Approved for Public Release; Distribution Unlimited
Case # 05-0247

Accelerating GEOINT Analysis: User Feedback to Improve Geospatial Place-name Services

Frank Linton
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102-7508
linton@mitre.org

Keywords: Usability Analysis, Information Sharing and Collaboration, Geospatial Information Systems, Geospatial Intelligence (GEOINT)

Abstract

Collaboration and information sharing across boundaries is accelerated when analysts in different organizations refer to the same place by the same name. Geonames, hosted by NGA, is the official US and Commonwealth repository of foreign place-names. Analysts access Geonames thousands of times per day. A re-engineered Geonames will be an operationally significant advance in technical support for GEOINT that provides considerably better access, for both humans and automated systems, to higher-quality place-name data. Using the general interview guide approach we surveyed users external to NGA to determine how well Geonames was working for them and what could be done to improve it. Informants were recommended by NGA staff embedded in external users' organizations. Informants have high expectations of Geonames and suggested extensive changes to Geonames data, user interface, search tools, and gazetteers. How well the re-engineered system addresses these users' needs will determine its success.

The Need for Improved Information Sharing

O wad some Pow'r the giftie gie us
To see oursels as others see us
It wad frae monie a blunder free us
(Burns, 1785)

Common sense suggests that intelligence analysis will improve with increased information sharing across boundaries. Indeed, the National Intelligence Reform Act of 2004 requires "sharing information in a form that promotes use in analysis, investigations, and operations". Collaboration and information sharing is accelerated when analysts in different organizations refer to the same place by the same name.

The GEOnet Names Server (GNS) http://earth-info.nga.mil/gns/html/index.html provides access to the National Geospatial Intelligence Agency's (NGA) database of foreign geographic feature names. The database is the official repository of foreign place-name decisions approved by the U.S. Board on Geographic Names (US BGN). For example, the BGN name for the city known variously as Leningrad, Petrograd, or Sankt-Peterburg is Saint Petersburg. The database also provides variant names, the feature type, the latitude and longitude of the feature, the JOG map number if it exists, name and feature identifiers, and other information.

The GNS (hereinafter Geonames) contains about 4 million features with approximately 5.5 million names. The DDCI/CM has issued Guidelines for using GEONAMES (2002). Geonames is available on both classified and unclassified networks and is accessed thousands of times per day by analysts in all US and Commonwealth government agencies with international geospatial interests.

Both advances in search across massive, heterogeneous, multilingual collections, and foreign language processing (including trans-lingual information retrieval, machine translation and machine assisted translation) will require cross-referencing of place-names, a service that only Geonames can provide. When a geographic location is indicated by a place-name, automated clustering, correlation, and fusion (including Multi-INT fusion) requires *indirect geospatial referencing*, an activity made possible by Geonames. That is, Geonames allows analysts or automated processes to convert place-names to location by coordinates.

The goal of the work reported here is a re-engineered Geonames system, an operationally significant advance in technical support for GEOINT that provides considerably better access, for both humans and automated systems, to higher-quality place-name data.

We began by doing a user survey and incorporating the results into Requirements and Conops design documents. The design documents also take account of results

from several other sources, including usage analysis from web logs, review of the state of the art of related technologies, interviews with internal business process owners, etc.

Value of Survey

NGA, the organization that supports Geonames, is one of the major users of Geonames data, yet more than half of Geonames users are from other organizations. To gain a better understanding of these users' current and future needs with respect to foreign geographic feature names, or placenames, the Analysis and Production Directorate of NGA, the business process owners of Geonames, decided to perform an external user needs assessment.

A survey of intelligence analysts outside NGA using Geonames might have taken several forms, an email to Geonames users, a questionnaire on the Geonames website, or interviews of Geonames users. We knew from prior experience that surveying users across classified networks with either email or web-based questionnaires posed numerous technical difficulties owing to security concerns, and that intelligence analysts are under time pressure and have little patience for interviews. However, NGA has numerous employees embedded in external users' organizations, and these, together with Geonames staff, were able to provide a selection of individuals who were frequent users of Geonames and who were willing to communicate their perceptions of its strengths and weaknesses. The interviewees were also aware of, and communicated, the perceptions of other, less-frequent users of Geonames in their organizations, as the interviewees were the people to whom these other analysts turned when they had questions or difficulties with Geonames.

The interviewer used the *general interview guide approach* (Patton, 1990) which involves outlining a set of issues to be explored with each respondent before interviewing begins. The set of issues was developed iteratively with those involved in re-engineering the Geonames system. During the interview, the issues were not discussed in any particular order and the actual wording of questions to elicit responses about those issues was not determined in advance. The interview guide simply served as a basic checklist during the interview to make sure that all relevant topics were covered. The interviewer adapted the wording and sequence of questions to specific respondents in the context of the actual interview. The general interview guide approach kept the interactions focused but allowed individual perspectives and experiences to emerge.

The interview guide was formatted as a set of questions. Two organizations distant from the interviewer's location (Washington DC area) responded to the questions by email. Their written responses were useful in that they corroborated the interviewees' perceptions and responses, but they tended to be much briefer than interviewees' responses.

Method

The Geonames customer needs assessment was designed to capture customers' perceptions of Geonames' strengths and weaknesses with respect to their place name needs. We asked users external to our organization: How well is this service working for you? What could we do to improve it? These customer perceptions were gathered in the context of the customer's products and work processes, both current and planned. The assessment compiled actual and desired Geonames features and capabilities.

The results of this needs assessment will better enable Geonames to share information in a form that promotes use in analysis, investigations, and operations, by developing products and services that more effectively meet its customer's needs. The information obtained will serve as input to conops and requirements documents, helping ensure that re-engineering Geonames results in a more-effective service from the external users' perspective. The survey helps identify specific features to enhance system success.

In particular, the goal of the survey is to gather and provide evidence that either supports or refutes current assumptions about customer needs with respect to geographic place names, to uncover any new needs that customers may express, and to determine the relative importance of the place name needs that customers have now or expect to have in the next decade.

Procedure

1. Conduct initial interviews with several individuals (1 to 6 per session) in each customer support organization representing selected NGA customer agencies. A general interview guide will ensure that all significant topics are covered while providing ample opportunity for unanticipated input from interviewees. The interviewer will take unclassified notes by hand on paper.
2. Write up the results of each interview immediately, permit the interviewees to review the draft report and make any desired changes; then make the report available to all interested parties in Geonames.
3. Synthesize the interviews into a briefing and draft report when the interviews are completed.

Survey Results

At the time of this writing we had surveyed six organizations, four by interview and two by email, as described above. The interviewed organizations were:

- State
    - INR/GGI (INtelligence Research, Office of the Geographer and Global Issues)
    - GGI/International Boundaries and Sovereignty Issues
- NSA: Geographic Resource Center
- DIA: Geospatial Analysis Division
- Map Library: Map Services Center

The organizations surveyed by email were:

- DIGO: Australia's Defence Imagery and Geospatial Organisation
- SOUTHCOM

Interviewees reported expecting more from the Geonames than its original intended purpose. They have three main uses of Geonames.

- Retrieve a set of BGN place-names with certain defined attributes to create a map or to label the significant features of an image.
- Find the location of a place-name mentioned, for example, in an intelligence report and spelled phonetically, i.e., to do indirect geospatial referencing.
- Obtain the BGN name of a variant name that has been obtained from some source, for example, a foreign language map or a passport application and that has been transliterated in one of any number of ways. This is the original intended purpose of Geonames.

There are two kinds of Geonames user, the professional geo-analyst, and the casual user, an intelligence analyst or government worker in some other specialty, who happens to need the official place-name of some feature. Users who make maps are frustrated by a lack of means of discriminating among names. They want to use the population of each named place as a means of selecting names suitable for maps of different scales.

In contrast, users who want, for example, to annotate photos, want to find the names of the populated places and other features in the photo. They want all the names within a certain area. They are frustrated by a lack of names for the features of interest, and by the lack of precise location of the named places in the database. Furthermore, they want Geonames input and output in the coordinate system they happen to be using at the moment. While the Geonames website cautions that the "Coordinates in the GEOnet Names Server are approximate and are intended for finding purposes only" these missing names and rough coordinates make their work harder.

Main Findings

Here we select and summarize a few of the main findings. For several of the findings a quotation illustrates the point in users' own words. A full report of the findings was used by the developers of the Conops and Requirements documents and will serve as a resource for the more detailed design documents.

Data

Add missing place-names (smaller cities, neighborhoods) and variants (as found, e.g., in the popular press).

> *We get asked about many place-names that are not in Geonames. We need all the place-names and their variants. We depend on Geonames as the authoritative source, but Geonames doesn't always have what we need.*

Add attributes, including population, name variants, and usage caveats; and local script, spelling, pronunciation, and ethnicity, of name.

> *The biggest beef with Geonames is: No attribute/value pair regarding the size or relative size of populated places. This is a big limitation on the usefulness of the data. A country will have lots of places with identical place-names, sorting them by size of the populated places would be very useful. For example, the roads, rivers, etc., of Iraq are clogged with place-names.*

Improve attribute quality, e.g., coordinates.

> *Biggest issue: Need accurate location data for place-names. The Administrative regions in Geonames are useless because their location in Geonames is [not correct]; what is needed is the polygon and its location.*

User Interface

Geonames is too hard to find and too hard to use. Many who should use it do not.

> *Prospective users have to follow too many obscure links to arrive at the Geonames page.*
> *The Geonames interface is so forbidding that analysts do not use it.*
> *Analysts see Geonames as too complicated, too confusing, and too slow to use. They need a more user-friendly interface.*

Make it easier to transfer data from Geonames to ARC GIS
Integrate Geonames into other GIS applications and portals, e.g., the NGA Portal, DGINet (a federated data sharing effort), etc.

Provide an interface that seamlessly unites Geonames with related GIS data, e.g., the Census Bureau's Population database

Provide direct links from place-names to further sources of place-name information such as the named feature in the feature database, the JOG, TPC, and DLA catalog.

Search

Improve search: add search in foreign languages, phonetic search, wildcards, fuzzy search, and faster search.

> *Big Issue: To search for a Cyrillic place-name, for example, with its various transliterations, an analyst may retrieve 300-500 names, and each requires a closer look in order to determine which place-name is the right one.*

An intelligence report will often have the place-name spelled phonetically. Analysts need to be able to put in the phonetic spelling of the name and get the accepted name.

Comment: The requirement to input a place-name in an English transliteration adds ambiguity to the name and to the search query, and may vastly increase the effort required of the searcher. Instead of a simple table lookup (Cyrillic input => English output), searching for the BGN name becomes a tedious guessing game, a game made more frustrating by the limited capabilities of the search interface.

Gazetteers

Need gazetteers in multiple media: online, in print, on CD, and rehosted.

The main thing that NGA could do to improve things at our organization with respect to Geonames is to arrange for us to replicate the Geonames database on our network.

Need ability to make custom gazetteers easily.

We print hundreds of copies of gazetteers for analysts who prefer hardcopy.

Comment: Many users find paper gazetteers more useful than Geonames online. They may have problems with connectivity, with security, or with the relative utility of the format for their purposes. They also need a means of keeping their paper gazetteers up to date.

Challenges

Improvements in Geonames' user interface, search, and gazetteer production are relatively straightforward to implement. In contrast, the challenge of providing up-to-date name information is not. The basic task is to provide the correct name of every place in the world in every language, dialect, script, and transcription system of the world. Add to this all the associated data that users desire, such as population, variant spellings, ethnic group providing the name, pronunciation, source quality, etc. and the task becomes overwhelming. Currently Geonames records are updated every decade or two. What is to be done?

First, completing the systems engineering and implementing the changes to make Geonames part of an integrated, federated, system is crucial. Neither Geonames nor any other organization could acquire and maintain all the data its users expect. It must pick a key portion of the task of providing feature names and leave the remainder of the task to others. Possible changes include modern database design and implementation, revised workflow, semi-automated multi-lingual data capture from open sources, an open API, an upgraded website, and integration with other geospatial applications. MITRE and contractors will develop a next-generation design based not only on customer needs but also on the effect of various tradeoffs (storage, bandwidth, etc.) within the given environmental constraints (funding, skills, etc.).

Second, the same technologies that depend on Geonames can also help feed names data into it. Entity extraction tools will improve the collection of geospatial names; at the same time, the existence of an excellent gazetteer improves the capabilities of a named entity extraction system. Question answering systems require the disambiguation of place-name references, and current research in this area uses Geonames to do so; and when a 'Where' question returns an answer that is not in Geonames, it signals the existence of a potential new record. Foreign language processing is key to automated and semi-automated acquisition of place-name information from publications in other languages; yet foreign language processors translate place-names by table-lookup, a service that Geonames must provide if translation results are to be consistent.

Questions and Conclusion

There are some questions remaining. First, how to capture the work of others, for example, once someone has figured out that a certain name in Cyrillic means Saint Petersburg in BGN English, how can that information be captured and made available to others? Second, how to implement a persistent means of obtaining user feedback, so that Geonames can address their issues and chart its improvements over time. Third, the information that Geonames provides reduces the effort and increases the quality of the work of others but how is Geonames compensated, or how does Geonames measure benefits to justify its costs? The users/ beneficiaries don't pay directly for the service, and may compete for budget dollars.

Improving services already in place, such as Geonames, may result in significant advances in intelligence analysis. The straightforward technique of performing a Customer Needs Assessment has revealed several ways to radically improve the GEOINT analysis process. The implication for practice is: Do something, anything!, to see your tool, service, or system from your users' perspective. Ultimately they will determine its success.

References

Burns, Robert. 1785. To a Louse: On seeing one on a Lady's bonnet at Church.

Patton, Michael. 1990. Qualitative Evaluation and Research Methods. London: Sage Publications.

The July 2002 DDCI/CM memorandum on Guidelines for the Uniform Spelling of Foreign Place Names and Standardized Use of Geographic Coordinates

Note

The complete last verse of *To a Louse*:
O wad some Pow'r the giftie gie us
To see oursels as others see us
It wad frae monie a blunder free us
An' foolish notion
What airs in dress an' gait wad lea'e us
An' ev'n Devotion

David Sibbald http://www.robertburns.plus.com/louse.htm
translates this Burns verse as follows:
Oh, that God would give us the very smallest of gifts
To be able to see ourselves as others see us
It would save us from many mistakes
and foolish thoughts
We would change the way we look and gesture
and to how and what we apply our time and attention.