International Workshop on Digital Gazetteer Research & Practice
7-9 December 2006

Convened by

National Center for Geographic Information and Analysis (NCGIA)
University of California, Santa Barbara
and
Redlands Institute
Redlands, California

Sponsored by
National Geospatial-Intelligence Agency (NGA)

Held at
Upham Hotel
Santa Barbara, CA 93101
International Workshop on
Digital Gazetteer Research & Practice
7-9 December 2006

Convened by
National Center for Geographic Information and Analysis (NCGIA)
University of California, Santa Barbara
and
Redlands Institute, Redlands, California
Sponsored by the National Geospatial-Intelligence Agency (NGA)

The workshop takes place at
Upham Hotel
1404 De La Vina St,
Santa Barbara, CA 93101
(805) 962-0058

AGENDA

Thursday 12/7
All day – Arrivals; poster setup
5:30 pm - Welcome Reception (hosted bar, enough eats to hold you until dinner)
Sponsored by go2 Systems, Inc. (Lee Hancock, President)
6:00 - 7:30 Presentations
Mike Goodchild & Linda Hill (UCSB): welcome and introductions
Jordan Henk (University of Redlands)
Beth Driver and Randy Flynn (NGA)
Chris Rewerts (US Army Corps of Engineers, Engineer Research and
Development Center -Construction Engineering Research Laboratory)
Lee Hancock (go2 Systems, Inc.)
Jim Frew (UCSB): Keynote presentation
7:30 - Dinner nearby (on your own: e.g. Louie’s in hotel, or walk to options on
State Street)
Saturday 12/9

8:00 - 8:30  Continental Breakfast, Coffee and Tea
8:30 - 9:00  Plan for the day (Mike Goodchild & Linda Hill)

SESSION 3: INTEROPERABLE GAZETTEER SERVICES
(Session chair: James Reid, EDINA, University of Edinburgh)

9:00 - 9:30  Greg Janée (University of California, Santa Barbara)
9:30 - 10:00  Discussion
10:00 - 10:30  BREAK
10:30 - 11:00  Ruth Mostern (University of California, Merced)
11:00 - 11:30  Discussion
11:30 - 12:00  Response/comments by two discussants: Paul Ell (Queen’s University, Belfast) and Tom Elliott (University of North Carolina)
12:00 - 12:30  General discussion and summation

12:30 - 1:30  Lunch (provided; can be taken into breakouts)
1:30 - 2:30  Breakout Session I
2:30 - 3:00  Reports from breakout groups
3:00 - 3:30  BREAK
3:30 - 4:30  Breakout Session II
4:30 - 5:00  Reports from breakout groups
5:00 - 6:00  Concluding discussion and future directions

7:00 -  Conference Dinner at Opals, 1325 State Street, (805) 966-9676

Sunday 12/10

Sunrise Hike (optional)

8:00 - 9:30  Continental Breakfast, Coffee and Tea
10:00 - 2:30  Condor Express wildlife cruise of the Santa Barbara Channel (optional)
# DGRP Research & Practice Workshop Participants

Santa Barbara, December 2006

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dave Anderson</td>
<td>MITRE Corp</td>
</tr>
<tr>
<td>David Bodenhamer</td>
<td>IUPUI</td>
</tr>
<tr>
<td>Allen Carroll</td>
<td>National Geographic Society</td>
</tr>
<tr>
<td>Larry Carver</td>
<td>UC Santa Barbara</td>
</tr>
<tr>
<td>Thomas Connelly</td>
<td>Biblioteca del Congreso Nacional, Chile</td>
</tr>
<tr>
<td>Helen Couclelis</td>
<td>UC Santa Barbara</td>
</tr>
<tr>
<td>Mike Dobson</td>
<td>Telematics</td>
</tr>
<tr>
<td>Beth Driver</td>
<td>NGA</td>
</tr>
<tr>
<td>Paul Ell</td>
<td>Queen's University</td>
</tr>
<tr>
<td>Tom Elliott</td>
<td>University of North Carolina, Chapel Hill</td>
</tr>
<tr>
<td>Peter Fisher</td>
<td>City University</td>
</tr>
<tr>
<td>Randy Flynn</td>
<td>NGA</td>
</tr>
<tr>
<td>John Frank</td>
<td>MetaCarta Inc</td>
</tr>
<tr>
<td>Mike Freeston</td>
<td>UC Santa Barbara</td>
</tr>
<tr>
<td>Jim Frew</td>
<td>UC Santa Barbara</td>
</tr>
<tr>
<td>Bruce Gittings</td>
<td>University of Edinburgh</td>
</tr>
<tr>
<td>Michael Goodchild</td>
<td>UC Santa Barbara</td>
</tr>
<tr>
<td>Steve Guptill</td>
<td>US Geological Survey</td>
</tr>
<tr>
<td>Lee Hancock</td>
<td>go2 Directory Systems</td>
</tr>
<tr>
<td>Jordan Hastings</td>
<td>UC Santa Barbara</td>
</tr>
<tr>
<td>Jordan Henk</td>
<td>Redlands Institute</td>
</tr>
<tr>
<td>Greg Hill</td>
<td>University of Colorado</td>
</tr>
<tr>
<td>Linda Hill</td>
<td>UC Santa Barbara</td>
</tr>
<tr>
<td>Jerry Hobbs</td>
<td>University of Southern California</td>
</tr>
<tr>
<td>Greg Janée</td>
<td>UC Santa Barbara</td>
</tr>
<tr>
<td>Krzysztof Janowicz</td>
<td>University of Muenster</td>
</tr>
<tr>
<td>Chris Jones</td>
<td>University of Cardiff</td>
</tr>
<tr>
<td>Ray Larson</td>
<td>UC Berkeley</td>
</tr>
<tr>
<td>Naicong Li</td>
<td>Redlands Institute</td>
</tr>
<tr>
<td>Inderjeet Mani</td>
<td>MITRE Corp</td>
</tr>
<tr>
<td>David Mark</td>
<td>University at Buffalo</td>
</tr>
<tr>
<td>Marc-Andre Morin</td>
<td>Defence R&amp;D Canada</td>
</tr>
<tr>
<td>Ruth Mostern</td>
<td>UC Merced</td>
</tr>
<tr>
<td>James Reid</td>
<td>University of Edinburgh</td>
</tr>
<tr>
<td>Chris Rewerts</td>
<td>Army Research Office</td>
</tr>
<tr>
<td>Susan Stone</td>
<td>UC Berkeley</td>
</tr>
<tr>
<td>Bjorn Svensson</td>
<td>ESRI</td>
</tr>
<tr>
<td>Will Tefft</td>
<td>Map Link</td>
</tr>
<tr>
<td>Waldo Tobler</td>
<td>UC Santa Barbara</td>
</tr>
<tr>
<td>Paul Veiszze</td>
<td>CA Office of Emergency Services</td>
</tr>
<tr>
<td>Howard Veregin</td>
<td>Rand McNally</td>
</tr>
<tr>
<td>Xiaobai Angela Yao</td>
<td>University of Georgia</td>
</tr>
<tr>
<td>May Yuan</td>
<td>University of Oklahoma</td>
</tr>
</tbody>
</table>
Biographical Sketch
David J. Bodenhamer

The Polis Center at Indiana University Purdue University Indianapolis
1200 Waterway Boulevard, Suite 100
Indianapolis, Indiana 46202
intu100@iupui.edu

Current Appointment (1989-)
Founding and Executive Director, The Polis Center, and Professor of History, Indiana University-Purdue University, Indianapolis. Responsibilities: Plan and direct activities of a self-supporting 25-person multidisciplinary center; recruit, supervise, and evaluate professional and support staff; plan and manage center and project budgets; secure external funding; develop and supervise grant and contract research; represent center, school, and university to external agencies and public. Annual budget: $3 million.

Selected Relevant Publications and Websites

Selected Recent Presentations

Other Publications and Presentations
In addition to the titles listed above, I also am author or editor of seven books and author or co-author of twenty-five book chapters/journal articles. I have made more than sixty presentations to professional meetings on four continents and have served on three editorial boards (one current).
STATEMENT OF INTEREST

TO: Organizing Committee
   Digital Gazetteer Research and Practice Workshop

FROM: David J. Bodenhamer
      Executive Director, The Polis Center at IUPUI, and Professor of History

RE: Workshop Application

DATE: September 26, 2006

Please accept this memo as a strong expression of interest in the December 7-9, 2006, workshop on digital gazetteers. The opportunity could not be more timely nor my need greater.

Over the past decade—and especially over the last five years—I have been involved in a variety of projects related to the digital humanities. Most of these activities have involved electronic cultural atlases and historical GIS, especially the development of Web-based projects (e.g., the North American Religion Atlas, www.religionatlas.org). This work, in turn, has led to what is currently a loosely federated effort with EU partners to create a conceptual framework for an initiative to link various national historical GIS efforts. It also has resulted in a parallel but related effort to use GIS to locate and link an array of UK digital historical resources in multiple data formats. This effort, which involves several colleagues from the UK, will rely heavily on the use of digital gazetteers, some extant and others that must be created. My compelling interest in this workshop is to ensure that we enter our work on both fronts fully aware of the advances made by groups such as the National Center for Geographic Information and Analysis.

As the director of a center that sues GIS as its technology of choice, I also have a strong desire to know about the potential value of digital gazetteer research for our larger agenda. The Polis Center at Indiana University Purdue University Indianapolis, established in 1989, has developed a strong record for applying GIS effectively at local, state, and federal levels. A completely self-funded unit with 25 FTE staff, we have developed a comprehensive community information system for the Indianapolis MSA (www.savi.org), an internal Web-based mapping and evaluation center for the National library of Medicine and its National Network of Libraries of Medicine, and, in a work-in-progress, a system of distributed and integrated Web services for mapping and spatial analysis for a variety of state agencies and counties in Indiana. We also work extensively with federal and state emergency management agencies on disaster planning, including significant involvement in training users nationally on HAZUS, a GIS software developed for pre-disaster mitigation efforts. Finally, over the past several years we have increasingly teamed with researchers in the IU School of Medicine, located on our campus, to incorporate spatial analysis in public health-related projects. In all of this work, it is becoming clear that digital gazetteers are important resources that we must learn to tap effectively.
Allen Carroll  
Chief Cartographer  
National Geographic Society  
1145 17th Street NW  
Washington, DC 20036  

Voice: 202-857-7799  
Fax: 202-429-5704  
Email: acarroll@ngs.org  

Employment history:  

1998-present: Chief Cartographer and Executive Vice President, National Geographic Maps  
1995-1998: Managing Director, National Geographic Maps  
1991-1995: Art Director, National Geographic magazine  
1989-1991: Director of Design, National Geographic Cartographic Division  
1986-1989: Associate Art Director, National Geographic magazine  
1983-1986: Assistant Art Director, National Geographic magazine  
1976-1978: Editor, Connecticut College Alumni Magazine  
1973-1976: Environmental Analyst, Connecticut Department of Environmental Protection  

Education:  

BA, Environmental Studies, Connecticut College, 1973 magna cum laude  

At National Geographic:  


Conservation projects: Instrumental in conceptualizing and launching the Conservation Geoportal (www.conservationmaps.org) with ESRI, The Nature Conservancy, and other partners. The site provides shared access to spatial data relating to conservation. In 2001, conceptualized and designed Wild World, a collaboration with World Wildlife Fund, NGS, ESRI, and Ford Motor Company that distributed 10 ecoregions maps to every K-12 school in the U.S. and produced an interactive Web-based map (www.wwfus.org/wild-world)  

Presided over the shift of the Society’s cartographic unit from a division of the Magazine to the Society’s new taxable subsidiary, National Geographic Ventures. Led expansion of NG Maps from its traditional role as service provider to the magazine and book divisions of the Society to a publisher and distributor of map products. With partnerships and acquisitions,
Position Paper
Digital Gazetteer Research and Practice workshop
November 17, 2006

**A Case-Study-In-Progress: How a Media Organization Tackles the Georeferencing Challenge/Opportunity**

Founded in 1888 to increase and diffuse geographic knowledge, the National Geographic Society has grown into a multi-faceted organization producing editorial material in multiple media and across numerous business units. The Society, like other media companies, is working hard to make its digital assets more versatile and accessible amidst a rapidly evolving, increasingly fragmented media marketplace. The growing popularity of Web-based mapping platforms and consumer-oriented GPS applications have spotlighted a particular opportunity, namely organizing NGS content for presentation in the context of location. The Society’s venerable brand, its renowned cartography, much of its editorial content, and the core of its mission are all about geography. For all these reasons, the notion of georeferencing Society content has been an easy sell.

The process of implementing georeferencing, however, has proven to be anything but simple. Challenges include the vast scope and variety of content at the Society, its archiving in scattered locations across multiple business units, its access via a variety of databases with little or no common standards or nomenclature, and widely varying rights restrictions. Despite these challenges, progress is being made in the creation of an enterprise-wide infrastructure, and in pursuing project-based initiatives fueled by specific business opportunities, most of which complement and strengthen the enterprise effort.

Overall goals of the enterprise-wide effort are to create an infrastructure for organizing and accessing content by geography, including archival content and new content as it is created in the field or archived at the office. The infrastructure must allow for a variety of media formats, several database and metadata systems, and a range of potential applications including print, Internet, video, film, and mobile/GPS. An additional goal is to integrate georeferenced content with National Geographic’s cartography, potentially providing a unique competitive advantage.

The key to georeferencing legacy content, of course, is the placename or names associated with that content. National Geographic Maps’ placenames database, which is essentially the same as the index to the Eighth Edition Atlas of the World, was identified as the basis for a “master gazetteer” against which other databases would be “harmonized” or cross-walked. Disadvantages of the atlas index are the small number of placenames (some 140,000) compared with other gazetteers, and the fact that the features those placenames are associated with are cartographic (adjusted for visual display) rather than truly spatial. Advantages include the high editorial quality of the placenames and their association with a
Resumé

Thomas John Connelly, MA

email: tconnelly@bcn.cl

Born October, 1946, Chicago, IL.

Professional experience

2000 - 2006: Research and georeferenced products development for Congress persons and citizens in the Library of the National Congress of Chile. Team leader for gazetteer/thesaurus development, sharing and use.

These responsibilities have entailed participation in on-going discussions about the management of georeferenced information with government agencies (such as, the Military Geographic Institute (IGM); Planning Ministry (MIDEPLAN); Commission for the Environment (CONAMA); Coordination Committee for Territorial Information (SNIT)) and the preparation and internet publication of the first digital gazetteer/thesaurus published in and for Chile.

1990 - 2000: Research and production of analytical reports for Congress persons in the Library of the National Congress of Chile.

1974 - 1989: Catholic priest, member of Saint Columban’s Foreign Mission Society, diverse posts in parishes of the Archdiocese of Santiago, Chile.


Education

1973: BD. Saint John's College and Seminary. Boston, MA.

1986: MA. Anthropology. University of Texas at Austin.
Digital Gazetteer Research and Practice

Expression of interest

20 Sept. 2006

Thomas Connelly
Researcher
Sistema Integrado de Información Territorial
Biblioteca del Congreso Nacional
Chile
tconnelly@bcn.cl

Chilean context

The benefits of georeferencing information have become increasingly evident over the last five or six years. Indeed, the team tasked with coordinating government agencies in this regard (SNIT: Sistema Nacional de Información Territorial) has recently been given a new, more specific action brief. This growing awareness of the importance of the adequate management of georeferenced data has created the opportunity to promote the adoption of the best criteria and practices of the international community. Thus, we face challenges on two fronts: how to learn those criteria and practices and how to best further their adaptation to Chilean reality.

The Library of the National Congress has pioneered the promotion of the design and implementation of national policy for the management of spatial information, focussing on gazetteers as key components in the creation of metadata records and offering a gazetteer-cum-thesaurus for use and comment via Internet.¹ In fact, the Library Web site offers the only place-name-based search tool available for Chile (apart from GoogleEarth, of course).² Although we do make maps for the specific purposes of Congress persons and our Web site allows user interaction with our databases to create cartographic visualizations of statistical data, the Library is chiefly a consumer of georeferenced data and, thus, a manager of territorial information.

We have come to realize that gazetteers are a powerful instrument to this end and participate in the Chilean inter-agency discussion on what their creation and maintenance imply and how to use them adequately.

Creating metadata records

¹ See: http://geoinfo.bcn.cl/ (in Spanish).
Michael W. Dobson, Ph.D.

Expertise and Key Information

Senior executive with broad business experience focused on technology. Expertise includes product and services development, technology planning and management, IT management, wireless telephony, software engineering management, geographic information systems, Location Based Services, Local Search, mapping, publishing, telematics, organizational planning, strategic planning, market development and business development.

Excellent leadership skills coupled with a demonstrated ability to: 1) build highly efficient, results oriented technology and software development organizations, 2) structure, design and implement new products, systems and related policies and procedures on a cost efficient basis and 3) manage multiple, complex development tasks simultaneously.

- A strategic thinker with an outstanding ability to form strong, cooperative interpersonal relationships coupled with excellent communications and presentations skills.
- Recognized expert with over twenty-five years of experience dealing with issues and applications of technology for use in the development of market leading products, tools and services.
- Past member of the Mapping Sciences Committee of the National Research Council of the National Academy of Sciences.

Experience

2002 -2006
TeleMapics, LLC
Laguna Hills, CA

Founder and President

Served as the president founder and principal consultant of TeleMapics, LLC (www.TeleMapics.com), a consulting practice focused on: Local Search, Location Based Services, Wireless Location Based Services, the product development/product management process, strategic technical consulting, strategy development and organizational change, and uses of geographic information in business and government.

2001 – 2002
go2 Systems
Irvine, CA

Chief Technology Officer and Executive Vice President of Engineering & Technology

Served as corporate officer responsible for technology, engineering, IT and product development at go2 systems, the leading provider and distributor of business directories designed for the wired and wireless Internet. go2 provided directory services to MSN Wireless, AT&T wireless, Sprint, Nextel, Verizon and other major wireless carriers in the US.

- Managed a technical team of 70 software developers focused on delivering products to the wireless handset and online markets for the presentation of business directory information including maps and directions.
• Established strategic partnership with Istituto Geografico de Agostini and served as a Director of the Company (DeAgostini/Rand McNally). The Company was an international joint venture providing cartographic and software development services.

Vice President of Industry Affairs and Chief Cartographer
Led corporate communications effort to expand Company's presence in scientific and consumer venues.
• Set corporate standards and practices for product development.
• Participated as co-manager in development of Rand McNally's New Media Division, a highly successful software business unit which grew to 20 million dollars in revenue in its first 18 months, based on the market leading, award winning Trip Maker and Street Finder software.

1987-1992  Rand McNally and Company  Skokie, IL
Vice President of Creative Services and Chief Cartographer
Managed all product development activities (Editorial, Cartography, Art and Design and Digital Cartography) for the Company.
• Responsible for approximately 300 staff at 8 locations nationwide.
• Implemented conversion to advanced technology and GIS based mapping software from manual cartographic methods.

1986-1987  Rand McNally and Company  Skokie, IL
Director of Cartographic Services
Conducted requirements analysis to determine best practices in converting mapping operations from a manual to a digital environment. Managed Company's mapping centers and technical support for Company's commercial trucking and routing business.

1985-1986  Greenborne and O'Mara  Greenbelt, MD
Director of Computer Mapping
Successfully led efforts to capture new Geographic Information System and spatial software-based contracts with the Agency for International Development, Federal Emergency Mapping Agency and other clients.

1981-1985  Spad Systems  Reston, VA
Senior Associate
Consulted for the US Geological Survey, Bureau of the Census, Rand McNally and various clients on the application of GIS methods and applications to extremely large spatial databases. Key player in the SPAD team that wrote the Requirements Analysis for the U.S. Census at the start of the development of what would come to be known as TIGER.

1972-1985  State University of New York at Albany  Albany, NY
Assistant/Associate Professor of Geography
Taught courses and conducted research related to mapping, Geographic Information and Spatial Data Handling, computer technology and software.
My experience with gazetteers has spanned my careers as an educator, executive of a mapping company and as a central component in my consulting practice (TeleMapics).

I have practical experience in creating geographic gazetteers, indexes and in the problems and potentials in deciding “appropriate” name forms and publishing policies governing their use at Rand McNally & Company. Rand’s business was international and I was exposed to the pressures of “commercially-based name preferences”, as well as politically charged “naming” issues across the world. During my tenure at Rand, I licensed data to Getty for their Geographical Thesaurus and had discussions with a number of companies interested in geographic names and their use for purposes far beyond the world of maps. Perhaps of more interest, I was with the company when the name systems were automated for our internal use.

In my consulting career, which is focused on Local Search and Location Based Services, I have found that geographic gazetteers are an area of interest to all of the major players in Local Search, as well as those focused on general Internet search. In the last year, Yahoo acquired WhereOnEarth, a UK based company that created a unique digital gazetteer capable of aiding search engines to disambiguate geographical terms. The company’s products are now being integrated with Yahoo’s search engines in attempt both to sharpen the relevancy of results to queries that include geographic tokens and to assist advertising buyers with “buying” and targeting of advertisements that could enhance their sales on a geographic basis (and increase Yahoo’s profitability in the process).

I have spent considerable time, recently, examining many online gazetteers and speaking to developers of these products. In a make versus buy project for a client, I spent considerable time comparing gazetteers and creating a work plan for creating one from scratch using both proprietary and public domain sources.
Beth H. Driver

Beth H. Driver is the Scientific Advisor for Data Bases at the National Geospatial-Intelligence Agency (NGA). She currently serves as Chief of the Outreach Division in the Office of Basic and Applied Research and is responsible for administering and coordinating academic outreach programs, including the Agency’s academic research programs.

Dr. Driver was selected to the Defense Intelligence Senior Level service in 1993. Since that time she has led and supported numerous programs to adopt and exploit modern system development and data management practices and technologies. A particular area of interest has been practical ways to incorporate and use various bodies of existing, often-conflicting data.

Prior to assuming her current position she served as Technical Executive for the Acquisition Systems Office and for its predecessor organizations. She advised the GeoScout Program Manager and program team on architecture and programmatic matters related to the new acquisition model envisioned for the program and on achieving the open, data-centric architecture that is one of NGA’s goals.

She led efforts to improve Intelligence Community support for precision targeting by adopting more flexible production strategies, more rigorous practices, and developing a knowledge base to minimize production and support overhead, and to make wider use of non-traditional data sources.

She has led several studies on how to extend and foster commercial production of data, how to organize data storage and data production, and on overall architectures for NGA production and data management systems.

Before joining the government, she managed various defense and intelligence programs in the private sector, encompassing both strategic planning and development of data-intense systems. Problems addressed included capture, storage, and full-text search of large volumes of scientific and technical reports, including full page images, and use of structured data base management systems for nontraditional kinds of data, including geospatial information, foreign-language data, and bibliographic information. She designed numerous studies to evaluate data handling and dissemination alternatives.

She holds a Ph.D. and M.A. from the University of Texas at Austin and an A.B., summa cum laude, from Ohio University, where she was elected to Phi Beta Kappa. She has published papers, served on national study groups for such organizations as the National Research Council, and served on ANSI/SPARC standards committees. She received the Defense Mapping Agency Distinguished Civilian Service medal.

Dr. Driver lives in Arlington, Virginia.
Statement of Interest

The Digital Gazetteer Research and Practice workshop interests me for two reasons: digital gazetteers afford the opportunity to expand both the nature and the extent of NGA’s support to our customers; in addition, work on gazetteer questions affords an outstanding opportunity to deepen our collective understanding of the nature of and limits to formal descriptions of geospatial data and of commonality between expressions of geospatial knowledge and natural language. The questions posed in the workshop description illustrate the potential for addressing both “applied” and “fundamental” questions within a single, focused body of research.

Components of gazetteer services

The topics identified as focus areas for the workshop are likely to reveal a host of related questions, starting with “what is gazetteer data and, in a digital world, how does it differ from ‘regular’ geospatial data?” Does the difference lie in the data, in the assumptions that we make about it, or in the kinds of services offered? Which of the following would be gazetteer data: township names; zip codes; telephone dialing codes; designated service areas for commercial services; areas of responsibility for military or non-military organizations.

It might be useful to discuss the expanded utility of gazetteer data provided in digital form, along with scenarios for using new capabilities. Not only will digital services render it faster and easier to perform searches that theoretically might be done with hardcopy gazetteers, new capabilities might support entirely new services and expand the community of users for gazetteers. An extreme case might be to compare an excerpt of spoken language to a list of place names whose various pronunciations are indexed.

Such indexing might require capabilities well beyond today’s soundex representations, e.g., application of phonological rules for multiple dialects, in addition to grammar and morphological rules. More straightforward uses might be to find spoken and orthographic variants of a place name (within a designated dialect or sub-dialect or across languages), to find alternative referents for the name and to find (or recognize) alternative names for the “same” place.

Georeferencing as a process

Many of the relationships between bits of linguistic data, and bits of geospatial data, are implicit and context-dependent; a better understanding of such relationships and how to exploit them is essential. Explicit identification of “all” relationships will fail for two reasons: first, the cost of capture and exploitation would be prohibitive; more importantly, we cannot enumerate all of the relationships of interest for a given bit of data.

Traditionally, we have looked for general rules and relationships in grammars and in sets of logical operators that can be combined in useful ways. Linguists impose an “adequacy requirement” on grammars that we do not use in computer science: a grammar should
Paul S. Ell
Curriculum Vitae

Address: Centre for Data Digitisation and Analysis
School of Geography, Archaeology and Palaeoecology
The Queen’s University of Belfast
Belfast
Northern Ireland
BT7 1NN

Telephone: University: Direct line +44 (0)28 90973408
University: My secretary +44 (0)28 90973883
E-Mail: paul.ell@qub.ac.uk

FAX: University +44 (0)28 90973943

Posts

1991 – 1993 University of Leicester Leverhulme Research Associate
1993 – 1998, The Queen’s University of Belfast, Research Fellow working with the Database of Irish Historical Statistics, Department of Economic and Social History.

Current employment

1998 – The Queen’s University of Belfast, Senior Research Fellow and then Director of the Centre for Data Digitisation and Analysis (CDDA), School of Geography, Archaeology and Palaeoecology.

Major Grants received

1. ESRC, Crop Yields, Environmental Conditions, and Historical Change, with Professor BMS Campbell, £91,912, 2005-6.
2. Leverhulme Trust, Early Career Fellowship awarded to Dr I.N. Gregory to be based under Dr Ell’s supervision, £46,000, 2004-6.
5. JISC, British and Commonwealth Census Project £900,000 with AHDS, 2003-7
7. AHRB, EPPI: Enhanced British Parliamentary Papers on Ireland, 1801-1922, £290,782 with the University of Southampton, 2002-4.
8. ESRC, A Spatio-Temporal Analysis of the Great Irish Famine, £29,000 with the Universities of Portsmouth and Essex, 2002-3.
11. British Academy, Completing the Jigsaw: Digitising parish level census tables for Great Britain, £4,950 with the University of Portsmouth, 2001-2.
12. New Opportunities Fund - Digitisation, A Vision of Britain through Time, £620,000 with the Universities of Portsmouth and Leeds, 2001-3.
This is a formal request for a place at the Digital Gazetteer Research and Practice Workshop to be held at UC Santa Barbara. In addition to this expression of interest I attach a two-page résumé.

I am director of the Centre for Data Digitisation and Analysis at Queen’s University Belfast. The Centre’s main interest is the development of scholarly spatially referenced historical electronic research resources, and the exploitation of these large datasets in research using GISc. I have been involved in many large-scale projects relating to the construction of e-resources and their utilisation including the Great Britain Historical GIS, where I was a co-applicant on its main Economic and Social Research Council funding award; the associated ‘popular’ Vision of Britain Through Time incarnation of the GBH GIS with a variety of multimedia materials all referenced by place; the development of the Database of Irish Historical Statistics, which resulted in a significant publication mapping and analysing the geography of the 1840s Irish Famine; two new projects digitising, and associating by place name, medieval British sources including the 1086 Domesday Survey and a multitude of later medieval materials; a project which digitised and spatially referenced the first map of the whole of Britain – the Gough Map dating from around 1350; and the construction of three text-based resources containing geographical information – a virtual library of materials relating to the Act of Union between Britain and Ireland, a large repository of British and Irish Parliamentary Papers, and an e-version of the debates held in the Northern Ireland Stormont Parliament from 1921 to 1972. What is common to all these e-resources is that they can be referenced, and associated, by location.

In addition, I am co-author of a Cambridge University Press book that will be published in 2007 examining the use of GIS in history and geography research. I am an active member of the UC Berkeley based Electronic Cultural Atlas Initiative chairing ECAI’s Scholarship and Content Committee and the E-Publications Committee. At Queen’s University I head the Spatial Technologies Research Forum, a body created by Senate to promote GIS across the humanities and arts as well as the traditional disciplines in which the technology is used. I am editor of Humanities Computing, which supersedes the Edinburgh University Press journal History and Computing. The re-launched journal will have a strong multimedia GIS bias.

My primary interest in gazetteers is the potential to use them to associate disparate e-resources, such as those briefly described above, in space and time. As such elements of the workshop which focus the components of gazetteer services and on interoperability are particularly relevant. I recently received funding from the Arts and Humanities Research Council in the UK to hold a workshop on ‘GIS e-Science’ to discuss whether GIS technology might form the basis for organising and interrogating the rapidly developing range of e-resources in the Arts and Humanities, e-resources that are always, or almost always, referenced in space. It was quite clear from the workshop that if GIS was going to make a contribution to e-Science with relation to the Data Grid an effective way of linking information by location was vital. There are obviously many ways of expressing location as, for example, a defined polygon composed of administrative boundaries (as with the Great Britain Historical GIS or the Database of Irish Historical Statistics), a set of co-ordinates forming an arc or point (roads, rivers and castles as in my Gough Map project), or, and most commonly, an amorphous poorly conceived sense of location normally referred to by a place name (as with the Domesday GIS, the Medieval British Isles ‘Domesday II’ project, the various Parliamentary Paper projects etc).

In the UK I am leading a consortium (including AHDS, AHDS History, UC Berkeley, IUPUI, Portsmouth University and Nottingham University) that is making application to the Arts and Humanities Research Council under their e-Science grant call. AHRC have focussed the call on the development of e-Science infrastructure or substantial primary research using e-Science methodologies. The project I am working on does both. It will develop a comprehensive hierarchal place name gazetteer and use the gazetteer to advance
Dr. Thomas R. Elliott, Ph.D.: Director, Pleiades Project (thomase@email.unc.edu)

Tom Elliott (b. 21 Oct 1967, Wichita, Kansas, U.S.A.) entered Duke University (Durham, North Carolina; 1985-89) on a 4-year U.S. Air Force Reserve Officer Training Corps scholarship. He graduated, cum laude, with a Bachelor’s of Science degree in Computer Science and a second major in Classical Studies. E. then served as a Communications and Computer Systems Officer, U.S. Air Force (Grissom Air Force Base, Indiana; 1989-92), with successive posts as Chief of Communications Maintenance and Operations, and Base Closure Planning Officer. Upon separation, E. received the Meritorious Service Medal and was promoted to Captain in the Inactive Reserve. As a computer scientist for AEGis Research Corporation (Huntsville, Alabama; 1992-95), E. led the development of the primary engineering and training simulator for the U.S. Army’s Avenger Weapons System.

A prestigious Morehead Fellowship facilitated E.'s entry into the graduate ancient history program at the University of North Carolina (Chapel Hill; 1995). As a graduate research assistant for the Classical Atlas Project (www.unc.edu/depts/cl_atlas; 1998-2000), E. prepared the alphabetical gazetteer and assisted with the Map-by-Map Directory for the Barrington Atlas of the Greek and Roman World (R. Talbert, ed., Princeton, 2000). He also founded the international EpiDoc Collaborative (epidoc.sf.net), which develops standards for the digital encoding of Greek and Latin epigraphical texts. In Aug 2000, he was appointed Founding Director of the Ancient World Mapping Center (www.unc.edu/awmc). As Director, he supervised all aspects of the Center’s activities, including a successful proposal to the U.S. National Endowment for the Humanities that secured startup funding for the Pleiades Project: an on-line workspace for ancient geography (icon.stoa.org/trac/pleiades). In Feb 2006, E. stepped down as Center director to assume full-time leadership of Pleiades.

Select Publications and Papers:


Numerous original map designs and digital map production for publication in monographs, journal articles, and educational games under the auspices of Classical Atlas Project and AWMC, 1999-2005 (www.unc.edu/awmc/mapsforstudents.html).

Epigraphic Evidence for Boundary Disputes in the Early Roman Empire, Ph.D. Dissertation (History), Univ. of North Carolina at Chapel Hill, 2004.


Tom Elliott, Ph.D.
Director, Pleiades Project
Ancient World Mapping Center
University of North Carolina at Chapel Hill

I am interested in issues of collaborative geodata creation and maintenance, and interoperability of geoservices for humanistic research and teaching. I am particularly interested in the tension between new community-based approaches to content creation and well-founded traditional concerns about accuracy, citation and verifiability. Two other areas of concern are: cross-project collation and georeferencing of toponymic gazetteers, and feature extraction and georeferencing from historical maps and cartifacts that cannot be rubbersheeted usefully.

As the director of the Ancient World Mapping Center’s Pleiades project, I preside over an effort to build a collaborative, web-based system for the perpetual update and diversification of a complex geo-historical dataset for Greek and Roman antiquity. The dataset itself derives from the work of the Classical Atlas Project (1988-2000). Comprising more than 50,000 named and unnamed geographic features, it was compiled and vetted by a 200-person international team of scholars and professional cartographers. Their initial goal was the creation of a scholarly reference atlas in print, and this goal was realized in 2000 with the publication of the Barrington Atlas of the Greek and Roman World (R. Talbert, ed., Princeton; see www.unc.edu/depts/cl_atlas). The Center’s core mission, from its foundation in 2000, has been the perpetual update, diversification and dissemination of this legacy.

With initial funding from the National Endowment for the Humanities (2006-2008), Pleiades combines open, community-based content development approaches with rigorous editorial review. Pleiades will soon enable anyone — from scholars to casual students of antiquity — to suggest updates to geographic names, descriptive essays, bibliographic references and geographic coordinates. Once vetted for accuracy and pertinence, these suggestions will become a permanent, author-attributed part of future publications and data services. These will include OGC-compliant Web Mapping and Feature Services (WMS/WFS), as well as a geocoding/gazetteer service sensitive to the full range challenges posed by our toponyms: fragmentary witnesses, scholarly hesitation in assignment, variant orthographies and scripts, etc.

To support the Pleiades “community of practice” we are customizing an enterprise-quality content management system (plone.org) by adding custom “content types” to handle structured spatial, toponymic and bibliographic records (beta site with early results: icon.stoa.org/pleiades-beta). Robust version control, document history and threaded comment-and-review mechanisms will facilitate not only granular and incremental updating of individual records, but also large-scale expansion and diversification of our holdings. A flexible, dynamic mapping tool will permit on-the-fly visualization of arbitrary subsets of the dataset (including query results). Working groups will facilitate collaboration on topics of group interest.

Plans are now in preparation for a multi-year, collaborative effort to leverage this dataset and its maintenance environment to establish a reliable digital infrastructure for Greek and Roman geography. This effort will involve a number of major digital projects that are cataloging and publishing documentary and archaeological resources (e.g., for inscriptions, papyrus documents, coins), as well as critical reference resources (e.g., for personal names and prosopography). Most such projects in the field of Greek and Roman history have limited budgets and staff and therefore
Where is that place? Modelling spatial footprints for a gazetteer

by Peter Fisher

Department of Information Science
City University, London EC1V 0HB, p.fisher@city.ac.uk

Historically gazetteers have been lists of place names which form the index of maps. Thus the classic gazetteer is bound as part of the atlas, sometimes as a single gazetteer for the whole Atlas and sometimes multiple gazetteers are bound in for different parts of the atlas whether single or multiple maps. The gazetteer is the principal means by which users of the atlas conduct guided searches to access the information in the atlas. Clearly another popular method for accessing the information is unguided search where the user has a general idea of the geographical region in which their place of interest is located and they use the map index sheets to find the correct map and then search on that map using local knowledge to assist refining the search. The only notion of space in these historical gazetteers is of a point or grid cell on a particular map; the point where the name of the place occurs or the cell which includes it.

A Gazeteer is becoming a list of the named features on a map and therefore is also an official or semi-official record of “place”. But in society “places” are created and destroyed. I do not mean physically destroyed although that of course can happen but I mean rather erased from the cultural record – the name goes out of use. Indeed physical destruction does not mean that the name goes out of use.

But how are gazetteer entries collected? For example, I know that the Ordnance Survey of Great Britain originally collected names systematically from local worthies who assisted drawing sketch maps of the extent of the features. This information has never been digitised and is now lost to all but researchers. These sketch maps were used in deciding which names to include on maps and the hierarchy. This information is now being stored in the National Archive at Kew.

If the names on maps are an official recognition of places then why do names on maps come and go and come back through multiple editions of maps within a short period of time? Why do different editions of maps have different hierarchies of placenames encoded? Why do the positions of places in that hierarchy vary among maps? How is a particular hierarchy decided? Ultimately why is a particular name on any particular map?

Searching gazetteers can produce ludicrous results for inquirers. If you look for the Marsall Islands in the gazetteer accompanying Google Earth, you do not see islands but an extent of water. If you search for Ilkley Moor in the Ordnance Survey electronic gazetteer, you are offered three different locations all within Ilkley Moor but in most interfaces you have to choose just one of them. Similarly if you search for a particular street in an A-Z atlas for the UK, it is referenced by a single cell on the map even if the street itself crosses multiple cells and even multiple maps. The same think happens on Multimap.com or Mapquest.com. The same problem was associated with the original GNIS where all rivers were located by a single location.

All these searches have one thing in common. A more flexible and appropriate response would not be to show the one index location for the feature in question, but a map that shows the full spatial extent of the feature. At the simplest the minimum bounding rectangle for the feature will provide this. To achieve this only requires that the every entry has two coordinate pairs instead of one.

I cannot be exhaustive, but I am aware that some modern gazetteers have gone further than this. The Getty Thesaurus of Geographic Names, for example, does not just have the name of inhabited places with a single grid reference but the hierarchy of named places in which they fall is stored.

Further enhancement of the gazetteer would be to have variable interpretations of what is meant by the term. So a query for “I want to see exactly the extent of the City of Leicester” would result in one response from a map server, while “Where is Leicester?” would result in a different map showing a more general context for the city. As long as the city is discernable on the map one that shows anything up to the full extent of the globe provides information to the inquirer. So should an
Randall E. Flynn

Biographical Sketch

Mr. Flynn serves as the Geographer of the National Geospatial-Intelligence Agency, a position he assumed in 1993.

Mr. Flynn holds the degrees of Master of Arts (1977) and Bachelor of Arts (1975) in Slavic Languages and Literature from the University of Virginia, and has also studied at Indiana University and as an exchange student at Leningrad State University (now Saint Petersburg State University, Russia). Through the Defense Mapping Agency's Long-Term Full-Time Training Program, he participated in the graduate-level program in Geodetic Engineering at Virginia Polytechnic Institute and State University, with a concentration in estimation theory (1985).

Mr. Flynn joined the Defense Mapping Agency in 1979 as a geographic names and boundary analyst and has since served in numerous capacities at the National Geospatial-Intelligence Agency and its predecessor organizations. From 1993 until 2005 he also served as the Executive Secretary for Foreign Names with the U.S. Board on Geographic Names. In 2005 he was appointed the principal Department of Defense member to the Board.

He has lectured on geographic names database design and national names standardization at the State Bureau of Surveying and Mapping in Beijing, China, the Main Administration for Geodesy, Cartography, and Cadastre in Kiev, Ukraine, the National Land Board in Riga, Latvia, and other national and international venues. He is a U.S. delegate to the United Nations Group of Experts on Geographical Names (UNGEGN), and is the UNGEGN liaison with the International Hydrographic Organization. Within UNGEGN, he has served as the convenor of the Working Group on Toponymic Data Files and Gazetteers. He has been an instructor in the annual PAIGH course on applied toponymy, having taught courses on geographic names in Bolivia, Honduras, Paraguay, and Peru. He is a member of the Federal Geographic Data Committee's Subcommittee on International Boundary and Sovereignty Data.

Mr. Flynn resides in Oakton, Virginia.
level to urban-level, and even to cadastral level 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 underway 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 (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 suffer 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
John R. Frank  
Founder and Chief Technology Officer

John founded MetaCarta in 1999 while beginning work on his Ph.D. in physics as a Hertz Fellow at the Massachusetts Institute of Technology. While studying the microclimates of forests, John encountered a need for a new way to view collections of documents: geographically. John's vision has led MetaCarta to its current position as the pioneer of unstructured geographic information retrieval. At MetaCarta, he focuses on new technology explorations and partnerships. He speaks to technical audiences and advises on implementation strategies.

MetaCarta's Office of the CTO (OCTO) develops new concepts for bridging unstructured information and maps. In 2005, OCTO started the OpenLayers project, which has become a widely used open source javascript library for combining geographic data layers from many sources. This and other OCTO projects can be found at labs.metacarta.com and in MetaCarta's commercial products.

John holds a B.S. in physics with distinction from Yale University where he led Yale's first solar car team to the top rookie seat of Sunyrace 97. In addition to his role as CTO at MetaCarta, John is continuing his graduate studies in condensed matter theory at MIT.
does not capture the essence of what people mean by locations like Aix-en-Provence or Rome or the Great Wall. One naturally thinks of history, political hierarchy, and other relationships that mankind has with that entity.

Is the oil platform known as "Jack" simply a longitude-latitude bounding box? Oil platforms' most striking features are their drills stretching into the depths and changing orientation during operation. Besides this geometric complexity, when such an entity enters the consciousness of more people, its name takes on new meaning. Just as Cazumel means "vacation" in a somewhat generic sense, the discovery of oil at Jack has taken on significance independent of its literal location – for some people.

In the process of communicating, people add attributes to location entities. As a location gains importance, communities wander away from asserting rigorous gazetteer definitions. Definitions escape. The line between defining and describing fades.

Locations often come into existence as bookkeeping conveniences. As they progress from mere gazetteer entries up the fat-tailed distribution of awareness, locations gain personality. Only natural language can capture the richness of these attributes.

What defines a location? The collective awareness of many people defines a location. Structured gazetteers cannot capture this – at least not without true artificial intelligence.

The emerging field of geographic information retrieval offers a pragmatic way forward. Instead of engineering more and more knowledge into gazetteers, we gain more ground by acknowledging the boundary where gazetteers stop and unstructured GIR takes over.

Gazetteers form a fundamental part of GIR. Articulating useful boundaries between these two types of tools will foster progress in both areas of effort. GIR offers a bridge toward cartography and information presentation, which connects gazetteer efforts to people whose awareness drives the organization of gazetteers.
MICHAEL FREESTON

Institute for Computational Earth System Science
University of California Santa Barbara
Santa Barbara CA 93106
Tel: +1-805-893-8589
email: freeston@icess.ucsb.edu

Visitng Professor
Dept. of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ
UK

Brief bio:
Michael Freeston has almost forty years experience of academic and industrial Computer Science, specializing in database and knowledge base systems, and is best known in the international database community for his work on multi-dimensional indexing methods. He began his computing career as a programming assistant to Prof. Sir David Bates at Queens University, Belfast in 1967. He became a lecturer in Computer Science at Southampton University, UK in 1971, where he also managed a start-up database software company for several years. In 1985 he joined the European Computer-Industry Research Centre in Munich, Germany, where he developed his long-standing interest in database indexing techniques and their potential application to deductive database systems. He moved to the University of California at Santa Barbara in 1996 to work on the development of the Alexandria geo-referenced digital library. In 1998 he was appointed Professor of Information Science at Aberdeen University, Scotland, but retained a research link to the Alexandria project. He returned full-time to UCSB in 2002 to work on the successor to the Alexandria project, ADEPT until the close of this project in 2005.

Over the past twenty years he has been a frequent member of international program committees for database and digital library conferences and a reviewer for the US National Science Foundation, the European Commission, and journal publications.

Education:
1963 B.Sc. Physics, Leicester University, U.K.
1970 M.Sc. Atomic and Molecular Processes, Queens University, Belfast, N. Ireland, UK.
1997 Ph.D. Computer Science, Southampton University, U.K.

Appointments:
2002-5 Project Coordinator, Alexandria Digital Earth Prototype Project (ADEPT), University of California, Santa Barbara.
2002- Visiting Professor, Department of Electronics and Computer Science, University of Southampton, UK.
1998-2001 Professor of Information Science, University of Aberdeen, Scotland.
1992-1996 Head of Technology Transfer and Integration, ECRC, Munich, Germany.
1990-1992 Head of Knowledge Bases Group, ECRC, Munich, Germany.
1985-1990 Research Scientist, European Computer-Industry Research Centre (ECRC), Munich, Germany.
1982-1985 Managing Director, Solent Information Management Systems Ltd.
1970-1985 Lecturer (Associate Professor) in Computer Science, University of Southampton, U.K.
As a computer scientist with a long-held interest in database technologies, my interest in gazetteers stems from my work over the past twenty years on spatial database systems and georeferenced digital libraries. — notably the Alexandria digital library. For much of this time, and certainly the past ten years, it has been a puzzle to me why the potential power of search by georeference has not been a more widely recognized and exploited paradigm.

In particular, it is at least at first sight surprising that neither the academic nor commercial database research community has so far developed adequate support for this paradigm. Within a database management system (DBMS), it is still hard if not impossible to represent the varying degrees of accuracy of georeferenced information, and there is certainly no direct support for an ellipsoidal frame of reference. Geospatial indexing and querying in database systems remains poor. The relative lack of commercial interest in this direction is however not hard to trace: a schism developed between the GIS and database communities as long ago as the 1960’s, when the GIS community found that database technology was not able to support the basic functionality it required. In particular the GIS focus on geospatial visualization was not seen as relevant to the requirements of business database systems of the time. Nevertheless, as GIS moved from batch processing to real time, and the scale of GIS data management increased by orders of magnitude, GIS systems developed ‘loose couplings’ to DBMSs. Even so, at the conceptual level at least, the representation, querying and visualization of geospatial objects remained — and remain today - a function of the GIS system rather than the DBMS. For example, the components of the representation of a spatial extent in ArcInfo can be stored in a loosely-coupled DBMS, and search operations on such extents utilize the standard database index methods (of which, until only ten years ago, none were inherently spatial). But the interpretation and manipulation of these components as parts of a single spatial object, and its visualization, remain within the GIS system.

The complete integration of GIS and DBMS therefore remains an unsolved problem. But the world has moved on, and now the problem has become much more challenging. Two particular aspects interest me:

1) the use of ontologies to express not only a wide variety of geospatial data types, but relationships between them;

2) the representation, manipulation and querying of geospatial data on a global scale;

It is now widely recognized that there is a need for the creation of an exhaustive domain ontology in almost every field of knowledge. In the case of geospatial information, the humble gazetteer has now grown into a fully-fledged ontological framework. Much work has already been done on defining a gazetteer content standard and access protocol. The challenge now is to integrate this framework, and a reasoning engine over this framework, into a DBMS.

However, this cannot be satisfactorily achieved without addressing the fundamental problem of how to represent and manipulate geospatial data within a DBMS. I am particularly interested in the development of a geospatial partitioning and addressing mechanism based on the recursive division of the surface of the Earth into a hexagonal grid. The objective is to find a way — or at least a satisfactory compromise between conflicting requirements — of representing geospatial extents accurately at different scales.
Digital Gazetteer Research and Practice: Position Paper
Bruce M Gittings, University of Edinburgh

Scotland and Gazetteers

"Next to a good dictionary, the most generally useful book is a good gazetteer"
-- W.G. Blackie (1855)

Through the 19th Century Scotland maintained a strong tradition of the production, and use, of quality gazetteers. These volumes described the geography of Scotland, Britain and indeed the World. In terms of Scotland itself, the culmination of the art was undoubtedly the *Ordnance Gazetteer of Scotland* compiled in 1885 by Francis Groome, which still remains the standard geographic reference text to be found in major libraries on Scotland. This work was updated in the 1890s to reflect significant administrative changes, with a final edition published posthumously in 1903. Since then, the descriptive gazetteer has largely been replaced by the tourist guide and the place-name list (*short-form gazetteer*), the latter being the form which is most familiar to those of us involved in geographical information science (GIS). The tourist guide is no replacement however, in that they often introduce unacceptable bias, either in terms of selection of places described or what is actually said. Yet, with likes of Google Maps and Microsoft local.live are finding that basic mapping and vertical aerial photography have their limits and are clamouring for information. Thus, the need for a richer geographical description (through imagery, text and enriched databases) is becoming a focus of attention. Equally a broader community interests in heritage, local history and genealogy are developing apace. The importance of family history tourism to the Scottish economy is becoming significant, reflected by government resources being out into initiatives such as Scotland's People (www.scotlandspeople.gov.uk). Also initiatives such as the Millennium-lottery funded Scottish Cultural Archive Network (SCRAN) has given rise to the digitisation of a range of heritage resources, including papers, artefacts, art-works, books and maps held in a range of private and public collections. What connects all of these together is geography and place. Yet, in the United Kingdom, we have no authoritative source for place-names information and, worse, librarians and others without a geographical understanding are beginning to propose 'solutions' which range from the unworkable to the bizarre.

The Gazetteer for Scotland Project

Thus, while high quality geographic information in the form of maps and images have gained widespread acknowledgement as an important resource, descriptive information is becoming rare - largely because it needs expert collation, writing and editing - and texts on Scotland were becoming increasingly out-of-date. The only topographical directories for Scotland produced this Century have been place-name lists published, for example, by the Ordnance Survey and the Registrar General for Scotland and the Johnston's Gazetteer of Scotland (last updated 1973) which is limited in content and now out of print.

Against this background, David Munro (now Director of the Royal Scottish Geographical Society) and I created *The Gazetteer for Scotland* in 1995. Initially intended to be published as a book, the project was soon directed towards the Web, which was a new medium at the time. This decision has had a number of advantages. As the magnitude of the project became clear it was increasingly obvious that the amount of information produced would be well beyond the limits imposed by a modern-day publisher. Equally, the web permitted the incorporation of a large number of georeferenced photographs, allowed draft text to be modified following exposure to a critical public and the inclusion of a variety of interactive facilities which make the information more accessible. Thus the Gazetteer for Scotland has been built as a database based on Oracle, making use of its
interpretation of place based on historical events. To allow public input / revision is one thing, but at some point the professionals need to take over.
Biographical Sketch
Stephen C. Guptill

Stephen Guptill is a Senior Research Physical Scientist at the U.S. Geological Survey. Guptill is internationally recognized for his contributions dealing with geographical information science. Currently he is directing the Center of Excellence for Geospatial Information Science which was established within the USGS Geospatial Information Office in 2006. In recent years he also has been conducting spatial analysis research with colleagues from the Centers for Disease Control and Prevention to determine the environmental influences on emerging zoonotic and vector borne diseases.

Guptill received a Ph.D. (1975) and M.A. (1974) in Geography, from The University of Michigan, Ann Arbor, Michigan and a B.A. (1972) in Chemistry and Geography, from Bucknell University, Lewisburg, Pennsylvania.

Guptill is a member of the Editorial Boards of the International Journal of Geographical Information Science (IJGIS) and Cartography and Geographic Information Systems. He serves as a member of the Steering Committee for the National Consortium on Remote Sensing in Transportation. In addition, he served as the Department of Interior member on the National Science and Technology Council Subcommittee on Health and the Environment.


Guptill is co-author of Elements of Cartography, 6th Edition, and co-editor/author of Elements of Spatial Data Quality. He has written over 80 articles and papers appearing in the natural science and public health literature.

Contact information:

U.S. Geological Survey
MS 810, National Center
Reston, VA 20192
phone: +1 703 648 4520
mobile: +1 703 608 3664
fax: +1 703 648 4603
email: sguptill@usgs.gov
Background

The U.S. Geological Survey developed the Geographic Names Information System (GNIS) for the U.S. Board on Geographic Names as the official repository of domestic geographic names data. Confusion and controversy about geographic names and their applications to places and features led President Benjamin Harrison to establish the U.S. Board on Geographic Names in 1890. That early Executive Order was based on recognition that conflicts in naming geographic features were, in fact, a serious detriment to the orderly process of exploring and settling this country. A later decision, in 1906, by President Theodore Roosevelt to extend the responsibilities of the Board to include standardization of all geographic names for Federal use was a far-reaching decision that, coupled with the Harrison order, forms the foundation for the present organization of the U.S. Board on Geographic Names established in Public Law 80-242, signed by President Truman in 1947.

GNIS is the official vehicle for geographic names use by all departments of the Federal Government and the source for applying geographic names to Federal electronic and printed products. GNIS can be accessed at http://geonames.usgs.gov. The GNIS contains information about physical and cultural geographic features of all types in the United States, associated areas, and Antarctica, current and historical, but not including roads and highways (under statute, the Board on Geographic Names has purview over road and highway names, but has chosen not to execute that authority). The database holds the Federally recognized name of each feature and defines the feature location by state, county, USGS topographic map, and geographic coordinates. Over 1.9 million features are represented in GNIS. Other attributes include names or spellings other than the official name, feature designations, feature classification, historical and descriptive information, and for some categories the geometric boundaries.

Components of Gazetteer Services

Geospatial Location:

Most features in GNIS are associated with one location (e.g. summit, structure, lake, populated place) even though that feature may have an areal extent. Streams have a location for the source and mouth. The extent of an areal feature can be estimated from a set of coordinates that identify a series of geographic tiles with one point per USGS topographic map containing the feature. More accurate geographic descriptions of the features are not maintained within the gazetteer, but are contained in thematic geodatabases. For example, the National Hydrography Dataset (NHD, http://nhd.usgs.gov), which is the USGS’s most comprehensive set of digital spatial data about surface water features, is closely integrated with GNIS. The hydrographic feature names contained in and displayed by the NHD are from the GNIS. When partners submit new data to NHD, feature names are validated against the GNIS. To meet legal and policy requirements of the Board on
Lee Hancock
Chief Executive Officer and Founder
go2

Lee Hancock is the CEO and founder of go2®, a leading mobile local search and content company in the U.S., with industry leading distribution through Alltel, Cingular, Sprint, and Verizon wireless carriers. He is one of the early pioneers in the mobile, local search and content industry, having started and launched go2, the world’s first mobile, local search directory in 1999. He is the co-inventor on ten U.S. patents with 1996 priority dates covering various aspects of automatic location based services and location referencing technologies.

From go2’s initial public launch in 1999, Mr. Hancock led go2 to become one of the most-trafficked mobile website networks in the U.S. go2 will soon announce that in November, 2006 it delivered its one Billionth page view of local search and information to mobile phone users in the U.S.

From 1989 to 1998, Mr. Hancock was a partner with the 200-person California law firm of Allen, Matkins, Leck, Gamble and Mallory. He was the partner in charge of the firm’s corporate law department from 1995 until he resigned to devote full time efforts to go2. Mr. Hancock specialized in the fields of tax, corporate securities, mergers and acquisition and the representation of venture financed technology companies.

Mr. Hancock received the Ernst & Young Entrepreneur of the Year award in Orange County, California in 2000. He is also a co-founder of inBuilding Systems Corp., a provider of high-speed Internet access for multi-tenant office buildings, which was successfully sold in 2000. He has served on the boards of directors and assumed key leadership roles in several charitable organizations and business associations, including the Orange County Community Foundation.

Mr. Hancock, who is also a CPA, received a B.S. in Accounting summa cum laude from Missouri State University in 1975 and a J.D. cum laude from Southern Methodist University School of Law in 1979.
In 1994 it became apparent to me that technology was going to cause a monumental increase in devices and systems that would be able to precisely determine location and to create, store, organize and provide access to incredible quantities of local information. It also occurred to me that location determining capabilities and devices (primarily GPS) would eventually be available to the masses but that their usability and adoption could be enhanced greatly by the development of new and user friendly geo-referencing and user interfaces.

My initial concerns focused on ensuring that a variety of consumer devices, including cellular phones, watches, etc., were optimized with ‘dial tone’ friendly, simple and easy user interfaces and referencing systems. It soon became apparent, however, that the acquisition, storage, organization, manipulation and dissemination of accurate and comprehensive local information were also problems in desperate need of addressing. The analogy I have often used is that while GPS and other technologies were clearly creating new and very powerful “engines” that could revolutionize local information capabilities for the masses, inadequate efforts were being made to create a corresponding increase in the “fuel” for these new and revolutionary engines. Accordingly, the new GPS and other location technologies were in desperate need of new systems and interfaces to fuel these new capabilities and allow them to more quickly achieve their full potential.

The original efforts in 1994 resulted in the filing of an initial patent in August, 1996, significant efforts to educate and evangelize to the GPS and GIS communities to address the issue in from 1998 to 2002, and the launch of go2®, the world’s first location-based directory available over mobile phones, in 1999. This latter product initiative and go2’s technology and business model ultimately achieved broad distribution across virtually all major wireless carriers in the U.S. (AT&T, Cingular, Nextel, Sprint, and Verizon) and attracted the attention of significant industry resources from SAIC, Verisign, and Amdocs as strategic partners and investors. Like many high technology companies started inside the dot-com bubble, however, go2 was unable to sustain its growth and had to retrench significantly during 2002.

Notwithstanding significant downsizings and various legal restructurings, go2’s mobile, local search and directory applications and services managed to survive to cross the proverbial market adoption ‘chasm’ to better times in 2005. In 2006, go2 was recently
PERSONAL SUMMARY
Accomplished practitioner, manager, and educator, with a diverse background of academic and public-sector experiences in management of geographic/scientific data.

The unifying theme in my career is geocomputation, which spans computer algorithms, database management, geographic information systems, scientific modeling and visualization. In addition, I am an experienced project manager, having led a number of database and GIS/mapping activities for DOE, EPA, NOAA, NSF, and USGS. I also have participated broadly in the university research and teaching community. My current research focuses on geodatabase designs for digital gazetteers and geologic maps.

EDUCATION
Ph.D. Candidate, University of California Santa Barbara, 2006 expected
M.S. Geography, University of Nevada, Reno, 1998
M.S. Computer Science, University of Colorado at Boulder, 1974
B.S. Mathematics, University of San Francisco, California, 1970

RELEVANT EMPLOYMENT / ACCOMPLISHMENTS
Staff Researcher, University of California, Santa Barbara, CA; part-time, 2006 to present
   Doctoral candidate; developing advanced gazetteer (place referencing) systems to support data mining, information retrieval and Web search applications
Interim Director, Info. Systems, Tahoe Regional Planning Agency, Stateline NV; 2004-2005
   Restarted IT operations after a mass resignation. Managed the Agency’s information systems and resources overall; prepared State budget requests; began planning for enterprise GIS
   Developed databases and software tools for a national library of geological maps; prototyped these systems in a collaborative GIS for the Lake Tahoe region
Chief Scientist / Research Director, go2 Systems, Inc., Irvine, California; 1999-2001
   Established company’s research unit in Santa Barbara, employing 6; implemented a novel, patented georeferencing system for use in Web and wireless location-based services
Research Assistant Professor, Computer Science, University of Nevada, Reno; 1996-1999
   Instructed courses in database management and geographic information systems; supervised four Masters students; conducted independent research with NSF and USGS
Assistant Research Scientist, Data Manager, Desert Research Institute, Reno, NV; 1992-1996
   Developed GIS/database applications for tracking nuclear cleanup at the Nevada Test Site; founding GIS/data manager for McMurdo (Antarctica) Long-Term Ecological Research site
Senior Analyst, Lockheed Engineering and Sciences Company, Las Vegas, NV; 1991-1992
Professional Research Assistant, University of Colorado/CIREs, Boulder, CO; 1987-1991
Position Statement
DGRP Workshop

I regard a digital gazetteer (DG) as the quintessential geographic information system (GIS) having just one kind of feature, place, with two required descriptive attributes, a free-text placename and a categorical placetype. The geospatial attribute, footprint, is interpreted directly by the GIS. Despite this apparent simplicity, a DG pushes GIS technology in many ways: footprints may be approximate, placetypes are often fuzzy, and placenames have important linguistic connotations. All three descriptors are time-dependent.

Moreover, a single place may be described by multiple placenames, placetypes and/or footprints concurrently. A critical – perhaps the critical – task in gazetteer construction, therefore, is place identification, i.e. determining when two (or more) different descriptions in fact apply to the same place, and conversely when the same descriptions apply to different places. Bigler, Daowaga, Tula Tulia, and Lake Tahoe are all names for the same place, whereas Lake Geneva might be a city in New York, or in Wisconsin, or a water body in Switzerland. Place identification also presents itself after the fact, in that need to detect and remove entries for “duplicate” places from an extant gazetteer, or alternatively to conflate them into a single entry. Finally, place identification is essential to federating and ranking results from distributed gazetteer queries.

Objects with indeterminate boundaries, differing names, fuzzy types, time dependencies, etc. appear in GIS applications other than gazetteers, of course. Coming to grips with these issues in the context of DGs is appealing for two reasons: 1) because of their relative simplicity as GIS, gazetteers cameo the underlying knowledge engineering and data management issues, without loss of generality; and 2) gazetteers are increasingly utilized in data mining, information retrieval and Web query applications, so correct place identification and conflation have practical importance.

[end]
Jordan Henk

Professional

Current Position: Director, Redlands Institute, University of Redlands

Previous Positions of Interest: Senior Technical Consultant; Environmental Systems Research Institute (ESRI); GIS Department Director; Lancaster County, PA

Education

M.S. Geography, Pennsylvania State University
B.A. Anthropology, Pennsylvania State University

Work

Principal Investigator for University of Redlands’ Army Research Office grant for integrating ecological science and geographic information science in support of the Ft. Irwin National Training Center.

Research Supervisor for University of Redlands work for the National GeoSpatial-Intelligence Agency (in collaboration with University of California – Santa Barbara), the NASA Jet Propulsion Laboratory, the U.S. Environmental Protection Agency, the CA Water Resources Control Board, and others.

Position

The University of Redlands involvement in the specialist workshop is related to collaborative research funded by the NGA. The Redlands Institute (at the University of Redlands) is working to design and develop application prototypes that integrate research products from UCSB and others with complementary information science technologies. Our grant sponsors and clients have an increasing need for tools that can synthesize knowledge from disparate, unstructured information sources – databases, file systems, document libraries, the internet, and enterprise GIS repositories.

We are currently exploring how ontology-based inference engines could enhance the functionality of a digital gazetteer service. We are interested in demonstrating how a gazetteer could be extended to by an ontologically-driven inference engine to facilitate the automated extraction of geographic information for a specific named location.
Extending Gazetteer Service in Geographic Information Retrieval

A Position Paper

Naicong Li, Jordan Henk
The Redlands Institute, University of Redlands
naicong_li@institute.redlands.edu, jordan_henk@institute.redlands.edu

Hill (2000) identifies three basic elements in a digital gazetteer: name, footprint, and type (or category). Gazetteer entry types are organized in a hierarchical classification system. The name and footprint properties of geographical features enable automatic association between non-spatial data (e.g. text documents) and the location of geographical features, whereas the feature type hierarchy provides automatic organization of the the spatially indexed data. Used in a geographic information retrieval system (GIR, Larson 1996), such spatial indexing makes it possible to answer thematic ("what") and locational ("where") queries. The "theme" in the query may be spatial, referring to a geographic feature (e.g. the City of Redlands) or to a category of such features (cities). It can also be non-spatial, referring to some property of the feature (e.g. the city’s population). While the name and footprint properties of geographic features are formally defined in a gazetteer, the typical set of properties associated with a particular feature category often are not. Similarly, the typical set of relationships among geographic features (other than the is-a relationship, such as the capital-of relationship) are often not present in gazetteers. GIR systems could potentially benefit from explicit definitions of these properties and relations by using them to automatically index and later retrieve relevant data, especially in the case of highly structured data such as GIS datasets.

The Redlands Institute at the University of Redlands is conducting applied research in GIR using a prototype platform known as the Geospatially Referenced Information Portal (GRIP). GRIP was designed to facilitate knowledge synthesis across a disparate set of information resources (sources such as internet engines, local file systems, digital libraries, databases, etc. - all stored in heterogeneous formats, possibly with non-standardized structures). GRIP (which is being extended to support knowledge management tasks for a desert tortoise science and habitat management project) currently uses a project gazetteer with a feature type hierarchy modeled after the ADL Feature Type Thesaurus (FTT). This gazetteer has over 12,000 features extracted from recent versions of the USGS GNIS gazetteer (for the states of CA, NV, AZ, and UT); USGS Federal Land polygons augmented with boundaries for states, counties, state parks, etc.; and further supplemented with project-based GIS data. Using this gazetteer, GRIP supports basic spatial searches and basic queries (including theme and location).

Our users are often interested in the exact value or range of values of a certain property of a geographic feature (e.g. the actual tortoise population density of a critical habitat unit), or the statistics on these values, or retrieving a subset of geographic features of a certain type based on some specification of their property values. Sometimes the data value already exists in the project GIS dataset(s), sometimes GRIP needs to perform spatial and/or non-spatial processing to derive the answer. To handle such queries, we have adopted an approach similar to those described in Fonseca et al. 2002 and Lutz et al. 2006. We are prototyping using a light weight (not heavily axiomatized) project ontology for our user community. This ontology contains project-based concept categories (both spatial and non-spatial). For each concept category, we define a set of relevant properties (e.g. the ‘area’ of a ‘habitat’), including definitions of the property value type and value range. We also define its relationship (of types is-a, partitive, causal, functional, associative, etc.) to objects of other categories when applicable. Data in the project database are linked to appropriate categories as instances of these categories. Such links are established by a process similar to “registration mapping” as described in Owers and Ludäscher 2004, and in Lutz and Klien 2006. The categories in this ontology thus can be used as an index into the data in the project geodatabase. The categories can also be used to index document metadata, and they may be used to annotate text documents (through recognition of their corresponding natural language expressions in the documents) to produce sub-document level metadata for various levels of discourse units such as word, phrase, sentence, paragraph, etc. This ontology is designed to import pre-established upper level ontologies including an ontology of spatial relations with definitions of topological relations, distance and direction. It also incorporates definitions of frequently used spatial and non-spatial operations which can be invoked when GRIP is processing the more complex spatial search
Summary


Recent Accomplishments

• Software architect and developer for ‘GRIPPER’ (Geographic Referencing for Investigating Phylogenetics and Pathogenicity). With GRIPPER, a researcher can answer questions like “At what temperature and humidity ranges did taxon ‘a’ mutate into taxon ‘b’? Were there farms with pigs or ducks within 10km of the mutation site? How would the answer change if considering only wetlands within China? GRIPPER is a web application built using a customized gazetteer containing at present close to 500 million worldwide features that include occurrence data, a catalogue of tiled raster layers, and geographic features. Users import occurrence and phylogenetic data for analysis; results include GoogleEarth visualization. AJAX, Struts, JSF, and deegree form core of the UI.
• Gazetteer expert for ‘Biogeomancer’, a multi-institutional collaboration for automated georeferencing of natural history collection records. PostGIS gazetteer built from delimited text files and from shapefiles. Source data includes TIGER, National Atlas, GNS, The Getty Names, Protected Areas, and a worldwide dataset of Administrative boundaries provided by collaborator UC Berkeley. Handled source data from a variety of international encodings through automatic character set detection and conversion.
• Designed and wrote ‘Island Net’, an application similar to GRIPPER, but specialized for Caribbean Island biodiversity, and including support for species distribution prediction.

Tools

• Platforms (including system administration): Windows, Unix.
• Communication: Excellent written and oral presentation skills. Skilled with natural languages (proficient in French, some German).
• Configuration Management: ClearCase, CVS (WinCVS), Subversion.
• Database Administration: PostGIS.
• Other: Please Enquire.

Education

• SUN Java Certification (Programmer Level), 2004.

In addition, completed the coursework and preliminary exams for a Ph.D. in Computer Science at the University of Colorado.
• **U. Colorado Natural History Museum, Volunteer, 10/04 – 1/05**
  Implemented a gazetteer lookup feature for MaPSTeDI (Mountain and Plains Spatio-Temporal Database Informatics Initiative, see http://geomusc.colorado.edu).

• **USGS, 06/03 – 09/03 (Volunteer)**
  Worked with ArcIMS, Java, Javascript, Apache Webserver, and Tomcat to implement a servlet returning correct metadata for extents when downloading a cursor-selected sub-region with the USGS’s National Map Server interface.

  Completed a student project on fire perimeter attributes, working with the GeoMAC project at the USGS, which serves information on wildfires.

*Please contact me for additional prior work experiences or for references.*

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, and developed as part of the Biogeomancer suite of georeferencing tools for natural history and biodiversity. The main client application is called GRIPPER. 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 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.

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), Geonet Names Server (GNS), Geographic Names Information Service (GNIS), 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.

1 http://www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/version3.2.html
2 http://www.biogeomancer.org/
3 http://www.getty.edu/research/conducting_research/vocabularies/tgn/
5 http://rhhd.usgs.gov/gnis.html (not all of this data has been processed at this time).
Linda L. Hill  
Geography Department, University of California, Santa Barbara (Emeritus)

Between 1996 and 2004, I was a member of the research group for the Alexandria Digital Library (ADL) Project. My role was that of a library/information specialist working with computer scientists, geographers, and library staff to develop the model for and prototype of a georeferenced digital library. I was responsible for user evaluations of early prototypes of ADL and developing the ADL Gazetteer Content Standard (GCS) and the ADL Feature Type Thesaurus (FTT) and recreating the ADL Gazetteer based on GCS relational database schemas and reclassification of the nearly five million gazetteer entries. I participated in the development of the ADL Gazetteer Protocol and the ADL Thesaurus Protocol, both spearheaded by Greg Janée, and in the development of a Content Standard for Computational Models with Scott Crosier, Terence Smith, and Michael Goodchild.

Mike Goodchild and I organized the Digital Gazetteer Information Exchange (DGIE) workshop funded by the NSF and held in 1999 at the Smithsonian Institution in Washington D.C. The reception for the workshop was hosted by and held at the National Geographic Society. In 2002, I organized a workshop on Digital Gazetteers: Integration into Distributed Digital Library Services which was held at the Joint Conference on Digital Libraries in Portland, Oregon.

Based on this and earlier experiences with georeferenced information, I wrote an introductory book entitled *Georeferencing: The Geographic Associations of Information* which was published by MIT Press in September 2006.

I was a prime mover of the Networked Knowledge Organization Systems/Services (NKOS) group which has organized workshops and hosted a discussion list since 1997 and which continues to be active in both the U.S. and Europe as a focus for efforts to enable knowledge organization systems (KOS)—such as classification systems, thesauri, gazetteers, and ontologies—as integrated services in networked information systems.

My PhD is in library science from the University of Pittsburgh; my Master of Library Science (MLS) is from the University of Michigan. With the MLS but before the PhD, I worked as head of a department of a public library, as head of an exploration and production library for a petroleum company, and as assistant director of the Petroleum Abstracts Service at the University of Tulsa which indexes the world’s exploration and production literature for the petroleum industry. After the PhD, I joined Information International Associates, Inc. as a consultant and worked on a work flow analysis project for NASA’s Center for AeroSpace Information (CASI). This was followed by another NASA-based project through the University of Maryland at College Park working with the interagency Global Change Data Management Working Group. This led to my getting involved with the ADL Project.
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.
JERRY R. HOBBS

Personal:

Address: USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, California 90292

Telephone: office: (310) 448-9140
fax: (310) 822-0751

E-Mail: hobbs@isi.edu

Web Page: http://www.isi.edu/~hobbs

Citizenship: USA

Education:

New York University, Ph.D., Computer Science, 1974
Thesis Advisors: J.T. Schwartz, Naomi Sager
New York University, M.S., Computer Science, 1970
Duke University, B.S., Mathematics, 1964

Honors:

Awarded Doctor of Philosophy honoris causa, University of
ISI Fellow, May 2006-present.
Elected Fellow, American Association for Artificial Intelligence, 1992

Professional Experience:

Senior Computer Scientist, Information Sciences Institute, University of Southern
California, 2002-present
Research Professor, Department of Computer Science, University of Southern
California, 2003-present
Principal Scientist, SRI International, 1994-2002
Program Director, Natural Language Program, Artificial Intelligence Center, SRI
Consulting Professor, Department of Linguistics, Stanford University, 1992-2002
Assistant Professor, Dept. of Computer Sciences, City College, City University of
New York, 1974-1977
Instructor, Dept. of Computer Science, Yale University, 1972-1974

Other Professional Activities:


**Selected Chapters in Books:**


**Selected Papers In Conference Proceedings:**


Brief Biographical Sketch

Krzysztof Janowicz is research associate and PhD student at the Institute for Geoinformatics, University of Muenster. He earned a Dipl. in Landscape Ecology from the University of Muenster (2003). His master thesis was about composition and interoperation of semantic-enabled web services. He started to work as student assistant at the Institute for Geoinformatics in 1999 and works as research associate since 2003. His PhD thesis is about a role-based approach to semantic similarity measurement for concepts specified in high expressive description logics and will be finished in 2007. The PhD is supervised by Werner Kuhn and Martin Raubal. During his studies Krzysztof Janowicz was working as network administrator, software developer and network security consultant within various companies and also as independent contractor. In 2001 he published a book about internet security at O’Reilly Press and a revised 2nd edition in 2005. Krzysztof has authored or co-authored over 10 articles in journals, books, and conference proceedings relating to GIS, computer science, and cognitive science. His research interests are in semantic interoperability, formal ontologies, similarity, analogy, context, spatial cognition as well as situated and role-based categories. He was an invited participant at the Dagstuhl Seminar – Towards Affordance-Based Robot Control and one of the organizers of the Workshop on the Potential of Cognitive Semantics for Ontologies at FOIS 2004. Krzysztof teaching activities focus on object oriented modelling and development.

Reasons for Wishing to Participate and Scope of Contribution

My work is focussed on developing a context-aware and meaningful notion of similarity between concepts and also individuals. Strictly speaking I am interested in similarity assessments between role-filler pairs (such as next to-province) used to determine the overall similarity of a described concept or individual. Beside applications in classical information retrieval and concept matching tasks, the developed theory can also be applied to compare (disambiguate) RDF resource descriptions. As part of my PhD work I have developed a theory for a similarity-based identity assumption service for historical place. The service uses thematic information as additional reference points to evaluate whether compared place descriptions point to the same real world place. Currently we started a joint project with a museum of biodiversity to implement the theory within a pilot application (for historical gathering sites) and combine similarity assessments with spatiotemporal and subsumption reasoning. Hence I consider my current work as well as my background in ontology engineering to be a useful contribution to the following points of your meeting agenda:

- Creation and sharing of category schemes for gazetteers
- Interpretation of prepositional referencing (e.g., near, next to, left of) in information georeferencing
- Evaluating if two pieces of gazetteer data are about the same place when names, types, spatial location, and time frame can all vary

Moreover I expect that the outcome of this workshop will influence our ongoing work on creating the identity assumption service and the underlying theory. Taking part in the meeting will also give me the possibility to present parts of my PhD to domain experts.
Proposal to participate in the NCGIA specialist meeting on 
Digital Gazetteer Research and Practice

Krzysztof Janowicz 
Institute for Geoinformatics, University of Muenster 
Robert-Koch-Str. 26-28, 48149 Muenster, Germany 
Email: janowicz@uni-muenster.de 
Webpage: http://ifgi.uni-muenster.de/~janowicz

Similarity-Based Identity Assumptions for Historical Places

The domain of cultural heritage is very heterogeneous; the themes or exhibits that museums and related institutions are concerned with range from history of science to various kinds of art and historical documents, and biodiversity. Accordingly, the number and type of preserved exhibits range from millions of collected organisms to a small number of valuable paintings. Creating and maintaining metadata about exhibits and historical facts in general gets increasingly important for scholars and curators in order to structure, manage, and query their own data. As long as metadata is used for internal workflows only (such as the preparation of an exhibition), each institution may develop and maintain their own schema and representation format; however, to refine and enrich their own knowledge base or to answer complex scientific questions, interchange with external sources becomes necessary. Cleaning up the local knowledge base is especially important because one needs to keep in mind that historical knowledge may be vague, incomplete, or even misleading. To support these tasks the Committee on Documentation (CIDOC) provides a well established and standardized core ontology (called CIDOC CRM; ISO 21127) intended to annotate heterogeneous cultural heritage information to make it available in a machine-readable format (RDF) and reasonable way for knowledge integration, mediation and interchange. The long-term vision is publishing all annotated datasets through web services and therefore create a shared network of interlinked historical information enabling automatic metadata harvesting. The CIDOC conceptual reference model can be regarded as the underlying semantic level that provides meaning within the intended cultural heritage data infrastructure (which can be seen analogously to a Spatial Data Infrastructure) by delivering a common metadata schema. Instead of trying to reach a community wide agreement on definitions for concrete entity classes (such as types of exhibits) the strength of CIDOC CRM lies in defining an abstract but interrelated vocabulary describing the fundamentals of historical facts, namely established links (relations) between places, actors, objects, and events.

To make use of external data sources, however, a common language is not sufficient. It must be guaranteed that the collected metadata refers to the same real world phenomenon (which could be a historical place, person, event, or object) as the local datasets. Global authorities (such as the Alexandria Digital Library Gazetteer Server) provide unique identifiers and annotated datasets for some common kinds of real world phenomena. Scholars can refer to these global identifiers in addition to (or instead of) their local identifiers and therefore reduce maintenance effort and redundancy on the one hand and enable data interchange on the other. If compared datasets refer to the same global identifier and the scholar decides to trust the global authority as well as the external party that linked their dataset to the specific identifier, it can be assumed that the same real world phenomenon is meant.

Nevertheless, so far most datasets do not refer to global authorities and scholars must decide as the case arises whether the harvested information is relevant for their own knowledge base. There are several reasons for this:
SHORT CURRICULUM VITAE: Christopher Bernard Jones

Present post (since 2000): Professor of GIS, School of Computer Science, Cardiff University, UK

Date of birth: 6 June 1951

Educational qualifications:
1977: PhD University of Newcastle upon Tyne (Dept. of Geophysics and Planetary Physics)
1972: BSc, Geology, University of Bristol

Previous employment:
1993-1994: University Lecturer in GIS, Department of Geography, University of Cambridge.
Fellow and Director of Studies for Geography, Fitzwilliam College, Cambridge
1983-1992: Reader (from 1990), Computer Studies, University of Glamorgan

Research interests
My main research interests are in geographical information systems (GIS), computer cartography and the retrieval of geographically-specific information on the web. For several years I have worked on the representation of digital map features at multiple levels of detail, resulting in the development of multi-scale spatial databases and of operators and optimisation techniques for map generalisation. Recent research on geographically-aware web search engines has been concerned with spatial indexing of web documents and with designing ontologies that represent knowledge about the terminology and form of geographical places. I have also worked on environmental change detection, map labelling, data integration and 3D modelling of terrain and of fossils.

Selected Examples of Research Grant Funding:
Ordnance Survey: Gazetteer research, 3 years 2007-2009;
EPSRC: Multiscale spatial databases and map generalisation, 1998-2001;
ESRI Research Contract on Automated map generalisation 1998-2003;
NERC: Environmental change detection 1996-1998;
HEFCW QW funding for research in GIS 1994-1996;
ESRC: Logic databases for geographical information systems 1992 1995;

Postgraduate research supervision and examining:
Gazetteers and Geographical Information Retrieval

Chris Jones
School of Computer Science
Cardiff University, UK

Gazetteers are coming to play an increasingly important role in geographical information retrieval on the web. They enable users of transport timetables, routefinders, yellow pages, web mapping services and geographical web search engines to employ place names when specifying the geographical context of their requirements. The use of gazetteers in this role has also served to highlight some of their limitations with regard to the needs of the user. In practice, many queries that specify place names fail. One of the prime reasons for this is that the user may employ an informal or vernacular place name that is in common use but which is not recorded in the available gazetteers. In the UK, examples of such names are the “Midlands” the “Chilterns” and the “Wye Valley”. The reason the name would not be recorded is that gazetteers tend to reflect an administrative view of the world with an emphasis upon places that have precise boundaries. Some gazetteers do record the names of topographic features such as mountains and valleys, but they are not usually accompanied by data that record an estimate of the spatial extent of the features. The existing gazetteers may also fail to recognise a name because they lack the required level of detail or geographical extent or simply because they are out of date.

There is a need therefore for richer gazetteers that reflect common knowledge of place names. Because of the high rate at which place names change or are introduced there is also a need to develop a system of interoperable web gazetteer services that reflect local and regional knowledge of places throughout the world.

For the purposes of geographical information retrieval it is possible to envisage an ideal situation in which there is a system of multilingual gazetteer services in which the content conforms to agreed methods for specifying: preferred and alternative names; the timeframe for use of names; an ontology of place categories; rich information on spatial context including geo-political and topographic hierarchies, coordinates in well defined reference systems, spatial relations to adjacent places, and spatial footprints at different levels of generalisation, with information on the nature of boundaries (precise / vague).

Vernacular names
The issue of incorporating vernacular names into gazetteers raises several challenges with regard to the source of the knowledge and methods for modelling and representing it. Where does knowledge of the names come from? There is a great deal of personal knowledge of place names that it is possible to envisage eliciting via some form of mass questionnaire conducted perhaps on the web. There is a considerable body of vernacular place name knowledge within textual documents. Many such documents are to be found on the web and web mining or web harvesting is therefore another route to knowledge acquisition.
CURRICULUM VITAE

Ray R. Larson

University of California, Berkeley
School of Information Management and Systems
102 South Hall #4600
Berkeley, CA 94720-4600
(510) 642-6046 (office)
EMAIL: ray@SIMS.Berkeley.EDU

CURRENT POSITION
Professor
University of California, Berkeley
School of Information Management and Systems

EDUCATION
University of California, Berkeley
1980-1986
PhD: Library and Information Studies.
Certificate: Library Automation and Information Science
M.S.L.S.: Library Science.
B.A.: English and Comparative Literature

California State University, Fullerton
1974-1976
1969-1974

WORK EXPERIENCE
University of California at Berkeley
2003-
Professor, School of Information Management and Systems
(Assoc. Dean 2003-2005)

University of California at Berkeley
1991-2003
Assoc. Professor, School of Library and Information Studies.
And School of Information Management and Systems

University of California at Berkeley
1986-1991
1984-1986
Asst. Professor, School of Library and Information Studies.
Acting Asst. Professor, School of Library and Information Studies.

University of California, Systemwide Administration, Division of Library Automation.
1978-1984
Programmer-Analyst and Production manager

Fullerton Public Library
1967-1978
Clerk

BIOPGRAPHICAL SKETCH

Prof. Larson specializes in the design and performance evaluation of information systems, and the evaluation of user interaction with those systems. His background includes work as a programmer/analyst with the University of California Division of Library Automation (DLA) where he was involved in the design, development, and performance evaluation of the UC public access online union catalog (MELVYL). His research has concentrated on the design and evaluation of information systems.
Enhanced Gazetteer Development for Multilingual
Geographic Information Retrieval of Natural Language Text

Ray Larson and Fredric Gey
University of California, Berkeley
ray@sims.berkeley.edu, gey@berkeley.edu

Geographic information retrieval (GIR) from text is the subject of active research in the information retrieval research community. GIR focuses upon search and retrieval with a geographic component, e.g. Find stories about cities near the Danube and Rhine rivers in Europe. There have been GIR research workshops in 2004 (SIGIR, Sheffield UK), 2005 (CIKM Hanover, Germany) and 2006 (SIGIR Seattle USA). GeoCLEF (http://ir.shef.ac.uk/geoclef/) is a component track in the European Cross-Language Evaluation Forum (CLEF) which evaluates performance of Geographic Information Retrieval on multilingual text by creating test topics in multiple languages which are run against document collections in those languages. For the GeoCLEF 2006 evaluation just concluded and results presented in Alicante Spain, the languages were English, German, Portuguese and Spanish (additionally, topics were translated into Japanese for cross-language search from that language). The document collections consisted of news stories from USA, Swiss and Germany, Portugal and Brazil, and Spain. The total number of documents being searched exceeds 1 million documents. GeoCLEF has emerged as the standard by which GIR for text research advances can be objectively evaluated.

Among the components of GIR search topics which differ from ordinary information retrieval are:

- Geographic challenge:
  <EN-title>Cities within 100km of Frankfurt</EN-title>
  <DE-title>Städte im Umkreis von 100 km um Frankfurt</DE-title>
  <PT-title>Cidades a menos de 100 quilómetros de Francoforte</PT-title>
  <ES-title>Ciudades a menos de 100 kilómetros de Fráncfort</ES-title>
  <JP-title>100kmの北東の都市</JP-title>

- Geographic location disambiguation for vaguely defined entities:
  <EN-title>Scientific research in New England Universities</EN-title>

- Geotemporal disambiguation for vague references
  <EN-title>Credits to the former Eastern Bloc aka the Warsaw Pact</EN-title>

- Approximate regional restriction:
  <EN-title>Forest fires in Northern Portugal</EN-title>

Research issues which have been explored by GeoCLEF participants and GIR researchers include named entity extraction in multiple languages, place name disambiguation, geographic hierarchy and expansion, as well as examining issues about the granularity of gazetteer information (e.g., when expanding queries using placenames derived from gazetteer lookup, should only major populated areas be used, or should all toponyms in
CURRICULUM VITAE

Naicong Li
The Redlands Institute, University of Redlands
1200 East Colton Avenue, PO Box 3080, Redlands, CA 92373-0999
Phone: (206)855-8792, Fax: 909-307-6952, Email: naicong_li@institute.redlands.edu

EDUCATION
B.A. in French, Peking University, China, 1982.

RELATED PROFESSIONAL EXPERIENCE
Senior GIS analyst, The Redlands Institute, University of Redlands, September 2005 – present.
Research in ontology-driven semantic technology to be applied in spatially enabled knowledge management system; application domain and project ontology design; urban growth modeling using agent-based modeling and GIS; collaboration with UCSB on NGA funded research project on spatial web, networked gazetteers.

GIS consultant, Environmental Systems Research Institute (ESRI), March 2005 – present.
Programmer for ArcSketch, an ArcGIS extension which offers intuitive sketching tools, symbol driven sketching, and automated feature class/field/value creation.

GIS Consultant, Jones and Jones Architects and Landscape Architects, May 2004 – March 2005. Developed ILARIS, a GIS model for evaluating landscape visual quality, which won the ASLA 2006 Award of Honor in Research.


Senior Programmer/Analyst at Countrywide Home Loans, Department of Artificial Intelligence, September 1998 – May 2000. Redesigned and implemented Mustang, the rule based expert system or loan compliance checking.

Senior Linguistic Analyst at Language Systems, Inc., February 1997 – August 1998. Designed and implemented a knowledge based translation system, including a concept network and the reasoning module, a module that builds semantic representation for the input sentence, and the module for target language generation.

Linguistic analyst at Language Systems, Inc. August 1991 – October 1995. Participated in a project for evaluating various natural language processing systems. Designed and implemented the morphological analysis module in LSI’s machine translation system. Developed the automatic testing capability for the system. Designed and implemented a discourse processing module for tracking discourse entities. Designed and implemented various tools for knowledge based text data analysis and information extraction from texts.
Extending Gazetteer Service in Geographic Information Retrieval

A Position Paper

Naicong Li, Jordan Henk
The Redlands Institute, University of Redlands
naicong.li@institute.redlands.edu, Jordan.henk@institute.redlands.edu

Hill (2000) identifies three basic elements in a digital gazetteer: name, footprint, and type (or category). Gazetteer entry types are organized in a hierarchical classification system. The name and footprint properties of geographical features enable automatic association between non-spatial data (e.g. text documents) and the location of geographical features, whereas the feature type hierarchy provides automatic organization of the the spatially indexed data. Used in a geographic information retrieval system (GIR, Larson 1996), such spatial indexing makes it possible to answer thematic (“what”) and locational (“where”) queries. The “theme” in the query may be spatial, referring to a geographic feature (e.g. the City of Redlands) or to a category of such features (cities). It can also be non-spatial, referring to some property of the feature (e.g. the city’s population). While the name and footprint properties of geographic features are formally defined in a gazetteer, the typical set of properties associated with a particular feature category often are not. Similarly, the typical set of relationships among geographic features (other than the is-a relationship, such as the capital-of relationship) are often not present in gazetteers. GIR systems could potentially benefit from explicit definitions of these properties and relations by using them to automatically index and later retrieve relevant data, especially in the case of highly structured data such as GIS datasets.

The Redlands Institute at the University of Redlands is conducting applied research in GIR using a prototype platform known as the Geospatially Referenced Information Portal (GRIP). GRIP was designed to facilitate knowledge synthesis across a disparate set of information resources (sources such as internet engines, local file systems, digital libraries, databases, etc. - all stored in heterogeneous formats, possibly with non-standardized structures). GRIP (which is being extended to support knowledge management tasks for a desert tortoise science and habitat management project) currently uses a project gazetteer with a feature type hierarchy modeled after the ADL Feature Type Thesaurus (FTT). This gazetteer has over 12,000 features extracted from recent versions of the USGS GNIS gazetteer (for the states of CA, NV, AZ, and UT); USGS Federal Land polygons augmented with boundaries for states, counties, state parks, etc.; and further supplemented with project-based GIS data. Using this gazetteer, GRIP supports basic spatial searches and basic queries (including theme and location).

Our users are often interested in the exact value or range of values of a certain property of a geographic feature (e.g. the actual tortoise population density of a critical habitat unit), or the statistics on these values, or retrieving a subset of geographic features of a certain type based on some specification of their property values. Sometimes the data value already exists in the project GIS dataset(s), sometimes GRIP needs to perform spatial and/or non-spatial processing to derive the answer. To handle such queries, we have adopted an approach similar to those described in Fonseca et al. 2002 and Lutz et al. 2006. We are prototyping using a light weight (not heavily axiomatized) project ontology for our user community. This ontology contains project-based concept categories (both spatial and non-spatial). For each concept category, we define a set of relevant properties (e.g. the ‘area’ of a ‘habitat’), including definitions of the property value type and value range. We also define its relationship (of types is-a, partitive, causal, functional, associative, etc.) to objects of other categories when applicable. Data in the project database are linked to appropriate categories as instances of these categories. Such links are established by a process similar to “registration mapping” as described in Owers and Ludäscher 2004, and in Lutz and Klien 2006. The categories in this ontology thus can be used as an index into the data in the project geodatabase. The categories can also be used to index document metadata, and they may be used to annotate text documents (through recognition of their corresponding natural language expressions in the documents) to produce sub-document level metadata for various levels of discourse units such as word, phrase, sentence, paragraph, etc. This ontology is designed to import pre-established upper level ontologies including an ontology of spatial relations with definitions of topological relations, distance and direction. It also incorporates definitions of frequently used spatial and non-spatial operations which can be invoked when GRIP is processing the more complex spatial search
Inderjeet Mani, Ph. D.  Senior Principal Scientist, The MITRE Corporation, Bedford, MA

http://complingone.georgetown.edu/~linguist/inderjeet.html

Education
1976  B.Sc. Physics, University of Delhi.
1979  B.Sc. Experimental Psychology, University of Sussex
1980  M.S. Computer and Information Science, University of Pennsylvania
1997  Ph.D. (with Distinction) Computational Linguistics, Georgetown University

Professional Experience
1992-Present  Senior Principal Scientist, The MITRE Corporation
2005-Present  Research Affiliate, Computer Science and Artificial Intelligence Laboratory, MIT
2005-Present  Visiting Scholar, Department of Computer Science, Brandeis University
2003-2006  Associate Professor and Program Head, Computational Linguistics, Georgetown University
1988-92  Technical Staff, Artificial Intelligence Laboratory, Microelectronics and Computer Technology Corporation (MCC)
1984-88  Technical Staff, Artificial Intelligence Laboratory, Texas Instruments, Inc.

Publications  3 books, and 56 papers, including 26 journal articles and book chapters.

Sample Publications


A Framework For Inferring Spatial Locations And Relationships From Text

Inderjeet Mani, Dave Anderson, and Janet Hitzeman

Systems that interpret spatial information in natural language text need to deal not only with 'absolute' references (e.g., “Rome”, “Rochester, NY”), but also relative references (“thirty miles north of Boston”, “an underpass beneath Pushkin Square”). Current approaches to extracting information from text have made excellent progress using a methodology of first developing an annotation scheme for marking up expressions of interest with various features, and then training machine learning algorithms to reproduce the annotation. Earlier research along these lines has yielded success in resolving absolute and vague time expressions in different languages using the TIMEX2, annotation scheme (timex2.mitre.org, Mani et al. 2005), and temporally situating text mentions of events using the TimeML annotation scheme (www.timeml.org, Mani et al. 2006). We have recently begun a 3-year project to apply such a methodology, for the first time, to the automatic interpretation of spatial expressions in natural language texts in English and Chinese. Here we describe aspects of our project, building on our work to date, that are relevant to the themes of the workshop.

SpatialML Markup Language
We are currently developing a markup language for spatial expressions called SpatialML that provides a semantically-based scheme for marking up spatial expressions. It is being applied to a variety of different types of texts (including news, weather forecasts, route descriptions, geographical descriptions, etc.), with a corpus with this markup (currently already marked up with place names disambiguated with gazetteer-related features) being distributed and used as training data by various machine learning algorithms. For example, in the case of “an underpass beneath Pushkin Square”, “underpass” would be tagged in SpatialML as a feature of a particular type based on an existing feature ontology, “beneath” would be tagged as a signal with a value for a topological relation feature, and “Pushkin Square” as a particular place with a value for a geo-coordinate feature.

Place Name Disambiguation
A common way of referring to space is of course in terms of proper names. Accurate disambiguation of place names in text in terms of points and regions on a map are dependent on gazetteers with geo-coordinates and geographic inclusion information. Large gazetteers increase the degree of ambiguity; for example, there are 1420 matches for the name “La Esperanza”, according to the GeoNames Database from the National Geospatial-Intelligence Agency (NGA). A recent study (Garbin and Mani 2005) on 6.5 million words of news text found that two-thirds of the place name mentions that were ambiguous in the U.S. Geological Survey’s GNIS gazetteer were ‘bare’ place names that lacked any disambiguating information in the containing text sentence.

Information extraction systems can use disambiguation rules based on human intuition as well as rules discovered by programs trained from disambiguated examples. Since it is expensive to generate adequate samples of training data, research has tried to trade off
Biographical Sketch
David M. Mark

Professional Preparation
Simon Fraser University (British Columbia, Canada) Geography B.A., 1965-1970
University of British Columbia Geography M.A., 1970-1974
Simon Fraser University Geography Ph.D., 1974-1977

Appointments
1981-present Department of Geography, State University of New York at Buffalo
Asst. Prof. (1981-1983), Assoc. Prof. (1983-1987), and Professor (1987-date)
1995-2000 Member, School of Graduate Studies, University of Maine
1978-1981 Assistant Professor, University of Western Ontario

David M. Mark is a Professor of Geography at the University at Buffalo (UB), the State University of New York, and is the Director of the Buffalo site of the National Center for Geographic Information and Analysis (NCGIA). Mark also is Project Director of the University at Buffalo’s two NSF-funded Integrative Graduate Education and Research Traineeship (IGERT) projects in Geographic Information Science, projects that together have supported almost 50 doctoral level trainees in seven academic departments. He also is a member of UB’s Center for Cognitive Science. Mark has written or co-authored almost 230 publications, including more than 80 refereed articles, 5 edited books, 30 book chapters, 69 conference proceedings articles, and more than 40 technical reports. He has made almost 230 academic presentations, almost three-quarters at professional meetings, and the others as invited talks at universities and government agencies. Previously, Mark served as Vice-chair (1987-88) and Chair (1988-89) of the Geographic Information Systems Specialty Group of the Association of American Geographers and Chair of the Technology Interest Group, Canadian Cartographic Association (1987-89). Mark was involved in the founding of the University Consortium for Geographic Information Science, and later served as President of the UCGIS (1998). He also has served on numerous international editorial boards and program committees, and was program co-chair for Auto Carto 10 (1991), COSIT’99, COSIT’05, and GIScience 2000 and 2002.

Mark’s research interests focus on many aspects of geographic information science, notably geospatial ontology, semantics of geospatial information, spatial cognition and language, history of geographic information systems, human-computer interaction, digital elevation models, and indigenous mapping. His research has been funded by the National Science Foundation, the National Institutes of Health, the National Imagery and Mapping Agency, and other agencies, and he has served on several advisory panels at NSF.

Selected Publications Most Closely Related to Gazetteers:


Project Director, IGERTs in Geographic Information Science, University at Buffalo. 1998-present.
These NSF funded programs provide innovative blends of education and research training to broadly
prepare doctoral students in the field of Geographic Information Science.
http://www.ncgia.buffalo.edu/igis/

Chair, Panel on Cognitive Models of Geographic Space, and Member, Executive Committee, Project
Varenius: The NCGIA’s Project to Advance the Science of Geographic Information, 1997-present.
http://www.ncgia.ucsb.edu/varenius/

Co-leader, NCGIA’s Research Initiative 2, Languages of Spatial Relations (1989-1990); NCGIA’s
Research Initiative 13, User Interfaces for GIS (1991-1994); and NCGIA’s Research Initiative 21,
Formal Models of Common Sense Geographic Worlds (1996-)
http://www.ncgia.ucsb.edu/research/research.html
to the particular entity being referenced.

Cultural and Linguistic Differences in Placenames, Categories, and (Perhaps) Footprints

It is clear that many of the above components vary by language and by culture. Most of us are familiar with different language-specific placenames. Firenze becomes Florence, London becomes Londres, and Cologne becomes Köln or Colonia; even though these are usually just symbol substitutions and refer to "the same thing", that will not always be the case. That is because categories also vary by culture and language, sometimes by splitting or merging (for example, English 'river' splits (more or less) into 'fleuve' and 'rivière' in French) but often in more complicated ways (Mark, 1993; Mark and Turk 2003; Mark et al., in press). Although geographic entity delimitation has not been studied cross-linguistically, it seems quite possible that the footprint of "the same" feature might be significantly different in different cultural and linguistic contexts, especially for features with graded boundaries such as mountains (Smith and Mark 2003).

The presentation will some examples and suggest some approaches to solving the multilingual and multicultural aspects of gazetteers.

References


MARC-ANDRÉ MORIN

Defence R&D Canada – Valcartier
2459 Pie-XI Blvd North
Quebec City, Quebec
G3J 1X5 Canada

Phone: (418) 844-4000 ext: 4558
Facsimile: (418) 844-4538
Email: Marc-Andre.Morin@drdc-rddc.gc.ca

EDUCATION

Computer Science Certificate, 2002
Laval University, Sainte-Foy, Quebec City, Canada

B. Sc. Geomatics Engineering, 2001
Laval University, Sainte-Foy, Quebec City, Canada

COMPLEMENTARY FORMATION

Semantic Web Technologies, CRIM, 2006
Advanced Object-Oriented Development using Design Patterns, Technologia, 2006
Advanced Web Application Development using Struts, CRIM, 2006
OGC and ISO Specifications, Natural Resources Canada, 2005
Advanced GML, Galdos Systems Inc., 2005
GEOINT 2004 Symposium, New Orleans, 2004
Open Source GIS Conference 2004 & Annual Mapserver User Meeting, 2004
LuciadMap and LuciadWMS, Luciad, 2004
BEA WebLogic: Developing Data solutions, BEA Education Services, 2004
Microsoft Visual C#.NET, Microsoft Canada Co., 2003

EXPERIENCE

Defence R & D Canada – Valcartier, Valcartier, QC
System Architect 2006/04 –

Project: Applied Research for Geospatial Information Management (ARGIM)
- Develop next generation capabilities integrating knowledge exploitation and geospatial technologies in support to gathering and analysis of geospatial intelligence from multiple sources;
- Develop a new GEOINT framework based on free and open source software, portal technology (JSR-168) and standards (ISO, OGC, W3C, etc.);
- Manage contracts with industrial partners.

Natural Resources Canada – Geological Survey of Canada – Quebec City, QC
System Architect / Analyst / Programmer 2005/09 – 2006/04

Project: National Groundwater Database (NGWD)
- Determining a database exchange standard based on Geographic Markup Language (GML), Observations and Measurements (O&M) and SensorML from the international OGC standards (Open Geospatial consortium);
- Implementing the architecture that allows the dynamic conversion of hydrogeological data from a local format distributed in a number of databases towards a common standard (H2O);
- Implementing the architecture for exchanging data using protocols derived for OGC standards, i.e. Web Mapping Service (WMS), Web Feature Service (WFS), Web Coverage Service (WCS) and Sensor Observations Service (SOS).
Marc-André Morin  
System Architect  
Defence R&D Canada  

Expression of Interest in the topics of the Meeting

Defence R&D Canada (DRDC) is an agency of the Canadian Department of National Defence. Its mission is to improve Canada’s defence capabilities, through research and development. My R&D area of expertise is the Knowledge and Information Management, specifically Geographic Information Systems (GIS). I am involved in a major applied research project: ARGIM, a GEOINT solution integrating knowledge exploitation and geospatial technologies.

ARGIM stands for Applied Research for Geospatial Information Management. The Project Leader is Dr Alain Auger, a Defence Scientist with expertise in computational linguistics. The ARGIM project leverages on both natural language processing and GIS expertise in order to develop a new GEOINT framework, or a wide toolbox, based on free and open source software to rapidly integrate, deploy and evaluate new Information & Knowledge Management concepts and technologies.

Designed and implemented by our team, the current prototype relies on Service Oriented Architecture (SOA). The portal includes:
1. Document Management Service for grabbing, structuring and sharing sources of information;
2. Natural Language Processing (NLP) services for semantic-based text search and analysis, including automatic annotation of geographically-related entities in unstructured documents;
3. GIS capabilities for indexation, retrieval, visualization and analysis of spatio-temporal information. Notice that specifications from the Open Geospatial Consortium (OGC) are taken into considerations.

NLP services include:
1. TerroGate service, a new information retrieval system dedicated to the terrorism domain recognition (tactics, weapons, targets, groups, etc.);
2. GRID service, a geoparser for geographic pattern-based recognition;
3. Geocoder for assigning geometries to representative geographic location terms contained in texts;
4. Finally, gazetteers for feeding all services described previously in their specific activities.
Ruth Mostern  
School of Social Sciences, Humanities and Arts  
University of California, Merced

Work Experience  
2004-present, University of California at Merced  
Assistant Professor of History and Founding Faculty, School of Social Sciences, Humanities and Arts  
Interim Associate Director, World Cultures Institute, 2005-6  
Head of Collections Development (Associate Academic Specialist), 2000-2004  
Administrative and Technical Liaison to China Research Team, 1998-2000

Visitorships  
Short Term Visiting Fellow, July-August 2006: University of Sydney Archaeological Computing Laboratory, Sydney, Australia  
Visiting Scholar, 1997-8: Tōyō Bunko (The Oriental Library), Tokyo, Japan  
Visiting Scholar, 1996-7: Institute of History and Philology, Academia Sinica, Taibei, Taiwan

Educational Background  
Ph.D., 2003: History, University of California at Berkeley  
(Committee: Lewis Lancaster, Chair; Frederic Wakeman, Jr. [deceased]; Michael Watts)  
Ph.D. Orals Passed with Distinction: Exam Fields: Social and Economic History of Tang-Song China (618-1276 CE); Social and Economic History of Ming-Qing China (1368-1911 CE); Mughal and Early Colonial India; Cultural Geography and Agricultural Intensification  
M.A., 1992: History, University of California at Berkeley  
B.S., 1989: Georgetown University School of Foreign Service, cum laude

Publications  

Articles  

Reports  
“The Virtual Silk Road Atlas: Exploring Culture in Time and Place,” The Silk Road Project: Arts
A Historian’s Perspective on Gazetteers and Interoperability

Position Paper for Submission to the International Workshop on Digital Gazetteer Research and Practice

If a historian wishes to study the political history of the region south of the Caspian Sea, she might begin with the Achmenaeid Persians, architects of the world’s first great empire; trace the history of the region through its dominion by the Greek speaking Selucid and Parthian empires; chart its return to Persian rule during the Sasanian Empire, and note its conquest by the Arab Umayyad Caliphate and their Buwayhid successors. In medieval times, she would observe that the region was occupied by the Turkish Khanate of Khwarizm, the Mongol Il-Khanate, and the Turko-Mongol Timurids before reverting to Persian rule under the Safavids in the sixteenth century. With the emergence of nation-states in the eighteenth century, she would note that the region south of the Caspian became known as Persia, a name it maintained until the founding of Iran in 1921. As imperial fortunes shifted, the boundaries around and within the region south of the Caspian Sea changed dramatically. And, as suzerainty over the region changed hands between overlords of many linguistic backgrounds, place names changed as well. Often, places in this region were known simultaneously by names in many languages at the same time; naming systems that could represent different and conflicting perspectives about power and dominion. The historical trajectory of Persian political geography is hard to map with detail and precision, but is reasonably easy to model in a gazetteer.

Considered simply in their basic function as indexes of named places, gazetteers are an essential tool of historical geography, and, by extension, a crucial element of any kind of historical research whatsoever. Much geographically focused historical research, particularly on eras prior to the nineteenth century, is based on sources that are saturated with place names, but ones that can be assigned geographical coordinates only with difficulty and a very high degree of ambiguity. By contrast with historical GIS per se, with its premium on place, gazetteers have the capacity to liberate historical geographers from a preoccupation with location. They allow scholars to save untold quantities of time, money and frustration attempting to precisely georeference historical places and their indeterminate boundaries, while allowing us to accomplish what we do best: investigating how place and space made meaning to people. In a way that would be very difficult for a historical GIS system, a gazetteer can:

- model how places were related to one another (as components of a hierarchy or as nodes in a network, for instance)
- associate multiple coexisting names with one another through time and in many languages
- structure information about the changes to names, locations, feature types and relationships for any place
- link all of this information to sources, in the tradition of rich attribution expected by scholars of history and culture.

Why, at that point, should historians and humanists be encouraged to create gazetteers at all, rather than writing monographs, as our disciplinary tradition and colleagues
Reid Short Bio

James Reid is Projects Manager and Business Development Leader for EDINA with over 10 years experience in the area of GI and project management. He was a founding member of the Association for Geographic Information (Northern Ireland Chapter) when working in the public sector GI arena. In addition to extensive management experience in public sector authorities (N. Ireland & Scotland) he has particular relevant experience in developing the UK academic Spatial Data Infrastructure (http://www.dlib.org/dlib/may04/05contents.html) and acts as technical advisor to a wide a range of GI project and service developments for UK academia. He is the Service Manager for the EDINA ‘UKBORDERS’ online national service and is also project manager on a range of GI projects including:

- GeoCrossWalk [http://www.geoxwalk.ac.uk/] – GB middleware gazetteer server
- GRADE [http://edina.ac.uk/projects/grade] a geospatial data repository

Education

1993 Master of Arts (distinction), Geographical Information Systems, University of Leeds.
1991 Master of Town Planning, University of Newcastle-upon-Tyne.
1989 Batchelor of Science (1st Class Hons.), Geography, Queens University of Belfast.

Selected publications


Subject Discussion Areas of Interest

A. Components of gazetteer services

Through our GeoCrossWalk project and associated GeoParser development we have specific interests in:

- database population and creation issues related to multiple scales and multiple data sources
- feature typing (esp in multilingual contexts)
- accommodation for the variations and repetitions of placenames on a worldwide basis, including exonyms
- treatment of temporal issues

B. Georeferencing as a process

EDINA has been endeavoring to build georeferencing into the process of resource management within the UK academic network infrastructure by providing resources and tools to assist in explicit georeferencing. These include:

- geoparsing technologies
- middleware services and servers
- spatial data discovery services including registries and catalogues

C. Interoperable gazetteer services

EDINA builds services and projects on open standards and interoperability principles. We are active Open Geospatial Consortium (OGC) members and are involved in a range of data harmonization and interoperability initiatives, including:

- SOA approaches
- distributed federations of authority name files
- authentication and authorization issues including geoDRM

EDINA also has an interest in Grid computing (eScience) and emerging relationships with OGC and other geospatial interoperability standards.
CURRICULUM VITAE

Chris C. Rewerts
Research Scientist, Land & Heritage Conservation Branch,
U.S. Army Corps of Engineers Engineering Research and Development Center
Construction Engineering Research Laboratory
CEERD-CN-C, 2902 Newmark Drive, P.O. Box 9005, Champaign, IL, 61826-9005
Phone: 217-373-5825, FAX: 217 373 7266, email: chris.rewerts@us.army.mil

EDUCATION

Ph.D. Agricultural Engineering, Purdue University, 1992
M.S. Computing Education, Purdue University, 1987
B.S. Agricultural Education, Purdue University, 1982

PROFESSIONAL EXPERIENCE

Agricultural Engineer (DB-IV) – U.S. Army CERL, Champaign, IL.
- Aug 1997 – present

Senior Principle System Development Specialist – Computer Data Systems, Inc.
(Contracted to U.S. Army CERL, Champaign, IL)
- Jan 1996 – Aug 1997

Research Fellow – Oak Ridge Institute for Science and Education Research Participation
Program administered by ORISE through an interagency agreement between the US
Department of Energy and US Army CERL.
- Feb 1993 – Aug 1997

PAPERS AND REPORTS

Rewerts, Chris, William Goran, and Pamela Sydelko. Ecosystem Knowledge Mapping to
Support Military Lands. In Proceedings of: 8th World Multi-Conference on
Systemics, Cybernetics, and Informatics, 18-21 July 2004, Orlando, Florida.

Little Data and Little Time. In Proceedings of: 2003 Joint Conference of the
Southern Mensurationists and Northeastern Mensurationist Organization, Virginia
Tech.(in press).

Dale, Virginia H., Chris Rewerts, Webb Van Winkle, Mark A. Harwell, Mike Vasievich,
and Steve Hodapp. Barriers to the Use of Ecological Models in Decision Making.
Chapter 6 in: Ecological Modeling for Resource Management. Virginia H. Dale,
Susan Stone  
1715 Rose Street  
Berkeley, CA 94703  
Phone: (work) 510-642-8595  
sstone@berkeley.edu

Education

M.L.I.S., University of California, Berkeley, School of Library and Information Studies, May 1988  
Summer language study in China, 1984, 1987  
M.A., University of California, Berkeley, School of Education, Program in Reading and Language Development, 1977  
Study abroad in Taiwan, 1974-1975  
B.A., University of California, Berkeley, Department of Oriental Languages, 1974 (Phi Beta Kappa)  
Attended University of Michigan, 1970-1971

Employment

Museum Informatics Specialist, Museum Informatics Project, Information Services and Technology, University of California, Berkeley, 1992-  
Asian Serials Librarian, University of Hawai'i Library, 1989  
Assistant Specialist, Institute of East Asian Studies and Center for Chinese Studies Library, University of California, Berkeley, 1988  
Freelance Editor and Indexer, 1988-  
Editor, University of California Press, 1985-1987  
Editor, Foreign Languages Bureau, Beijing, China, 1985  
Editor, Institute of East Asian Studies, 1983-1984  
Research Assistant to Professor Frederic Wakeman (Chinese History), University of California, Berkeley, 1980-1988 (part time)  

Languages Studied

Chinese (Mandarin), Japanese, Thai, French, Spanish, German, Russian, Mongolian (proficiency varies greatly)
Expression of Interest in Workshop on Digital Gazetteer Research and Practice

Susan Stone, University of California, Berkeley

My interest in gazetteers is a result of my longtime interest in two different areas of information science and digital libraries research. Since studying for my master’s degree in Information Science at Berkeley in 1988 and throughout my work at Berkeley in the Information Server Project in the early 1990s and later the Museum Informatics Project, I have been interested in information system interoperability. My second interest is in language issues and multilingual data. This interest stems from my bachelor’s degree in Chinese and my first master’s degree in an interdisciplinary program involving sociolinguistics and education, where I specialized in writing systems and scripts.

In the context of the gazetteer workshop, I am most interested in issues involving multilingual and international gazetteer content. In the area of interoperability, I am interested in distributed interoperating gazetteers, gazetteer search and retrieval protocols, and integration of gazetteer data from multiple sources. In my professional role, I am interested in incorporating gazetteer lookup into Berkeley campus collection databases and other information systems. I am also interested in looking at gazetteer development in the context of new technologies and architectures for collaboration.

I first learned about spatial indexing in about 1992, when Linda Hill spoke at a conference of the American Society for Information Science and published a paper in the conference volume, which I edited with Michael Buckland. When I began working the Museum Informatics Project at Berkeley soon after, we used thesauri, such as the Getty Thesaurus of Geographic Names, for standardizing locations in databases of museum collections. We also began to collaborate with the campus GIS center with an eye toward mapping specimen data and other data in campus collections. It was as a participant, representing the Museum Informatics Project, in ECAI, the Electronic Cultural Atlas Initiative, where I actually met Linda Hill and became involved with gazetteers. I participated in ECAI’s NSF-funded gazetteer project, which sought to extend the ADL standard to accommodate historical, cultural, and multilingual data. Later, I was part of a research group at Berkeley with Ruth Mostern (then of ECAI, now of UC Merced) and a graduate School of Information Management and Systems student, Melanie Feinberg, to model time period data to complement location data in gazetteers.

In September 2001 through August 2002, NSF funded the project titled "A Multilingual Gazetteer System for Integrating Spatial and Cultural Resources," in which ECAI worked with Academia Sinica of Taiwan to extend existing gazetteers standards to accommodate the cultural and historical resources that ECAI and its collaborators have been developing. For our prototype, we worked on a gazetteer of Taiwan. We worked with location data collected by Academia Sinica field workers, along with the data for Taiwan in the ADL gazetteer and Taiwan data from the Getty Thesaurus of Geographic Names. Data were in Chinese and English, and I developed a testbed in which we were able to search and deliver records from the combined data using the ADL protocol and code from UC Santa Barbara. Part of our work in this project involved investigating a suitable feature type thesaurus for the types of locations researchers in the humanities were most interested in, one that could
RESUME

Björn Svensson

Years with current company: 6
Nationality: Swedish

KEY QUALIFICATIONS
An internationally accomplished Geographic Information Systems (GIS) expert with more than 12 years of experience in the United States, Africa, Europe, Southeast Asia, Central America and the Caribbean. Experienced in project management, software development, consulting, training, data conversion, design and development of very large enterprise-wide as well as smaller project or departmental databases. Worked in both the private and public sectors, including donor organizations such as the IDB and USAID, with clients ranging from the world's largest mining company to local city governments in developing countries.

Versatile technical and analytical skills in GIS and information technology as well as a comprehensive knowledge of hardware systems, network protocols, and peripheral devices. Extensive experience with a wide variety of GIS and Relational Database Management Systems (RDBMS) software products such as ArcMap, Oracle, Spatial Database Engine (SDE), GRASS, SPSS, dBASE. Extensive programming skills in languages such as Perl, php, C#, SQL, and c-shell. Familiar with multiple operating systems, including UNIX, Linux and Windows.

Experienced trainer with a proven ability to transfer technical skills in international environments, including developing countries where limited resources often favors a creative mind to achieve effective skills transfer. Trained local staff as well as other consultants in a wide variety of GIS software, procedures, and equipment.

Fluent in English and Swedish. Good working knowledge of Spanish, German, Norwegian and Danish.

EDUCATION
MA, Geography, University of California, Santa Barbara, 1994
BS, Social and Economic Geography, Lund University, Sweden, 1991
Expression of Interest

Björn Svensson

Name: Bjorn Svensson
Current Position: Project Manager
Company: Environmental Research Institute (ESRI).
Years with current company: 6

In 2002 I collaborated with Linda Hill and Greg Janée on the design of the ADL Gazetteer Protocol, but my professional interest in gazetteer and placenames probably started in 1997 when working on the South African National Spatial Infrastructure Framework. As part of setting up the first African FGDC clearinghouse the seemingly simple concept of geographic metadata and place and theme keywords turned out to be an intriguing matter that still fascinates me.

I work as a project manager at ESRI’s ArcWeb Services group where, among other things, I’m in charge of the “ESRI World Gazetteer” which is used in different ESRI products. The ESRI World Gazetteer is one of several gazetteer and gazetteer-style databases that are provided through ArcWeb Services. It is used both by desktop GIS software like ArcMap and ArcGIS Explorer, as well as web browser applications, for example MapMachine by National Geographic and Where’s Yours by Nature Valley. The ESRI World Gazetteer is available for free for certain non-commercial usage.

Currently we use our gazetteers primarily as a means to find a location (preferably with a footprint) and to quickly zoom in to the location in question. This could be either in web applications or desktop software. In this context, a “location” is primarily used to be able to zoom in a map to a certain area (using the location footprint), or to find certain data about a specific placename. This location finding functionality can be divided into two main categories:

1. Traditional gazetteers – type in a placename
   - returns Placename, with Type, and Spatial Footprint
     - ESRI World Gazetteer
     - Special Gazetteers for U.S. landmarks, World Postal Codes etc. Note that the main reason to keep these as separate gazetteers is for royalty purposes.

2. Addresses – supports not just a “street address” but also an IP address, a domain name or a “phone number” – all of which can be thought of as placenames/locations.
   - Returns “placename” with point location and other attributes where available (but no bounding box)
     - ArcWeb currently support street address geocoding for Australia, New Zealand, many European countries, United States and Canada.
     - Phone Numbers for the United States
     - IP addresses
26 October 2006

Michael Goodchild
UCSB
Santa Barbara, CA

Dear Dr. Goodchild,

Thank you for squeezing me in to the Digital Gazetteer workshop at such short notice.

As you know, Map Link is unique. We maintain an inventory of thousands of map and atlases from every corner of the globe. In the process of buying and re-selling this material Map Link must collect and manage metadata for every item. This means every sheet in a set of topographic maps. We collect basic data such as: scale, sheet size, format, date of edition, sheet name, a simple LC classification; as well as data relating to acquisition, cost and sales information, and a thumb-nail image of each map for display purposes. We have metadata for over 90,000 different maps and atlases.

Each of the three workshop components is an area Map Link has had at least some experience that may prove interesting, if not valuable to the discussion.

“Interoperable gazetteer services”

Map Link has tested a variety of database management tools to effectively store and manage map metadata internally. In the last few years our Internet presence has come to offer the full range of our holdings. As you might imagine, this presents us with many challenges, so our approach was to limit direct sales to resellers. These resellers presumably knew how to lookup the products they needed. Yet, I think we always knew the key to sales is in replicating the inquiry of a good map dealer, or librarian, in getting to the needed item quickly and efficiently. Our commercial demands occasionally mirror the conditions other managers of spatial data may face. Indeed, I believe we have a lot to learn from each other, particularly in the way users approach the data. Gazetteers, properly designed and deployed, are one key tool for efficient data lookup.

“Components of gazetteer services”

We have experimented with some advanced indexing and metadata collection techniques. When we first began to collect data for the Internet it was for a Geosystems (now Mapquest.com) retail web store in 1997. In addition to the foregoing metadata, we also started recorded lat/long coordinate values for each sheet. As many of the maps we sell use an unmanageable or detectable grid system, we also began to set down a list of important place names for these sheets. This rapidly became an untenable prospect, and we backed off to simply record the already mentioned metadata.

“Georeferencing as a process”
Professor Emeritus, Geography, University of California, 1994-present.
Research Professor, Geography, University of California, 1994-1996
Professor of Geography, University of California, Santa Barbara, 1977-1994
Professor of Statistics, University of California, Santa Barbara, 1985-1994
Assistant Professor to Professor, University of Michigan, 1961-1977.
Visiting Professor, University of Minnesota, Spring 1969.
Visiting Professor, Technical University of Vienna, Spring 1993.
Invited Lecturer, Australia and Peoples Republic of China, Fall 1982.
Distinguished Lecturer, University of Alberta, Spring 1983.
Marschak Lecturer, UCLA: 1979, 1996
Ph.D., University of Washington at Seattle, 1961
Doctor, honoris causa, University of Zurich, 1988.

Member:
- National Academy of Sciences of the United States.
- Association of American Geographers.
- Western Regional Science Assn.
- Assn. Pacific Coast geographers

Formerly:
- Royal Geographical Society, London.
- Regional Science Association.
- American Congress on Surveying and Mapping.
- Association for Computing Machinery; SigGraph.
- IGU Commission on Mathematical Models
- Institute of Electronic and Electrical Engineers, Affiliate Computer Society
- Honorary Fellow, American Geographical Society;
- Meritorious Contributor Medallion, Association of American Geographers, 1971;
- Andrew McNally Award, 1986;
- ESRI Lifetime Achievement Award, 1999.
- AAG Microcomputer Specialty Award, 1993.
- Phi Beta Kappa, Sigma Xi, Phi Kappa Phi.

Senior Scientist, Natl. Center for Geographic Information and Analysis, 1988-1994
Mathematical Social Science Board, 1971-1974; Chairman 1974
Committee on Behavioral and Social Sciences & Education, NRC, 1983-1986
Board on Earth Sciences, National Research Council, 1986-1987
Former Editorial Boards: International Journal of Geographic Information Systems;
The American Cartographer; Geographical Analysis; Mappemonde

Courses taught include:

Migration
Navigation
Location Theory
Regional Analysis
Computer Graphics
Analytic Cartography
Computer Cartography
History of Cartography
Introductory Cartography
Cartographic Transformations
Geographic Information Systems
Mathematical Models in Human Geography
Digital Gazetteer Research & Practice Workshop: Candidate Participant Resume

1PAUL M. VEISZE
GIS Manager
California Governor’s Office of Emergency Services
3650 Schriever Avenue
Mather, CA 95655
916-845-8542

ACCOMPLISHMENTS

Chaired California Advisory Committee on Geographic Names, 2000-2002
Chaired national conference of federal, state, local agencies on geographic names, 2003
Chaired first statewide GIS conference for Department of Fish and Game, 2001
Co-Organized workshop on shoreline and marine boundaries, datums, and policy, 2005
Coordinated multi-agency grant proposal for coastal remote sensing, 2004
Founding Manager, Department of Fish and Game Marine Region GIS Program, 1999
Managed GIS and CERES catalog for California’s first urban Conservancy, RMC, 2000
Coordinated first CERES marine data catalog (w/CSUMB and DFG MR), 1998
Integrated first GPS & video technology for airborne and seaborne DFG research, 2001
Prepared first documentation of California watershed boundaries to federal standards, 2004
Managed coastal watershed GIS databases via DFG and SWRCB interagency agreement, 1997

POSITIONS

Interim Working Group Chair, Homeland Security Data Model Committee, CA GIS Council
Research Manager II (GIS), CA OES, Mather, CA, February, 2006 – present
Research Manager I (GIS), CA Dept. Fish and Game, Sacramento, 1999 – 2006
Project Manager, OSPR Spill Response Internet Map System, SONS Drill, 2004
Director for State Government, California Geographic Information Assn., 2004-05
Fish and Game Representative, CMCC (State), FGC3 (Federal) & Regional GIS Councils
Project Manager, CRA Rivers & Mtns. Conservancy GIS, 2000 – 2002
Member, CA Resources Agency Advisory Committee on Geographic Names, 1994-2006
Advisor, CA Resources Agency Advisory Committee on Geographic Names, 2006-present

EDUCATION

Bachelor of Science, Renewable Natural Resources, University of California Davis, 1978
Master of Forestry, University of California Berkeley, 1985
Candidate Participant Interests in Digital Gazetteer Research and Practice Workshop
December 7 – 9, 2006, Santa Barbara
Paul Veisze, GIS Manager, California Governor’s Office of Emergency Services (OES)

General

• Digital gazetteers in georeferencing applications

Effective management of California emergencies depends on accurate, accessible location information, delivered in forms applicable to local needs. Emergency Managers must have the means to answer “Where is it?” and to extend that answer to the communities they serve in terms understandable by all.

• Collaboration and advancement of a research and practice agenda

OES is working with the University of California Office of the President in support of the California Hazards Institute (Rundle and others, 2006), a multi-campus initiative to leverage University resources for statewide emergency management. Placenames are key.

Core elements of gazetteers

• Placenames

I have a professional passion for the “cross-disciplinary data compression” that placenames offer: language, history, geography, sociology, technology, policy…

• Place categories

California’s administrative complexity requires parity in the complexity of reference systems, particularly in the hyper-sensitive arena of public safety. I would like to launch a census of administrative names (Ranger Districts, Water Districts, and the like) that would lead to their comprehensive encoding in the USGS Geographic Names Information System (geonames.usgs.gov).

• Geospatial locations

This workshop is fertile ground for engaging debate on the merits of the National Grid (aka Military Grid Reference System) for emergency management applications.
Biographical Sketch
Howard Veregin

I am Director of Operations for Geographic Information Services at Rand McNally & Co., in Skokie, Illinois, in the Chicago metro area. I lead a team of about 30 cartographers and geographic information specialists. We produce all of Rand McNally’s small-scale and mid-scale products, including the Road Atlas line, regional road atlases, specialty road atlases, the Commercial Atlas, Goode’s World Atlas, international atlases, and educational publishing atlases and maps. We contribute to the development of numerous electronic products. I am also editor of Goode’s World Atlas, currently in its 21st edition.

Prior to Rand McNally I was an Associate Professor of Geography at the University of Minnesota – Twin Cities. There I taught courses in GIS, cartography, and remote sensing.

I received my PhD in 1991 from Santa Barbara. Most of my academic research concentrated on geographic data quality, error propagation and error modeling, and I retain an active concern with data and its quality components.
Duplication and Apparent Duplication

In this country, place name duplication is the norm, not the exception. One reason is the relationship between intersecting administrative units, the most obvious example being municipalities (cities, towns, villages) and MCDs ("towns" and townships). Coupled with this is the existence of various official agencies, like the Census Bureau, which create areas for statistical reporting at various levels of the geographic hierarchy.

As a result, Groton, CT – an incorporated place with a population of about 10,000 – co-exists alongside Groton, CT – an MCD ("town") with a population of about 40,000. The MCD has a larger population because it covers a larger area. While these two entities have the same name (ignoring for now the issue of how the generic term “Town” is handled) they have different legal status, which is important in reference mapping due to the need to ensure that product specifications are being adhered to. If nothing else, given the different populations of these two entities, choosing one over the other would affect the dot and text size on the map.

A similar type of confusion occurs between CDPs and other delineations of unincorporated places. While CDPs are constrained along the boundaries of census enumeration units, there is no reason why this must be so. Landscan and similar population grids allow us to define population clusters more arbitrarily, which can give rise to multiple views of the same population cluster, all potentially valid relative to definitions and criteria.

A more difficult problem is the inverse – places that are truly duplicated in the database, but have not been identified as duplicates because their names are different. It is relatively easy to de-dup Goldengate, IL and Golden Gate, IL. More difficult is Branson West, MO, which is also known as Lakeview. Most places in the US have multiple place names they inherit from common usage or official status, or that have been designated by the US Postal Service or Census Bureau.

The goal of having one entry for each named place oversimplifies reality – or at least the reality of the US populated place landscape. Multiple records for named places is a common occurrence. If de-duping of records is to occur, it needs to be based on more than just name and location, as this will undoubtedly degrade the richness of the data to some degree. For reference mapping purposes it is more desirable to retain multiple records if there is some logical reason for them, such as different legal status, as long as the gazetteer provides the necessary attribution to differentiate these records, and as long as the alternate names problem is properly handled.

Geographic Hierarchies

Explicit information about hierarchical relationships between geographic features is a necessity for map production tasks. If these relationships were not available cartographers would not be able to perform even simple functions like sorting places by state to create a state-by-state index, or performing feature selection to customize symbology or naming conventions in different regions.

In the US, hierarchical relationships can be complicated, and simple spatial point-in-polygon rules often fail to provide expected results. Places can have multiple affiliations, such as Buffalo Grove, IL, which crosses the Cook/Lake county line. The same is true at the MCD level. To correctly associate places with their MCD and county “parents” it is necessary to compare polygonal representations, which can be tricky given the slivers and gaps that result from different spatial representations and levels of generalization.

Even with polygonal representations, errors can occur. In about half of the “township states” (itself a fuzzy set) there is no overlap between municipal and MCD governments. An example is Wisconsin, where municipalities and MCDs are mutually exclusive from a legal point of view. In the other half of the township states, some (but not all) of the municipalities within the state operate within territory that is also served by an MCD government. Some governmental functions are the responsibility of the municipality and some are the responsibility of the MCD. In Illinois, for example, all municipalities are within an MCD except Chicago, Cicero, and those municipalities in counties that have no MCDs. If product specs call for differentiation by governmental authority, it may be necessary to rely on other sources (e.g., the Census of Local Governments) if accurate information is not available in gazetteer format.

Or consider Hawaii, which has no legally-defined cities at all, except Honolulu, which has a single municipal government exercising control over the entire island of Oahu. The Census Bureau recognizes a CDP called Honolulu that is much smaller in size that the city/county of Honolulu. One issue is whether smaller unincorporated places on
Xiaobai (Angela) Yao  
Curriculum Vitae

Room 204, Department of Geography  
University of Georgia  
Athens, GA 30602 -2502

Assistant Professor  
Department of Geography, University of Georgia  
08/2002- present

EDUCATION

PhD, Geography (GIS and Urban Regional Analysis), 1997-2002  
University at Buffalo, the State University of New York

MS (with Distinction), GIS for urban applications, 1995-1997  
International Institute of Aerospace Survey and Earth Sciences (ITC), the Netherlands

BS (with Top Honor), GIS for urban planning and management, 1986-1990  
Wuhan Technical University of Surveying and Mapping (WTUSM) – now Wuhan University, China

RESEARCH INTERESTS

Qualitative GIS ; Urban Land Use and Transportation Modeling ; Spatial Data Mining

SELECTED PUBLICATIONS

Yao, X. Reasoning Qualitative Spatial Relations in GIS. In M. Madden (ed) Manual of Geographic Information Systems. 1st edition. ASPRS publisher. (Forthcoming)


Dear Professor Goodchild:

I am writing to apply for participation in the Digital Gazetteer Research and Practice Workshop which will be held from December 7 though December 9, 2006. With this letter, I am also applying for financial support to cover the travel and accommodation costs, in terms as described in the open call for participation that was sent to me via the UCGIS mailing list. I am an assistant professor in geography department at the University of Georgia. My primary research in the past has focused on representing and analyzing qualitative information in GIS, which fits very well with the focuses areas of the workshop. I am really excited to see the organization of this event. It is my sincere desire to be able to contribute to and benefit from the workshop.

In the past few years, I have been envisioning the “next-generation” GISs that can georeference, integrate, and analyze qualitative geographical information. I am particularly looking at qualitative information such as places names, qualitative spatial relations (near, north, in, on etc...), and qualitative modifiers (very, a little, etc.). As human beings often have incomplete and/or inexact knowledge of the environment, it is crucial for GIS services to be receptive to qualitative inputs. Qualitative descriptions such as place names are very often used in people’s daily life, and they also exist in many text-based databases. Contemporary GISs do not make much use of them, neither can they represent or analyze them. With this envision of a new generation GIS, I started my research with the definition of a new concept, qualitative location, as “the reference of locations using their qualitative descriptions and/or qualitative spatial relations with other features” (Yao and Jiang 2005; Yao and Thill 2006). I first explored strategies to query and visualize qualitative locations in GIS. I then focused on the proximity spatial relations (such as near and far). A close examination of the qualitative spatial relations revealed the research challenges brought by two innate characteristics of the qualitative spatial relations: context-contingency and vagueness. These two characteristics are present not only in the proximity spatial relations, but also generally in many other qualitative descriptions and qualitative spatial relations. These characteristics bring about some very interesting and often unavoidable research issues when we try to interpret qualitative geographic descriptions. I have proposed two approaches, a neuro-fuzzy inference approach and a statistical approach (Yao and Thill 2005; 2007), to account for context factors in the translation/interpretation process. Particularly, the neuro-fuzzy inference approach deals with both context contingency and fuzziness.

Based on my prior studies, I am developing two research projects/ideas concerning the referencing and analysis of qualitative location information with the aid of GIS and other state-of-the-art technologies. The projects/ideas are in line with the first two focus areas as identified in the workshop call, and are marginally related to the third focus area. First,
May Yuan
Edith Kinney Gaylord Presidential Professor

PROFESSIONAL PREPARATION

B. S. (Geography) the National Taiwan University
M. A. (Geography) the State University of New York at Buffalo
Ph.D. (Geography) the State University of New York at Buffalo.

APPOINTMENTS

1990-1994 Graduate assistant and teaching assistant, Department of Geography, the State University of New York at Buffalo, Buffalo, New York and the National Center for Geographic Information and Analysis.
1994-2000 Assistant Professor of Geography, The University of Oklahoma, Norman, Oklahoma
2000-2005 Associate professor of Geography, The University of Oklahoma, Norman, Oklahoma
2005- Professor of Geography, The University of Oklahoma, Norman, Oklahoma

SELECTED 10 PUBLICATIONS


Open Source and Open Environment to Enrich Digital Gazetteers and Facilitating Georeferencing Processes

A position Paper to the Workshop on Digital Gazetteer Research and Practice by May Yuan, University of Oklahoma

Place names (or toponyms) are highly variant and dynamic. Multiple places may have the same name across different administrative areas at a higher order. For example, Miami in Florida is a metropolitan city, but Miami in Oklahoma is a rural town. While both places share the same spelling, their pronunciation is quite distinct: m-ai-a-mi vs. m-i-a-m-ai. Furthermore, a place may have multiple names and local variants, especially for place names that are translated from one language to another. The English spelling for China's capital city can be Peking (an earlier version) or Beijing (the current official spelling). New communities and streets are developed with new names. Existing communities and streets may experience name changes over time.

In addition to city names, names for geographic features (such as mountains and rivers) can change, and as geographic features evolve, locations and geometries associated with these names will change accordingly. Volcanic eruptions or landslides can quickly alter the correspondent morphological and geometrical (shape and spatial extent) associated with these geographic features. New toponyms are given to new geographic features developed by natural or man-made processes (e.g. lagoons, retention ponds). Besides names, relationships among places and/or geographic features can change as well. These relationships may be containmentship, intersection, distance, and other non-spatial or spatial cases. The highly variant and dynamic nature of place names makes it challenging to build a comprehensive digital gazetteer of the world by any one or group of organizations. Hence, the position paper promotes the use of open source information in an open environment to enrich digital gazetteers that take the advantages of rich information from different places, and broad-based local knowledge on the World Wide Web.

Open source information can be on-line or off-line. Off-line open sources are documents or records open for inquires and browsing, such as unclassified government documents, newspaper, books, and other academic literature, etc. While there may be charges to access these off-line open source materials, the information is in general available for the public. The growth of cyberinfrastructure democratizes further publications and dissemination of facts, information, and knowledge. Quality of internet posting, however, varies, but a range of mechanisms has been used to build credibility and reliability. E-commerce and wikipedia are two examples of great success. E-commerce takes user feedbacks to build a reputation for sellers, buyers, or products. Wikipedia, on the other, provides a collaboratory open environment in the cyberspace to build the most comprehensive encyclopedia with the broadest and most diverse author communities. Both e-commerce (reputation established by peer feedbacks) and wikipedia (broad-based authorship) models offer new thoughts to the use of open source and open environment approaches to enriching digital gazetteers. A board-based authorship allows extensive and intensive incorporation of local knowledge that is critical to address the variant and dynamic nature of place names, while peer feedback mechanisms provide a measure of credibility to the authorship and local knowledge. Moreover, a board-based authorship promotes the opportunity to supply historical and geographic contexts to individual place names, most of which have historical, cultural, geographical, or social traces.
essence of semantic, temporal, spatial components of a place and help us to understand places with historical, geographic, and social contexts.