

Statement of interest for the Digital Gazetteer Research and Practice Workshop, December 2006, by Gregory B. Hill, University of Colorado.

My interest in gazetteers stems from recent experiences developing web based spatially enabled gazetteers and applications that depend upon them. The gazetteers themselves are distributed instances with a common schema, inspired by the Alexandria Digital Library Gazetteer Content Standard<sup>1</sup>, and developed as part of the Biogeomancer<sup>2</sup> suite of georeferencing tools for natural history and biodiversity. The main client application is called GRIPPER<sup>3</sup>. One of the distinguishing features of GRIPPER is its use of a gazetteer to catalogue tiled raster layers. Via this means, gazetteer queries can be enhanced through the application of filters based upon environmental conditions or species occurrence data. By way of example, a query upon wetlands within Labrador could be further refined by specifying an average temperature range, or a species identifier. GRIPPER provides a flexible and powerful query building interface<sup>4</sup> and the ability to modify feature data in a user-specific gazetteer instance. It also allows features returned to be added to a selection layer; selection layers in turn are visualized using an integrated map portal<sup>5</sup>.

The following paragraphs describe in more detail the work I have been doing. The workshop will be an excellent opportunity for me to present these developments to others working in the field and to learn from their questions and comments as well as to interact with researchers and developers who have been working and thinking about related challenges.

Gazetteer data has been ingested from various sources and normalized to WGS-84. The placenames are run through automatic character set detection during ingest and converted to UTF-8.

Current datasets ingested include the Getty Thesaurus of Geographic Names (TGN)<sup>6</sup>, Geonet Names Server (GNS)<sup>7</sup>, Geographic Names Information Service (GNIS)<sup>8</sup>, U.S. Roads, Worldwide Protected Areas (reserves), and Worldwide Administrative Boundaries. The feature count exceeds 100 million, and will likely exceed 500 million in the future. Most features include MULTILINE or MULTIPOLYGON spatial descriptions. There is no hard limit on the maximum number of features; a distributed design with multiple gazetteer instances keeps the maximum virtual gazetteer size open ended and accessible. Each instance corresponds to a dataset and can reside on an arbitrary server.

---

<sup>1</sup> <http://www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/version3.2.html>

<sup>2</sup> <http://www.biogeomancer.org/>

<sup>3</sup> Geographic Referencing for Investigating Phylogenetics and Pathogenicity. R&D demonstration build running at <http://biogeobox.colorado.edu:8080/gripper>, please enquire for demonstration.

<sup>4</sup> For example, queries can be run in batch mode, and counts for the number of results can be obtained prior to running the queries. For queries returning many results, it is possible to iterate through result sets for a given query. Queries can be held in memory, modified, and re-run. Results are displayed in a UI incorporating recent advances in asynchronous technologies such as AJAX.

<sup>5</sup> Currently using the deegree WFS/WMS/WCS portal, see [www.deegree.org](http://www.deegree.org)

<sup>6</sup> [http://www.getty.edu/research/conducting\\_research/vocabularies/tgn/](http://www.getty.edu/research/conducting_research/vocabularies/tgn/)

<sup>7</sup> <http://earth-info.nga.mil/gns/html/index.html>

<sup>8</sup> <http://nhd.usgs.gov/gnis.html> (not all of this data has been processed at this time).

Coordinated use of different classification schemes is handled via hierarchical relationships that are captured in a simple database table that is transparent to the gazetteer server code. For example, if a TGN classification term bears a child relationship to an ADL classification term, any search for the ADL term will also return features matching the TGN term. Similarly, if an ADL term bears a child relationship to a TGN term, searches on the TGN term will also return features classified using the ADL term. Equivalency is indicated in a similar fashion. Searching of equivalent and child terms can be turned on or off. This does not resolve all issues involved with multiple classification schemes, but it is backwards compatible: a dataset developed with a given classification scheme can still be searched as before using the assigned terms.

The geometric content is comprised of 0, 1, or 2 dimensional representations, a radius calculation based upon the concept of 'spatial fit'<sup>9</sup>, a GML 3.1<sup>10</sup> representation, and several other forms designed to be used as data sources for the map portal. The taxonomic and raster extensions do not affect regular gazetteer content. For example, the raster filter capability is implemented not by modifying the content standard, but by adding raster tiles as features to a special gazetteer instance.

When queries are run, results can be edited. Classifications, alternate names, and the display name can be modified, added, or removed. The geometric representation can be translated. When these edits are committed, they go into the user's local database, not to the released databases used for searching. A secure, administrative logon-based procedure is used to commit modified records from user databases to the shared released databases, but users can immediately take advantage of the results of modifications to their own gazetteer instance. In the ordered list of search databases for each user, the user's gazetteer instance is first in the list, so their edits are returned in the first set of matches.

This open source software architecture is primarily based upon AJAX<sup>11</sup> (zk<sup>12</sup>), JSF<sup>13</sup>, Struts<sup>14</sup>, Java, deegree (web map client supporting WMS/WFS/WCS), PostGIS<sup>15</sup>, and other leading open source technologies.

---

<sup>9</sup> <http://bgdev.berkeley.edu/?q=node/568>

<sup>10</sup> <http://www.opengis.net/gml/>

<sup>11</sup> Asynchronous Java XML

<sup>12</sup> [zk1.sourceforge.net](http://zk1.sourceforge.net)

<sup>13</sup> Java Server Faces: [jsf.sun.com](http://jsf.sun.com)

<sup>14</sup> [Struts.apache.org](http://Struts.apache.org)

<sup>15</sup> [postgis.refrains.net](http://postgis.refrains.net)