

The PATTON Program

Gazetteer Evaluation Project

The PATTON Alliance (Preparing Analytic Tools and Technology for an Operational Network) substantially accelerates the transition of information technologies from government research to the field. One of the early projects of the PATTON Alliance is to evaluate the previously ARDA-funded Integrated Gazetteer Database (IGDB) and identify the gaps between this R&D effort for ARDA and the wider intelligence community (IC) needs for a gazetteer. A gazetteer is a list of geographic names, location information, and descriptive information that can be used in text base searches for location identification as well as in geographic information systems (GIS).

One of Sandia's roles in the PATTON Alliance is to review technologies and perform independent review of the available technologies and research to provide an unbiased evaluation. This is the approach we are taking with the Gazetteer Evaluation Project. We are evaluating gazetteer data sets, gazetteer data services, and research into gazetteers to provide an overall view of the "state of the art" in gazetteer research and gazetteer resources.

In addition, we are involving the National Geospatial-Intelligence Agency (NGA) and the United States Geological Survey (USGS) in our research. The USGS Geographic Names Information System (GNIS) is the US Government's authoritative source for geographic place names within the boundaries of the United States. The



Place names near Kabul, Afghanistan

NGA and the GeoNames Server (GNS) are the primary source for foreign place names for the federal government. In addition, these two sources of place names are the core data sets of the majority of the gazetteers we have been evaluating.

The main requirements for gazetteers from the IC have these characteristics:

- Is free or low cost to federal agencies
- Can be easily updated
- Can easily ingest new sources for gazetteer data
- Can be used for both geographic queries within a GIS as well as for text-based queries
- Contains both "official" and local/colloquial/tribal place names
- Can contain foreign place names in English and native language
- Can connect to other sources of ancillary data such as population, CIA FactBook, etc.
- Can manage historical names as well as place name changes over time

We are using these main requirements as part of our evaluation process. Each gazetteer investigated is evaluated against these criteria.

The PATTON Gazetteer Project will be completing the initial gazetteer evaluation work in the summer of 2007, and we look forward to working with the end-user community to develop a gazetteer solution that will meet their long-term needs.

For more information Contact:

Denise Bleakly, Gazetteer Project Lead
 Phone: 505-284-2535
 Email: drbleak@sandia.gov



Table of Candidate Gazetteer Resources – as of February 2007

US GOVERNMENT GAZETTEER	URL
National Geospatial Intelligence Agency (NGA) GEONet Names Server	http://earth-info.nga.mil/gns/html/index.html
NGA - Names file of Selected Countries	http://earth-info.nga.mil/gns/html/cntry_files.html
United States Geological Survey (USGS) Geographic Names Information System (GNIS)	http://geonames.usgs.gov/domestic/index.html
Research Gazetteers	
Alexandria Digital Library (ADL) Project Gazetteer Development	http://www.alexandria.ucsb.edu/gazetteer/
Getty Thesaurus of Geographic Names (TGN)	http://www.getty.edu/vow/TGNSearchPage.jsp
Electronic Cultural Atlas Initiative (ECAI)	http://ecai.org/projects/gazetteer/
Academia Sinica	http://www.sinica.edu.tw/main_e.shtml http://pgis.sinica.edu.tw/en/research.html
US Gov. Funded Research on Gazetteers	
MITRE Corporation	http://www.mitre.org/
Commercial Gazetteers	
MetaCarta	http://www.metacarta.com/
Columbia Gazetteer of the World on line	http://www.columbiagazetteer.org/
Geonames	http://www.geonames.org/
Other Web-based Gazetteers	
Falling Rain Gazetteer	http://www.fallingrain.com/world/
Earth Search	http://www.earthsearch.net/
MultiMap World Map & Placename Index	http://www.multimap.com/index/
Information Commons Gazetteer	http://www.maya.com/web/what/papers/maya_infocommons_gaz.pdf
Worldwide Intex – Tageo.com	http://www.tageo.com/index.htm
Country Specific Geographic Data	
EuroGeoNames (EGN)	www.eurogeonames.org
National Sultanate of Oman	http://www.nsaom.org.om/english/services/
Qatar - On line mapping service	http://www.gisqatar.org.qa/MapServer/Default.htm
Electronic Cultural Atlas Initiative - Iraq	http://ecai.org/iraq/

DATA SOURCES	STANDARDS
This is the authoritative source for the US government for foreign names	Not clear what standards they use, however this is the "Official" source for foreign geographic names for the US Government
These are the GeoNet Names Server entries for specific countries.	Not clear what standards they use, however this is the "Official" source for foreign geographic names for the US Government
This is the authoritative source for the US Government for domestic names	Not clear what standards they use, however this is the "Official" source for domestic geographic names for the US Government
USGS Geographic Names Information System (GNIS); NGA GeoNames (GNS)	Yes - Created the Gazetteer Content Standard and Relational Database Model for Gazetteer information
Not clear from reference materials	This is a thesaurus, compliant with ISO and NISO standards for thesaurus construction.
They wanted to extend the Alexandria Digital Library Project Gazetteer	Thesaurus
Unknown	Unknown
USGS GNIS, NGA GeoNames, TIPSTER, NMSU Arabic Names; Harvard U Chinese Names, Context information from CIA Fact Book	Unknown
USGS GNIS; NGA GeoNames (GNS), and proprietary sources	Unknown
Built on the data contained in the Lippincott Pronouncing Gazetteer, the Columbia-Lippincott Gazetteer of the World, USGS GNIS and other academic sources	Unknown
Is an internet project (not clear on who is sponsoring it) to create a "free" gazetteer based on the "wiki" concept. Many sources are listed on their website http://www.geonames.org/about.html	Unknown
Unknown. A "wiki" type gazetteer	Unknown
CIA World Fact Book, NGA Earth-Info; USGS GeoNames	Unknown
Unknown	Unknown
Getty Thesaurus, Alexandria Digital Library Gazetteer, NGA GNS, USGS, GNIS, World Gazetteer	Proposing a "Universal Database Architecture" a peer-to-peer system based on emerging internet data sharing standards. Open Standards.
Unknown. A Database of 2,667,417 cities in the whole world., a database of geographic coordinate information – wiki concept	Unknown
Objective is to connect 5-10 national databases together.	EGN will establish a European Infrastructure of geographical names data and set up and interoperable Web (Gazetteer) service compliant to open standards - Open GIS Consortium (OGC) standards for web Feature Services - They are aiming to connect 5-10 national databases over the life of the project.
Unknown	Unknown
Unknown	Unknown
Unknown	Unknown

PATTON ALLIANCE Gazetteer Evaluation

Working Bibliography – as of February 2007

Introduction

This working project bibliography was assembled as part of the initial gazetteer project task, and allowed us to quickly understand the breath and depth of the current research as well as some of new trends in research and development (R&D) with regards to the use of gazetteers in electronic systems. This is not a comprehensive bibliography, and it will continue to evolve as the project evolves.

Discussion

The bibliographic entries are grouped by category and then alphabetically. The categories were derived from the progression of topics discussed in the literature.

Working Bibliography

Alexandria Digital Library Gazetteer. 1999- . Santa Barbara CA: Map and Imagery Lab, Davidson Library, University of California, Santa Barbara. Copyright UC Regents.
<http://www.alexandria.ucsb.edu/gazetteer> (ADL Gazetteer Development page with links to various clients and protocols that access the ADL Gazetteer).

General Gazetteer Research

Axelrod, Ammittai E.,(2003). On Building a High Performance Gazetteer Database. Workshop on the Analysis of Geographic References. May 31, 2003. NAACL-HLT 2003, Edmonton, Alberta Canada. <http://people.mokk.bme.hu/~kornai/NAACL/WS9/paper.html>
Proceedings of the HLT-NAACL2003 Workshop on Analysis of Geographic References, Alberta Canada.

Brauner, Daniela F., Casanova, Marco A., and Milidu Ruy L, 2006. "Towards gazetteer Integration Through an Instance-based Thesauri Mapping Approach. In proceedings of the VIII Brazilian Symposium on GeoInformatics (GeoInfo 2006), Campos Do Jordao, Brazil. Downloaded 11/6/06 from :<http://geoinfo.info/geoinfo2006/papers/p74.pdf>

Goodchild, M.F. The Future of the Gazetteer. Presented at the Digital Gazetteer Information Exchange Workshop, Oct 13-14, 1999. Transcribed and edited from audiotape. http://www.alexandria.ucsb.edu/~lhill/dgie/DGIE_website/Perspectives/brandt.htm.

Hill, L. L. (2004, May). Georeferencing in digital libraries (guest editorial). D-Lib Magazine, 10 (5). <http://www.dlib.org/dlib/may04/hill/05hill.html> (All of the articles in this issue refer to the role of gazetteers in georeferencing applications).

Hill, L. L. (2000). Core elements of digital gazetteers: placenames, categories, and footprints. In J. Borbinha & T. Baker (Eds.), Research and Advanced Technology for Digital Libraries : Proceedings of the 4th European Conference, ECDL 2000 Lisbon, Portugal, September 18-20, 2000 (pp. 280-290). Berlin: Springer. PDF file.

Hill, L. L., Frew, J., & Zheng, Q. (1999). Geographic names: The implementation of a gazetteer in a georeferenced digital library. D-Lib (January 1999). Downloaded 8/22/2006 from <http://www.dlib.org/dlib/january99/hill/01hill.html>.

Hill, L. (1999). Gazetteer and collection-level metadata developments. In R. T. Kaser & V. C. Kaser (Eds.), Metadiversity. The Grand Challenge for Biodiversity Information Management through Metadata. The Call for Community. Proceedings of the Symposium sponsored by the U.S. Geological Survey Biological Resources Div. & the National Federation of Abstracting & Information Services (pp. 141-145): NFAIS. HTML file.

Hill, L. L., & Zheng, Q. (1999). Indirect geospatial referencing through place names in the digital library: Alexandria Digital Library experience with developing and implementing gazetteers. Proceedings of the American Society for Information Science Annual Meeting, Washington, D.C., Oct. 31- Nov. 4, 1999, pp. 57-69. PDF file.

Irie, R. and Sundheim, B. (2004). Resources for Place Name Analysis. In Proceedings of the 2004 International Conference on Language Resources and Evaluation - LREC2004, pages 317-320.

Kornai, Andras, and Sundheim, Beth. (2003). Workshop on the Analysis of Geographic References. May 31, 2003. NAACL-HLT 2003, Edmonton, Alberta Canada. <http://people.mokk.bme.hu/~kornai/NAACL/WS9/paper.html>. Downloaded 9/27/2006. Conference website: <http://www.kornai.com/NAACL>

Landcaster, Lewis, (2001). "A multilingual Gazetteer System for Integrating Spatial and Cultural Resources." Electronic Cultural Atlas Initiative, UC Berkley. http://ecai.org/projects/gazetteer/nsf_multisys_proposal.html. Downloaded 9/27/2006.

Linton, Frank, (2005), "Accelerating GEOINT Analysis: User Feedback to Improve Geospatial Place-Name Services" MITRE Corporation White Paper.

http://www.mitre.org/work/tech_papers/tech_papers_05/05_0247/index.html

Mardis, Scott and Burger, John, (2005). Design for an Integrated Gazetteer Database: Technical Description and User Guide for a

Gazetteer to Support Natural Language Processing Applications. MITRE Corporation White Paper.
http://www.mitre.org/work/tech_papers/tech_papers_06/06_0375/06_0375.pdf

Rugg, Bob. (1999). Emerging Geographic Information Standards and Gazetteer Data. Presented at the Digital Gazetteer Information Exchange Workshop, October 13-14. Transcribed and edited from audio tape. http://www.alexandria.ucsb.edu/~lhill/dgie/DGIE_website/Perspectives/rugg.htm
(Date accessed: 8/22/2006).

Smith, David A. (2002). Mining gazetteer data from digital library collections. In proceedings of the Networked Knowledge Organization Systems (NKOS) Workshop, July 18, 2002, Portland, Oregon. Joint Conference on Digital Libraries (JCDL 2002) on Digital gazetteers. Downloaded 10/26/2006 from <http://nkos.slis.kent.edu/>

Uryupina, Olga, 2003 "Semi-supervised Learning of Geographic Gazetteers from the Internet." In Proceedings of the HLT-NAACL Workshop on the Analysis of Geographic References, Edmonton, 2003.

Ourioupina, Olga. Extracting Geographical Knowledge from the Internet. In Proceedings of the ICDM-AM International Workshop on Active Mining, Maebashi, 2002.

Semantic Web Applications

A.Arampatzis, M. van Kreveld, I. Reinsbacher and other SPIRIT researchers (2004). "Web-Based Delineation of Imprecise Regions" In: the Proceedings of Workshop on Geographic Information Retrieval, SIGIR, 2004. Downloaded 11/06/2006 from: <http://www.geo-spirit.org/publications/GeoImprecise2004.pdf>

Chavez, Marcirio Silveira, Sliva, M.J., and Martins, B. 2005 "A Geographic Knowledge Base for Semantic Web Applications" from the the Proceedings of SBBD-05, the 20th Brazilian Symposium on Databases. Downloaded 9/27/2006 from: http://www.cs.cmu.edu/~ehn/JAVELIN/papers/javelin/lrec2006_jko.pdf

Clough, Paul, Sanderson, M., and Joho, H. 2004. "Spatially Aware Information Retrieval on the Internet (SPIRIT)" European Union IST Programme Contract Number: IST-2001-35047.

C.B. Jones et al (2002). "Spatial Information Retrieval and Geographical Ontologies: An Overview of the SPIRIT project". In: 'SIGIR 2002: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval August 11-15, 2002, Tampere, Finland', ACM Press, pp.387 - 388. Downloaded 11/6/2006 from: http://www.geo-spirit.org/publications/SPIRIT_project_overview.pdf

Lucas, Peter, Balasubramanya, M., Widdows, D., and Higgins, M. 2006. The Information Commons Gazetteer: A Public Resource of Populated Places

and Worldwide Administrative Divisions. 2006 Language Resources and Evaluation Conference (LREC), May 24-27, 2006, Genoa, Italy. Downloaded 9/25/2006 from: www.maya.com/web/what/papers/maya_infocommons_gaz.pdf

Ko, Jeongwoo, Hiakumoto, Nyber, Eric. 2006. Exploiting Multiple Semantic Resources for Answer Selection. 2006 Language Resources and Evaluation Conference (LREC), May 24-27, 2006, Genoa, Italy. Downloaded 9/25/2006 from: http://www.cs.cmu.edu/~ehn/JAVELIN/papers/javelin/lrec2006_jko.pdf

Souza, LA, Davis Jr, CA, Borges, KAV, Delboni, TM & Laender, AHF (2005), 'The Role of Gazetteers in Geographic Knowledge Discovery on the Web', Proceedings of the 3rd Latin American Web Congress, Buenos Aires, Argentina. Downloaded 11/6/06 from <http://ieeexplore.ieee.org/iel5/10609/33518/01592372.pdf?tp=&arnumber=1592372&isnumber=33518>

WordNet and Gazetteers

Duscaldi, Davide, Paolo Rosso, and Emilio Sanchis Arnal, 2006. WordNet as a Geographical Information Resource. Third International WordNet Conference (GWC-06). Downloaded 10/30/06 from <http://nlp.fi.muni.cz/gwc2006/proc/index.html>.

G. Fu, C. Jones, and A. Abdelmety. Bulding a geographical ontology for intelligent spatial search on the web. In Proceedings of the IASTED International Conference on Databases and Applications, 2005. Downloaded 11/06/06 from:

Lita Lucian Vlad, Hunt, Warren A. and Nyberg Eric, 2004. Resource Analysis for Question Answering. In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004), July 21-26, Barcelona, Spain. Downloaded 11/6/06 from : <http://www.cs.cmu.edu/~llita/papers/lita.resources-acl2004.pdf>

Sundheim, Beth M., Mardis, Scott and Berger, John, 2006, Gazetteer Linkage to Word Net. Third International WordNet Conference (GWC-06). Downloaded 10/30/06 from <http://nlp.fi.muni.cz/gwc2006/proc/index.html>.

Toral, Antonio, Munoz, Rafael, 2006. "A proposal to automatically build and maintain gazetteers from Named Entity Recognition by using Wikipedia" Workshop on New Text, 11th Conference of the European Chapter of the Association for Computational Linguistics. Trento, Italy, April 2006. Downloaded 11/6/06 from: http://www.sics.se/jussi/newtext/working_notes/09_toral_munoz.pdf



The PATTON Alliance Mission

The central mission of the Patton Alliance is to create a channel for deploying mature, federally funded technologies to government users, while avoiding the discontinuity and subsequent cost and loss of capability that often occurs between prototype development and system deployment. The PATTON Alliance will provide feedback to its sponsors on the operational utility of all candidate technologies that are tested and applied toward operational problems. Furthermore, it will actively seek matches between mature technologies and interested stakeholders. In its efforts to find a home for new technologies in operational networks, the PATTON Alliance will enhance the technical capability of the supported communities. The motto of the Alliance, then could be phrased, "We don't make the tools you use; we make the tools you use better."

To become a Patton Alliance member or for more information visit the Patton Alliance web site at <http://pattonalliance.sandia.gov> or contact:

Judy Moore, Ph.D., Program Manager

Phone: 505-845-9415

Email: jhmoore@sandia.gov

Rolf Carlson, Ph.D., Technical Lead

Phone: 505-844-9476

Email: recarls@sandia.gov

Denise Bleakly, Gazetteer Project Lead

Phone: 505-284-2535

Email: drbleak@sandia.gov



**Sandia
National
Laboratories**

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.