
Changing Boundaries: Gazetteers, Information Retrieval and Data Browsing

This paper examines the role which historical gazetteers can play in web-based catalogues and data delivery systems. A gazetteer is a list of geographic names, which includes locational and other descriptive information. In this paper, the term 'historical gazetteers' is used specifically to describe gazetteers that incorporate both historical and modern geographical perspectives. In order to handle changed and changing geographical boundaries these gazetteers need to hold a wide range of information about geographic names, units, and hierarchies. This paper explains why gazetteers of this type are crucial for effective information retrieval and data browsing. In particular, it uses the History Data Service as a case study to describe how gazetteers of this type can be used to improve access to data via web-based catalogues and data delivery systems. This paper does not aim to describe the actual process of constructing and populating gazetteers (see Harper 1997, Hill *et al.* 1999, Moss *et al.* 1998).

The History Data Service (<http://hds.essex.ac.uk>) is funded by the Joint Information Systems Committee (<http://www.jisc.ac.uk>) of the UK Higher Education Funding Councils to collect, preserve, and encourage the re-use of digital resources which result from or support historical research and teaching. The History Data Service is part of the UK Data Archive and is the Arts and Humanities Data Service (<http://ahds.ac.uk/>) service provider for the historical disciplines.

The History Data Service collection covers a wide range of historical topics, and brings together over 450 separate data collections transcribed or compiled from original sources. The data collections cover a time period from the late tenth century to the mid-twentieth century, and the vast majority of data collections are either explicitly or implicitly geographically referenced. It is for this reason that the History Data Service is interested in developing and using gazetteers.

Explicitly and implicitly geographically referenced data correspond to a maze of complex geographies, which include administrative, electoral, census, and ecclesiastical geographies. These geographies are composed of a multiplicity of geographical unit types, which include

*by Cressida Chappell**

amongst many others counties, wards, registration districts, and parishes. Because of this complexity, gazetteers are crucial for effective information retrieval and data browsing. This holds true both in the context of an historical service provider like the History Data Service, and in the context of the wider social sciences and humanities community.

Gazetteers are needed to make sense of this maze of complex geographies for three main reasons. Firstly many geographic names have a number of variant forms; secondly there are many incompatibilities between different geographies which means that boundaries do not align; and thirdly geographic names, units and hierarchies have changed in the past, and will continue to change. These problems are greatest with historical data, which are often associated with geographic names that have changed, or with geographical units that no longer exist, or with geographical units whose boundaries have changed significantly. It hardly needs saying that the disparity between modern and historical geographies increases with time.

Gazetteers improve information retrieval and data browsing by standardising geographic names and providing a controlled vocabulary of current and historical names within a system of preferred and non-preferred names. By linking disparate and changing geographies, gazetteers can help to integrate geographically referenced data collections, and deal with some of the incompatibilities when boundaries do not align. For example, gazetteers can make it easier to construct time series and other comparative data series by helping to identify those geographical units which, to a greater or lesser extent, correspond in different geographies.

If gazetteers are to be used to improve information retrieval and data browsing, it is essential that we understand the needs and requirements of users. The History Data Service has an active and ongoing policy of consulting with actual and potential users, and we have established that many users from the historical community require web-based catalogues and data delivery systems which will allow them to perform sophisticated geographical searches in a fairly automated manner. Users would like to be able to search

for data that cover a given place at a sufficient level of detail. For example, a user searching for the county of Essex would like to recover not only data that are indexed by the geographic name Essex, but also data collections that contain Essex county-level data but which are indexed by a higher level geographical unit such as England. They might also wish to extend the search to include data that are indexed by geographical units within Essex. Users would also like to be able to search for any data that can be analysed at the level of a specified geographical unit. It is self-evident that a reasonably complex gazetteer, which holds information about geographical units and hierarchies, would be required if these types of geographical searches were to be supported.

The History Data Service is working to improve and enhance access to its collection and a comprehensive UK historical gazetteer will be central to this work. Historical gazetteers are attracting an increasing amount of interest from data providers, research projects, and traditional archives. In consequence, the History Data Service would like to develop a comprehensive UK historical gazetteer in collaboration with other services and projects.

The History Data Service would use a comprehensive UK historical gazetteer both in web-based catalogues and data delivery services. It would use the gazetteer in web-based catalogues to support the types of geographical searches that users would like to be able to perform. Information about the History Data Service collection is made available through three different catalogues, the UK Data Archive's information retrieval system BIRON, the CESSDA Integrated Data Catalogue and the Arts and Humanities Data Service Gateway; however, of these only BIRON even adequately supports geographical searches.

In BIRON geographical searching is facilitated by the geographical hierarchies in the Humanities and Social Science Electronic Thesaurus, HASSET (Data Archive, 1998). The geographical hierarchies in HASSET have been built up over time by the UK Data Archive and the History Data Service, but they are not by any means comprehensive; the historical hierarchies in particular have been developed only as when they have been needed. The geographical hierarchies in HASSET handle changing geographical boundaries by including geographical units in multiple hierarchies where necessary. The UK Data Archive and the History Data Service have increasingly recognised that the geographical hierarchies in HASSET cannot fully support the types of geographical searches that users would like to be able to perform, and that in consequence a more complex and comprehensive UK historical gazetteer is needed.

The History Data Service would also use a comprehensive UK historical gazetteer to help users to browse a web-based tree-structure, which will provide users with an alternative

means of accessing information data. This will allow users to adopt a drill-down approach to locating data in addition to the more sophisticated geographical searching offered by web-based catalogues.

In web-based data delivery services the History Data Service would use a comprehensive UK historical gazetteer to support geographical data subsetting. A geographical subsetting service has been developed for a large collection of nineteenth and twentieth century statistics assembled by Humphrey Southall as part of the Great Britain Historical GIS Programme (Southall and Gregory, 1998). The Great Britain Historical Database Online (History Data Service, 1998) allows users to search across 30 tables simultaneously to retrieve a geographical subset. Users can select which variables are included in the subset, and they can also access online documentation. Because the data collection included all the necessary gazetteers it was easier to develop a geographical subsetting service as part of the Great Britain Historical Database Online. However, a comprehensive UK historical gazetteer is essential if the History Data Service is to extend this type of service to a wide range of other geographically referenced data.

The History Data Service would also like to use a comprehensive UK historical gazetteer in web-based data delivery services to provide integrated access to historical data and appropriate digitised boundary data, which users could then utilise in a GIS. The History Data Service and the UKBorders service, located at the Edinburgh Data Library, have been discussing the possibility of developing a joint interface which would provide integrated access to digitised boundary data held by UKBorders and attribute data held by the History Data Service (such as the Great Britain Historical Database Online). It hardly needs saying that it would not be possible to develop this type of service without a fairly comprehensive UK historical gazetteer.

The History Data Service is confident that a comprehensive UK historical gazetteer can be developed in collaboration with other services and projects. We believe that it will enable us to respond to user needs and develop web-based catalogues and data delivery services which allow users to perform sophisticated geographical searches, and we believe that its use will result in improved information retrieval and data browsing.

References

Data Archive, 1998. Humanities and Social Science Electronic Thesaurus: Version 2. [online] <http://biron.essex.ac.uk/services/zhasset.html> (14 May 1999).

Harping, P., 1997. *The Limits of the World: Theoretical and Practical Issues in the Construction of the Getty Thesaurus of Geographic Names*. ICHIM 97/EVA 97, September 1997, Paris, France.

Hill, L., Frew, J. and Zheng, Q., 1999. Geographic Names: The Implementation of a Gazetteer in a Georeferenced Digital Library, *D-Lib Magazine*, 5, 1. [online] <http://www.dlib.org/dlib/january99/hill/01hill.html> (14 May 1999).

History Data Service, 1998. Great Britain Historical Database Online. [online] <http://hds.essex.ac.uk/gbh.stm> (14 May 1999).

Moss, A., Jung, E. and Petch, J.R., 1998. The Construction of WWW-Based Gazetteers Using Thesaurus Techniques. Proceedings of International Symposium on Spatial Data Handling (SDH), July 1998, Vancouver, Canada.

Southall, H. and Gregory, I., 1998. Historical GIS Project Home Page. [online] <http://www.geog.qmw.ac.uk/gbhgis/> (14 May 1999).

* Paper presented at the IASSIST Conference, May 19, 1999, Ryerson Polytechnic University, Toronto, Ontario..
Cressida Chappell, History Data Service, UK Data Archive, University of Essex