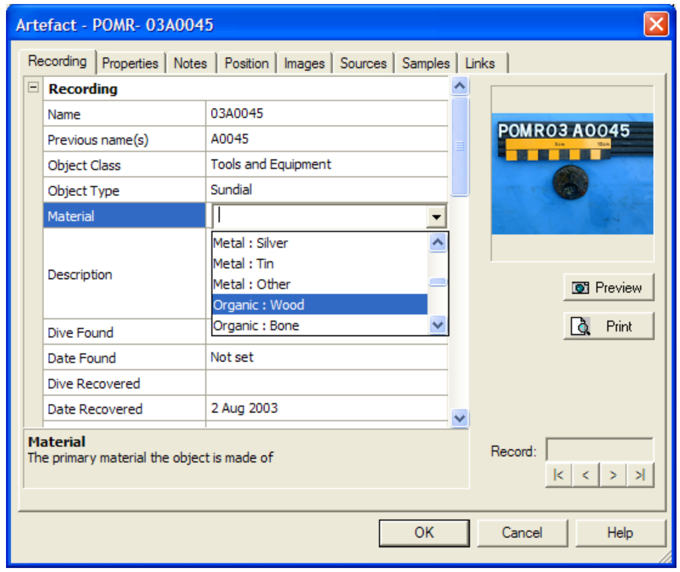


Figure 11: Artefact record from Site Recorder 4

Example standards include:

- INSCRIPTION, a collection of wordlists maintained or recommended by the Forum on Information Standards in Heritage (FISH) (http://www.fish-forum.info/inscript.htm)
- The MDA Archaeological Objects Thesaurus (http://www.mda.org.uk/archobj/archint.htm)
- The UK National Monuments Record Thesauri (http://thesaurus.english-heritage.org.uk/frequentuser.htm)

6.3 Documenting the Archive

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

The documentation should include:

- 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 geo-referencing 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 filenames 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

Derived from Archaeology Data Service Guidelines for Depositors (http://ads.ahds.ac.uk/project/userinfo/deposit.cfm).

6.4 Describing the Data with Metadata

To make best use of all of our hard work we need to ensure that the data we record gets used as often and as widely as possible. To help make this happen we need to make it easy for others to get a copy of the dataset but first they have to be aware that the dataset exists. The potential users of the data need to be able to find out what information the dataset contains, for this we need to use extra properties in the archive known as metadata. Metadata is used to describe what information is recorded, how and when it was recorded, who did the work and who owns the information. Metadata is data about data. The metadata provides a useful summary and can be used to see if the information is of interest without having to inspect all of the information first. People can use metadata information to help identify data they are interested in but greater power comes from publishing datasets online and allowing computer search engines to use the metadata during their searches.

All sorts of extra metadata could be added to the dataset in many different ways but we need to ensure that what is added is as useful as possible. Ideally we would add all of the metadata that potential users would expect to find and in the format in which they would like to use it. Fortunately there is an existing standard set of metadata elements we can adopt called the Dublin Core. The Dublin Core was developed by the Dublin Core Metadata Initiative (DCMI), an organization dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources that enable more intelligent information discovery systems. Dublin Core specifies a standard list of metadata elements such as:

Name	Definition
abstract	A summary of the resource
accessRights	Information about who can access the resource or an indication of its security status
created	Date of creation of the resource
creator	An entity primarily responsible for making the resource.
format	The file format, physical medium, or dimensions of the resource.
identifier	An unambiguous reference to the resource within a given context.
issued	Date of formal issuance (e.g., publication) of the resource.
publisher	An entity responsible for making the resource available.
title	A name given to the resource.
type	The nature or genre of the resource.

Note that this is just a sample of the many elements in the latest specification, for the full list see http://dublincore.org/documents/dcmi-terms.

6.5 Exporting and Publishing Data

The recording system may only be suitable for collecting data rather than in-depth analysis or public-friendly publication. As we want our data to be used for analysis and as wide a publication as possible we need to make sure that the data in the system can be extracted easily and reliably.

Although any format for the data could be used, more widespread reuse would happen if data were in a standard and widely used format. The ideal mechanism for data interchange is now available as the eXtended Markup Language (XML) which is a non-proprietary open 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 computer - something which was such a problem in the past. Another benefit is the ability to automatically validate the information before use and so avoid corruption of crucially important data.

6.6 Archiving

The data contained in the recording system is likely to have been expensive and time consuming to capture and may be the only surviving record of a site. The data may well be priceless from both financial and cultural standpoints so it is essential that the data is well looked after. 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, soaked, or fried 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. If multiple copies exist then it is essential to be able to tell which is the master and which are the copies. Each dataset that is copied should have a version number that is unique to that version so older copies should have a lower version number than the master. Using the datestamp on a file for version control is not recommended as the date is easily altered.

The dataset can be copied to any media large enough to hold it. At the time of writing the favoured media include external hard drives or DVDs for large projects, CDs or USB memory sticks for small projects. This leads neatly into the problem of media survival as we must ensure that our copies can survive on the media they are written on. It was not that long ago that the floppy disk was the media of choice for archiving yet now it is hard to find a floppy disk drive on a computer. Disk media such as DVDs and CDs may also have a surprisingly short lifespan so any archive policy must include periodic refreshing of the datasets by copying to new media of the same type or moving to a new, non-obsolete type. 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 long term 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.

Technical Note: Guidelines for Depositors

A detailed and comprehensive guide to depositing an archive can be found in the Guidelines for Depositors (2008) from the ADS (http://ads.ahds.ac.uk/project/userinfo/deposit.cfm). The guidelines include recommendations for documenting the dataset, metadata to include, preferred data formats, file naming strategy and delivery method.

6.7 Existing Standards

Wherever possible, it is recommended that recording systems are developed so that they comply with any existing national and local data standards so any information gathered can be used directly with no translation. By adhering to existing standards you allow more ready exchange of information between your 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 would wish to record.

The MIDAS Heritage published by English Heritage would be a good example of a suitable standard, MIDAS Heritage is the UK data standard for information about the historic environment. MIDAS Heritage sets out an agreed list of the 'units' of information that should be included in an inventory or other systematic record of the historic environment.

7 Data Sources for each Project Phase

7.1 The Phases of a Project

We can now look at the types of information and data sources that are relevant to each phase of an archaeological project. Projects go through a number of phases during their lifetime, from initial inception through to 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 archaeology project are:

Planning and Assessment

Search

Survey and Monitoring

Excavation

Conservation

Analysis and Interpretation

Publication and Archiving

During each phase a different set of data sources is encountered. The team member responsible for data management needs to ensure that the correct versions of any datasets are available at each stage so the information can be used to assist decision making.

7.2 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 see what is missing.

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

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, geodesy
- Scan and link images
- Scan and link documents, copyright
- Create a timeline of events

7.3 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 would be 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.

Typical sources of information include:

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

Typical tasks include:

- Import target lists
- Import raw survey data and reprocess to create target lists
- Import sidescan sonar mosaics and multibeam images

7.4 Survey and Monitoring

Once a site has been located then 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 excavation work; this is done more carefully as the work cannot be repeated at a later date. High accuracy survey work 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, relative movement of structure as a way of monitoring gradual collapse or the measurement of sediment depth in and around a site.

Typical sources of information include:

- Primary, secondary and detail control points
- Distance, depth, height, offset, ties and radial measurements
- Surface position measurements from GPS
- Subsea position measurements from an acoustic positioning system (APS)
- Drawing frame drawings
- Photomosaics
- Point positions from 3D Photogrammetry
- Tide measurements for correcting depths
- Survey measurements defined above
- Sediment depth measurements

Typical tasks include:

- Design the survey control point network
- Add survey points
- Add survey measurements
- Process measurements
- Scan, import and digitise drawing frame drawings
- Import and georeference photomosaics
- Import point positions from 3D photogrammetry
- Correct depth measurements for the effects of tide

7.5 Excavation

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.

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

Typical tasks include:

- Add Artefact, Feature, Trench and Sample records
- Add linked images
- Add Dive Logs
- Scan, import and digitise artefact drawings or photographs and add to the site plan

7.6 Conservation

During conservation work further information is recorded about the artefacts being conserved. The artefacts will need to be cleaned then recorded before conservation, conserved then recorded afterwards to note any differences. The conservation treatments applied to each artefact will also need to be recorded; the artefact may need further treatment at a later date so a history of its treatments should be recorded.

Typical sources of information include:

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

7.7 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 from the analysis of samples taken during fieldwork.

Typical sources of information include:

- Sample analysis reports

7.8 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.

Part 2 – Data Management using Site Recorder

8 Site Recorder

8.1 Introduction

From the data sources identified above we can create a list of objects that need to be captured in our recording system. This schema can be used to record things that exist or happen, most usually in or around an archaeological site. The collection of objects and the properties to be recorded about each one are collectively known as a 'schema'. The schema we will now look at has been developed for use within the Site Recorder 4 program and is known as the Maritime Archaeological Recording Schema (MARS).

Here we have created a list of things or 'objects' to be recorded such as an Artefact or a Sample. The objects are grouped together in a hierarchy so we have some objects that are used to collect together other objects, Layers and Projects and Sites are used for this. Along with the objects and their collections the schema also defines the metadata used to describe the contents of the recording system.

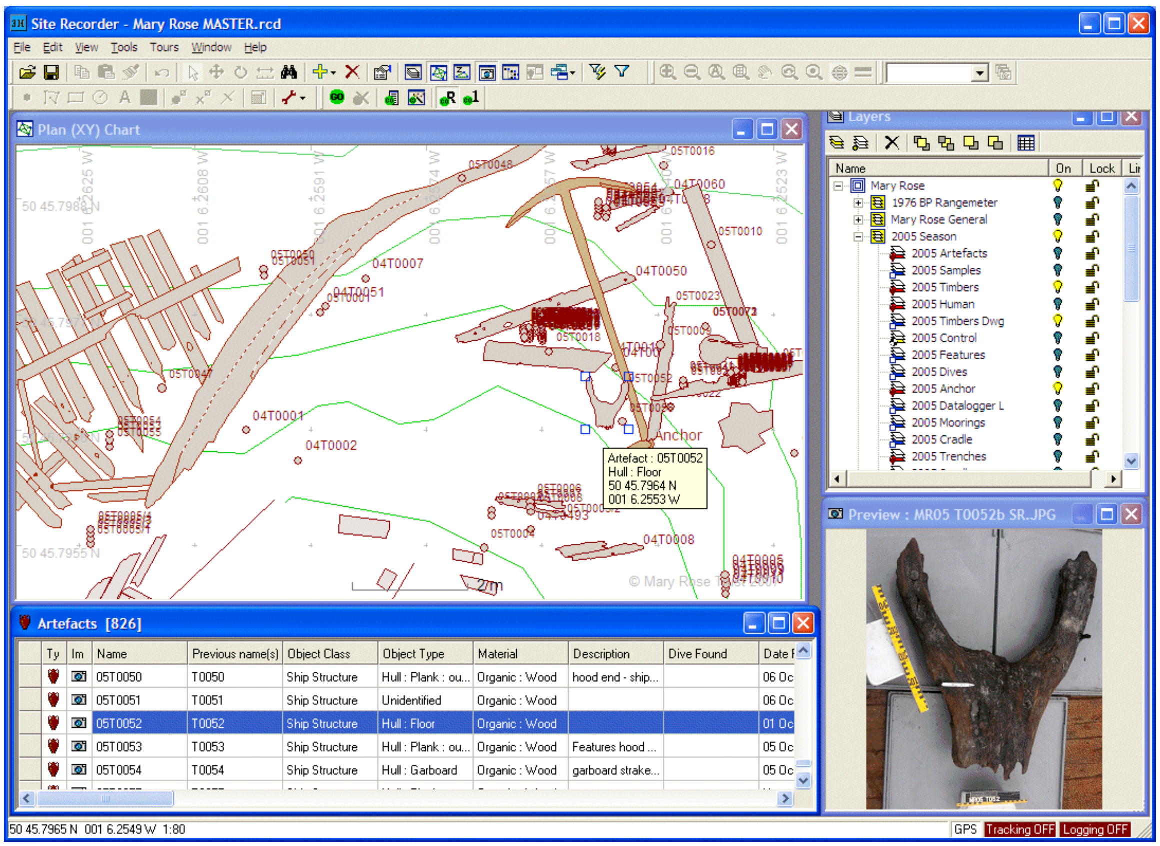


Figure 12: A typical screenshot from Site Recorder 4

8.2 About Site Recorder 4

Site Recorder 4 is a fully-featured and versatile Geographic Information System (GIS) specifically designed for use in maritime and intertidal archaeology. Site Recorder is powerful yet easy to learn and has been designed by archaeologists for archaeologists.

With Site Recorder you can collect together all of the information about an underwater or inter-tidal site in one place. Separate pieces of information can then be associated with one another allowing easy location, analysis and interpretation of the information. Unlike most other GIS programs, Site Recorder is designed for collecting information not just displaying it. Site Recorder has been designed to replace the separate surveying, drawing, finds handing and reporting programs usually used on site with one single program. All data can be geo-referenced and time-stamped allowing true 4 dimensional analysis.

Typical applications for Site Recorder 4 include:

- Excavation planning and recording
- Site survey planning and recording
- Geophysical survey post-processing
- Search planning and post-processing
- Resource evaluation and management
- Site publication and reporting
- Archiving site data

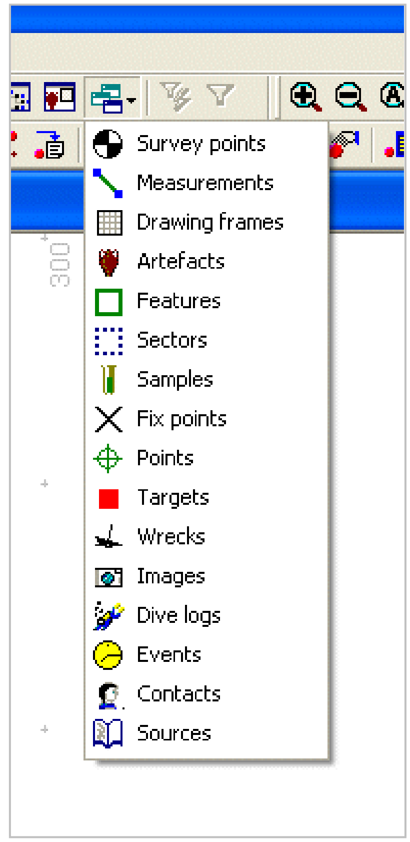
With all of the information about a site captured in a Site file it can be used as a basis for creating reports or exported for further processing in other programs. Complete sites and associated image and source files can be published on CD or DVD along with a free reader program called Site Reader. With the entire Site information in one place and linked together this also forms an ideal method for archiving site data.

Since its first field trials in January 2004, Site Recorder has been used on projects in more than 18 countries around the world. The history and development of Site Recorder are described in the paper *Development of an Object-Oriented GIS for Maritime Archaeology - Motivation, Implementation and Results* (Holt 2007a).

8.3 Object Types

8.3.1 Artefact

The Artefact object is used to record information about a single artefact or find. All of the information about that one find is recorded within one single object.



8.3.2 Feature

Feature objects are used to record features and contexts found on the site. Features are commonly used to record information about concretions that contain other Artefact objects.

8.3.3 Sector

Sector objects are used for defining an area of the site such as a trench.

8.3.4 Sample

Sample objects are used to record information about samples taken for environmental evidence or type identification. Environmental samples can give clues about the potential survival of artefacts and the detail of an object's burial environment can give clues to conservation required.

8.3.5 Survey Point

Survey Point objects are used to represent the control survey points and detail survey points used for positioning on site.

8.3.6 Measurement

Individual measurements made between these points are recorded in separate Measurement objects, supported measurement types include Distance, Depth, Offset, Ties, Radial and Position.

8.3.7 Image

Image objects are used to record information about any image object including photographs, drawings and video clips.

8.3.8 Image Basemap

Image Basemap objects are used to record information about base map pictures such as side-scan sonar traces, geo-referenced multibeam echo sounder (MBES) images, scanned site plans or photomosaics.

8.3.9 Dive Log

Dive Logs are used to record information about dives as these are often used to tie together objects that have been recovered with location information.

8.3.10 Contact

Information about people associated with the site such as archaeologists and divers can be recorded in Contact objects

8.3.11 Event

Event objects are used to record things that have happened on site or to the site, such as survey work, excavations and site assessments.

8.3.12 Source

Sources such as documents, reports and letters can be recorded along with a link to the source if available in a digital form.

8.3.13 Target

Target objects are used to record positions and information about items found during a geophysical survey.

8.3.14 Wreck

Information about known shipwrecks and reported wrecks can be recorded in a Wreck object.

8.3.15 Logbook

Logbooks can used to record day-to-day events, ideas, interpretation and thoughts as text documents.

8.3.16 Drawing Object

Drawing objects include points, lines (polylines), rectangles, circles and text. Drawing objects can be used to draw such things as maps, trench outlines, contours and artefacts.

8.4 Object Hierarchy and Metadata

As well as recording information about objects encountered in a site archive the schema also specifies ways of grouping objects into hierarchies and ways of defining relationships between objects. This allows the data to be ordered in a structured manner making it easier to evaluate and navigate through the archive; generally the grouping is ordered by object type. Stratigraphic grouping is made feasible with all georeferenced objects by specification of position and orientation for each object.

The schema objects that define hierarchy include the Site, Project and Layer. A Site can contain a number of Projects and each Project can contain a number of Layers. The Layers are used to contain georeferenced objects such as Artefacts and Survey Points.

The Site is the main object that represents the ship, monument or structure being recorded. The Site contains one or more project objects and these are usually used to group together data from individual seasons or from self-contained sets of work such as a geophysical survey. Each Project can contain one or more Layers, the layers are used to collect together objects that are associated with each other, most usually by type.

The objects are brought together into collections or families which can be 'owned' by another object, this is a handy way of collecting together similar objects or ones related to each other. The hierarchy of objects forms a 'tree' shape with the Site at the root of the tree and all the many other objects branching out from it.

Information relating to the site as a whole is also contained within the site object. This includes archive version, metadata based on the Dublin Core and geodetic information.

Technical Note: MARS Specification

The specification for the Maritime Archaeological Recording Schema (MARS) is available online from http://www.3hConsulting.com/Research/research_schema.htm (Holt 2007a).

Technical Note: Fixed Schema Recording

Site Recorder 4 (SR4) uses a fixed schema for recording information so the fields or properties of each object are fixed within the system and they cannot be modified or added to. Other systems have been developed where the fields can be manipulated or created from scratch, such as IDEA (Madsen 1998).

This restriction within Site Recorder may at first sight seem to be a limitation however there are a number of good reasons for its introduction. Deciding what to record and how best to record it is not a simple task so with SR4 the user is relieved of that problem.

The schema is the end result of work on many underwater sites from simple survey tasks to major excavations so contains the sum of experience from all of these sites.

Setting up a database is also difficult and requires particular expertise that may not be available within the project team; in fact Eiteljorg dedicates an entire 50 pages to it in *Archaeological Computing.*

Customised recording systems works directly against the idea of data sharing as users may record different information or worse, the same information in different ways.

The fixed schema was implemented in SR4 initially as an experiment to see if a standard schema could be applied to any maritime archaeological site. To date, this experiment has been successful and the program has been used unchanged on all the projects its been applied to.

9 Data Management using Site Recorder 4

9.1 Introduction

We can now look in detail at using SR4 as the data management system for a typical project, identifying the tasks that need to be completed at each phase of the project. From a data management perspective, the project phases include:

1. Project Start-up
2. Pre-Fieldwork Planning
3. Fieldwork
4. End of Fieldwork
5. Post-Fieldwork
6. End of Project

The process of setting up a new project in Site Recorder is a simple matter of creating a new Site file then adding whatever relevant information is to hand. In the first instance the Site file may just contain a point on the chart that identifies the approximate position of the site, but as the work progresses the Site file becomes the repository for all information about the site and its size grows accordingly.

The same process is undertaken when migrating an existing project archive to Site Recorder, this is discussed at the end of this section.

9.2 Project Start-up

9.2.1 Introduction

The start of any project is always an exciting time, often the planning work is done before the fieldwork season has started and it hints of the adventures and discoveries to come.

9.2.2 Project Name

The first step is to decide a name for the project, this task should not be a decision taken lightly as the name will live forever in the archive. Avoid names that have been used before, those that are supposed to be humorous or those that are just too long. The name you use at the start will often remain in use even if the name is changed at a later date. Remember that you may need to introduce the project at a conference in front of a packed audience.

9.2.3 Location

The next step is to decide on the co-ordinates of a central location for the work and the extents of the area to be investigated. Where the project is centred on a known site this is relatively straightforward as the position of the site will be known and an estimate of the size can be made. For a project that involves searching for a site a planned search area has to be defined.

At this point it is important to decide on the co-ordinates to be used for mapping. Positions can be recorded using your own co-ordinates local to that site or in real world co-ordinates taken from GPS or similar. In the short term it may seem easier to set up your own local grid co-ordinate system but this becomes a problem if you want to integrate charts, aerial photographs or geophysical survey data.

The next step is to decide which real world co-ordinate system to use, unless this has already been defined for you by the client. In general it is better to use a commonly used co-ordinate system than something specific defined for the local area. In general, mapping should be done using the WGS84 datum as it is used by GPS receivers and the Universal Transverse Mercator (UTM) projection as it is a common standard.

If the site to be recorded is a shipwreck then a this stage it should be possible to add a Wreck object at the best estimate of position for the site, information about the wreck can then be added to the Wreck object. For other site types a simple Point object can be used to mark its position.

9.2.4 Site Code

Although not essential at this stage, it is useful to define a site code for the site, this is a unique identifier that is used to identify the site you are working on. The name should be unique as it is used as part of the name of every object on the site so you can differentiate artefact ‘04A0423’ from one site from one of the same name on another. See section 5.5.8.

9.2.5 Sharing and Archiving

As soon as the Site file has been created for a project then you need to consider how and where the files will be archived and the ownership of the information.

Archiving may not seem important in the early stages of a project but as the information contained within the Site file grows it becomes essential. Just consider how long it would take to re-create the information from scratch if it were lost. Ensure that the archive is backed up on CD or DVD and ensure that copies are made after each significant addition to the archive.

Ownership of the information needs to be decided. Decide who is to get a copy of the archive and ensure that they get an updated copy after each significant addition to the archive. Ensure that the metadata is set correctly before passing on any copies as this will identify the version of the archive as well as ownership.

9.3 Pre-Fieldwork Planning

9.3.1 Introduction

With the Site file created we can start to add any information we have about the site and plan for the first season of fieldwork. At this stage it is often useful to collect together all the information that is known about the site to help with the pre-fieldwork planning. As well as confirming what is known about the site this should also highlight what is not yet known particularly if it affects how the fieldwork is to proceed.

9.3.2 Importing Existing Information

If a position for the site is known then charts and maps of the area can be added to the Site file. Charts and maps help to put the site into context and may give clues about how the site formed and where other parts of the same site may lie. Previous results from geophysical surveys can be included; this may establish a more accurate position for the site or may help define the site extents.

- Typically you may want to import coastline information in vector form, when imported this appears as a collection of drawing objects on one or more Layers.
- Charts can be scanned and added to the site plan as Image Basemap objects
- Targets or anomalies detected during geophysical surveys can be imported as Target objects or added by hand.
- Sidescan sonar mosaics and multibeam (MBES) images can be added as Image Basemaps.
- Previous site plans or sketches can be scanned and added to the plan. In some cases the plans can be digitised and used as the starting point for the main site plan.
- Documents relating to the project and site can be added as Sources

- Located or recorded shipwrecks can be added as Wreck objects
- A timeline of important events relating to the site can be created using Event objects

In general, each dataset or image should be added to its own Layer with image basemap Layers added to the last or lowest Project so they always appear under other objects on the chart.

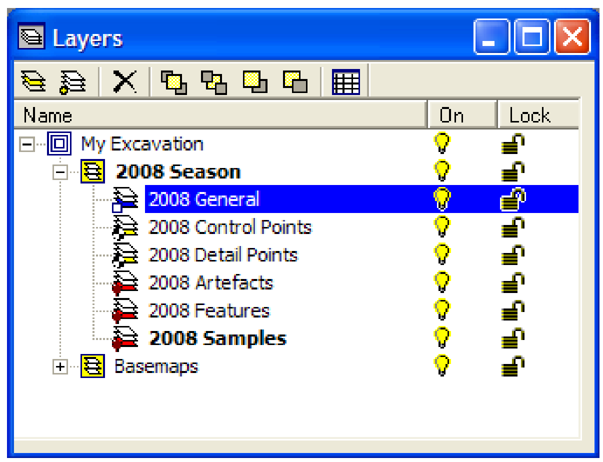
It is important to obtain permission to use any images or chart data before they are added to the dataset. Additional permissions may also need to be sought when publishing or archiving the dataset.

9.3.3 Setting Up the Layers and Projects

For any project it is worth setting up a sensible set of Layers and Projects right from the start. The reason for this is that the lifetime of the project is often not known at the outset and small projects can turn into larger ones. If the Layers and Projects are right from the start then they will not have to be changed if the project grows.

In general, there should be a separate project for each season's work. The name of the Project should include the year and a description such as '2008 Season'. Within each Project, Layers can be added for each object type as required. For example, the Layers needed on a typical project in 2008 include:

2008 General *(for drawings)*
2008 Control Points *(for survey control)*
2008 Detail Points *(for survey detail points)*
2008 Artefacts
2008 Features
2008 Samples



A separate Layer should also be used for each set of geophysical targets or raw dataset.

In practice a Site file will contain many Layers within each season's project, each containing a separate type or category of object.

9.3.4 Survey Methodology

It is necessary to agree the survey methodology before any fieldwork starts so the correct equipment and tools can be sourced and the team trained in the methods to be used.

For most underwater sites it is the humble tape measure that is the tool of choice. For simple assessment surveys it may be desirable to use the more primitive survey methods over the whole site such as offset, ties and radial measurements. For surveys in advance of excavation, important sites that may be at risk or where time allows then a full 3D trilateration survey should be completed as the quality of the results is so much better. For deep water projects or those that need to be efficient the survey work can be completed using a high-accuracy Acoustic Positioning System (APS) such as the Sonardyne Fusion LBL APS.

Figure 13: Diver with surface supply equipment and Sonardyne Fusion APS

The task of survey processing can often become the responsibility of the project's data manager, not because of any inherent skill in surveying but simply because it's a job to be done on the computer. Fortunately, SR4 contains the tools needed to process all type of survey measurements underwater so making this task relatively straightforward.

9.3.5 Object Names

The object naming policy to be use on the project should be agreed in advance, documented and anyone who will be adding data to the system should be trained in its use.

9.3.6 Wordlists

Wordlists and thesauri are used to help maintain consistency within the recording system so need to be set up in advance of fieldwork. Users adding information can be limited in their choice of words to use, for example this helps make sure that all iron objects are described as 'metal : iron' rather than 'iron metal', 'iron' or any other variant.

The standard wordlists used in SR4 can be adapted to suit your particular project. The default wordlists are ideal for 17th century shipwrecks so would need adapting for a project on a classical wreck.

9.3.7 Adding Contacts

At this stage an interim list of people involved in the project can be added as a list of Contacts. This will be necessary when fieldwork starts as the Contacts will be needed for Dive Logs, but it is also useful to have team members' contact details readily available and SR4 is a good place to keep them.

9.4 Fieldwork

9.4.1 Introduction

By the time that the fieldwork starts the Site file for the project should contain all of the information known about the site at that time. In the field the recording system is used as the repository for all of the information gathered during each dive but it is also used for day to day planning and decision support.

Technical Note: Recording Underwater

Although digital recording systems would seem to do away with the need for recording on paper, the practical aspects of the work mean that paper, or plastic film, is still needed. Measurements and observations made underwater should be recorded on plastic film or mylar sheet taped to an A4 or A3 slate. Standard forms should be used where possible to help maintain consistency of the data. On completing the dive the forms should be removed from the slate, gently rinsed in fresh water and left to dry. Once dry the text and information should be transferred into SR4, if the form includes a sketch then this should be scanned, imported into SR4 and linked to the dive log. Once the information has been retrieved from the form it should be filed.

The forms that went underwater are very important as they are part of the primary archive. Forms should not be cleaned and reused and the plastic slates themselves should not be used for recording.

9.4.2 Adding Artefacts, Features, Sectors and Samples

The bulk of the work may be in the addition of Artefacts for finds recording. Each Artefact is added to the site plan in a position derived from the survey work happening at the same time. The details of each Artefact are added either directly or from a finds recording form, the form is used as a way of separating the 'clean' computer from the wet and dirty finds.

In-situ or registration photographs can be imported into SR4 then digitised to scale so a 2D drawing of each find can be added to the site plan. Other photographs and video clips can also be linked to each Artefact.

With more than one person adding information it is essential that the data is verified for consistency. The use of wordlists limits the scope for mistakes with some objects' properties but free form text properties can be a source of error. Training and the application of working standards is the most efficient guard against mistakes but it may be necessary for any new object records to be independently verified by a suitable person. It is far better to find and correct errors and omissions in the field than it is to attempt to do it once the fieldwork is over.

9.4.3 Adding Dive Logs

Dive logs should be added on the same day that the dive was completed and any backlog should be avoided. The logs will need to be linked to the Artefacts found or recovered during that dive or for any measurements recorded. Sketches or drawings associated with the dive should be scanned and linked to the Dive Log

9.4.4 Survey Processing

A backlog in survey processing should also be avoided as the positioning of Artefacts will be dependant on the timely processing of detail survey point positions. Artefacts can be temporarily positioned on the site plan in estimated positions but they should be positioned correctly as soon as possible.

9.4.5 Data Security

With large quantities of information being added to the recording system as fieldwork progresses, the importance of backing up cannot be stressed enough. Version control is also of paramount importance, especially if the main site file is being copied and shared regularly.

It is essential that the archive is kept free from viruses and other malware. All copies of the archive should be separately checked for viruses when they are created.

- Copies of the entire Site archive should be copied to an external hard drive or CD at the end of each day
- Ensure all copies are clearly marked with their contents, version and date of creation using permanent ink
- Where possible, a separate external hard drive should be used for the online backup (see note) so if the main computer hard drive fails the online backup should survive.
- The data recovery procedure (to be used after data loss) should be tested periodically to ensure that it works in practice
- File Archiving should be used for all important projects (see note)
- The Site file version should be updated each time the file is backed up
- All computers being used to add to the archive must have up-to-date antivirus software installed

Technical Note: Online Backup and File Archiving

SR4 automatically saves two copies of the main site file even if file archiving is disabled. The two files are saved in the **Backup File Folder**, this can be set in the **Site Properties** in the **Files & Versions** page. Where possible, a different hard drive should used for the backup files than is used for the main and archive files.

When a Site file is saved with File Archiving enabled, the older Site file is moved to an archive folder before the new file is saved. This means that all of the previous saved files are kept so in the event of the loss of the main Site file the previous files can still be used. The archive folder should be cleaned up regularly to remove the oldest files.

9.4.6 Large Project Operation

When working on large projects it may become necessary for more than one person to add information to the recording system at the same time, fortunately this can be done by using the merge tool available within SR4. Another user can be adding information to a new Layer in another site file while the main Site file is being updated and when complete the new Layer can be merged into the main Site file.

9.5 End of Fieldwork

9.5.1 Introduction

At the end of a season or the end of the project the information in the archive should be verified, published and archived.

Most of the problems should have been corrected in the field but it is usual for there to be a few mistakes and omissions remaining in the dataset so these should be corrected before it is archived. All metadata should be checked and updated.

Copies of the dataset should be given to the recipients agreed at the start of the project. If required, the primary dataset should be made public on CD or via the Web using the Package Tool in SR4 to create a self-installing electronic publication.

A copy of the archive should be placed in long term secure storage and a separate copy kept by the project team.

9.6 Migrating an Existing Archive to SR4

New life can be given to an existing archive that was recorded on paper or in an electronic form that is difficult to publish. The archive can be moved or migrated to SR4 so it can be published more widely. The process of migration is similar to that described above for a fieldwork project; the difference is that the site data exists already.

- Paper site plans can be scanned, imported, georeferenced and digitised to form the basis of the new electronic site plan.

- Where information about survey points and measurements exists they can be added by hand or imported as a CSV file created by a spreadsheet program such as MS Excel. The measurements may need to be reprocessed before the survey points can be used.

- Finds lists can be scanned and converted to electronic documents using OCR software. The new documents can be imported to a spreadsheet program, formatted then exported to a CSV file that can be imported into SR4.

- Finds photographs can be scanned and added to SR4 in bulk then linked to the appropriate Artefact object.

10 Summary

The earlier sections have highlighted the importance of archaeological archives as the means for long-term preservation of archaeological information. We have seen the advantages of digital documentary archives and have discussed the problems associated with them. The significance of the quality of recording fieldwork has been highlighted as has the need for widespread reuse of the captured information. We can now approach the next archaeological recording project with this in mind in the hope that the resulting archive will be fit for purpose.

11 References

Andresen J. & 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., 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* [online], Center for the Study of Architecture, http://archcomp.csanet.org/ [Accessed May 2008]

Greene, K. 2002, *Archaeology: an Introduction* (4th edn.) Routledge: London.

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 [online], CAA2007 Southampton

Holt, P., (2007b) The Site Recorder Database Schema [online], Available from: http://www.3hConsulting.com/Research/research_schema.htm [Accessed May 2008]

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

Madsen T., 1998, Design considerations for an excavation recording system, Design and use of field information systems, CAA-NL98

Richards J., Online Archives, Internet Archaeology 15 [online], http://intarch.ac.uk/journal/issue15/richards_index.html, [Accessed May 2008]

12 Resources

12.1 Related Books and Papers

Andresen, J., & Madsen, T., 1996, IDEA Integrated Database for Excavation Analysis
English Heritage, 2002, *Centre for Archaeology Guidelines – Environmental Archaeology*, English Heritage Publications
Gawronski, J., 1987, *Amsterdam Project Annual Report*, ISBN 90 7169002 4
Gawronski, J., 1992, *Hollandia Compendium*, Elsevier, ISBN 0 444 89415 2
Hildred, A., 2003, *Mary Rose Excavation Report 2003*, Mary Rose Trust
Martin, C., 2003, Duart Point Shipwreck Excavation 2003
Mathers, W., 1993, Conception site report
Nautical Archaeology Society, 1994, Finds Handling Workshop
Nickerson, S., 1994, A Site Information System (SIS): CADD/Database Integration for Field Use, APT Bulletin 24(1):56-62
Quinn, R., (in press 2001), The Assimilation of Marine Geophysical Data into the Maritime Sites and Monuments Record, Northern Ireland, *Historical Archaeology*
Richards J., Online Archives, (http://intarch.ac.uk/journal/issue15/richards_index.html)

12.2 General Archaeology

Barker P., *Techniques of Archaeological Excavation*, Batsford Ltd., ISBN 0713427396
Green J., 2004, *Maritime Archaeology: A Technical Handbook* 2nd Ed., Elsevier Academic Press, ISBN 0122986326
Hester T., Shafer H & Feder K, 1997, *Field Methods in Archaeology*, Mayfield Publishing Company, ISBN 1 55934 799 6
Kipfer B., 2007, *The Archaeologists Fieldwork Companion*, Blackwell Publishing
Milne G., McKewan C., Goodburn D., 1998, *Nautical Archaeology on the Foreshore*, RCHME, ISBN 1 873592329
Museum of London, 1994, *Archaeological Site Manual*, Museum of London, ISBN 0904818403
Roskams S., 2002, *Excavation*, University of Cambridge, ISBN 0521355346
Watkinson D. & Neal V., 2001, *First Aid for Finds*, British Archaeological Trust

12.2.1 MoRPHE Project Managers' Guide

The Management of Research Projects in the Historic Environment (MoRPHE) has been produced following a review of an earlier guidance document, Management of Archaeological Projects (English Heritage 1991), widely known as 'MAP2'.
(http://www.english-heritage.org.uk/upload/pdf/MoRPHE-Project-Managers-Guide.pdf)

12.2.2 MoRPHE Project Planning Note 3 Archaeological Excavations

Management of Research Projects in the Historic Environment – Project Planning Note 3 Archaeological Excavations
(http://www.english-heritage.org.uk/upload/pdf/MoRPHE_Project_Planning_Note__3_Archaeological_Excavation_v1.0.pdf)

12.3 Recording Standards and Guides

Brown D., Archaeological Archives – A guide to best practice in creation, compilation, transfer and curation, 2007, Archaeological Archives Forum, ISBN 0948393912

Archaeology Data Service, Geophysical Data in Archaeology: A Guide to Good Practice (http://ads.ahds.ac.uk/project/goodguides/geophys)

Archaeology Data Service, 1990, GIS Guide to Good Practice, Oxbow Books, ISBN 1900188694

Archaeology Data Service, 2008, Guidelines for Depositors
(http://ads.ahds.ac.uk/project/userinfo/deposit.cfm)

12.3.1 IFA Standard and Guidance for Nautical Archaeological Recording and Reconstruction

(http://www.archaeologists.net/modules/icontent/inPages/docs/codes/NARR2007interim.pdf)

12.3.2 CIDOC CRM

The CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives. (http://cidoc.ics.forth.gr/index.html)

12.3.3 MIDAS Heritage

MIDAS Heritage is the UK data standard for information about the historic environment. It states what information should be recorded to support effective sharing of the knowledge of the historic environment, and the long-term preservation of those records. (http://www.english-heritage.org.uk/server/show/nav.18041)

12.3.4 Object ID

Object ID is an international standard for describing cultural objects. It has been developed through the collaboration of the museum community, police and customs agencies, the art trade, insurance industry, and valuers of art and antiques. (http://www.object-id.com)

12.3.5 OpenGIS

The Open Geospatial Consortium, Inc (OGC) is an international industry consortium of companies, government agencies and universities participating in a consensus process to develop publicly available interface specifications. OpenGIS® Specifications support interoperable solutions that "geo-enable" the Web, wireless and location-based services, and mainstream IT. (http://www.opengeospatial.org/ogc)

12.3.6 Archaeology Data Service standards in Archaeology

A list of standards and resources for use in archaeology
(http://ads.ahds.ac.uk/project/userinfo/standards.html)

British Colombia Shipwreck Recording Guide
(http://www.tsa.gov.bc.ca/archaeology/docs/shipwreck_recording_guide/index.htm)

12.4 Wordlists and Thesauri

12.4.1 INSCRIPTION

(http://www.fish-forum.info/inscript.htm)

12.4.2 National Monuments Record Thesauri

(http://thesaurus.english-heritage.org.uk/frequentuser.htm)

12.4.3 British Museum Materials Thesaurus

(http://www.mda.org.uk/bmmat/matintro.htm)

12.4.4 MoRPHE Project Planning Note 2 Developing Controlled Terminology

Management of Research Projects in the Historic Environment - Project Planning Note 2 Developing Controlled Terminology.
(http://www.english-heritage.org.uk/upload/pdf/MoRPHE_Project_Planning_Note_2_Developing_Controlled_Terminology.pdf)

12.5 Metadata and Data Exchange

12.5.1 Dublin Core

Dublin Core Metadata Initiative (DCMI) http://dublincore.org)

12.5.2 The Open Archives Forum

The Open Archives Forum supported projects and national initiatives with an interest in using an open archive approach to interoperability. It provided a focus for European initiatives implementing the recently released Open Archives Initiative (OAI) metadata harvesting protocol. (http://www.oaforum.org/index.php)

12.5.3 Resource Description Framework

The Resource Description Framework (RDF) integrates a variety of applications from library catalogues and world-wide directories to syndication and aggregation of news, software, and content to personal collections of music, photos, and events using XML as an interchange syntax. The RDF specifications provide a lightweight ontology system to support the exchange of knowledge on the Web. (http://www.w3.org/RDF)

12.5.4 GEMINI

The UK GEMINI Discovery Metadata Standard is a defined element set for describing geo-spatial, discovery level metadata within the United Kingdom.
(http://www.govtalk.gov.uk/schemasstandards/metadata_document.asp?docnum=903)

12.6 Archiving

12.6.1 Digital Archives from Excavation and Fieldwork

Archaeology Data Service, 2000, Digital Archives from Excavation and Fieldwork: Guide to Good Practice, Oxbow Books, ISBN 1 900188 73 2
(http://ads.ahds.ac.uk/project/goodguides/excavation)

12.6.2 Archaeological Archives Forum

http://www.britarch.ac.uk/archives/

12.6.3 MoRPHE Technical Guide 1 Digital Archiving and Dissemination

Management of Research Projects in the Historic Environment - Technical Guide 1 Digital Archiving and Dissemination.
(http://www.english-heritage.org.uk/upload/pdf/MoRPHE_Technical_Guide_1_Digital_Archiving_and_Dissemination.pdf)

12.6.4 The Digital Preservation Coalition

The Digital Preservation Coalition (DPC) was established to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK. (http://www.dpconline.org/graphics/index.html)

12.6.5 Archaeoinformatics.org

Archaeoinformatics.org is established as a collaborative organization to design, seek funding for, and direct a set of cyberinfrastructure initiatives for archaeology

12.6.6 Access to Archives

A2A allows you to search and browse for information about archives in England and Wales, dating from the eighth century to the present day. (http://www.a2a.org.uk)

12.7 Electronic Publication

12.7.1 Internet Archaeology

Internet Archaeology is the first fully refereed e-journal for archaeology and publishes articles of a high academic standing which utilise the potential of electronic publication. (http://intarch.ac.uk)

12.7.2 Nabonidus

Nabonidus is a web application designed for Archaeological Excavation data storage, sharing, manipulation and analysis. It aims to revolutionize the way we as Archaeologists collect, analyze and interpret excavation data. (http://www.nabonidus.org)

12.7.3 The VENUS Project

The VENUS project aims at providing scientific methodologies and technological tools for the virtual exploration of deep underwater archaeology sites. (http://piccard.esil.univmed.fr/venus/)

12.7.4 The OASIS Project

The overall aim of the OASIS project is to provide an online index to the mass of archaeological grey literature that has been produced as a result of the advent of large-scale developer funded fieldwork. (http://ads.ahds.ac.uk/project/oasis/)

12.7.5 OCHRE

The Online Cultural Heritage Research Environment (OCHRE) is an Internet database system for research on cultural heritage in all its forms. It is intended for researchers and students who are engaged in archaeological, anthropological, linguistic, and textual investigations of many different kinds. (http://ochre.lib.uchicago.edu/)

12.7.6 Open Context

Open Context is a free, open access resource for the electronic publication of primary field research from archaeology and related disciplines. Open Context provides an integrated framework for users to search, explore, analyze, compare and tag items from diverse field projects and collections. (http://www.opencontext.org/)

12.8 Map Sources

12.8.1 NOAA

http://rimmer.ngdc.noaa.gov/mgg/coast/getcoast.html

12.8.2 GenMaps

http://freepages.genealogy.rootsweb.com/~genmaps/index.html

12.8.3 SeaZone

http://www.seazone.com/index.php

12.8.4 USGS Maps

http://nationalmap.gov/gio/viewonline.html

13 Appendix 1: Glossary

Archaeological Archive: All parts of the archaeological record, including the finds and digital records, as well as the written, drawn and photographic documentation (Brown2007: 3).

Archaeological Project: Any programme of work that involves the collection of information about an archaeological site, assemblage or object (Brown 2007: 3)

Artefact: an object made or modified for use by humans (Greene 2002: 280)

Association: An association is any defined relationship between artifacts or features. An association must be defined. An association is not something that is found.

Context: a neutral term for any deposit or structure recorded during an excavation; sometimes described as a 'unit of stratifiction' (Greeme 2002: 280)

Database: A database is a structured collection of records or data, most often in a computer.

Database Schema: See Schema

Documentary Archive: Documentary records associated with the project, part of the Archaeological Archive

Geographic Information System (GIS): Any system for capturing, storing, analyzing and managing data and associated attributes which are spatially referenced to Earth.

Information Management System (IMS): A system capable of capturing, storing, managing, reporting and publishing information

Material Archive: All of the recovered finds and samples, part of the Archaeological Archive

Object: Something that can be recorded using an object-oriented database.

Primary Record: Original, unprocessed recorded information

Record: A collection of items of information relating to one thing, entity or Object.

Relational Database (RDBMS): A relational database is a database that conforms to the relational model, and refers to a database's data and schema (the database's structure of how those data are arranged).

Resource: Something that can be used, in this case more sources of information relating to this document

Secondary Record: Processed information derived from primary records

Schema: How the data is arranged and recorded in a Database.

Stratification: (Stratrigraphy): by analogy with geological strata, deposits on archaeological sites may be arranged in a sequence (with the oldest at the bottom) that may be dated with the help of any diagnostic artefacts they contain (Greene 2002: 286)

Thesaurus: A hierarchical Wordlist that includes word equivalences and preferred terms.

Wordlist: A list of terms that can be used for a Field or Property, each with a precise and unique meaning.

Index

