

## **Position paper: M. Freeston**

**Digital Gazetteer Workshop    Santa Barbara, CA    7-9 Dec 2006**

As a computer scientist with a long-held interest in database technologies, my interest in gazetteers stems from my work over the past twenty years on spatial database systems and georeferenced digital libraries. – notably the Alexandria digital library. For much of this time, and certainly the past ten years, it has been a puzzle to me why the potential power of search by georeference has not been a more widely recognized and exploited paradigm.

In particular, it is at least at first sight surprising that neither the academic nor commercial database research community has so far developed adequate support for this paradigm. Within a database management system (DBMS), it is still hard if not impossible to represent the varying degrees of accuracy of georeferenced information, and there is certainly no direct support for an ellipsoidal frame of reference. Geospatial indexing and querying in database systems remains poor. The relative lack of commercial interest in this direction is however not hard to trace: a schism developed between the GIS and database communities as long ago as the 1960's, when the GIS community found that database technology was not able to support the basic functionality it required. In particular the GIS focus on geospatial visualization was not seen as relevant to the requirements of business database systems of the time. Nevertheless, as GIS moved from batch processing to real time, and the scale of GIS data management increased by orders of magnitude, GIS systems developed 'loose couplings' to DBMSs. Even so, at the conceptual level at least, the representation, querying and visualization of geospatial objects remained – and remain today - a function of the GIS system rather than the DBMS. For example, the components of the representation of a spatial extent in ArcInfo can be stored in a loosely-coupled DBMS, and search operations on such extents utilize the standard database index methods (of which, until only ten years ago, none were inherently spatial). But the interpretation and manipulation of these components as parts of a single spatial object, and its visualization, remain within the GIS system.

The complete integration of GIS and DBMS therefore remains an unsolved problem. But the world has moved on, and now the problem has become much more challenging. Two particular aspects interest me:

- 1) the use of ontologies to express not only a wide variety of geospatial data types, but relationships between them;
- 2) the representation, manipulation and querying of geospatial data on a global scale;

It is now widely recognized that there is a need for the creation of an exhaustive domain ontology in almost every field of knowledge. In the case of geospatial information, the humble gazetteer has now grown into a fully-fledged ontological framework. Much work has already been done on defining a gazetteer content standard and access protocol. The challenge now is to integrate this framework, and a reasoning engine over this framework, into a DBMS.

However, this cannot be satisfactorily achieved without addressing the fundamental problem of how to represent and manipulate geospatial data within a DBMS. I am particularly interested in the development of a geospatial partitioning and addressing mechanism based on the recursive division of the surface of the Earth into a hexagonal grid. The objective is to find a way – or at least a satisfactory compromise between conflicting requirements – of representing geospatial extents accurately at different scales.