

Workshop on Digital Gazetteer Research and Practice  
Santa Barbara, California, 7-9 December 2006

National Geospatial-Intelligence Agency and United States Board  
on Geographic Names

Statement of Interest:

Dr. Beth Driver: [Beth.H.Driver@nga.mil](mailto:Beth.H.Driver@nga.mil)

Mr. Randall Flynn: [Randall.E.Flynn@nga.mil](mailto:Randall.E.Flynn@nga.mil)

The United States Board on Geographic Names (BGN) is the U.S. Government body that has been appointed, through act of Congress, as the national names authority for geographic names within the United States. It has also been given authority to set standard forms of foreign place names for use across the whole U.S. Government.

The National Geospatial-Intelligence Agency (NGA) is one of the sixteen agencies that make up the United States Intelligence Community (IC); since 2003 NGA has been continuing the critical national security missions of its predecessor organization, the National Imagery and Mapping Agency (NIMA).

NGA provides staff support to the BGN and maintains the Geographic Names Data Base (GNDB), which is the official repository of standard place name spellings for foreign place names for use throughout the U.S. Government. The GNDB currently holds more than five million place name entries for more than three million unique geographic features. Its coverage is worldwide, with the exception of the United States and Antarctica. Domestic features and geographical names are stored in a separate database called the Geographic Names Information System (GNIS) maintained by the BGN Domestic Names Committee and staff at the U.S. Geological Survey.

The GNDB was originally designed in support of military cartography in areas of the world where the United States had active conventional conflicts and generally covers scales of between 1:250,000 and 1:50,000. Recent changes in the nature of warfare towards asymmetric conflicts, as well as increased centrality of geospatial intelligence (GEOINT) within the IC, have forced corresponding alterations in the nature and content of the BGN/NGA gazetteer.

First, there has been a dramatic increase in the scale of information requirements from IC and DoD customers, from county-level to urban-level scales, and even to cadastral-level scales

in some areas. This demand for increased detail has led to challenges in collection, storage, and dissemination. An as-yet undetermined question is the proper amount of cross-references (variant entries) to add to the gazetteer: a delicate balance must be found between improved searchability and increased noise. Features generally found only on very large-scale maps have revealed previously unknown problems of competing names authorities; for example, discussions are under way between NGA's Office of Global Navigation and the BGN over who has ultimate authority to set standard names for lighthouses.

Increased focus on joint operations with multinational partners leads to the need for multilingual gazetteer services. To support this requirement, the GNDB became fully Unicode-compliant in 2005, and non-roman script names began to be added (about 300,000 non-roman script names to date). Enriched metadata has been found necessary to distinguish amongst the new classes of entries to the GNDB: language of the name, dialect, script, orthography, and transliteration system. Each of these domains of metadata suffers from challenges of definition; for example, does "language" refer to the etymology of a geographic name, or to the language in which it is currently used? Each new metadata domain as well poses difficult challenges of identifying and reconciling existing, and sometimes competing, standard enumerations.

Alterations to the GNDB's current name typing schema, which encodes name provenance and authoritativeness, have also been proposed, in response to the increased complexity of the entries in the enriched GNDB. A renewed interest in time-variant aspects of geographical names has led to lengthy discussions about how to encode historical change within the GNDB: at the feature level or the name level.

The explosion of third-party digital gazetteer data has revealed a major technological and resource bottleneck for maintaining the GNDB: how to conflate external gazetteers into the GNDB. Several independent research efforts are currently testing algorithms based on record-linkage methodology, which will give a measure of automation to the current manual conflation process. Beyond conflation, federated architectures and distributed gazetteers promise to ease the burden of maintaining gazetteer data on a global scale.

In conjunction with these changes, the BGN Gazetteer has been recognized as an important enterprise-wide service that needs to be available across NGA and the IC. This realization comes from

statistics that point out the critical role of place names in organizing information. For example, Microsoft and Yahoo recently reported that between 12 and 16% of all user queries on their general search sites contain at least one geographic name.<sup>1</sup> Similarly, a recent U.S. Government report notes that geographic location is a key feature of 80-90% of all government data.<sup>2</sup> The challenges and opportunities of geographic information retrieval (GIR) lead to questions of interoperability; more than ever there is a need for greater coordination between national and international standards organizations, such as the Open Geospatial Consortium, the International Standards Organization, and the Digital Geographic Information Working Group, for example.

The integration of the GNDB into the wider NGA enterprise platform has entailed another painful process: the transition in the gazetteer away from a strictly point-based geometry towards one where each entry may have an arbitrary geometry. Increased complexity of gazetteer geometry representations has meant increased cost of ingesting and maintaining locational information, and difficult choices are being made about costs and benefits of collecting various types of locational references for various features.

A final area of interest to highlight is the relationship of gazetteers to ontologies. What level of ontological information is most appropriate to store in a gazetteer? How can gazetteers be made more interoperable with the proliferation of geospatial ontologies? Current NGA/BGN efforts in this vein include the mapping of the GNDB feature catalogue to the NGA Feature Catalogue (NFC), as well as the federated linkage of the GNDB and several other NGA data stores to the Cycorp Upper Ontology.

Dr. Driver & Mr. Flynn look forward to discussing these and all the other topics of the Gazetteer workshop. Dr. Driver is the Chief of the University Outreach and Partnership Branch within NGA's InnoVision Directorate and is responsible for coordinating all NGA collaborations with the academic community. Mr. Flynn is the Executive Secretary for Foreign Names of the United States Board on Geographic Names, as well as the NGA Geographer, and is responsible for oversight of the BGN/NGA GEONet Names Server Gazetteer.

---

<sup>1</sup> SIGIR 2006, Workshop on Geographic Information Retrieval, Seattle, Washington, August 2006.

<sup>2</sup> Federal Geographic Data Committee, <http://www.fgdc.gov/library/whitepapers-reports/white-papers/homeland-security-gis>.