Do Geospatial & Heritage standards work and do they work together?

Glen Hart,

University of Nottingham

November 2014

Summary

This research compared the ability of three geospatial and heritage standards to meet the needs of a heritage organisation concluding that the standards in either isolation or combination did not fully meet the requirements and that the nature of the standards made it difficult for them to work together. The work recommends the development of micro-standards to overcome these difficulties.

KEYWORDS: Standards, Heritage, GML, CIDOC-CRM, MIDAS

1 Introduction

The efficient exchange and reuse of data is vital to the future growth of the digital economy but is still a major challenge. Standards are important elements in addressing this issue by the geospatial and heritage communities among others. It is therefore worth asking how helpful, compatible and mutually supportive standards written by different communities are. Qualitative research was conducted through studying the Royal Commission on the Ancient and Historic Monuments of Scotland (RCAHMS), and with a specific focus on data standards, rather than standards involved with delivering services and APIs. The work was funded by the Horizon Digital Economy Institute at the University of Nottingham.

2 RCAHMS and Standards

RCAHMS Identifies, surveys and analyses the historic and built environment of Scotland. It obtains information on the historic environment through its own surveys and investigative work, and from archaeological consultancies. Data is made available to professional bodies and the general public in a number of formats including ESRI Shapefile, CSV and Excel but not standards such as GML. The design of the RCAHMS Database has been heavily influenced by MIDAS, a heritage standard created in the 1990's.

RCAHMS' interests centre around three main concepts: Monument or Sites, Event and Collection. Of these the first was the main focus of the research. Monuments are the primary interest and represent things such as historic buildings, statues, gardens, standing stones, earth works, battlefields, and find spots. A monument's location is important as are dates, monument (place) names and classification. Events record monument investigations, such as a site survey using Lidar or architectural assessment. Collections are largely documents and photography concerning monuments.

RCAHMS' systems are not perfect; RCAHMS recognises the deficiencies and is working to improve them. For example monument names are inefficiently held and represented making it difficult to query them, and where a monument has multiple names there is no way to differentiate or classify them as

preferred, official, historic etc. Dates are not consistently implemented, there is no mandated way of representing location in terms of geometry, and there is also only a fairly limited way to reference one monument as being part of another. A greater adoption of standards could help to resolve the current deficiencies and promote greater data interoperability. Three standards were examined as being appropriate to RCAHMS: Geography Mark-Up Language (GML), CIDOC-Common Reference Model (CRM) and MIDAS Heritage.

GML is an important OGC standard for geospatial data transfer and modelling. GML is implemented as XML and has schemas that support topics such as geometry, topology and temporal models. The geometry and topology models are hugely influential having been used to implement spatial extensions to (o)rdbms' and triplestores including their query languages (SQL and SPARQL). GML is rather complex and non-governmental users often use popular less formal standards such as GeoJSON.

Produced by the CIDOC Documentation Standards Working Group within the International Council of Museums, CRM is a domain standard that promotes the exchange of heritage information. CRM's scope "can be summarised in simple terms as the curated knowledge of museums." (Le Boeuf et al, 2013) and "is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to." Like GML, CRM provides a language or vocabulary to enable a user to model their system, rather than be very specific about absolute structure. CRM is a complex standard that has most recently been expressed as an OWL ontology. The high-level abstract structure and depth of hierarchy make for a standard with "sometimes bewildering complexity" (Carlisle et al, 2014). It does nonetheless provide a good description of the informational needs of the heritage community.

MIDAS is a heritage standard developed by English Heritage in conjunction with many other UK heritage organisations. Published in 1998 it was revised in 2012 as MIDAS Heritage. MIDAS has been influenced by CRM. MIDAS does not tie itself to any particular implementation technology or philosophy. MIDAS is categorised into a number of Information Themes, which in turn are broken down into Information Groups. The main themes identify the primary interests in monuments and artefacts, what has happened to them and how they are documented. The supporting themes then provide the where, when and by whom. This is very similar to CRM, but much clearer and easier to digest.

3 Discussion

MIDAS provides the best description of the Heritage community's data needs. CRM provides greater detail in terms of structure. CRM and MIDAS lack specificity in some areas such as positional information, place names and time. GML provides well-defined geometry types and a formal way to specify coordinate reference systems. In certain areas all the standards lack specificity, for example no detail is given as to how digital identifiers, place names or classification should be represented. In all cases what is missing is not so much a mandate that says that this classification system should be used, or these approved place names are the ones, but structurally how they should be represent and the semantics associated with those structures.

Ideally using these standards together as well as supplementing them with the missing detail would provide a good solution. However the heritage standards were not designed to work with GML and vice versa. MIDAS and CRM are essentially compatible and it is possible to map MIDAS onto CRM. GML and CRM are competing implementations, both designed to provide "the data model". Most sensibly would be for CRM to provide the domain framework and to utilise those aspects of GML where CRM is deficient such as geometry. This is more or less what Arches, an open source solution

for managing heritage data, does. It implements the CRM model but also supports GML as an import and export format (along with a number of others) and represents geometry using the OGC standard (Well-known text). Arches hides unwanted complexity rather than address the problem at source and a considerable amount of the GML standard is effectively redundant. The simple aspects of GML will be reusable but other aspects are not as transferable. The way GML defines a Period (such as the Ice Age) is too simple for heritage organisations where a period is not simply a tag associated with a time range, but a more complex interaction between culture, geography and time. The temporal model is more complex than currently required by RCAHMS and where the complexity is required by others does not always meet expectations resulting in the model being ignored (Parcero-Oubiña, 2012) (Plumejeaud, 2010) or modified (Siyuan et al, 2007) (Quak & Vries, 2005).

These observations point to improvements to the way standards are developed. There is a need to identify and standardise the way that some basic things such as place names and classifications are exchanged. Secondly, we should avoid wherever possible constructing large complex standards. This will be harder to do for domain standards where the domains themselves are necessarily complex; but here MIDAS addresses this complexity through sensible modularisation, conversley CRM supports a complex and unnecessarily deep hierarchy. GML is attempting to provide supporting 'horizontal' standards to 'vertical' domains such as heritage, but does so in an all-encompassing package. It might be better if a trend in ontology development were followed where there is a move away from large 'upper' ontologies such as DULCE and SUMO in favour of lightweight micro-ontologies (Hart & Dolbear, 2013) which could be mirror by moving away from large horizontal standards in favour of smaller micro-standards. Hence the well-defined and popular aspects of GML such as geometry could be disaggregated into compact micro-standards that can be easily adopted by domain standards. Lastly, it would be beneficial if such standards are not tied to particular technologies such as XML but instead have an abstract definition accompanied by several technology profiles.

4 References

Carlisle, P. K., I. Avramides, A. Dalgity, and D. Myers. 2014, "The Arches Heritage Inventory and Management System: A Standards-Based Approach to the Management of Cultural Heritage Information." Paper presented at the CIDOC Conference: Access and Understanding – Networking in the Digital Era, Dresden, Germany.

Hart, G., Dolbear. C. 2013, Linked Data: A Geographic Perspective, CRC Press, ISBN-10: 1439869952, ISBN-13: 978-1439869956

Le Boeuf P., Doerr M., Ore C. E., Stead S. (Eds), 2013, Definition of the CIDOC Conceptual Reference Model, Version 5.0.1(draft), http://cidoc-crm.org/official_release_cidoc.html

Lee E. (Ed.), MIDAS Heritage, 2012, The UK Historic Environment Data Standard, V1.1,http://www.english-heritage.org.uk/publications/midas-heritage/

Parcero-Oubiña C, Fábrega-Álvarez P, Vicent-García M, Uriarte-González A, Fraguas-Bravo A, del-Bosque-González I, Fernández-Freire C, Pérez-Asensio E, 2012, Conceptual basis for a cultural heritage data model for INSPIRE, Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, ISBN: 978-90-816960-0-5

Plumejeaud C, Mathian H, Gensel J, Grasland C, 2010, Spatio-temporal analysis of territorial changes from a multi-scale perspective, International Journal of Geographical Information Science, 25:10, 1597-1612

Quak, CW & Vries, ME de. Topological and temporal modelling in GML. In J Drummond (Ed.), 2005, Proceedings of the topology and spatial databases workshop (pp. 1-8). Glasgow: University of Glasgow

Siyuan F, Griffiths A, & Paton N, 2007, GML for Representing Data from Spatio-Historical Databases: A Case Study, Transactions in GIS, 11(2): 233–253