

Linda L Hill
University of California, Santa Barbara (emeritus)

Digital Gazetteer Research and Practice Workshop, December 7-9, 2006

In 1999, Mike Goodchild and I co-chaired a workshop on gazetteers—the Digital Gazetteer Information Exchange (DGIE) workshop, funded by the National Science Foundation. The workshop report includes a number of specific recommendations for future research and development activities. The highlighted conclusions were that there was an “immediate opportunity and requirement to coordinate the building of shareable digital gazetteer data in the interest of digital earth applications,” that the temporal aspects of gazetteer data should get more attention, and that a gazetteer service protocol to support distributed gazetteer services should be developed.

In planning for this DGRP workshop, I see the developments that have taken place in the intervening seven years and the situation today against the backdrop of the DGIE report. Now there are more services based on gazetteer data, including interpreting geographic references in text (geoparsing), placename orientation of map views, placename access to online geospatial data, and communities developing shared gazetteers for their own purposes. There are more online gazetteer datasets, some with open contribution through web applications. The gazetteer data model developed by the Alexandria Digital Library (ADL) Project has been adapted for special gazetteer applications and a group at UC Berkeley has adapted it for representing named time periods, but no further work has been done (that I know of) to develop similar models for other geospatial and temporal data sets, such as meteorological events (e.g. hurricanes). The only networked example for gazetteers is the one based on the ADL Gazetteer Protocol giving access to both the ADL Gazetteer and to the ESRI Gazetteer. The OGC has published a gazetteer service protocol as a specialization of its Web Feature Service specification; the protocol is based on the model of gazetteer data formalized in the ISO TC 211 standard for *Geographic Information – Spatial Referencing by Geographic Identifiers*. There is a growing understanding of the role of gazetteers as components of information retrieval systems—gazetteers as a type of knowledge organization system (KOS)—but text-based information management and retrieval systems are still largely without a geospatially-enabled access mode and translation from placenames to coordinates. Some papers on gazetteers have been published, some conference presentations on gazetteer related research have been made, and with the publication of my book on *Georeferencing: The Geographic Associations of Information* by MIT Press there is now a substantial chapter on gazetteers available in book form.

Gazetteer research and practice draws from the expertise of diverse professions because placenames, classification schemes, and geospatial representations are all key components. The chief division is between text-based systems and geographic information systems. I think it is fair to say that currently those who work with text-based information resources and information retrieval systems are slowly integrating their placename georeferencing methods with geospatial representations and those who work primarily with geospatial information are realizing that their treatment of placenames and place classification would benefit from a more formal approach. Projects outside of the official toponymic authority agencies are realizing the complications and cost of building and maintaining quality

gazetteers. Such projects are very interested in using gazetteer data already collected and documented and developing ways to collect new place description information from knowledgeable sources.

The primary gazetteer research interests that I bring to this workshop are:

Components of Gazetteer Services:

1. An information retrieval test environment where footprint generalizations and similarity calculations can be tested for performance for given tasks, answering, for example, when bounding boxes are sufficient are for geospatial information retrieval.
2. Analysis of cross-walking options for feature type classifications, including automatic methods derived from the placenames themselves, for a ‘gazetteer classification advisory service’ that can be used to support gazetteer search interactions and gazetteer creation.
3. Development of software for gazetteer creation and maintenance based on community standards that can be customized for individual and organizational purposes.
4. Modeling of the temporal and spatial components of gazetteer data for applications in which the temporal aspects are on equal footing with the spatial.

Georeferencing as a Process:

1. Establishing college curricula and internships to educate GIS and LIS students in role of gazetteers in information services.
2. User studies for gazetteer services.

Interoperable Gazetteer Services:

1. Conflation of placename data from multiple sources for one place. This is a complex problem because all attributes of a place can vary: it may be known by different names, different terms may be used to represent its feature type, and different representations of its coordinate location may be used due to source, scale, and time period.
2. A test environment for gazetteer service interoperability, testing gazetteer service protocols and the suite of services needed to support discovery, search and retrieval. This would include a network of gazetteers accessible by a common gazetteer protocol and methods to obtain comparable performance data.

As a result of this workshop, I would like to see an organized and focused research effort to address gazetteer data and gazetteer service issues and specific efforts to teach the fundamentals of gazetteers and the services built on them through our professional educational systems.