Data supply and Reconciliation between NRHE and HERs
(EH 6953)
Notes from meeting with English Heritage 12th September 2014

Project Meeting 12-09-14, Stonehenge Room, English Heritage, Swindon

Present

Paul Adams (EH)
Dave Batchelor (EH)
Nick Boldrini (Co Durham HER)
Kieran Byrne (EH)
Carlton Carver (EH)
Nick Davis (EH)
Crispin Flower (exeGesIS)
Jane Golding (EH)
Gill Grayson (EH)
Sarah MacLean (EH)
Dan Miles (EH)
Andrew Minting (Wiltshire Council)
Martin Newman (EH)
Graham Orbell (EH)
Matt Reynolds (EH)
Graham Tait (Devon HER – Chair)
Ben Wallace (Warwickshire HER)
Chris Webster (Somerset HER)
Helen Winton (EH)

Content of Summary

This summary deals with the questions arising from the demonstrations made of each system during the meeting and the subsequent discussion. It makes no attempt to summarise the content of the demonstrations themselves.

GT prefaced the meeting by explaining the background to the project and emphasised that the main focus of the meeting was to be to achieve greater understanding of the systems involved.

Summary

National Record of the Historic Environment (Kieran Byrne & Martin Newman)

Data is captured and managed in the AMIE database which is linked to deskGIS.

Q: Are links between the NRHE and the NHLE good? (GT)

Links are to be found in 'other identifiers'. Those regarding Listings are less comprehensive and up-to-date (MN).

Q: Are parent/child links used extensively within the AMIE database? (NB)

They are used to interlink very complex records, for example where components are physically dispersed (such as the conduits towers at Hampton Court Palace) (MN).

The convention used when recording monuments from aerial photographs has been that the link should be conceptual, not geographical (HW).

Q: Does a record of all related children reside with each parent record? (AM)

Yes, within the background menu (KB).

Somerset HER makes widespread use of parent/child relationships. These have been found to be useful when records are viewed on the website (CW).

Q: What goes in the summary? (NB)

The latest up-to-date interpretation. The long text records the history over a period of time s it is incremental (KB).

Q: What is the system regarding multiple entries within individual records? (CF)

These can be tabbed through on screen (KB).

Q: In cases with multiple addresses, are these dated? (CF)

No. (KB)

Q: Are post codes indexed on the record? (DB)

They are not held within AMIE data but can be accessed through GIS. (KB)

Definitely want to improve entry of geographical information as this is one of the main issues with AMIE (MN).

How an address is recorded within AMIE has been shown to have ‘knock on’ effects in other areas. For example, if no address is recorded the record will not appear in print-outs (HW).

Q: Does the system allow an NLPG link? (AM)

This has been investigated (via a GIS link) as part of the GINA project. However, it proved too complex to introduce (MN).

AMIE records appear rather disjointed and it seems that GIS data will be very important in this project. How robust are the cross references between AMIE and GIS? (CF)

The GIS is enhanced by the same recorders responsible for updating the monuments record (KB). (But not where AP recorders are concerned (HW)).

Q: Are forced links employed? (CF)

The creation of links is closely integrated with the process as a whole (KB).

Q: Are related polygons automatically deleted with their associated monument? (CF)

They should be. GIS deletions are carefully monitored (KB).

Q: In the case of maritime records is the county/district/parish (CDP) information automatically generated? (NB)

Other than in the case of very inexact accounts CDP data can usually be deduced from named locations (KB).

Q: Is the CDP look up information used regularly updated? (BW)

This is revised annually by EH Data Standards Unit (DM).

It should be remembered that CDP and Grid References are not automatically linked within the system (GO).

Q: How is the AMIE Phase Prerogative likely to be easy to accommodate within the exeGesIS system? (MN)

Types and periods are tightly bound within the system. Site types are visually more prominent. These are then attached to period categories. The NRHE data could probably be exported without major issues as the differences relate to presentation and implementation. (Whilst there are issues in period categorisation this is unlikely to be one of them) (CF).

Inputting AP data into the AMIE system is difficult to do efficiently as it requires multiple periods to be attached to the same monument type entry but the system is good at managing complexity, e.g. for buildings recording (HW).

The migrated data may initially look a little ‘clunky’ in HBSMR due to this element of repetition. It will, however, work in terms of searching. Linking between evidence and site type, as it is stored in AMIE, may prove more problematic since it seems to have been used irregularly (CF).

Maritime records can have complex indexing, particularly for casualties. We don’t prioritise indexing as we want maximise retrieval (KB).

For AP data there are national standards and detailed recording guidelines. QA is also in place (HW).

AP data within AMIE <u>should</u> carry link numbers. Not all recorders have been aware of this, however (HW).

Q: Is the 'Quantity' field used? (CF)

Neither the quantity nor certainty fields are used. Details are captured elsewhere in the record where relevant (KB).

Q: How are finds recorded within the system? (GT)

Land-based stray finds are no longer recorded, although some can be found amongst older records. Maritime records do, however, record finds individually that come from the Receiver of Wreck and BMAPA. Land-based hoards may be recorded but this has not been done consistently. The only finds which are now consistently recorded are those which are held to be indicative of a site. This would appear as a 'find spot' monument record. Exceptionally significant finds would also be indexed with an artefact type (MN).

Q: Have definite conventions been employed regarding the lumping and splitting of sites during recording? (BW)

When the site is one physical entity (rather than separate entities simply occupying the same physical space) they will be lumped together as one record. For example a barrow which was later used as a windmill mound would be recorded as a single record. Approaches to barrow cemeteries, on the other hand, vary. Some are covered by a single record: 'Barrow Cemetery'. Others are grouped using a parent/child system. Some outlying barrows may be independently recorded (although the extent of physical isolation required to merit this will often vary from recorder to recorder). The record has a long history and approaches have changed over the years (MN).

It has varied over time depending on staff resources too (CW).

For NMP we tend to discuss the unit of record with the HERs at the start of a project (HW).

The findings of an analysis by the EngLaid project how shown great variation between HERs on the question of lumping and splitting. Experience in Devon suggests that legacy data can also contain recording practices at variance to those currently being used (GT).

Maybe the whole of issue of lumping and splitting needs to be re-examined. Is there a role for Informing the Future of the Past too? (SM).

The key point concerning lumping/splitting is that none of the systems involved adhere to any strict system. This has to be accepted and worked with. Lumping is fine until something gets misrepresented (for example where only part of a monument is designated) (CF).

Wjhere it does matter is when you want to move data (GT),

Q: Another potentially problematic issue is that of the material in bibliographies and how this is represented in the notes field. This is not atomically organised. Have systematic conventions regarding numbering and spacing been employed here? (CF)

Forced conventions have not been applied (KB)

It might be possible to break this material down and re-integrate it using the bracketed numbers given to source references in the notes field (CF)

There are some anomalies in the source numbering in which sub-references within the free text are numbered 2a, 2b, 2c etc. These are corrected by the numbering used in the sequence field (HW). .

There is also the possibility of a mismatch within air photographic recording in that referencing can vary as to whether a specific AP or a sortie is used (HW).

Within monument records cross referencing to events will frequently involve Excavation Index records. These links have not always been consistently made (ND).

Q: Regarding the 'Associated Archives' fields: do these, in reality, actually constitute another source reference? (NB)

This area does deal with sources but is exclusively devoted to the paper and hard copy material held by EH. These fields will also include the relevant identifying numbers (HW).

Is the LBS number available? (AM)

This is now being indexed. Earlier records should have been automatically updated (MN).

HER numbering changes will not necessarily have been updated within the NRHE record. Also there are evident disparities in the way that the numbers appear within the various databases (although these may be stored differently in background files) (CF).

Q: Is there referential integrity between parent/child records? (CF)

Yes, all links are reciprocal and automated (KB).

The content of the People table is rather disorganised because of the lack of controlled entry (KB).

The Durham HER uses a validation field to indicate that a record has been checked and cleaned and is fit to be presented to enquirers. Does the NRHE have anything comparable? (NB)

No, previous data-cleaning exercises have employed fields of this type but these were subsequently dropped (MN).

Devon County Council HER (Graham Tait)

Devon CC uses the HBSMR system.

Text is held in the Monument Source table and is, therefore, linked (GT).

Q: Is the 'Sources, Additional Reading' list made up of indexed sources? (MN).

Yes (GT).

Listed Buildings form a large part of the record. Is this a potential vulnerability if designation data isn't kept up to date? (MN)

This potential does exist but most HERs will try hard to keep this material updated (GT).

Monument Type data in the Devon record is taken from look up rather than being text based (CF).

It seems evident that Devon legacy data creates individual records down to a very detailed level (eg a church organ) (CF).

Where cross references to other records are held within databases the correspondence is not always clear cut. For example 'NMR HOB UID' would not necessarily appear in a list of cross references as a standard convention. Devon HER also incorporates a Pastscape number which should usually correspond to an NMR HOB UID (CF).

OS.cards are sometimes recorded within the HER as source records (thus creating a source reference from a source reference) (GT).

Landscape – Soils – Geology fields: These may have been used more extensively before the advent of GIS. Data of this kind exists in legacy records within the Durham HER but it is no longer recorded (NB). This is also the case in Warwickshire where HLC has also replaced land use recording (BW).

The Devon HER also has a validation field (see Durham above) which flags up 'cleaned' data (GT).

Q: How standard is the HBSMR format as it is currently used by HERs across the board? (NB)

The biggest differences are to be found in the front description. Aside from this the system conforms to a pluggable, flexible model which allow tabs (for

example maritime) to be switched on or off according to the HER's requirements (CF).

The relationship between NRHE and HER records will not always be one to one. (In the case of the sample monument used for the demonstration, Devon has four records relating to the site compared to the NRHE's one) (CF).

Q: How are maritime deposits (rather than wreck sites) recorded? (DB)

These would be recorded as monuments but in a maritime location (GT).

Clarification of roles in relation to maritime data is necessary (GT).

If you could match records, the AMIE data could be held under another tab. The challenge is concordance (NB).

Somerset County Council HER (Chris Webster)

This is a simple system initially derived from Superfile and written by CW. The design philosophy was influenced by the fact that it does not have to be used by a large number of people. The system is not well documented for other users (CW).

Listed Buildings are currently indexed as a monument category (and not as a designation) (CW).

In order to ensure website functionality some data fields have to be entered (CW).

Q: Can the website do hierarchical searches? (MN)

Yes, however, the system does not use HER data directly but metadata drawn from the HER and restructured to allow its presentation (CW).

Discussion

A significant stumbling block seems to be the non-standardisation of source records (CW).

This doesn't mean that some means of matching them can't be found (BW).

The MIDAS Heritage data standard may have some relevance here (CF). Possibly also the British Library cataloguing system (MN).

The Devon Record contains numerous duplicate source records, possibly tens of thousands (GT).

Are sources actually a priority? Wouldn't matching records be a bigger issue? (DM)

It would be one of the aims of this phase of the exercise to establish whether a feasible method of satisfactorily matching of sources can be found. Perfect concordance in this area may not be possible (GT).

If all that comes out of this project is a swapping of reference numbers, this would be a really useful outcome (NB).

It may be that we need a completely different approach to making data accessible (GT).

The project is also highlighting the need to agree roles and responsibilities and protocols (GG).

The discussion touched upon some high level statistics which, it was thought, might have some relevance to the project's scoping:

- Devon HER contains around 70,000 monuments (about 12.9 per square kilometre).
- Chris Gosden's input into the TACOS initiative (suggesting that (whilst the recorded details may differ) in 40-60% of cases a record for a site will exist in both the NRHE and the relevant HER).
- The findings arrived at by trans-regional research done as part of the Englaid project.

It was concluded, however, that caution should be shown regarding anything that could be deduced from such general figures. Any overlaps would undoubtedly vary from case to case, as would the practicalities involved in addressing them.

Members of the BHUG group seem, in the main, to be anticipating a significant element of manual entry. Not everyone with a 'bespoke' database had CW's level of familiarity with its workings. This was a point which should be factored into the methodologies explored (NB).

The role of the forthcoming workshop will be to pick up some more definite principles as to what might be done and then develop methodologies accordingly (GT).

What might be explored are gradations of concordance, beginning with the matching of core data and then moving out to more complete levels of integration (DM).

Note of caution: when AMIE data has been offered previously, it has often ended up in backlogs and nothing has been done with it (MN).

The nature and perceived viability of any methodologies arrived at would inevitably depend on the level of commitment that EH was investing in the project. Has a definite decision been made that the NRHE will no longer be maintained after the proposed migration or will it be a case of national updates being forwarded to the relevant HER? (GT)

This, in turn, will also have a bearing on the nature of any replacement for the AMIE database. If this record is to receive little or no management will a replacement be seen as viable? (HW)

We need to be clear about where the data is going. If it is needed by HERs it should find a home there. Some data will, however, need to be held and maintained nationally (GG).

Is this a significant issue? Hasn't the viability of the proposed methodologies got to be gauged against a model in which both records remain available? (MN).

Also, it may not be worth doing this concordance in terms of the resource required to achieve it (MN).

A prominent aim of the project would seem to be looking at the figures for individual HERs and attempting from this to extrapolate an overview of the situation as a whole. How do the figures for undertaking migration compare with estimates for replacing AMIE? (DM)

The non-duplication of data has been advanced as one of the key aims of the Heritage Information Access Strategy. If this is the case it would seem to have a significant bearing on the ultimate decision (GT).

This is a situation which has to be viewed within the context of the development of the Heritage Information Access Strategy. By examining practicalities and methodologies the Data Supply project has the potential to shape this overarching initiative (GG).

If equivalent records continue to be enhanced by both parties after migration it will render the exercise largely pointless. If national record was to be shut down completely, on the other hand, it would greatly raise the priority and technical ambitions of the project. It would make the systematic integration of data essential (CF)

The question of interoperability has been discussed for years, together with the question of what the national record consists of. In real terms, however, it is unlikely that NMP commissions could continue without a national database (HW).

Through the process of exploration it might be concluded that resources would be better spent elsewhere and that more radical solutions will be necessary (GG).

Viewed at the 'big picture' level (making a workable system) the project is not so difficult. Complexities mainly reside in the minutiae of individual cases (CF).

Whilst it can be acknowledged as possible (with a lot of work both pre and post migration) the comparatively isolated BHUG user will tend to see the exercise as a 'big scary thing'. Support would definitely be required (NB).

Without the development of key protocols: dos and don'ts; what we do and don't want to do, there is a real risk that no-one will want to attempt the work at all (GT).
Is more guidance needed to underpin interoperability on a potentially much broader scale? There are, after all, possibly as many as 87 HER recording manuals presently being employed (SM).

If we are moving to a virtual record, the application of data standards is even more critical .This is certainly something which had to be set at the national level (CF).

Heritage Gateway highlights areas of problems with data standards (CF).

The Heritage Gateway seems to suggest that most systems are at least broadly compatible (GT). (It should also be acknowledged, though, that the Gateway has, in itself, acted as a catalyst for greater standardisation (SM)).

A range of advantages and disadvantages are evident when reviewing the concept of a virtual national database which is edited remotely (CW).

A virtual national record can be a distributed database or one managed centrally. Have 86 HERs and AMIE so 87 different recording manuals (CW).

The development of protocols has to take into consideration whether this is seen as a 'once off' concordance (and it is unlikely that anyone would ever want to do this again). Is the envisaged outcome a situation in which all parties are actually maintaining the same record? (GT)

Air photographic recorders probably create the most records overall and there is a recognised need here for greater conformity and efficiency. The current system amounts to trying to push a national 'square peg' into a varied range of local 'round holes' (HW).

The key things for NMP are efficiency, the use of national standards and methods and the ability to view and analyse the data nationally and regionally (HW).

Is this something which might undermine HER charging potential? What implications does the creation of Historic England have for the project? (HW)

The creation of Historic England and NHPP2 should be seen as opportunities to address long-standing issues and develop a new shared strategy (GG).

Will English Heritage create its own system? (NB)

EH site management systems already exist. Spatial management data will continue to be provided by Heritage Data Management according to a service level agreement. The products of survey, investigation and research will, therefore, be fed into the national record (MN).

A shared service agreement is being developed which should cover every element of this interchange (GG).

Within this scenario doesn't EH become akin to an organisation like the National Trust? Arguably, management systems like that held by the Trust do not constitute full HERs. Isn't this an opportunity to develop a 'proper' national record allowing all evidence to be linked to a single URL? (GT)

This would be misguided. Differing interpretations mean that there can be no such thing as one 'correct' record (CF).

Wouldn't this point still have validity in respect of data accessibility? (GT)

Individual data requests and the resulting data provision are nearly always unique in terms of levels of information involved (SM).

Next steps:

It will soon be possible to export AMIE data in MIDAS XML format (the result is likely to be imperfect but still usable) (KB).

A good starting point would be to view the evidence through GIS. The significant statistics would be for the overlap between records (CF).

Would the EngLaid statistics be useful in this respect? (GT)

These can be drawn upon but the project itself needs to study the various possible methodologies and their implications (GG).

It would also be useful to review the practicalities to find out how long it takes to do things. This would allow time estimates to be generally extrapolated across the project (NB).

This estimate could vary according to the amount of automation involved, both in terms of the different methodologies and in terms of the availability of technology to HERs (MR).

Actions

AMIE: MN to circulate AMIE powerpoint presentation and the link to the Internet Archaeology article on events.

Englaid: GT to check whether the trans-regional research report can be shared with the group.

Trial sampling: (by HER Officers – NB, GT, BW, CW; AMIE data provision KB, MN, MR)

- ***Select 10km grid square (ideally an area not covered by EngLaid) .***
- ***Concord HER records against AMIE records.***
    - ***Automatic check to find existing cross-references***
    - ***Manual check to establish how many records should have cross references.***
    - ***Establish number of records that only exist in one of the records.***
- ***Explore and contrast methods of undertaking concordance.***
- ***Identify helpful features.***
- ***Document time requirements.***

Vortex study of key technical options (in abstract): (by CF)

- ***Explore technical options.***
- ***Explore possibility of alternative export options.***

Timetable:

- ***AMIE data provided by 26-09-14.***
- ***Results of analyses to be available before workshop 06-11-14.***

N J Davis 18 September 2014

Heritage Information
Partnerships Supervisor
English Heritage

G Grayson 25 September 2014
Head of Heritage Data Management
English Heritage

If you require an alternative accessible version of this document (for instance in audio, Braille or large print) please contact our Customer Services Department:
Telephone: 0370 333 0607
Fax: 01793 414926
Textphone: 0800 015 0516
E-mail: customers@HistoricEngland.org.uk