

## HISTORIC ENVIRONMENT INFORMATION RESOURCES NETWORK (HEIRNET)

Minutes of an HEIRNET meeting, held on Thursday 28 January 2009 at the British Museum, Great Russell Street, London. starting at 12.30pm.

Those present: David Baker OBE (Chair), Roger Bland OBE (PAS), Cat Cload (EH), Dan Hull (CBA), David Thomas (RCAHMW), Dan Pett (PAS), Stuart Jeffrey (ADS), Peter McKeague (RCAHMS).

ACTION

### 1 Welcome and introductions

David Baker welcomed everyone on behalf of the PAS who hosted the meeting.

### 2 Apologies for absence

Apologies were received from Prof Julian Richards (ADS), Catherine Hardman (ADS), Dr Mike Heyworth (CBA), Ian Shepherd & Bruce Mann (ALGAO Scotland), Paul Gilman (ALGAO England), Nigel Clubb & Vikki Fenner (EH NMR), Hilary Malaws (RCAHMW).

### 3 Minutes of the last HEIRNET meeting

The minutes of the last meeting were AGREED with no corrections. Dan Hull informed us that following discussion at the last meeting about how best to replace David Dawson on the committee, Nick Poole from the Collections Trust has now been invited to attend.

### 4 Report on current HEIRNET services:

#### i. HEIRNET Register (<http://ads.ahds.ac.uk/heirnet>)

Stuart Jeffrey reported that there had been two updates since the last meeting. Although the HEIRNET Register will be maintained as a simple list of services and their urls, the primary focus now is on establishing the HEIRNET Registry – a web services registry maintained by the ADS on behalf of HEIRNET, with a ‘T-Model’ (an associated web services description document) owned and maintained by FISH (the Forum on Information Standards in Heritage). See below for further detail:

#### ii. HEIRNET Registry

The registry mechanism is complete and functional and has been demonstrated at the workshop following the last HEIRNET meeting in York in July. FISH agreed at their last meeting in December to own the T-Model. The ‘front end’ for the registry will shortly be completed by the ADS. Stuart Jeffrey and Dan Hull will be speaking about

ADS to continue registry develop-

this in a paper at the IfA conference in April this year. In terms of services signing up with the Registry, ideally we would 'capture' the online HERs for England as they joined the Heritage Gateway. David Baker enquired whether IHBC were aware of developments. Dan Hull responded that IHBC may indeed not be aware as a IHBC representative has not attended an HEIRNET meeting since 2005. It was agreed that we should make special efforts to update them of developments.

ment  
work

iii. HEIRPORT (<http://ads.ahds.ac.uk/heirport>)

DJH to specifically invite IHBC to attend the next meeting.

Stuart Jeffrey reported that though the ADS now regard this largely as a finished project, the cross-searching mechanism is still in place and largely functional (though some targets drop out of use from time to time). C.1250 searches per month are carried out.

The ARENA project, a similar Z39.50-based service in its first iteration is about to move into its second phase with a new project. This will 'wrap' Z39.50 targets in a web services layer. This development may be of advantage to HEIRNET in that this experimental work may provide a precedent for a future version of HEIRPORT.

## 5 Report on the HEIRNET Web Metrics Survey

Dan reported that although this survey had been live now since August 2008, still only 13 organisations have responded. Cat Cloud agreed to help in promoting the survey to online HERs, and it was generally agreed that we should continue the survey for a further month in order to gain some further responses. Stuart asked if additional questions about (a) referring sites and (b) whether organisations using Google Analytics are intending to place a disclaimer on their website. Dan said these could be added into the survey, though we'd have to make it clear that responses so far have not included these questions. Dan presented interim results so far:

Dan to amend survey and circulate for a further month. Cat to circulate to online HERs which are in-touch with the Heritage Gateway

- 90% collect results as part of a regular cycle (most quarterly, some monthly, one weekly)
- 30% page impressions
- 30% 'hits'
- 40% 'other' (logins, database queries, some a combination)
- Almost all collect unique users, 50% on IP addresses, smaller number on registered users, cookies or a combination
- Almost all collect metrics on where their users are: varies between country, domain, ISP location, user registration stats
- Less than half collect metrics on the operating system used
- Over 80% collect information on session duration
- A high percentage collect information on what the user is doing on the website, with most interested in file downloads, but also searches, adding user comments, and carrying out a transaction
- Over half don't collect information on what terms a user employs to get to a website, and less than half are interested in how a user got to the site at all

- And over half don't collect info on search terms employed once users are within a resource either
- Less than 20% are interested in user pathways
- How are the stats collected? More than 80% stated that they collect them in-house
- Roughly half and half split between information officers collecting them themselves and a specialist IT team elsewhere in the organisation doing so
- All organisations said that they don't have a specific budget for collecting and analysing stats
- Only 10% said they use a browser-based approach (employing Google analytics, for example); nearly 90% on the basis of web logs
- A wide range of software is used to gather and analyse web metrics, including Nettracker, sitemeter.com, SmarterStats, Analog, Google Analytics, and functionality that is part of Oracle
- Most (80%) said they either didn't collect, or didn't know whether they collected, stats on robot and/or spider activity
- Organisations stated that they were intending to select a range of further indicators in the future, including total number of visitors, the numbers who stay longer than seven minutes, 'user satisfaction' and other user feedback, and different forms of searching.
- Some also said they used non-metrical methods such as web surveys, help desk comments etc

## 6. Updating and development of the HEIRNET website

At the last HEIRNET meeting, we had a discussion about the future direction of the HEIRNET website. It was agreed that there is still a need for one, and that a site which is broadly the same as the current one would be fine. Dan offered to carry out some basic consistency changes (harmonising style sheets) and ensuring that contact details are up-to-date. This has now been done, though it is unclear what direction the 'technical advice' section should take as much of it is out-of-date and really needs to be deleted or updated. This was discussed, and though it was agreed that this section should not be exhaustive as better sources of information exist elsewhere, once the HEIRNET Registry is created Stuart will provide some brief text explaining the nature of Service Oriented Architecture, Web Services Description Models etc. It was also agreed that Dan will circulate the HEIRNET url to members and asking for further comments on the existing site.

Stuart Jeffrey to provide some basic technical advice on SOA and WSDLs for the website

Dan to invite further comments on the website

## 7. Future Challenges and Opportunities

Stuart Jeffrey explained that the origins of this agenda item lie with a discussion between Mike Heyworth (CBA), Catherine Hardman (ADS and Chris Puttick (Oxford Archaeology) at a recent grey literature meeting. The issue was raised whether there may be (or at least may have been at the time – in the autumn of 2008) opportunities for

the historic environment sector to approach a prominent IT provider for substantial funds to help further develop our information infrastructure. This has provoked further thought about which are the primary issues which such an opportunity (whether it now exists in the current financial climate or not) might be used to address. Following on from this, and with relevance to HEIRNET, what are the major information management issues which we all face, and how could we best use HEIRNET as a forum to address them? Stuart Jeffrey led this discussion by presenting a set of four major issues which, as far as he sees it, are the major ones we should all continue to scrutinise and compare notes on:

- i) Museums: HEIRNET could provide a forum to promote greater information sharing with museums. A recent Research Information Network study ([www.rin.ac.uk](http://www.rin.ac.uk)), entitled 'Discovery Aids', highlights the variation within the museum sector of approaches to the issue of information sharing. Many do not, and cannot, provide access to their collections online, in spite of a demonstrable demand among researchers for such information. Promoting the sharing of such information both online and directly with HERs, for example, would be of enormous benefit to the sector. There are real challenges and barriers to this, but HEIRNET could play a greater role in framing these issues and sharing information about the latest developments and opportunities. David Baker reported that the issue of uniting museums data with other data sets is of concern, for example, in Greater London. David Thomas commented that from the Welsh experience, obtaining good geographical location data for all data sets is an important starting point, and that this may require 'champion' examples to lead the way. Applying geographical tags to large data sets which don't already have them could be a real logistical and financial difficulty - community tagging (along the line of the Aerofilms Collections, for example) could be one way of achieving this. The 'Gathering the Jewels' project may have important lessons for us too. Peter McKeague commented that in Scotland such linkage has proved challenging due to the very different nature of the respective catalogues, and the underlying/original intentions of each catalogue. Dan Pett suggested we keep in touch with the National Museums Online project.
- ii) What is a 'trusted digital repository' - what definitions are now appropriate? Do we have criteria and methods for assessing the long-term reliability of particular repositories? Are such criteria adequate, clear and transparent? David Thomas informed us that there has been much SWISH planning on this issue - there is now a recognised certification process.
- iii) Who has primary responsibility for grey literature? Local authorities, the originators of the data, the national monuments records, or others? Associated with this issue, as access to (and aggregation of) grey literature from different sources becomes increasingly possible through a web services model, do we have a joined up strategy for grey literature unique identifiers? Who assigns these and what form do they take? David Baker pointed out that the issue of minimum standards for grey literature is also a problem - should unique identifiers only be assigned to those that reach basic standards (including metadata standards). David Thomas suggested that archiving

Dan H to make special efforts to invite Nick Poole (from Collections Link) and Gail Boyle (from the Society of Museum Archaeologists) to the next HEIRNET meeting

procedures are important to this ‘minimum standard validation, and Peter pointed out that security of an archive is also relevant. Stuart reminded us of the OAIS standards model, and that this represents an important benchmark for us to seek to achieve as information providers. David Baker suggested that the IfA has a role here through the Registered Organisation validation process. Dan H suggested that the Digital Preservation Coalition may be able to provide guidance about developments with archiving standards and validation processes. It was agreed that the new director of the DPC, William Kilbride, could be invited along to a future meeting to give us the DPC position.

- iv) What forms of threat does the financial downturn pose for historic environment information? Of particular relevance, and concern, here is the fact that with the very real possibility of archaeological contracting units going out of business that important data will remain un-archived and, potentially, lost. Also of concern is that with understandable financial cutbacks being made, specialist information managers may lose their jobs, thus endangering further the careful formation and archiving of databases, grey literature reports and other information sets. Do the national agencies, for example, have a strategic position on this issue? Are they willing to act as the archive of last resort for endangered data sets?

Stuart to consult the IfA on their position/ Thinking on this issue

It was suggested, and widely agreed around the table, that these four issues should remain on the agenda for discussion at each HEIRNET meeting until further notice, so that all partners can contribute information and share concerns.

### **8. Proposal for an HEIRNET ‘Heritage Information’ Handbook**

Dan circulated a proposal for a book, potentially produced as a CBA Practical Handbook which would be a short, readable and accessible summary of the forms of heritage information which exist and how to get access to them. The publication could have three main elements: a printed practical handbook, a marked-up web publication, and a summary pamphlet for widespread distribution. It is proposed that the CBA take the lead in compiling and editing the text, though with contributions requested from each of the main resource providers themselves. A sub-panel of HEIRNET representatives could provide additional guidance to the editors, reading through text and making further suggestions where necessary. The CBA would make a contribution to this book from core resources through the provision of staff time. In terms of print production expenses, it is hoped that funds could be contributed from each of the HEIRNET members.

All agreed with the general concept of this book, its structure and proposed format(s). David Baker suggested we approach the Heritage Lottery Fund for a grant. It was pointed out that for various of the HEIRNET partners to contribute time to this some formal planning will have to take place so that appropriate staff time from the various partners is reserved. It was also expressed that the differences nuances and details relating to the constituent nations are important to get right.

Dan suggested that the next stage could be to incorporate these comments into the book proposal and take it to the CBA Publications Committee for further scrutiny. If accepted, the CBA can then circulate the appropriate correspondence to HEIRNET partners to take forward the 'panel' who will guide the book's content.

Dan to report back on CBA Pubs Comm comments

### **9. Presentations on current projects and plans from HEIRNET partners**

**PAS:** Dan Pett updated us on progress with the new PAS website and Content Management System. Google Maps will be used extensively as an aid to presenting and interrogating PAS data, and data will be provided as a feed in a variety of formats: KML, GRSS and others. An Amazon feed will be employed for relevant books. Hoping for a February 2009 completion and then testing from then onwards. Roger added that the PAS has part-funded a HBSMR programme to provide PAS data directly to HERs (other forms of HER systems are possible in the future).

**RCAHMS:** Peter informed us that the Canmore database has now been rebuilt using a web services model and released in beta. There will be user-generated comments and customizable backgrounds. The 'Scotland's Places' service cross-searches the National Archives and Canmore: this is scheduled to be ready for the October 2009 'Coming Home' conference. An assessment of polygon data among the SMR community continues. A full audit of SMRs (now HERs) is about to take place, revisiting the previous survey (conducted by David Baker) ten years ago. A C14 resource is currently being developed jointly with the ADS,

**ADS:** Stuart informed us that the new ADS website would go to beta testing in March 2009. The new faceted classification is to be used. The Archaeotools Natural Language Processing project (conducted jointly with the computing department at the University of Sheffield) is reaching its final phase. The ADS is planning to use the EDINA GeoXwalk facility to generate geographical data, though there is some uncertainty currently about the application of an OS licence outside Higher Education. The ARENA 2 project (called DARIAH) will enter its second phase next week: each of the European partners will be discussing the current state/future development of their web services provision. The new Guides to Good Practice series (funded by the Mellon Foundation) is about to begin.

**EH NMR:** Cat Clod informed us that the NMR Access Programme continues: the EH Images business-to-business service has now been live for around six months. The new Heritage Explorer resource has now been relaunched, and is carefully designed to conform to the curriculum levels in England. PastScape now produces full NMR description (a development which has benefited Heritage Gateway too). The usability testing of PastScape has proved very useful. Heritage Gateway will also undergo some usability testing, and has had its interface streamlined somewhat – a little more redesigning is on the way. A new CMS is now being used. The numbers of datasets being consumed as services continues to rise, with 24 planned by the end of March 2009. The Parks and Gardens UK resource is set to join imminently Further usability testing

is planned also for Heritage Explorer and the Aerial England websites.

**CBA:** Dan H reported that the CBA is planning to launch the new biab online service for beta testing very soon. As well as an improved data model, the service will enable registered users to save and refine searches, to use a much-improved interface design. The full AIP grey literature bibliographic data up to 2006 will be included, with discussions underway between AIP-OASIS-ADS-CBA about the sharing of grey literature data via web services in the future. The new CBA website was launched in December 2008 with an improved design, revised content and news aggregator service. The CBA has also now received funding for an upgrade to the TORC service. The data model for this is almost complete, with the website to follow. Educational courses, events and excavation (Briefing) as well as contacts (Archaeology Online) will be served up via a single interface. Funding has been received from the Headley Trust for an upgrade to the Community Archaeology Forum, with Suzie Thomas about to embark on a research project consulting the voluntary sector about the support and facilities the new CAF should provide.

**RCAHMW:** David Thomas informed us that the new version of Coflein is now live (using the same web services as Canmore). The mapping of data is to be re-addressed soon. The Historic Wales project has had some technical difficulties, arising especially from the very different nature of the data sets, but all HER data is now available there. Listed Building descriptions will also go up there soon. The new HER database (for all four Welsh trusts) is now live and working internally. There are plans afoot to look at making this live to the public soon. The People's Collection, a project aimed at enabling the cross-searching of RCAHMW, the National Museum and the National Library data sets is also underway.

## 10 DPC membership

At the last meeting, Julian described the Digital Preservation Coalition ([www.dpconline.org](http://www.dpconline.org)) and their mission – they are a very high level, cross-sectoral body for organisations concerned about issues of digital preservation. The ADS was formerly a member through JISC. Might it be useful for HEIRNET to have a route through to it? The DPC will, in the future, have a role in the validation of digital archives. Membership rates are expensive (£10,000/annum Full; £2,000/annum Associate). It was agreed that HEIRNET should join together as an organisation, at the 'Associate' level, with partners contributing according to their means. The CBA offered to act as the administrator for this transaction, and to approach each of the HEIRNET partners directly now for contributions.

The CBA to discuss with each partner directly for contributions.

## 11 Other business

Peter alerted us to current ambiguity over the use of OS coordinates over Google Maps. There is some concern that this may not legally be possible.

Members to continue to share

inform-  
ation as  
this  
develops

## 12 Dates/venues of future meetings

It was agreed that six monthly meetings are of benefit to HEIRNET partners, and that we should schedule another for July, and to agree a venue.

Dan to  
circulate  
date  
options  
for  
confirm-  
ation