

The Decennial Award is a special award offered to mark the tenth anniversary of the DPC. It will be presented to the project, initiative or person that in the period 2002-2012, the judges assess to have made the most outstanding contribution to ensuring our digital memory is available tomorrow. Four finalists have been selected for this prestigious prize. Here the nominees describe their motivations, their projects and their impact.

### **Archaeology Data Service at the University of York**



Archaeology is unusual in that the creation of knowledge results from the physical destruction of primary evidence, making access to data all the more critical in order to test, assess, and subsequently reanalyse and reinterpret both data and the hypotheses arising from them. Over the years, archaeologists have amassed a vast collection of fieldwork data archives, a significant proportion of which remain unpublished. Furthermore, much fieldwork data is increasingly born digital, making it all the more precious. Access to data, even those which are published, is often difficult or inconvenient at best.

The Archaeology Data Service (ADS) was established in September 1996, as one of five discipline based service providers within the Arts and Humanities Data Service (AHDS). The ADS developed from a successful bid to the AHDS made by a consortium of university Departments of Archaeology and the Council for British Archaeology, led by the University of York. From an early stage the ADS also began to receive external funding from a variety of

other organisations, such as English Heritage, reflecting the diverse nature of the archaeological sector. It developed a charging policy which is based on a one-off digital deposit charge, and provides a sustainable financial model for digital archiving. Its innovative approach to charging was widely welcomed and ADS has worked with Charles Beagrie Ltd on the Keeping Research Data Safe projects to share best practice.

The ADS works with national and local archaeological agencies and those research councils involved in the funding of archaeological research, to negotiate deposition and secure archiving of project data. This includes data derived from fieldwork as well as desk-based studies. The types of data involved include: text reports, databases (related to excavated contexts or artefacts, for example), images (including aerial photographs, remote sensing imagery, photographs of sites, features and artefacts), digitised maps and plans, numerical datasets related to topographical and sub-surface surveys and other locational data, as well as 3D reconstructions. In March 2011 the ADS was accredited with the Data Seal of Approval, an international 'kite-mark' for digital repositories, becoming only the second UK repository to gain this recognition, after the UK Data Archive at the University of Essex. In 2012 it was accredited by MEDIN as a Data Archive Centre for the Marine Environment Data and Information Network.

Over a decade before the current push towards Open Data the ADS made the data it preserves available for re-use. On 15th September 1998 the ADS launched the first version of ArchSearch its online catalogue. All data sets are freely available for download and re-use; in 2002 it launched HEIRPORT, the first interoperable gateway for the historic environment sector. In 2005 it completed ARENA, the first portal to European cultural heritage and digital archives. In 2011 it launched TAG, the first Transatlantic Archaeological Gateway. ADS is now the UK supplier of 3D heritage data to CARARE, a Europeana best practice network, and a lead member of ARIADNE, a new EU Research Infrastructure project. Between 2007-9 ADS worked on Archaeotools, a JISC-EPSRC-AHRC eScience project to undertake Natural Language Processing of its collections and to implement a faceted browse interface, now available via ArchSearch. The ADS has also worked on Open Linked Data as another means of providing access to its online holdings, via the STAR and STELLAR projects.

ADS now preserves over 17,000 grey literature reports and over 500 data rich digital archives, derived from archaeological research projects and primary fieldwork. The archives represent some of the most important sites in British Archaeology. The grey literature has become the basis of other research projects, most recently a Leverhulme grant awarded to Professor Mike Fulford to study rural settlement in Roman Britain, and an ERC senior investigator award to Professor Chris Gosden, who is looking at Regional Identities in Britain. The significance of the digital heritage preserved is immeasurable. In an ongoing JISC commissioned survey on the Impact of ADS 74% of users said that ADS was very or extremely important for their academic research, 64% said it was important for their private research, and 55% said it was important for

their learning and skills development. 48% of depositors reported that not being able to provide data to ADS would have a severe or major impact on their work.

ADS has worked extensively on data standards. In 1999 ADS published the first Guides to Good Practice, and published 6 titles in its first 6 years. In 2006 it completed work on 'Big Data' for English Heritage, on the "Preservation and Management Strategies for Exceptionally Large Data Formats". In 2011 it launched the new online second editions of the Guides, freely available and including new areas, such as standards for Underwater Archaeology, derived from the EU-funded VENUS project. Keith Kintigh, Past President of the Society for American Archaeology writes that the ADS "is an enormous asset to the UK's archaeological community - within and outside academic settings. It provides both secure preservation and rich access to the irreplaceable records of archaeological investigations, thereby allowing this information to be effectively used in archaeological research and cultural heritage management.

ADS has been a key player internationally in advancing initiatives concerned with the preservation and dissemination of cultural heritage information and has, indeed, been a model of a sustainable and productive digital archive of archaeological data and documents. In our own multi-institutional effort to develop a digital archive for archaeological data in the US, ADS has not only served as a valuable model, its staff have provided critical advice and assistance. Further, ADS has been a major driver of international efforts to establish interoperability of digital repositories facilitating the sharing of archaeological information. Looking beyond archaeology, with its long (for a digital repository) history of success, ADS has also been a widely cited exemplar of a successful disciplinary repository."

The ADS has been awarded two British Archaeological Awards for Innovation, in 2008 and 2012, and it featured extensively in the successful nomination of the Department of Archaeology at the University of York for a Queen's Anniversary Prize for Higher and Further Education in 2011.

**Note - ADS (nominee) and University of York (Host) may not vote for ADS; JISC (initial funder) may not vote for ADS**

### **PREMIS Data Dictionary for Preservation Metadata**



PRONOM and DROID from The National Archives



DROID supports batch processing of large numbers of files. It is freely-available to download under an Open Source license, and is written in the platform-independent programming language Java. It provides both a graphical user interface and a command-line interface. DROID provides comprehensive reporting on collections of digital records, including formats, extensions, PUIDs, filepaths, and check sums, the latter offering a quick method of finding duplicate files even when the files may have different filenames. All reports can be saved and exported as spreadsheet files for detailed analysis.

Since DROID connects via web services to the PRONOM registry, users always have access to the latest available file format signatures. Users may also develop and implement their own signature files and we have produced detailed and freely available information on how to achieve this, meaning that individuals and institutions are not tied to The National Archives'

research alone.

In addition to appointing a full-time File Format Signature Developer, we have invested in continuous development of both PRONOM and DROID. DROID 5 introduced scanning of archive formats, such as .zip, meaning that DROID now reports on the contents of these files without the need for manually opening each archive file. DROID 6 provides container signatures for the first time, enabling accurate identification of compound formats, such as OLE2 used by Microsoft. DROID 6.1, to be released in autumn 2012 provides further stability and a more efficient command-line identification option. DROID 7, the next major release, has an openly available wiki for interested parties to submit their own requirements.

The next development we are planning for PRONOM is to make the entire registry available following a Linked Data approach. We have already made available via our website a prototype of Linked Data PRONOM.

The achievements of PRONOM and DROID are clear: they have stimulated debate and further thought on the subject of digital preservation; PRONOM was the first publicly available technical registry for file format information and both PRONOM and DROID have inspired a number of similar tools. PRONOM provided a significant amount of data to the Unified Digital Format Registry recently launched by the University of California Curation Center and the California Digital Library. DROID is embeddable within digital preservation workflows and systems conforming to the Open Archival Information System model, for example it is fully embedded within Tessella's Safety Deposit Box system for which we won a Queen's Award for Enterprise in 2011.

PRONOM and DROID are 10 and seven years old respectively and throughout their lifetimes they have contributed significantly to the field of digital preservation. The National Archives remains wholly committed to these tools, which have not only helped to drive the success of our own digital records infrastructure, but have also been recognised and adopted worldwide.

**Note - The National Archives (nominee) may not vote for PRONOM and DROID**

**International Internet Preservation Consortium**



