November 16, 2006

JORDAN T. HASTINGS

Position Statement
DGRP Workshop

I regard a digital gazetteer (DG) as the quintessential geographic information system (GIS) having just one kind of feature, place, with two required descriptive attributes, a free-text placename and a categorical placetype. The geospatial attribute, footprint, is interpreted directly by the GIS. Despite this apparent simplicity, a DG pushes GIS technology in many ways: footprints may be approximate, placetypes are often fuzzy, and placenames have important linguistic connotations. All three descriptors are time-dependent.

Moreover, a single place may be described by multiple placenames, placetypes and/or footprints concurrently. A critical – perhaps the critical – task in gazetteer construction, therefore, is place *identification*, i.e. determining when two (or more) different descriptions in fact apply to the same place, and conversely when the same descriptions apply to different places. Bigler, Daowaga, Tula Tulia, and Lake Tahoe are all names for the same place, whereas Lake Geneva might be a city in New York, or in Wisconsin, or a water body in Switzerland. Place identification also presents itself after the fact, in that need to detect and remove entries for "duplicate" places from an extant gazetteer, or alternatively to conflate them into a single entry. Finally, place identification is essential to federating and ranking results from distributed gazetteer queries.

Objects with indeterminate boundaries, differing names, fuzzy types, time dependencies, etc. appear in GIS applications other than gazetteers, of course. Coming to grips with these issues in the context of DGs is appealing for two reasons: 1) because of their relative simplicity as GIS, gazetteers cameo the underlying knowledge engineering and data management issues, without loss of generality; and 2) gazetteers are increasingly utilized in data mining, information retrieval and Web query applications, so correct place identification and conflation have practical importance.

[end]