CHAPTER 13
THE INTEGRATION OF DATA SOURCES

Diana Murray
Royal Commission for the Ancient and Historical Monuments of Scotland

There is no possible archaeological heritage policy without the systematic control and inventorying of the resource

Valetta

1. INTRODUCTION

1.a. Data integration Within ARM

Archaeological Resource Management today can draw on more data than has ever been previously available for heritage management policies and research agendas. Information technology is well-suited to the requirement of systematic control and inventoryisation of the resource and has been widely adopted by heritage professionals to hold and analyse archaeological data. This data is not only being collected faster and more comprehensively, but a significant proportion is being acquired in digital form directly from fieldwork or research.

Archaeological resource managers are being challenged to:

• create modern inventories not only of sites, but also of a comprehensive set of relevant data relating to them, not all of which may be archaeological
• assemble the information in a way that allows it to be accessible and informative for the construction of heritage policy and research agendas, and
• to increase public awareness of the cultural heritage including the provision of direct on-line access to the inventory.

The large amount of data that is being gathered, and the pace at which it is being assembled is overwhelming if it is not organised in a rational and accessible way. It is also wasteful of resources if work is duplicated or undertaken in ignorance of previous relevant research or field data.

The most effective way of managing both the quantity of data and the scope of information, the accessibility and the wide spectrum of users and an integrated heritage policy is to operate an integrated data system.
1.b. Data integration within the National Monuments Record of Scotland (NHMRS)

In order to describe how an integrated data system for inventorisation can operate, this paper draws mainly on the example of the National Monuments Record of Scotland where such a system is in place (Figure 13.1).

The NHMRS grew from a number of different origins including the Inventories published by the Royal Commission on the Ancient and Historical Monument of Scotland (RCAHMS) from 1908 to 1994. Since 1994, the database of the NHMRS, which is part of the RCAHMS, has been regarded as the Inventory, while publications are directed to more detailed analysis of the archaeology and historic buildings of Scotland. The records and responsibilities of the Archaeology Division of the Ordnance Survey (OS), the UK mapping agency, were transferred in the NHMRS in 1983 and have heavily influenced both the structure of the inventory and the current methodology of the field survey undertaken by RCAHMS. The NHMRS also maintains collections of primary archive material resulting both from RCAHMS field and aerial survey work and from items deposited in the national record. This includes not only archaeological material, but also a large and internationally important collection of architectural drawings, historical photography and aerial photography, both vertical cover and targeted obliques. (Further information about the work of the RCAHMS and the National Monuments Record of Scotland can be found at www.cultural-geography.co.uk which also includes access to Canmore, an interface to the NHMRS database, designed to allow the public to browse NHMRS data.)

The idea of integration is a strategy that has to be established and developed to cover all data. In the NHMRS, the decision was taken early on to ensure that the indexes and catalogues to all the material held in the archives, however diverse, was integrated so that users could cross-search or easily navigate between different elements of the data (Harley, 1992). Today, compatibility of systems is no longer a real hurdle and compatibility of data and standards for data integration is the real debate.

Data integration can only function as a model that has the facility for the user to assemble data in useful and meaningful ways from a number of different entry points. Therefore there is no clear starting point from which to describe integrated systems in a linear fashion. There are, however, two main ways of integrating archaeological data. One is by using unique identifiers that describe data elements or entities common to two data sets; the other is by using geo-spatial referencing. Although either separately or together, these integrating mechanisms ensure that data held in disparate forms is linked in a way that provides cross-data functionality.

2. GEO-SPATIAL REFERENCING

2.a. Maps and GIS as integrating mechanisms for data

One of the most frequently used means of integrating heritage data is to use the concept of the map. Maps are cartographic expressions of geo-spatial reference and, have been used for many years to integrate data by physically plotting information from a number of sources onto paper maps, either as indicative points, boundaries of extent or drawn out depictions.

2.b. Basic elements for geo-spatial integration of data

2.b.1. Metadata

Technological presentation can, however, give spurious authenticity to the data. Metadata such as information about accuracy, method of capture, date of capture and reason for collection should also be available for any data set to ensure that the user can know both how comprehensive and how accurate the source data is, and provide vital information about the reliability of the use of the data for secondary analysis and management purposes. Transparency concerning the level of reliability of the data is important because, as information becomes more widely available and transferable, it will be used for purposes for which it was never collected in the first place.
farms, townships and their associated agricultural remains were poorly represented in the national database and, as a result of the project, it is now possible to offer guidelines for any policy of study, survey or preservation (Stevenson, 1996, RCAHMS et alii, 2002).

2.3. Historic Landuse Measurement

Landscape characterisation is now accepted as a method of analysing the landscape made possible by the use of GIS technology. The Historic Landuse Assessment Project at RCAHMS combines data from a number of different sources to identify the origins of the component parts of the historic landscape and the elements of earlier historic land- scapes surviving within it (Figure 13.3). Using aerial photographs, land cover data compiled by the Macaulay Land Use Research Institute, Ordnance Survey maps (both current and nineteenth century editions), and previously recorded areas of archaeological landscape, a mosaic is produced in which each piece (at least one hectare in extent) has a Current Landscape type and may also contain one or more Relict Landscape Types. The end product is a set of inter-related GIS layers with related attribute tables. Interrogations can be made of how much of the historic landscape of any one period is still extant or to provide a broad overview of the historic landscape and an analysis of the forces of change that have acted upon it. Such analysis is extremely valuable in demonstrating the dynamic nature of the landscape and the pressures on surviving historic elements (Dyson-Bruce et alii, 1999, RCAHMS et alii, 2000, 2001).

This data is geo-spatially referenced in the GIS and therefore it can be used in combination with other more detailed data relating to the historic environment, such as that resulting from the FESP project or with aerial photographic data.

2.d. Data sources

2.d.1. Aerial photographs

Aerial photographs are a fundamental source of archaeological information, particularly cropmark photography which reveals information on sites that are otherwise unknown. The photographs can be listed and described in a database and this has been the method used by most inventories. Some unrelated databases have been devised to analyse the content and form of cropmarks in great detail in England (Edis et alii, 1989; Slooertz, 1997). Effective use of these methods can be, it is important to be able to see the photograph to both verify and understand the information that has been extracted. Using digital techniques, transcriptions of aerial photographs can be made (Brown, 1999) and the image can also be rectified to true position (Figures 13.4, 13.5 and 13.6). RCAHMS undertakes an annual programme of aerial survey capturing between 2,000 and 5,000 air photographs each year. The results are selected for transcription at an accuracy of 1:2,500, and about 10% of all known cropmark sites have now been rectified and loaded into the GIS. No analysis of the form of the cropmarks has been undertaken systematically in Scotland, but could be carried out using GIS analytical tools.

NRHS contains a collection of 1.5 million vertical air photographs dating from 1940s to 1990s. These are not systematically digitised at present, but are referenced by sortie plots tracing the path of the aircraft as photographs are taken. The sortie plots have
been digitized from hand-plotted traces and are linked to a separate database describing information about the flight (date, time of flight, etc.). These are used for the management of the collection, but can also be used in the GIS with GPS captured data from RACAHMS survey flights. Air photography can be discovered in three ways using the GIS through a flight sortie index, through rectifications of the images, or through transcriptions of the ornamental interpreted by an archaeologist, all of which can be set against map background or combined with other data of choice.

2.2.2. Field Survey Data

Non-intrusive survey of standing monuments and buildings has been the work of the RACAHMS since 1908 (Sanderson, 2000; Halliday, 2001; for full list of publications see www.racahms.org.uk). Surveys can be very detailed drawings of carved stones at large scale such as 1:20, accurate plans and elevations of buildings with interpretations of the building phases, plans table surveys producing hachured plans of archaeological features at 1:2000 or 1:500 and archaeological or industrial landscapes at mapping scales of 1:1,250 (1:500). These latter surveys are created at the same scale as the GIS maps and form the reference points. Surveys carried out at mapping scales are migrated directly into the GIS (Figures 13.7 and 13.8). All survey data can now be accessed digitally in the field and processed in house using digital mapping and reproduction software.

Larger scale surveys of archaeological features are scanned and related geospatially to position, using mapping detail captured at the time of survey, and even the most detailed drawing can be incorporated into the GIS as an image related to a point.

From a data management point of view, the GIS becomes the active catalogue to the data. Working interactively with the database, the detailed surveys, whether aerial photographic transcription or survey data, no longer have to be indexed as entries in the archive catalogue because they exist within the structure of the integrated system.

Similar digital survey techniques are now common to archaeological work and are used on excavations, geophysical surveys and contour surveys among others. All organizations that create archaeological archives and maintain inventories are now facing the problem of how to look after and access this data. Archiving the conversion of digital data is an issue in itself, but there is also the question of how such information can be deposited with and integrated into inventories. A trial carried out by RACAHMS has migrated some digital data into GIS layers. A contour survey of the carvings at Cleaven Dyke in Perthshire (Figure 13.9), was done by C. Burgess and G. Barclay, recorded in penmap and processed in DFX (Barclay et al., 1998). The information was deposited with NMRs, migrated into the NMRs integrated into the NMRs integrated into the NMRs integrated into the NMRs integrated into the NMRs. However, it is also the case that this catalogue to digital form includes plotting the location of the buildings on screen in the GIS using the OS seed point for each building and feeding the information back to the database. This produces a much more accurate location in relation to the map on the screen than can ever be achieved by feeding grid references into the database and then basic them to the GIS. It would not be cost effective to digitize every building outline, especially as the outlines are already present on digital mapping, which can often be used in conjunction with the data. Once the information is recorded in the inventory for each building may be minimal (eg, name, location) or can include extensive information that is structured in a number of tables, all of which are linked by unique identifiers.

For example, spatial referencing and display of data has brought immediate insights into the cataloguing of architectural drawings in another RACAHMS project. The Scottish Architects Papers Preservation Project was conceived following an assessment of all architects' practices in Scotland (Bailey, 2000). Funded by the HES, the project is cataloguing all 503 drawing sets. The project was conceived as a collection based project because the collections catalogue is part of the NMRS integrated system, it was agreed that...
it was essential to include a component in the project to record the geographical location of the architectural works that were the subject of the architect's drawings. Architectural designs for buildings that were never built can also be included in the catalogue, even though there may be no location because they can be catalogued in the collections table without being related to the site location table. Each architectural collection is also related to a collection essay giving an overview of the architect, their practice and the collection of material and is also associated with a table that provides a short biography of individuals such as artists and craftsmen associated with the work (Figure 13.11).

This provides a flexible approach to the data both for data capture as each table of information can be built together or independently, and for the user by locating each building represented in the collection, while the GIS related to the database can be used to visualise the work of an architectural practice.

Because this distribution is created from detailed data held in the GIS, the user can zoom in to discover the exact location of each individual building represented, and from there request all the information recorded in the system for that building from the architectural practice drawing to a relevant aerial photograph.

3.b. Relationships between data

This entity relationship, based on unique identifiers, provides the links that operate within the data tables held in the NMRS database. The information tables are a) directly related by holding fields in common e.g. the building/site table and the text table hold the unique NUMLINK in common which is a numerical unique identifier or b) indirectly related by means of a link table e.g. between the building/site table and the collections table, that concords two separate identifiers (Figure 13.12). This has the advantage that not every entry must be linked (e.g. building designs that were never built and consequently have no location, and it can be used as a means of linking external databases, e.g. data tables relating to scheduled monuments and listed buildings which are held and maintained by Historic Scotland and linked by means of a landline link. Similar relationships can now operate over the Internet.

3.c. Alphanumeric and image data

The information contained within the database includes:

i) text describing the sites

ii) collections—which provides a catalogue in summary and in detail of all the drawings and photographs and manuscript material held both in the NMRS collections and sometimes references to material held in other collections. Linked to this are associated tables providing information on accession, copyright, conservation treatment for each item, associated people e.g. archaeologists or architects and collections essays relating to a group of items. A programme of digitising images from the collection is underway and each scanned image is separately catalogued in this section of the database. Once this is in place the integrated system will allow the images to be selected through any of the enquiry points e.g. by people, copyright, collection, date, location or by area using the GIS as a starting point

iii) bibliography—which provides a catalogue in summary and in detail of all bibliographical references related to the entries in the descriptions held in the database and a catalogue of books held in the NMRS library. These can be similarly used to retrieve information across the whole spectrum of the integrated database if required

iv) maritime data—which extends the database to sites offshore, both wrecks and underwater sites that are located and those that are recorded as ships lost at sea. The structure of this database has to be different to include data that is not recorded for land-based sites, such as latitude and longitude, but by including a unique identifier can be integrated

v) a table recording events is planned for the NMRS. The structure of events, monuments and archive is one that is now well established in local inventories in the UK and provides a means of recording a chronological dimension to the data in addition to spatial and a history of the events that have taken place such as excavation, survey, change of use or date of construction and demolition. There will also be a link built to reference finds and their location in museums and an index relating to specialist reports and data sources is planned.

3.d. Site entities

The NMRS uses data entities, including unitary monuments that are a legacy of the former database system, to enable the attachment of data for data-management purposes, but the system has to be flexible enough to put the whole landscape back together as well as being able to dissect and analyse it. Therefore site and data entries are not rigidly defined, but adapted to suit each particular data set and to aid retrieval. Data entities can therefore relate to an item of archive e.g. a photograph, or to a unitary monument, or to an individual element in the monument e.g. a hut-circle within a prehistoric settlement. These can easily be reassembled through the database using the relationship of the unique identifiers and through the GIS by using spatial relationships. The definition of area of the archaeological site or building is therefore determined by ease of retrieval through the system rather than by any archaeological definition of entity.

3.e. External integration

Entities such as legal boundaries for monument protection and listed buildings only define the area that is protected, and these can be integrated using point references or by drawn areas. The use of the integrated system means that these legal boundaries can be viewed combined with other data so that it becomes more obvious that they only represent a small portion of the total historic landscape. The links with legal areas in Scotland were built initially using related unique identifiers in the database and these are still in operation although areas of scheduled monuments are now incorporated as a layer in the GIS. The link operates over a landline to Historic Scotland through a concordance table:
TABLE 13.1. NMRS Links table.

<table>
<thead>
<tr>
<th>Site name</th>
<th>NUM/LINK</th>
<th>Scheduled Monument no.</th>
<th>Listed Building no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elgin Cathedral</td>
<td>244</td>
<td>90142</td>
<td>5073 173</td>
</tr>
<tr>
<td>Elgin Cathedral Bishop's Palace</td>
<td>246</td>
<td>90141</td>
<td>5073 173</td>
</tr>
<tr>
<td>Elgin Cathedral Chapter House</td>
<td>247</td>
<td>90142</td>
<td>5073 173</td>
</tr>
<tr>
<td>Elgin Cathedral Meikle Cross</td>
<td>252</td>
<td>90142</td>
<td>5073 173</td>
</tr>
<tr>
<td>Elgin Cathedral Pictish Symbol Stone</td>
<td>253</td>
<td>90142</td>
<td>5073 173</td>
</tr>
<tr>
<td>Dirleton Castle</td>
<td>259</td>
<td>90096</td>
<td></td>
</tr>
<tr>
<td>Skara Brae</td>
<td>207</td>
<td>90276</td>
<td></td>
</tr>
</tbody>
</table>

These are one to many relationships so that a single entity recorded by Historic Scotland as a Scheduled Monument may be represented in the NMRS record as a number of separate entities. The relationship works effectively when queried from either end.

Linking heritage data to non-specialist data sets is now feasible using the Internet and can be achieved based on the principle of unique identifiers. Land and property registers are now being digitised and heritage data will be one of the sets of spatial data that will be available through services such as Scotlis (www.scotlis.com/) and NLis (www.nlis.org.uk).

3.6. Ordnance Survey Pilot Project

There is also the exciting possibility of links to data held in the national databases through the new OS digital mapping – Mastermap, due to be implemented towards the end of 2001. A pilot project, initiated in Scotland, is being undertaken with Ordnance Survey by the three national records in Scotland, England and Wales. The project associates data from the NMRS databases, using the unique identifier with the layers in the OS Mastermap through the OS unique identifier (TOID – Topographic Identifier). Users will have the facility when using the OS map data, not only to be able to search the "Heritage theme", but to have links built into the digital map that lead directly through to the data held in the three NMR databases. This provides an exciting means of dissemination and is being used by OS as a pilot for associating data from a number of different sources. It is also likely that this will include a layer of ortho-rectified air photography and indexes to "historic" photography held in locations such as NMRS.

4. AND FINALLY, PUBLIC ACCESS

The challenge for modern inventories is to capture and present data in a way that can be made compatible with other data sets to make strategic planning easier for the implementation of policy decisions and research agendas, and to be presented in a user-friendly way for public use, both in terms of ease of use and accessibility of language in terminology and description. This challenge is being met through on-line access to datasets including CANMORE, the interface to the NMRS. The next major development will be the release of a GIS web interface – CANMAP – that will complement the online database and complete the use of GIS as integrating technology for public access through CANMORE.

5. REFERENCES


McKeague, P. (2001): “Where we are understanding data and positional accuracy issues at RCAHMS”, Paper presented to GIS e-conference http://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind01i10&L=fs1h0=D&F=n&6=P=8190
http://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind01i10&L=feh0=D&F=n&6=P=8138.


Stevenson, J.B. (2000): “Rozpoznanie dziedzictwa archeologicznego Szkoły prace Królewskiej Komisji ds. Starożytnych i Historycznych Zabytków Szkoły (Royal Commission on the Ancient and His-
ACKNOWLEDGEMENTS

I should like to thank Peter McKeague, without whose help this paper would not have been completed. Thanks are also due to Jack Stevenson, Dave Cowley and Georgina Brown for their assistance and to all my colleagues whose work has helped to develop the NMRS system and who continually strive through highly skilled work to maintain it, to ensure the quality of the data and who are always coming up with ideas for improvement and further development.

CHAPTER 14
MANAGING THE SPATIAL DIMENSION OF THE EUROPEAN ARCHAEOLOGICAL RESOURCE. TRENDS AND PERSPECTIVES

DAVID W. WHEATLEY
University of Southampton, United Kingdom
LEONARDO GARCIA SANJUAN
University of Seville, Spain

I. INTRODUCTION

In the late 1960s Swedish archaeologist C.A. Moberg claimed that research and management of the European archaeological resource were entering a new stage characterised by the super-abundance of information, which in turn produced the challenge of an ‘archaeographic crisis’. The accumulation of data and archival material from rescue excavations and the explosion in the volume of scientific data in archaeological research were the two main factors leading to such an unprecedented increase in information (Moberg, 1987:13). Almost thirty five years on, it is hard not to be impressed by the predictive sharpness of Moberg’s statement. Both curatorial and research-oriented interventions (excavations or otherwise) have continued to be carried out at an increased rate, as the evidence discussed throughout this book shows. Archaeological Resource Management (ARM) and scientific organisations responsible for such interventions have become firmly established in many more European countries and regions (as they have world-wide). At the same time, archaeological data analysis practice now comprises techniques that derive from a vast range of scientific disciplines ranging from physics to chemistry, geology, soil science, medicine, biology, etc. This trend towards a significantly expanded concept of what constitutes archaeological data, only incipient at the time Moberg made his forecast, has now become mainstream. A survey carried out in the early 1990s, before the rapid growth of Internet access, suggested that an average of 3000 books on archaeological subjects were being published in the world per year, just in English (Runnels, 1994:358).

Archaeological data have continued to accumulate at a faster and faster pace in more and more places while expectations about the accessibility and usability of those data have also risen dramatically. The extension of computer networks in the last decade, particularly the Internet, has created a true culture of readily available, retrievable and usable data which is quietly permeating all areas of society. Moreover, the very existence of Internet, which as Kilbride points out (this volume), is not only expanding but doing so at an increasing rate, is acting as an stimulus for a further expansion in the amount of information available for the user. Put simply, the more readily available users expect information to be (faster data transfer, more user-friendly interfaces, etc.), the larger the amounts of information that will be made accessible to them. Archaeological organisations and users are already beginning to feel the effects of this paradox. In Chapter 12 of this volume, Hansen and Dam have shown how, as soon as the Danish national database of archaeological sites went on-line in 1997, the system became so popular among profes-