

# **geoXwalk – A Gazetteer Server and Service for UK Academia**

James Reid

GeoServices Delivery Team, EDINA, Edinburgh University Data Library,  
George Square, Edinburgh, EH8 9LJ, Scotland,

Tel: +44 (0)131 651 1383;

Fax: +44 (0) 131 650 3308;

Email:james.reid@ed.ac.uk

## **Biography**

UKBORDERS Service Coordinator, Project Manager geoXwalk, Business Development Officer and Senior Geoservices Officer at EDINA. Research interests include GIS, digital gazetteers and spatial databases.

## **Introduction**

This paper will summarise work undertaken on behalf of the UK academic community to evaluate and develop a digital gazetteer server and service which will underpin geographic searching within the UK distributed academic information network. It will outline the context and the problem domain, the rationale behind the project and reports on the issues investigated and findings to date. Lastly, it poses some unresolved questions requiring further research and speculates on possible future directions.

## **The Context**

The Joint Information Systems Committee (JISC) is a strategic advisory committee working on behalf of the funding bodies for further and higher education (FE and HE) in England, Scotland, Wales and Northern Ireland. The JISC promotes the innovative application and use of information systems and information technology in FE and HE across the UK by providing vision and leadership and funding the network infrastructure, Information and Communications Technology (ICT) and information services, development projects and high quality materials for education. Fig 1 provides an architectural summary of how JISC manage these areas within what is referred to as the JISC 'Information Environment' (JISC IE). The JISC IE provides access to heterogeneous resources to academia, ranging from bibliographic, multimedia and geospatial data and associated materials.

The geoXwalk project was conceived as a development project to build a shared service which would service the JISC IE by providing a mechanism for geographic searching of information resources. This would complement the traditional key term and author type searches that have been supported. 'Geo-enabling' other JISC services would provide a powerful mechanism to support resource discovery and utilisation within the distributed IE.

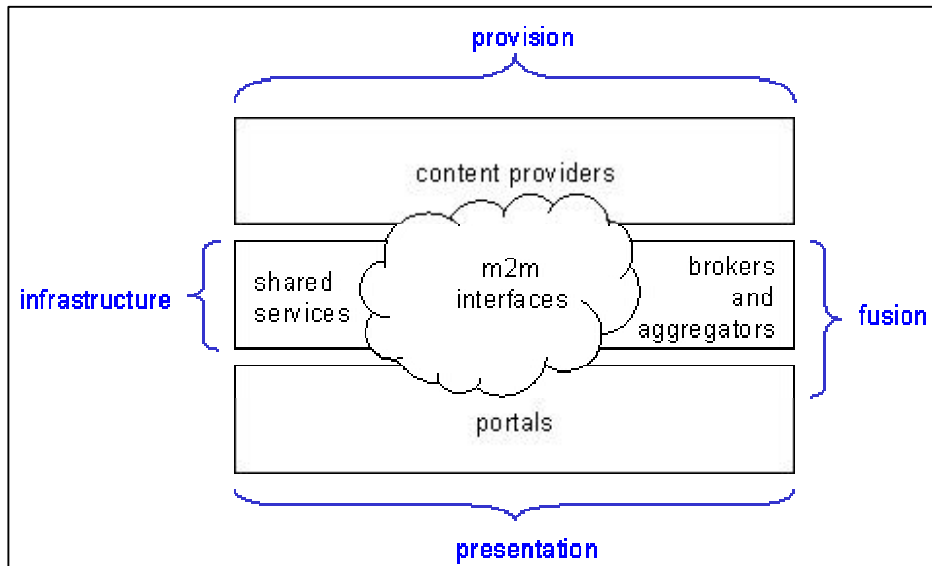


Fig 1. The general architecture of the JISC Information Environment.  
Source: Information Environment: Development Strategy 2001-2005 (Draft).

## Problem Domain

Geographic searching is a powerful information retrieval tool. Most information resources pertain to specific geographic areas and are either explicitly or implicitly geo-referenced. The UK's National Geospatial Data Framework (NGDF) estimates that as much as eighty per cent of the information collected in the UK today is geo-referenced in some form. Geography is frequently used as a search parameter, and there is an increasing demand from users, data services, archives, libraries, and museums for more powerful geographic searching. However, there are serious obstacles to meeting this demand.

Geographic searching is often restricted because geographic metadata creation is excessively resource intensive. Accordingly, many information resources have no geographic metadata, and where it exists, it usually only extends to geographic names. Search strategies based on geographic name alone are very limited (although they are a critical and often the only access point). An alternative is to geo-reference the resources using a spatial referencing system such as latitude and longitude or, in the UK, the Ordnance Survey National Grids (of Great Britain and Northern Ireland).

The existence of a maze of current and historical geographies has created a situation where there is considerable variation in the spatial units and spatial coding schemes used in geographic metadata. Many geographic names have a number of variant forms, the boundaries in different geographies do not align, and names, units and hierarchies have changed in the past, and will continue to change. In 1990, the UK Data Archive (University of Essex) identified ninety different types of spatial units and spatial coding schemes in use in their collection and more have since been added. By way of illustration, the Resource Discovery Network (RDN) ResourceFinder (<http://www.rdn.ac.uk/>) does not currently have a specific geographic search function, instead it simply searches using a text based mechanism, thus to find information

referring to a particular place the place must be referred to by name in the description field of the resource's accompanying metadata. Using geoXwalk, the placenames in the metadata can be turned into other geographical identifiers enabling searching by e.g. minimum bounding rectangle, postcode, county.

Clearly, no single system of spatial units and coding will suit all purposes as people conceptualise geographic space in different ways and different servers deploy differing geographic naming schemes. Ideally, users should not be forced to have to explicitly convert from one 'world' view to another. However, it is impractical for most service providers to support more than a few geo-referencing schemes, or develop the means to convert from one to another. A comprehensive gazetteer linking a controlled vocabulary of current and historical geographic names to a standard spatial coding scheme (such as latitude and longitude and/or the Ordnance Survey National Grid(s)) would be necessary to perform these translations (or 'cross-walks' – hence the name) and would therefore provide the capacity to be, what might be referred to as 'geographically agnostic' – see Fig 2. This is what the geoXwalk project aims to deliver.

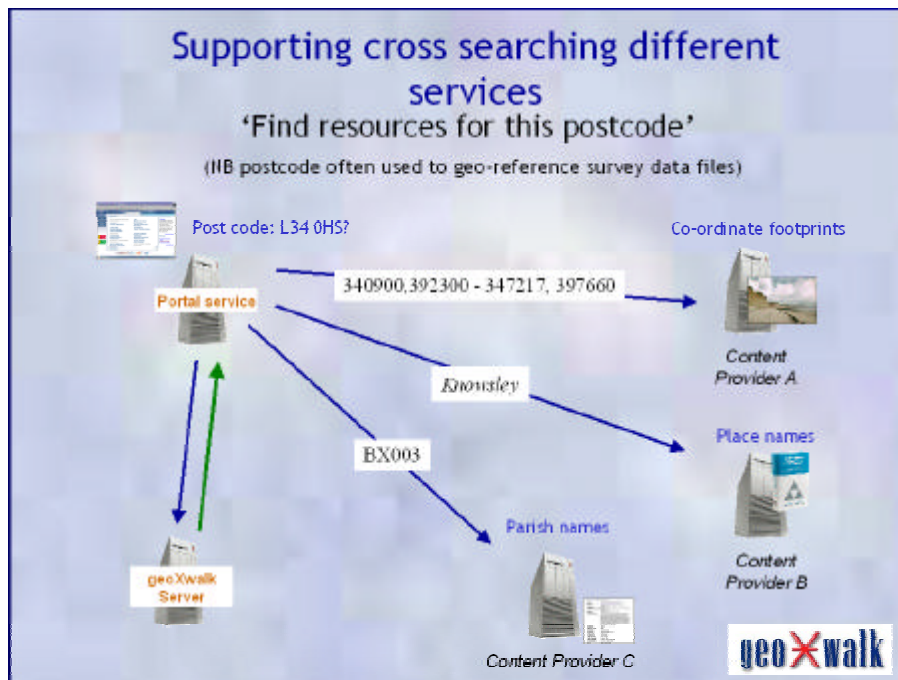


Fig 2. Example case of geoXwalk to 'cross-walk' different geographies.

### **Background And Rationale Behind Geoxwalk**

geoXwalk was funded by the JISC as a two Phase development project, the principle aim of which was to assess the feasibility of developing and providing an online, fast, scaleable and extensible British and Irish gazetteer service, which would play a crucial role in supporting geographic searching in the JISC IE. The project was a joint one between EDINA, Data Library, University of Edinburgh, and the History Data Service, Data Archive, University of Essex.

The general aim of the project was to investigate the practicability of a gazetteer service, (a network-addressable middle-ware service), implementing open protocols, specifically the Alexandria Digital Library Gazetteer Protocol (ADL 1999, Janée and Hill 2001) and the Open GIS Consortium's (OGC) Filter Encoding Specification (OGC 2001) and the Z39.50 protocol to support geographic searching for other information services within the JISC IE. Firstly, by supporting geographic searching and secondly, by assisting in the geographic indexing of information resources. As the project has progressed it has become clear that there is also a requirement for it to act as a general reference source about places and features in the UK and Ireland. Phase I was conducted as a scoping study to determine the feasibility and the requirements for such a service while Phase II which commenced in June 2002 aims to develop an actual working demonstrator geo-spatial gazetteer service suitable for extension to full service within the JISC IE.

The full paper will cover aspects of :

- (i) the technical implementation of the gazetteer server and it's architecture, issues related to data quality and management ,
- (ii) tools developed to support the semi-automatic geographical indexing of non-indexed resources (a *geoparser*),
- (iii) the major challenges identified to date and possible solutions, and
- (iv) concluding remarks on future developments.

## References

JISC 2001. Joint Information Systems Committee, *Information Environment: Development Strategy 2001-2005*.

ADL, 1999. *Alexandria Digital Library Project*. [online] <http://www.alexandria.ucsb.edu/> (01 October 1999) and ADL. *Alexandria Digital Library Gazetteer Development Information*. [online] <http://www.alexandria.ucsb.edu/gazetteer/> (01 October 1999).

Janée G. and Hill, L., (2001), The ADL Gazetteer Protocol, Version 1. Available online: < <http://alexandria.sdc.ucsb.edu/~gjanee/gazetteer/specification.html> >

OGC, 1999. Open GIS Consortium Inc., Filter Encoding Implementation Specification Version 1.0.0, (19 September 2001). Available online at: <<http://www.opengis.org/techno/specs/02-059.pdf>>

geoXwalk Phase I documentation available at: [www.geoXwalk.ac.uk](http://www.geoXwalk.ac.uk).