

Chapter 15

The Development of Thesauri by English Heritage

Christopher Tweed

15.1 Context

The computer-resident National Monuments Records (NMR) thesauri developed by English Heritage evolved from a paper-based list developed by the Royal Commission on Historical Monuments of England and have since been expanded by various means, such as one-off projects as well as continuous adjustment following their use. English Heritage (EH) has been involved in the development of several important thesauri that serve various purposes. The remainder of this case study will focus on the largest of the English Heritage thesauri, the National Monument Records Monument Type thesaurus.

15.2 Purpose and Aims of the Ontology

The main aim in developing the thesaurus is to standardise the terms archaeologists use to refer to monuments. This is intended to guarantee consistent use of terms within a number of archaeologically related disciplines. A secondary aim is to use the thesaurus to classify buildings and other structures that are listed under the English conservation legislation. This process operates in two directions: the existing monument types are applied to instances that have been erroneously classified or if the type is missing, but the thesaurus is also updated to accommodate building types that are missing.

C. Tweed (✉)

BRE Centre for Sustainable Design of the Built Environment, Welsh School of Architecture, Cardiff University, UK.

e-mail: TweedAC@cardiff.ac.uk

15.3 Scope

The thesaurus contains definitions of monument types including infrastructure (bridges, etc.), sites, and buildings. The thesaurus contains types that are found in England and restricts the terms to those used in England. The word ‘rath’ (used in Ireland to identify an ancient fort) is not found in the thesaurus. The thesaurus can contain a term for any period of history, though much of its contents could be described as archaic and historical.

15.4 Actors

Development of these thesauri involves many stakeholder organisations and individuals. English Heritage led the development but others participated in suggesting and approving candidate terms as well as revisions.

EH does not develop the software in which the thesauri are embedded. This is carried out by a company called *exeGesIS* that sells a HBSMR (Historic Buildings, Sites and Monuments Record) database which encapsulates the English Heritage definitions. The database is tied into a GIS and is marketed as a tool for the management of Historic Environment Records (HERs).

Other stakeholders are those who make use of the thesauri in English Heritage and in other organisations that have an interesting built heritage, including the National Trust, conservation bodies and local authorities.

15.5 Methods of Development

The thesauri in their present form were developed following the conversion of the Royal Commission on Historical Monuments of England from paper to computer. One part of that project involved examining lists of listed buildings and checking individual entries against types available in the thesaurus. This first project produced 200–300 new candidate terms for the thesaurus, often for buildings that had not been classified previously. New candidate terms emerge as scholars discover new types in the course of their research. The rate is roughly two or three per month, but can be more frequent than that.

Occasionally, specific projects will be carried out that can alter the overall shape of the main thesaurus, such as a study of the defence of Britain, which generated many new terms. These were considered to be too specialised and so were not added to the main thesaurus. Instead, a separate micro-thesaurus was created in which the top level terms, more or less map on to the bottom level terms in the main thesaurus. This means that if it was considered necessary the two thesauri could be merged fairly easily in the future.

English Heritage developed its own tools for creating the initial thesaurus. In the current development of a multilingual thesaurus, English Heritage is using Microsoft Excel with XML and XSLT, having tried a range of bespoke thesaurus building software packages.

The thesaurus supports three different relationships that are applied in the following order:

- equivalence—two or more terms are linked because they are considered as equivalent, one of which will be designated ‘preferred,’ the others classed as ‘non-preferred’;
- hierarchical—preferred terms are arranged in hierarchies following a ‘type-of’ relationship, such that, for example, farmhouse and shepherd’s hut appear at the same level in the hierarchy and are both immediately below agricultural dwelling;
- associative—related terms can be associated even though they are not otherwise linked, thus making it easier for someone to find similar (but not equivalent) terms.

It is worth noting that the thesaurus has no top terms as items are grouped under classes that are not part of the thesaurus. For example, under the class name ‘domestic’ it is possible to find the terms: backyard, cooking pit, kennels, etc. Clearly, the term ‘domestic’ is not part of the thesaurus. The thesaurus is poly-hierarchic in that it will allow terms to appear under more than one class. So, for example, ‘castle’ appears under the class name ‘domestic’ as well as under the class name ‘defence.’ Future development is moving towards an ontological basis in which terms can appear in several places, but concepts can only appear once. In the previous example, ‘castle’ could continue to appear in several places in the thesaurus but the concept of ‘a fortified building with towers’ could only appear in one place.

15.6 Content of the Ontology

The complete thesaurus contains more than 6,500 terms and can be viewed on the English Heritage website at:

http://thesaurus.englishheritage.org.uk/thesaurus.asp?thes_no=1

A snippet is shown in Fig. 15.1 below.

15.7 Usability

English Heritage publishes its thesaurus on the Web as a freely available resource. The thesaurus also forms a central component of a larger database system developed by *exeGesIS*, as shown in Fig. 15.2. The thesaurus is now informing the

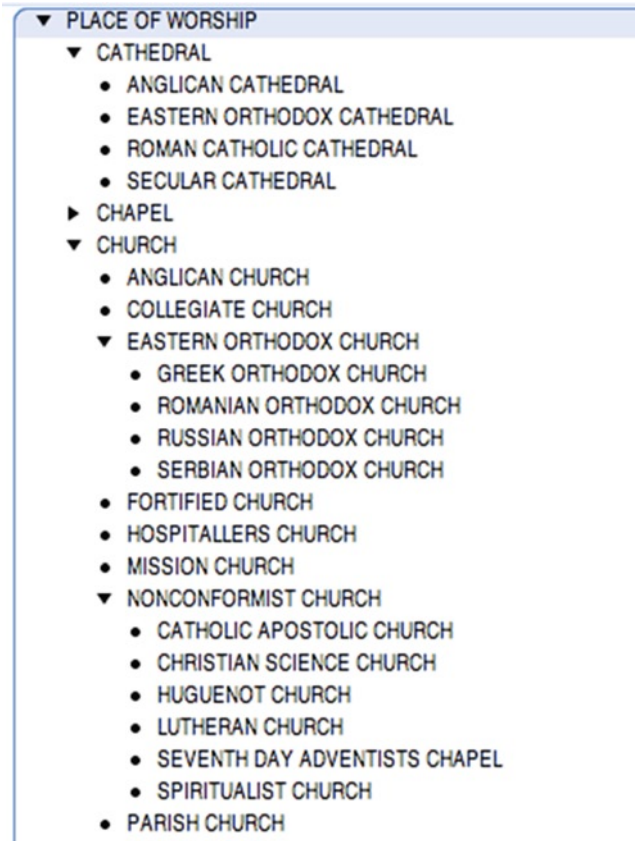


Fig. 15.1 A small part of the English Heritage NMR thesaurus

development of a multi-lingual thesaurus through the HEREIN project, and it is intended this will be made available via the web when it is finished.

End-users can modify the thesaurus once they have signed the licence agreement, though they are not allowed sell it on as a new product.

15.8 Benefits

The thesaurus offers benefits to those working in the area of conservation and archaeology as it provides a common reference point that allows a wide range and large number of organisations to remain consistent in the terms they use.

The mechanisms controlling the addition of new terms also seems to be sufficiently flexible and fluid to allow the thesaurus to evolve as new information comes to light.

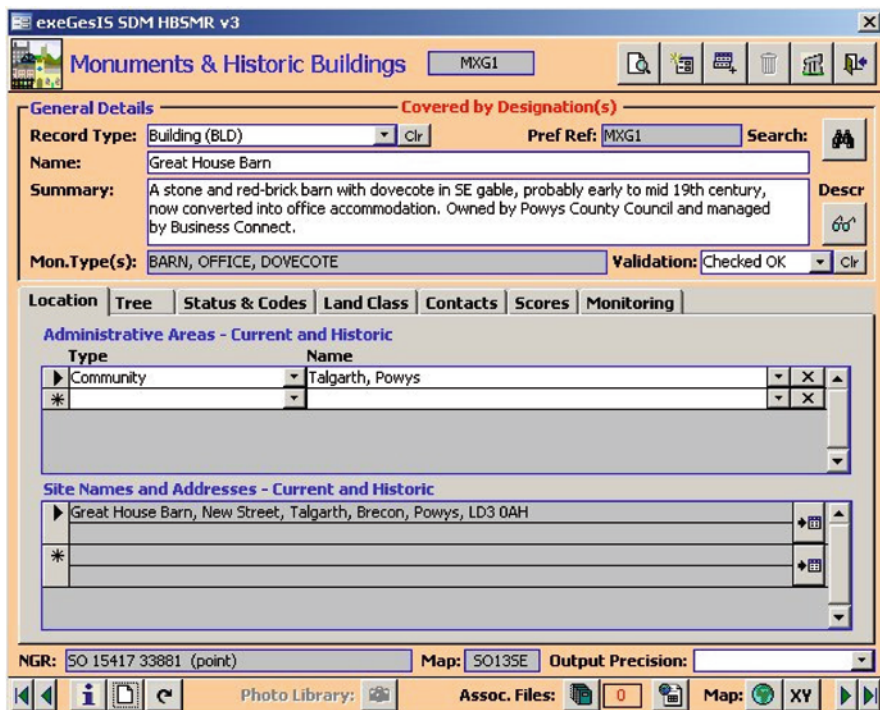


Fig. 15.2 Screenshot of the *exeGesIS* software, which uses the English Heritage thesaurus

15.9 Lessons Learned

The long period of development and refinement has allowed English Heritage to accumulate significant experience in dealing with the problems of surrounding the use of thesauri both in-house and by others. The main points are summarised here.

Ontology development always starts from some previous position, which can rarely be ignored, and must accommodate the legacy of pre-existing conceptual structures. It is rarely possible to start afresh.

Even when an ontology starts from a ‘clean slate’ it will invariably get “messy” over time as it gets extended and revised by its users.

Ontologies seem to work best when the user has a choice of how much she wishes to incorporate the terms. The English Heritage thesaurus works well because its users can choose how much of it they want to include.

There are always implementation issues following a change to the thesaurus. The software will need to allow for disambiguation changes, such as when the term ‘axe’ is divided into ‘axe-tool’ and ‘axe-weapon.’