

GeoCrossWalk

Briefing Paper for Shared Infrastructure Services Workshop, 2007-06-26
James Reid, EDINA, JISC National Data Centre

Aim

GeoCrossWalk is a middleware, shared service providing a range of services for support of geographic referencing and searching within the JISC Information Environment (IE). It aims to assist other applications, such as portals, virtual learning applications or research services, to discover and exploit materials that have a spatial reference (e.g. placenames, postcodes). The objective is to make access to resources by searching on 'where' as transparent as searching by 'what', 'who' and 'when'. At its heart is an enhanced gazetteer built from pre existing licensed JISC data sources (JISC OS/Digimap collections). It is referred to as a terminology service within the IE technical architecture, and is a 'Service in Development' project currently funded until April 2008.

The rationale behind the GeoCrossWalk project is that there is currently no unified entry point to assist in geographic searching within the existing academic network as each information provider/service adopts different geographic coding conventions (some use postcodes, others placenames, some grid references etc.). GeoCrossWalk is designed to make geographic searching transparent by 'crosswalking' these different geographies.

GeoCrossWalk is more than just a simple lookup facility however as every geographic feature stored in the gazetteer has its detailed coordinate geometry stored with it (i.e. a city would be stored as a polygonal footprint (coordinates), a river as a linear footprint etc.). Holding the geometry as an integral attribute of the feature enables complex spatial searching based on relationships *between* features e.g. is feature A within a distance of feature B?; what features are contained within feature C?; what features does feature D intersect? and so on.

The ability to derive the relationships between features implicitly by geometric computation is significant and provides more accurate results than can be ascertained by simple lookups based on hierarchical thesauri methods, as is traditional in gazetteers. Furthermore, geography in the UK is very complex and geographic boundaries in particular do not always nest, for example, postcode geography does not nest with census geography.

GeoCrossWalk obviates the problem of variable geographic naming by coding geographic features based on a **persistent** and **consistent** coding convention - national grid references. Again, by use of the implicit relationships that can be inferred from their geometries, it is possible to 'crosswalk' place names to postcodes ...or electoral wards... or health authorities etc...

GeoCrossWalk Services

GeoCrossWalk has several different facets:

A gazetteer database 'GeoCrossWalk' – a detailed database of geographical features, names and their associated geometrical representations. The database is a value added hybrid sourced from a range of Ordnance Survey datasets.

A middleware layer – an access layer that allows other services to query the gazetteer

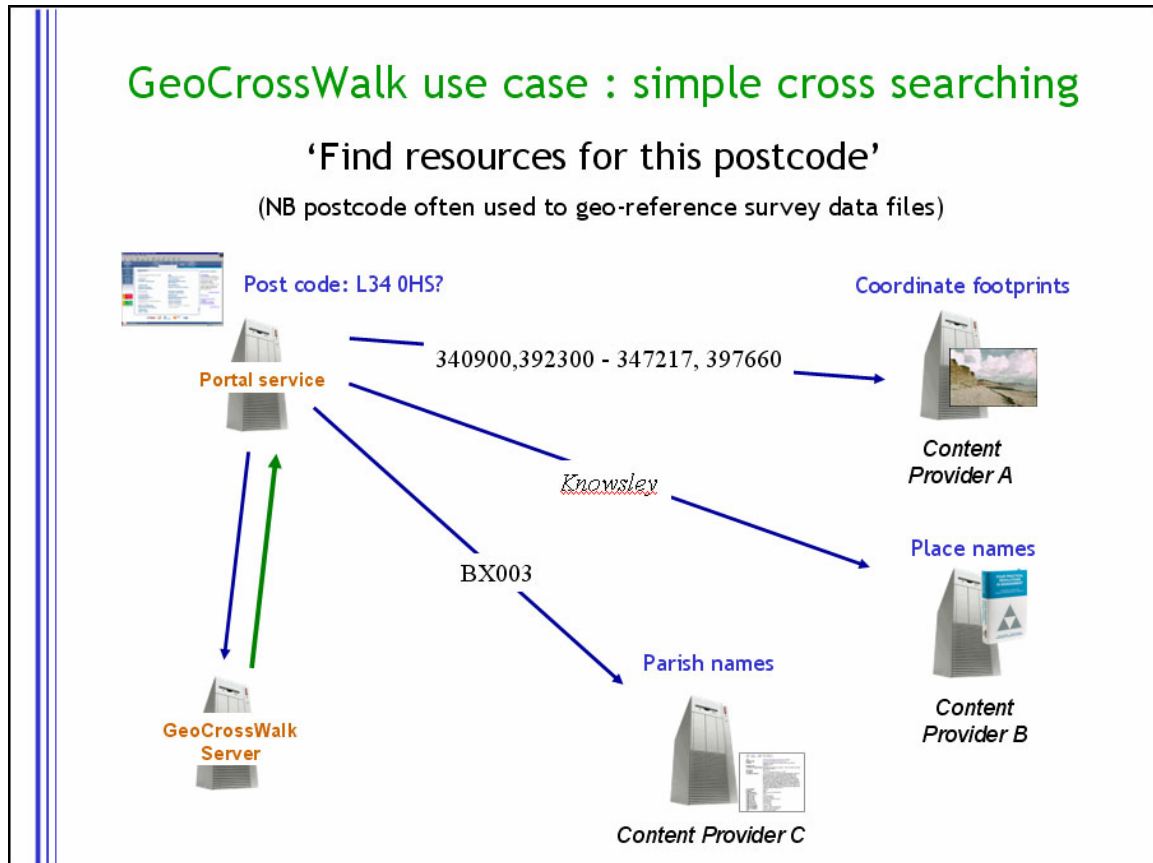
A geoparser – a natural language entity recognition service that can parse resources (text, html,xml) for geographic placenames and relate these to the gazetteer (the geoparser can use different gazetteers if available but defaults to GeoCrossWalk). The geoparser provides a mean to translate implicit geographically referenced resources into explicitly georeferenced ones.

Data Dissemination

GeoCrossWalk returns the results of queries in XML format conformant to a *de facto* gazetteer protocol (the Alexandria Digital Library [<http://www.alexandria.ucsb.edu/gazetteer/>]). Work is currently underway to provide lightweight REST based interfaces.

Using GeoCrossWalk

The over-arching use expected of GeoCrossWalk is shown in the figure below. A consumer service interposes GeoCrossWalk to resolve geographical terminology and native services can be searched using the resource appropriate terminology.



Alternatively, a service might exploit the geoparser (and implicitly the gazetteer) to georeference resources e.g. to add geographical coordinates to a resource deposited in a repository.

Current Users

GeocrossWalk is currently used by the geospatial portal Go-Geo! and by the AHDs ArchSearch service (for postcode searching).

www.gogeo.ac.uk

<http://ads.ahds.ac.uk/catalogue/search/basic.cfm>