ARCHIVING ARCHAEOLOGY: INTRODUCING THE GUIDES TO GOOD PRACTICE

Jenny Mitcham **Kieron Niven** **Julian Richards**
Archaeology Data Service

ABSTRACT

This paper looks at some of the domain specific preservation challenges faced by the Archaeology Data Service and how we work with these in order to maximise the re-use potential of the data that we archive. It looks in particular at one of the mandatory responsibilities of an Open Archival Information System (OAIS) and how we try to ensure that the data that we present to our designated community is 'independently understandable'. The paper introduces the collaborative 'Guides to Good Practice' project which aims to provide data producers with the guidance that they need in order to create data that is well documented and thus suitable for archiving and re-use. This Mellon Foundation funded project carried out in association with Digital Antiquity in the United States is now in its final stages and includes comprehensive and practical advice for data creators plus a number of case studies which demonstrate the real practical application of the Guides.

1. INTRODUCTION

The Archaeology Data Service (ADS) was founded in 1996 for the purpose of preserving digital data produced by archaeologists based in the UK, and making it available for scholarly re-use. The ADS was initially established as part of the Arts and Humanities Data Service (AHDS), with sister services covering other disciplines within the arts and humanities. Data are archived to ensure long term preservation, but they are also made available free of charge for download or via online interfaces to encourage re-use.

2. ADS AND OAIS

The digital archive at the Archaeology Data Service was established several years prior to the acceptance of the Open Archival Information System (OAIS) model as an ISO standard. ADS archival procedures and policies have evolved over time as the organisation itself and the wider world of digital archiving has grown and matured. We have now adopted the OAIS model and retrospectively tried to map our archival practices to it, looking in particular at data flows and at the six mandatory responsibilities. This has been an interesting process. Some of the OAIS mandatory responsibilities are easier to comply with than others. The ones which we have found most challenging are (perhaps unsurprisingly) the ones which we have the least control over. In particular where they relate to how the data producers create their data and prepare it prior to archival deposition with us.

OAIS states that an archive should:
"Ensure that the preserved information is independently understandable to the user community, in the sense that the information can be understood by users without the assistance of the information producer." [1]

This of course is not just up to the archive itself but will inevitably involve some input from the data producer as they are the ones who have the greatest understanding of the data in question and are best placed to provide suitable metadata and other crucial contextual information. Metadata isn't always something which can be generated in retrospect. In many cases it is essential that the metadata is created while project data is being actively generated and processed. It is at this point that creators have the clearest idea of what information each file contains, where it was collected, how it was collected and how it was subsequently processed.

3. THE DISAPPEARING SPIRAL

Take for example a project in 2004 to look for an elusive 'spiral' reportedly carved into the rock on one of the stones of the ancient stone circle at Castlerigg in Cumbria, England. The project team from the Universities of Durham and Bristol used the relatively novel techniques of 3D laser scanning and ground based remote sensing in order to reconstruct the 3D surfaces with millimetre and submillimetre accuracy [2]. These techniques can produce high quality images which can be analysed with a much higher level of objectivity than more traditional rock art recording methods such as wax rubbings and scale drawings. The team didn't find the spiral, suggesting that perhaps if it ever existed it was painted rather than carved on to the rock. In terms of research, this negative result is just as valid as a positive identification and the resulting point cloud and surface model data was archived with the ADS so that future

archaeologists can make use of it in whatever way they like.

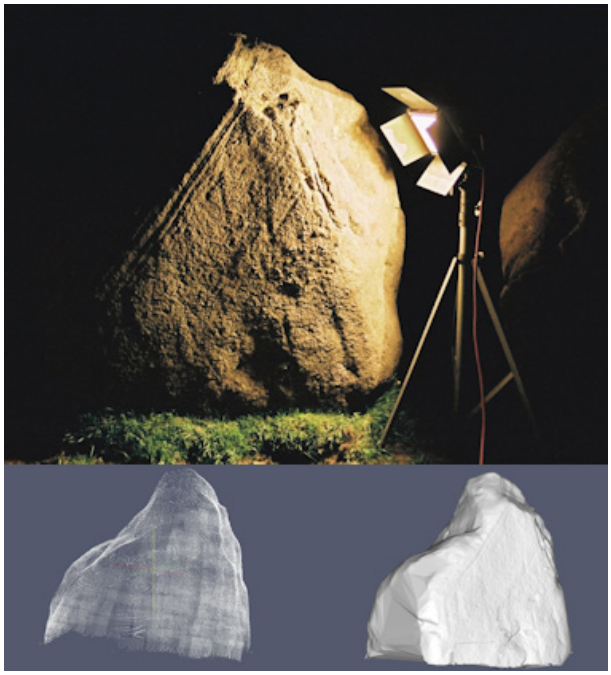

Figure 1. Fieldwork in progress on stone 11 of Castlerigg Stone Circle (top), point cloud and solid model created from the laser scanning data and archived by the ADS (bottom) © University of Durham

Perhaps a researcher some years down the line will want to return to the Castlerigg data files and continue the search for the 'spiral'. In order to fully assess the data from the original fieldwork and reprocess it they would need to know exactly how the 2004 fieldwork was carried out: what equipment was used, the point density on the object, which processing routines were carried out and what software was used. Even information about the date and time of the scan and the weather and light source could be useful. This is the sort of information we should be receiving as part of a Submission Information Package (SIP) so that we can ensure the data has enough contextual information alongside it to make it both understandable and useful. But how can we ensure that we always get what we require?

4. QUESTIONS AND CHALLENGES

Another and perhaps one of the biggest domain-specific challenges that we face as an archive for archaeological data is the range of file types that we are asked to ingest into our archive. A number of the projects we archive (such as that described above) feature cutting edge research using new and innovative technologies. As well as standard file formats that can be found in the majority of archives (documents, images, spreadsheets), we also have to deal with a diverse range of project outputs (maritime and terrestrial geophysics, geographic information systems (GIS), photogrammetry, lidar, virtual reality and more). The resulting files are often large in size and can come in a huge variety of proprietary and binary data formats. Finding ways of preserving these sorts of data can be a challenge. How do we get people to submit data in formats suitable for preservation? Which file types are we able to deal with and what levels of metadata need to be supplied in order to make the data 'independently understandable' to our designated community and thus suitable for re-use?

5. BACKGROUND TO THE PROJECT

These are questions we have been trying to address over the past few years, through projects such as the English Heritage funded 'Big Data' project[1] and the European funded VENUS (Virtual ExploratioN of Underwater Sites) [2] and also through our previous 'Guides to Good Practice'[3] publications aimed at data producers.

These Guides were published by the ADS from 1998 to 2002 and were available in hard copy and also free of charge as static on-line publications. They focused on subjects such as excavation, geophysical datasets, GIS, Computer Aided Design (CAD) and virtual reality, providing practical advice on the creation, preservation and re-use of digital resources and all including useful sections on metadata creation. They had been well received by the archaeological community at the time, but were in need of an update in order to keep up with the latest methods, techniques and technologies in use in these fast moving fields.

Figure 2. One of the original ADS Guides to Good Practice, *Archiving Aerial Photography and Remote Sensing Data* (both on-line and hard copy versions)

[1] http://ads.ahds.ac.uk/project/bigdata/
[2] http://ads.ahds.ac.uk/project/venus/
[3] http://ads.ahds.ac.uk/project/goodguides/g2gp.html

6. THE GUIDES TO GOOD PRACTICE

Building on these existing 'Guides to Good Practice' we have, over the last two years, been working with archaeologists in the US to refresh and enhance this resource. The current project is predominantly being carried out in support of the Digital Antiquity initiative, a Mellon Foundation funded US-based project with teams working at the University of Arkansas and Arizona State University.

Through this new, collaborative project we are in the process of updating and restructuring the original Guides, making them available in an on-line wiki environment[1] to allow easy and quick collaboration and also more frequent future updates. In order to keep pace with the wide range of techniques that archaeologists use, we are also including new subject areas such as 3D laser scanning, lidar and photogrammetry (Table 1).

Updated Guides	New Guides
Aerial Survey	Marine Remote Sensing
Geophysics	Laser Scanning
Geographic Information Systems (GIS)	Photogrammetry
Computer Aided Design (CAD)	Ground Penetrating Radar (GPS)
Virtual Reality	Polynomial Textual Mapping (PTM)

Table 1. The updated and new data types and technologies covered in the new Guides to Good Practice series

As well as these technology-specific guides, we have also concentrated on a set of 'Basic Components'. These are the common digital objects that often appear in an archive that is deposited with us, regardless of the nature of the project or the technologies used – primarily textual reports, digital photographs, databases and spreadsheets and occasionally digital audio or video files (Table 2). As these basic components are ones which the majority of data producers will need some guidance on, they have been separated out and are linked to from appropriate places in the other Guides.

Basic Components
Documents and Texts
Databases and Spreadsheets
Raster Images
Vector Images
Digital Video
Digital Audio

Table 2. The 'Basic Components' covered by the new Guides

[1]http://guides.archaeologydataservice.ac.uk/

In order to create these guides we have invited the original authors (all specialists in their particular fields) to review and update the content. New authors from both the UK and US have also been drafted in to contribute. Once the Guides have been updated, they will undergo wider review by a panel of experts.

The wiki format of these new guides has a number of obvious benefits. Several authors may work on the material simultaneously with the results being made immediately available to all. The wiki allows for page-level privilege control – so authors will have the ability to edit only those sections that they have permissions to author. For each wiki page it is possible for the editor to view the 'page info' in order to see all the edits that have been carried out. This allows them to keep track of all changes that have been made and view all previous versions.

Figure 3. Editing page content in the wiki environment

The Guides will provide data producers with a comprehensive and peer-reviewed set of guidelines explaining how to create data that is suitable for long-term preservation and how to package it up with the correct metadata to ensure it is 'independently understandable'. Different chapters of the Guides target different technologies or groups of files, so users will be able to quickly and easily find the section that is most relevant and useful to them. The wiki format also allows a high degree of interlinking between relevant sections of the Guides making them into a far more interactive resource than previously possible.

Unlike the original Guides to Good Practice series, this new wiki-based publication is not being produced in hard copy form. In recognition of the fact that some archaeologists may want to take a section of the Guides out into the field with them where they have no internet access, and that other users simply may not want to read large quantities of text from a screen, there will be a pre-

prepared PDF of each Guide allowing users to download and print-on-demand.

Although the Guides have clearly been written with archaeologists in mind, they do have wider application. Much of the advice contained within them, for example that relating to significant properties, suitable file formats and metadata, will be also be applicable to practitioners in other disciplines. As the project reaches completion the whole wiki will be open and freely available to all.

7. CASE STUDIES

A key component of the Guides to Good Practice project is the inclusion of a number of case studies. These case studies demonstrate how the Guides could be used by archaeologists to promote best practice in data creation and produce outputs that are suitable for long term archiving. The case studies will be used to illustrate the archiving of some of the specialist data types in archaeology, from creation and ingest through to dissemination. In this way, the real practical application of the Guides will be apparent.

The electronic nature of the new guides allows an integrated approach to these case studies. From each guide we will be able to link through to an exemplar archive which will illustrate the workflows that have been followed for the numerous data types and demonstrate what the final archived dataset might look like. This will be of particular value to archaeologists who are actively producing data, allowing them think about how their own data might look to other researchers once their fieldwork is complete.

All case studies are drawn from real and current projects in the academic and commercial worlds of archaeology.

8. TRENT-SOAR RIVER CONFLUENCE

The first of the case studies that we are working on is a study of the landscape surrounding the confluence of the Trent and Soar rivers in the East Midlands, England carried out by Birmingham Archaeology. Previous archaeological work on British river floodplains has suggested that river confluences can provide a focus for human activity through the ages. The distribution of archaeological remains in these regions is closely linked to the configuration of the landscape within the floodplain, both in terms of the original locations of sites and the level of preservation of the physical remains today. Attempting to accurately record and map this landscape was therefore a crucial goal of this project [1]

In order to achieve this aim a number of different techniques and technologies were employed by the project team – aerial photography, lidar, geophysics (including GPR), GPS survey and GIS. This diverse and complex dataset is relevant to several of the Guides to Good Practice and can serve both to test the guidance and illustrate best practice in creating and submitting data that is suitable for long term archiving.

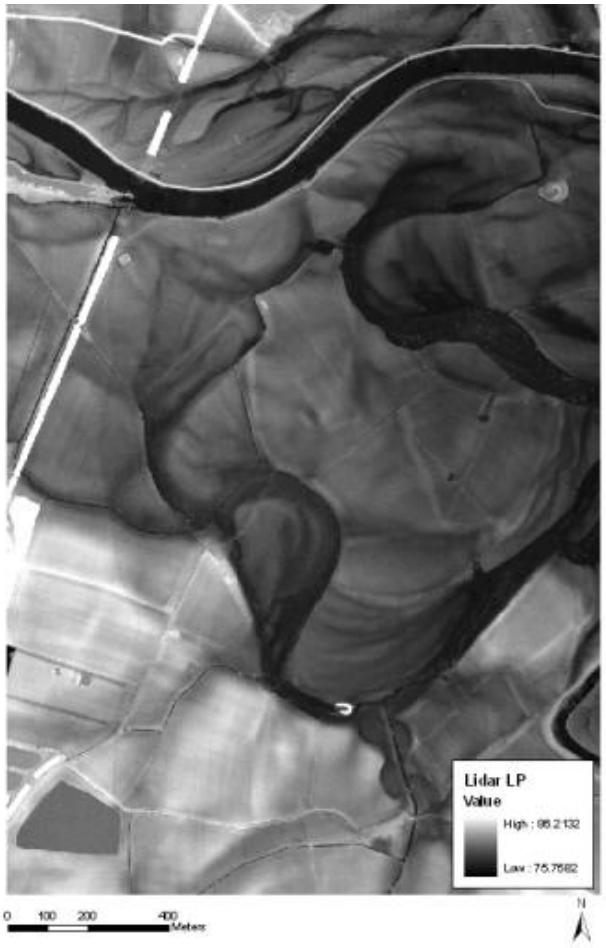

Figure 4. Lidar last-pulse (LP) surface model of the Trent-Soar confluence. Image © University of Birmingham. Lidar data © Infoterra Global Ltd

9. THE FUTURE

We have been working with our data producers for many years now, trying to ensure that the SIP we receive from them is adequate in terms of the types of files they send and the level of metadata attached to it. This however has never been a particularly easy job. We need to encourage data producers to think about digital archiving from the very earliest stage of their project in order to ensure that they create their data in the right way with the right documentation. Alongside systems we already have in place such as on-line guidelines for depositors and metadata templates[2], we will soon be able to point people

[1]See the following resource for phase I and phase II reports from this project which have been archived by the ADS http://ads.ahds.ac.uk/catalogue/resources.html?trentsoar_eh_2008

[2]http://ads.ahds.ac.uk/project/userinfo/deposit.cfm

to these new 'Guides to Good Practice' from the outset of a project. The net result being a better, more complete SIP, and data that is 'independently understandable' to our designated community.

10. REFERENCES

[1] Consultative Committee for Space Data Systems *Reference Model for an Open Archival Information System (OAIS)*. CCSDS 650.0-B-1 Blue Book, 2002.

[2] Díaz-Andreu, M., Brooke, C., Rainsbury, M., & Rosser, N. "The spiral that vanished: the application of non-contact recording techniques to an elusive rock art motif at Castlerigg stone circle in Cumbria", *Journal of Archaeological Science*, Volume 33, Issue 11, pages 1580-1587, November 2006.

11. ACKNOWLEDGMENTS

The authors would like to acknowledge Keith Kintigh and Francis McManamon at Digital Antiquity, Arizona State University and Fred Limp at the University of Arkansas for their work on the Guides to Good Practice project. We would also like to thank the following funding bodies: Andrew W Mellon Foundation, English Heritage, and the European Commission VENUS project.