MEDIN sponsored meeting: heritage vocabularies

Held at NIEA, Belfast 10 February 2015

Attendees: Ceri Binding, Hypermedia Research Unit, University of South Wales; Karl Brady, Underwater Archaeology Unit, National Monuments Service, Dublin; Philip Carlisle, Data Standards Unit, English Heritage; Peter McKeague, Data and Recording, RCAHMS; Rory McNeary, Marine Environment Division, Department of the Environment; John Murphy, NIEA: Historic Environment Division; Paul Walsh, National Monuments Service, Dublin; Kieran Westley Centre for Maritime Archaeology, University of Ulster.

Apologies: ADS, David Thomas (RCAHMW)

Background
The development of Linked Open Data for the national heritage agencies requires investment in research and development time beyond the day-to-day resources of these organisations. English Heritage, the Archaeology Data Service and the Royal Commissions on the Ancient and Historical Monuments of Scotland (RCAHMS) and Wales (RCAHMW) were fortunate to work with domain experts, Doug Tudhope as Principal Investigator with Ceri Binding as Research Fellow, at the University of South Wales on the AHRC funded SENESCHAL project to develop Linked Data resources for key vocabularies, published through the http:heritagedata.org website.

The meeting
The project identified the need overlap in the vocabularies held by the individual agencies and the need to bring in additional reference vocabularies from partners and potential partners. The MEDIN funded meeting in Belfast provided the opportunity to introduce the Linked Data approach to colleagues from the Northern Ireland Environment Agency: Heritage and Marine Divisions, the Centre for Maritime Archaeology, University of Ulster and from the National Monuments Service and Underwater Archaeology Unit in Dublin who share common seas and heritage across the Irish Sea and the North Atlantic. The purpose of the meeting was to gauge interest from participants to share and potentially align existing vocabularies to enable publication of definitive reference vocabularies and take a step towards improving interoperability across datasets.

Following an introduction to MEDIN and the aims of the Forum for Information Standards in Heritage (FISH: http://fishforum.weebly.com/) there was a presentation on work on Linked Data heritage vocabularies to date. The reasons for developing Linked Data were explained and the tools for converting vocabularies to Linked Data formats and the potential for data alignment through the use of the Levenshtein algorithm. The

algorithm uses string matching to highlight the likely hood of a match between terms in separate vocabularies greatly reducing the manual checking of terms.

There was a consensus from participants to work collaboratively and share vocabularies through a Linked Data approach particularly as British military ships in Irish waters were still owned by the UK. Participants from both Belfast and Dublin emphasised more active participation would be restricted by current financial constraints. Participants recognised the limitations of organisations to commit resources (both curatorial and technical) to implement change within internal systems but recognised the value in establishing reference vocabularies. There was also agreement to involve colleagues in the Isle of Man in any potential project.

The key vocabularies are the maritime craft terms for each organisation which describe the wrecks and documented losses in each jurisdiction and the monument type thesauri which include artificial constructions from the inter–tidal zone such as Kelp grids and Kelp kilns and submerged archaeological monuments and landscapes. English Heritage would also provide thesauri for maritime fittings and cargo.

Data provider	Existing vocabulary	New vocabulary	Data alignment
English Heritage	Maritime Craft		Y
RCAHMS	Maritime Craft		Y
RCAHMW		Maritime Craft	Y
NIEA		Maritime	Y
Underwater Archaeology Unit		Maritime	Y
Isle of Man		?	?
English Heritage	Monument type		Y
RCAHMS	Monument type		Y
RCAHMW	Monument type		Y
NIEA		Monument type	Y
Underwater Archaeology Unit		Monument type	Y
Isle of Man		?	?
English Heritage		Maritime Fittings	N
English Heritage		Maritime Cargo	N

Table 1: Potential linked Data vocabularies and candidates for data alignment

OR

Action

It was agreed that RCAHMS (as part of the MEDIN Historic Environment Data Archive Centre) would lead on putting the case to MEDIN for funding, in line with their workstream objectives, to take forward a project to develop and align data resources across our common seas. The Linked Data work would be undertaken by the University of South Wales who would need to cost out the work with Phil Carlisle from English Heritage Data Standards Unit advising on the terminologies. Partners would need to provide time to assess the results of the data alignment exercise and resolve mismatches. It is anticipated that the bulk of this work would be done by email. There would need to be a final meeting, possibly in Dublin, between partners to review the project outcomes.

If there is agreement to share our vocabularies, we need to apply to MEDIN for funding. As part of the HE DAC, I will put together the application but will require costings from USW for undertaking the work (previously estimated at about £5k) plus meeting time (possibly in Edinburgh or Douglas).

MEDIN are unlikely to fund core staff time at Heritage organisations but will support the research and development time by USW and cost for meetings (in part or full) and we can make the case for a researcher to validate the data alignment. The timescale for submitting a bid is potentially quite tight- May 2015.

What's next?

By 5 March: Peter McKeague to submit this report to MEDIN

26 March: Peter McKeague to report to MEDIN Data Archive Centre working party

1st week April: Ceri Binding, Phil Carlisle and Peter McKeague (and ADS?) to discuss with Professor Doug Tudhope, the possibility of a taking forward a funding application to MEDIN. Doug to provide guidance on time and costings required.

May (?) submit proposal to MEDIN