Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints

Linda L. Hill

Alexandria Digital Library Project
University of California, Santa Barbara
lhill@alexandria.ucsb.edu

Abstract. The core elements of a digital gazetteer are the placename itself, the type of place it labels, and a geographic footprint representing its location and possibly its extent. Such gazetteer data is an important component of *indirect geographic referencing* through placenames. Based on the gazetteer development work of the Alexandria Digital Library, this paper presents the nature of placenames, and the process of assigning categories to places based on the words in the placenames and other information, and discusses the nature of georeferencing places with geographic footprints.

1. Introduction

This paper concerns digital gazetteers, which can be defined as *geospatial dictionaries of geographic names* with the core components of

- a name (could have variant names also)
- a location (coordinates representing a point, line, or areal location)
- a type (selected from a type scheme of categories for places/features)

Digital gazetteers have new roles to play in information architectures for geo-locational access to information and data, where the placenames we use in everyday discourse to identify a location can be used to get (from computer systems) directions for driving there or finding information about an area. Structures for the interoperability of digital gazetteers hold promise for networked access to diverse sets of gazetteer data to maximize availability of global, local, and specialized place/feature descriptions and their spatial locations.

First, background information on the development of digital gazetteers and digital gazetteer information exchange (DGIE) will be given as a foundation for what follows. The next sections focus on the three required elements of a digital gazetteer: the nature of the placenames themselves; the process of assigning type categories to places/features; and a discussion of some of the issues of representing feature locations with geospatial *footprints* (e.g., latitude and longitude coordinates). This paper builds on previous papers describing the Alexandria Digital Library gazetteer development [1-3] and on online resources that are referenced herein.

2. Current Status of Digital Gazetteers

The U.S. Board on Geographic Names (BGN) supports the provision of two gazetteers; both of them are accessible through web services. They are: the U.S. Geological Survey's (USGS) Geographic Names Information System (GNIS) http://www-nmd.usgs.gov/www/gnis/ for domestic features, and the National Imagery and Mapping Agency's (NIMA) Geonames Server http://www.nima.mil/gns/html/. There are other well-established gazetteer web services, for example, the Canadian Geographic Names Service http://GeoNames.NRCan.gc.ca/english/ and a growing number of new ones. There are many other placename services that do not call themselves *gazetteers* but serve similar purposes. These include *yellow page* services for commercial locations, and placename access to online mapping and routing services and digital atlases. There are web sites that provide pointers to online gazetteers; a good starting point for such an exploration is the USGS GNIS web site.

The Getty Thesaurus of Geographic Names (TGN) http://shiva.pub.getty.edu/tgn_browser/ is a gazetteer of a different nature. It is one of three "structured vocabularies that can be used to improve access to information about art, architecture, and material culture…. Linked to the record for the place are names, a 'parent' or position in the hierarchy, other relationships, geographic coordinates, notes, sources for the data, and 'place types,' which are terms describing the role of the place (e.g., 'inhabited place' and 'state capital')." Names are in various languages and syntactic forms. One name is flagged as the preferred name. TGN does not require that records include geographic coordinates, but rather it contains a mix of entries with and without geographic footprints.

A characteristic of these gazetteer and gazetteer-like services[1] is that there is *no* standardization of any aspect of the record format, the content, or the service interfaces. Data cannot be easily shared among these resources. Standardization for official gazetteer efforts has been focused on establishing authoritative name forms for non-commercial places only. Official gazetteer agencies have not been involved with establishing authoritative boundary representations for the extents of features; their purposes have been satisfactorily served with point locations that disambiguate one place from another. Each gazetteer and gazetteer-like service also has created its own set of categories for placenames and its own record format and description details.

Implementations of digital libraries and online georeferenced collections and services have now created new roles for digital gazetteers. Placenames that are both geospatially referenced and categorized with a common scheme are the key to *indirect geospatial referencing*, meaning the use of a placename to identify a geographic location. For example,

- Using *Lisboa* or *Lisbon* instead of a coordinate expression (e.g., a point location of 9°7'59"W and 38°43'0"N) to find geospatial information about that particular location on our planet. A query such as "What satellite images cover the area of *Lisbon*?" can be translated into a query with the latitude and longitude coordinates for *Lisbon* by using a digital gazetteer. This query can be sent to a

[1] *Gazetteer-like* services and collections include the placename-location data inherent in GIS databases, the commercial yellow-page listings of named locations and spatial references, the Census address-location files, and all listings of named features that include spatial location representation.

collection of such objects where the only georeferencing is by coordinates or other geospatial referencing system - not by placenames.

- Using *Santa Barbara* to establish a point location on Earth so that sky maps and satellite locations can be oriented to the particular user's location. This method is used in the Heavens Above website from German Space Operations Center http://www.heavens-above.com/.
- Using placenames to identify point locations for navigation purposes, such as the GO2 system of wireless navigation services where users will be able to enter a short name identifier for, say McDonald's, and find the nearest McDonald's locations to the user's current location www.go2online.com.

These new roles for digital gazetteers result in new requirements for gazetteer data. Some applications (e.g., digital libraries such as the Alexandria Digital Library (ADL)) require geographic footprints giving spatial extent, not just point locations. Some applications require a mix of commercial and public places, of natural and manmade features, and of local and global locations. Some are based on street addresses and postal codes. Knowledge about historical places is of great interest to cultural and humanities activities and to genealogists. Information about geographic features and their names and characteristics is held in many ways by many organizations and individuals. What we need are ways to make this information shareable through search and retrieval protocols designed for gazetteer data, exchangeable record formats, defined elements of description, and common schemes for categories. Without a framework for generating and gathering and accessing gazetteer data, we will fall short in the areas of comprehensive placename knowledge and richness of representation and description.

3. Alexandria Digital Library Gazetteer

The ADL Project at the University of California, Santa Barbara, has included gazetteer development in its research from the beginning of its NSF-funded project in 1994. Details of this development are described elsewhere [1, 3] and current information can be found on the ADL website www.alexandria.ucsb.edu.

This gazetteer was built by loading available sets of gazetteer data. The datasets loaded into the ADL Gazetteer include those of the two U.S. federal agencies (U.S. National Imagery and Mapping Agency (NIMA) and the U.S. Geological Survey) plus U.S. topographic map areas, California earthquake epicenters, volcanoes from the Smithsonian Global Volcanism Program, and additional sets of bounding boxes for administrative areas.

The ADL Gazetteer data is described according to the ADL Gazetteer Content Standard and the ADL Feature Type Thesaurus. The ADL Gazetteer is available both as a web service www.alexandria.ucsb.edu/gazetteer/gazserver and as a component of the Alexandria Digital Library (the operational ADL). The operational ADL includes a catalog of georeferenced holdings of the Map & Imagery Laboratory (UCSB Davidson Library) with more than 2 million items as well as the ADL gazetteer with over 4 million entries; access to the operational ADL through either UCSB or through the California Digital Library is restricted to connections from the UC campuses only (as of April 2000).

This paper does not go into detail about the ADL Gazetteer Content Standard. Please consult previous papers and the ADL Gazetteer Development web page www.alexandria.ucsb.edu/gazetteer for the Content Standard itself, its relational database implementation, and the XML DTD version of it. There is also a link to information about the two-day workshop on Digital Gazetteer Information Exchange (DGIE), sponsored by the National Science Foundation (NSF) and hosted by the Smithsonian Institution, which was held in October 1999.

4. Nature of placenames and categories of places

ADL gained experience with the typing of placenames through the merging of the two U.S. federal gazetteers and the development of a Feature Type Thesaurus that draws on the terminology of existing typing schemes. An inventory of existing classification schemes for features identified the following representative schemes:

Table 1. Pointers to representative type schemes in current online gazetteers[2]

Organization	URL
U.S. Geological Survey	http://mapping.usgs.gov/www/gnis/features.html
U.S. NIMA	http://www.nima.mil/gns/html/FD_Cross_Ref.html
Australian Geographic Names	http://www.environment.gov.au/database/MAN200R.html
Canadian Geographical Names	http://GeoNames.NRCan.gc.ca/english/cgndb.html
Australia's National Mapping Agency	http://www.auslig.gov.au/mapping/names/names.htm
Getty Thesaurus of Geographic Names	http://shiva.pub.getty.edu/tgn_browser/about_tgn.html "Place types are indexing terms chosen from the structured vocabulary of the AAT"
ADL	http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/index.htm

There is no shared typing scheme among those listed in Table 1. The ADL Feature Type Thesaurus (FTT) was developed with the idea that it could provide a bridge among the various typing terminologies. To the greatest extent possible, all of the terminology discovered in the other schemes was added to the FTT as lead-in vocabulary, pointing to the preferred term for that concept in the FTT if it is not a valid term itself. Frequently occurring *surnames* in placenames were also added to the FTT (*surnames* are described below). The resulting thesaurus has close to a 5:1 ratio of lead-in vocabulary to preferred terms (209 preferred terms and 978 lead-in terms).

When the U.S. federal gazetteer data was loaded into the ADL Gazetteer, conversion programs were written in Perl that assigned an ADL Feature Type to each entry based on (1) the content of the placenames and (2) the classification of the originating agency. The most difficult set to convert was the GNIS from the U.S. Geological Survey. The difficulty was caused primarily by their use of broad categories that lumped discrete types together. For example, in the GNIS PARK category, ADL analyzed the placenames and found the following types of places, based on the placenames themselves: archeological sites, historical sites, hydrographic structures, forests, sports facilities, reserves, monuments, lakes, cemeteries, and museum buildings.

---

[2] Classification schemes for commercial activities, such as the U.S. Standard Industrial Classification (SIC) system and the North American Industrial Classification System (NAICS), are not listed in the table because they are designed for a different purpose. However, they are also a type of classification for sites that can be geo-located.

Placenames often contain information about the types of places they are labeling. Consider the examples in Table 2. *Surname* refers to the end position of the placename. In the ADL application, *surnames* can be either single words or two-word phrases. *Keyword* refers to a type indicator that is found in other positions in the placenames, as one entry in Table 2 illustrates; the keyword "correctional" appears in combination with a variety of other words in placenames and the word itself is a reliable indicator of the type "correctional facilities." The use of *keyword* terminology must be carefully done to avoid mis-assigning types to placenames. A general rule for such *keyword* assignments is that if the *surname* (the end word or words) is recognized from the list of surnames, then its assignment has precedence over other *keywords* in the name. For *Railroad Station Café*, the category assignment would be based on *Café*, not on *Railroad* or *Station*.

Table 2. Placename examples where type is represented in the name

Placename	1-Word Surname	2-Word Surname	Keyword
Perth Airport	Airport		
Baldwin County	County		
Admiralty Oil Seep		Oil Seep	
Jar Qudug Gas Field		Gas Field	
Sussex Correctional Institution			Correctional

Some placenames don't carry enough information within the name to distinguish a type category. For these, the classification must be provided by an outside source; for example, classifications assigned by an official gazetteer agency, from local knowledge, or from associated description. Note that many common placenames do not include their category, including countries ("Mexico," not "Country of Mexico"), U.S. states ("California," not "State of California"), and many cities ("Boston," not "City of Boston").

Some terms used in placenames are applied to various types of locations and therefore cannot be used alone to assign a type. The surname 'Field' is used as an example in Table 3.

Table 3. Examples of ambiguous type term in placenames

Placename	Classification
Kindley Field	airports
Bilate River Field	volcanoes
Jervis Field	agricultural sites
Lance Field	populated places
Eastern Fields	bars (physiographic)
Titas Field	oil fields

Add to this complexity the tendency of humans to be cute when they name shopping centers, housing developments, and so forth, and the use of historical names for buildings that are now used for other purposes, and you have a set of names that will be mis-typed if the names alone are used as a source of evidence for choosing a type category. Reliable manual typing is the best solution for these cases - and for all cases.

Our task was to convert the GNIS gazetteer to our ADL Gazetteer. To do this, programs were written in Perl to analyze the GNIS data in sets corresponding to the GNIS categories. The programs are of this form:

Step 1. Normalize the placenames by removing parenthetical phrases, numbers, directions (e.g. "North"), and trailing character strings from the ends of the names; normalize spacing and punctuation; and convert to lower case.
Step 2. Extract {two-word, one-word} surnames from sets of GNIS data, one set for each of their categories; keep surnames that meet a frequency of occurrence threshold; search the placenames for the occurrence of these surnames at the beginning and anywhere within the placenames.
Step 3. Compare these surnames to the ADL FTT Thesaurus; extract matches to valid and lead-in type terms.
Step 4. Create reports for two-word and one-word surnames, including the matching FTT terms that were found.
Step 5. Sort lists by frequency of occurrence of surnames and manually review.
- a. Accept automatic assignment of term(s) from FTT or change them
- b. Modify FTT as necessary
- c. Add type assignments for terms that did not match to FTT
- d. Set treatment of surnames to "keyword" if the surnames occur in other positions in the names with the same meaning.

The actual conversion process included monitoring to catch as many anomalies as possible while still processing large sets of data. Where type classification could not be reliably based on placename analysis, typing was based on a default mapping from the GNIS category to the ADL type. Where the GNIS category was too broad for a default mapping (e.g., their OTHER category) and the placenames were ambiguous, ADL noted the entries as *uncategorized*; .5% of the ADL Gazetteer entries are uncategorized.

We know that this automated process of assigning categories to places is not totally accurate and that mis-typing has occurred despite our best efforts. There is no substitute for direct knowledge of the places when assigning categories. We would like to be able to enlist the help of groups and individuals with local knowledge to review and revise the category assignments as necessary, as well as to contribute additional data and information. To this end, we have developed an XML version of the ADL Gazetteer Content Standard and the processes to export and import data in this format.

Geographic Name *
- **Name *** [the primary name for feature in a particular the gazetteer application]
- **Name Source**
- **Etymology**
- **Language** (default is English)
- **Pronunciation**
- **Transliteration Scheme Used**
- **Character Set (default is ASCII)**
- **Current / Historical Note *** (default is Current)
- **Beginning Date**
- **Ending Date**
- **Time Period Note**
- **Source Mnemonic**
- **Entry Date**

Variant Geographic Name (R)
[same attributes as for primary name]

*** Required (R) Repeatable**

Fig. 2. Attributes for placenames in ADL Gazetteer Content Standard

We have plans to mine the ADL Gazetteer of over 4 million entries using the surname extraction programs to produce a frequency distribution of surnames per category and categories per surname. This data will be interesting on its own, but it can also be used to provide automated support for assigning categories to new gazetteer entries and to support query formation in user interfaces.

All of these examples are in English and reflect Anglo-Saxon naming conventions, English-language characteristics of word construction, and perhaps Anglo-Saxon sets of categories. Extension to other languages will require customization to accommodate naming conventions and other character sets. There is an interesting possibility that a gazetteer that includes placenames in multiple languages would be able to bridge language differences. This could be done by analyzing the co-occurrence of non-English language keywords in the placenames to type categories based on the English language placenames. The ADL Gazetteer Content Standard and some existing gazetteer datasets (e.g., the NIMA gazetteer) include a language attribute for placenames (Figure 2) to support this type of category assignment across languages.

Street addresses and coding schemes such as postal codes and the U.S. Federal Information Processing System (FIPS) codes have not been included in this discussion because assigning categories to them is not complicated. The merging of retail sites with the types of place/features typically found in gazetteers will necessitate an analysis of the cross-over of commercial classification systems, such as the U.S. Standard Industrial Classification (SIC) system and the North American Industrial Classification System (NAICS), to gazetteer classification systems like the ADL Feature Type Thesaurus in the future.

5. Georeferencing places

A geographic *footprint* representing the location of a named place is the other requirement of a digital gazetteer, as defined here. This footprint, in latitude and longitude coordinates, can be one of the following (Table 4):

Table 4. Types of geospatial representation

Type of Representation	Description
point	single pair of latitude & longitude coordinates
bounding box	double pair of coordinates representing the maximum and minimum of latitude and longitude extent
line	set of points that do not enclose a space
polygon	set of points that do enclose a space
grid representation	grid references to a location according to an identified grid referencing scheme

A place can have multiple footprint representations:

- of different types: a point, bounding box, and a polygon, for example
- from different sources
- for different time periods (for example, the extents of cities change through time)
- suitable for varying resolutions (e.g., more detailed vs. more generalized)

Such a set of footprints bound to a named object (a *placenamed location*) would support a number of useful operations:

- viewing of footprint changes through time[3]
- comparison of footprints from different sources
- transformations between footprint representations

Detailed Spatial Geometry Representation (R) [set of points; dependent on system capabilities and requirements; can represent set of non-contiguous areas]

- **Detailed Spatial Geometry Representation *** {point, bounding box, linear, complex object}
  - **Number of Points ***
  - **Points Order** *
  - **(Longitude, Latitude)** *
  - **Current / Historical Note** *
- **Beginning Date**
- **Ending Date**
- **Time Period Note**
- **Measurement Date, Beginning Date**
- **Measurement Date, Ending Date**
- **Method of Measurement**
- **Accuracy of Measurement**
- **Source Mnemonic**
- **Entry Date**

*** required element** (R) repeatable

Fig. 3. Specification of location in ADL Gazetteer Content Standard

Placenames are inherently imprecise. Goodchild described this well in his presentation to the Digital Gazetteer Information Exchange workshop:

> We work in a world with words like "hot" and "cold", which have very vague meanings, and our vernacular geography is also vague. I come from Santa Barbara, and I don't worry about the precise definition of the footprint of Santa Barbara. In fact, I don't come from the municipality of Santa Barbara, if you wanted to be fussy about it, but you know what I mean when I say that I come from Santa Barbara. On the other hand, our GIS's are exact and "scientific"; they are in a very different, precise world. In many ways a gazetteer provides the linkage between those two worlds. It provides the linkage between the vernacular terms we use to talk about the world and the precise coordinate systems we use when we need to be exact…. [I]f one wants to combine vague and precise, then one has to deal explicitly with accuracy. The accuracy of a placename specification has to be an explicit part of the future gazetteer. We need to know how vague "vague" is. [4]

Each footprint used to describe a place is to some degree an approximation, which should be documented from the supplier to the end user. Often we depend on the nature of points and bounding boxes to convey a sense of approximation without further documentation. *Best practice*, however, would include documenting all footprints as specified by the ADL Gazetteer Content Standard (Figure 3) or in accord with the metadata standards for geospatial data [5].

Documentation of accuracy can be used to guide the appropriate use/interpretation of placenames and footprints, which is highly dependent on scale, ranging from local to global contexts. To quote Goodchild further:

> When location is specified, scale often provides a great deal of context. So when I say "Santa Barbara" at a global scale, you attach a different meaning to that than if I use the term at a local scale. "Santa Barbara" at a local scale probably means the municipal

[3] There has been some discussion about whether or not the gazetteer structure would be appropriate for named events like hurricanes where the temporal dimension is the "run time" of the event and the footprints are rapidly changing.

boundaries; "Santa Barbara" at a global scale probably means something much more vague - some blot on the map of Southern California. [4]

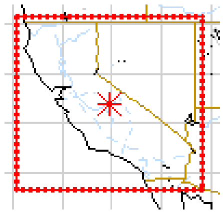
Sources of georeferencing include existing gazetteers, GIS datasets, numerous specialized listings, cartographic products, environmental analysis documents, research publications, measurements with Global Positioning Systems (GPS) units, and more. The problem isn't that these footprints don't exist, but rather that most of them are not easily accessible as gazetteer data. The key to making them accessible is a system of consistent naming (labeling) linked to the various footprint representations. That is, geographic names (location names) become primary keys to the spatial representations, and the geographic names themselves are *well known* in the sense that they can be referenced to an authoritative gazetteer source. Authoritative sources should (but they don't always) eliminate duplicate entries for the same places to minimize confusion.

Satisficing is a criterion that needs to be applied to the choice of gazetteer footprint representation for various purposes.

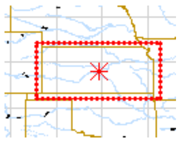
> *Satisficing* has been defined by Simon (1979) as a psychological form of the law of diminishing returns – we do not seek optimal solutions to certain problems because the costs are too high, so we settle for solutions that are satisfactory given the cost. [6]

Intuitively, we know that a point location is suitable for locating a person in space, for example, and for representing point-like entities like wells or monuments, or for representing large entities from a great distance (scale dependencies). Points for linear features like streams are satisfactory for the limited purpose of disambiguating one stream from another, but are not satisfactory for representing the extent of the feature or locating neighboring features.

We have less intuitive understanding of whether bounding boxes are *satisficing* specifications of locations in information retrieval systems. We can be assured that the extent of a bounding box encloses the extent of the feature but we cannot anticipate the similarity of the bounding box to the shape and coverage of the feature. This similarity depends on the characteristics of the shape of the feature. Figure 4 shows two examples of bounding boxes for U.S. states. It is obvious that the bounding box for Kansas is a better fit than the bounding box for California. We can measure this similarity for each instance but we don't have a formal way to specify in general the relationship of shape to bounding box and thus what we can expect from a general application of bounding box representations to geographic features. We have to estimate sufficiency of representation for bounding boxes by how they perform as a set for a particular purpose.



California



Kansas

Fig. 4. Examples of bounding boxes for U.S. states

For spatial matching operations in information retrieval systems, bounding boxes have an efficiency advantage over more detailed polygons; spatial *overlap*, *contains*, and *is-contained-in* operations can be efficiently performed on bounding boxes. In what circumstances is it satisfactory then to rely on bounding box generalizations in spatial information retrieval systems? If generalized polygons that more closely fit actual feature shapes are used instead of bounding boxes, what is the best balance of point density to efficiency and sufficiency for information retrieval? Hill [7] finds evidence to support the premise that ranking retrieval sets by spatial similarity to a reference region (e.g., a query region) "is relatively insensitive to differences in the exact shape and size of the [footprints]." This premise needs to be tested with more formal methods for spatially enabled information retrieval systems and other applications.

6. Conclusions

Three elements of digital gazetteer data have been the focus of this paper: the placenames themselves, the process of assigning type categories to places/features, and the representation of geospatial location. These three elements form the core descriptive requirements for digital gazetteers; optional elements include descriptions, links to related gazetteer entries, associated data, and links to online resources about the place/feature. Many communities have a stake in building and accessing digital gazetteers. The key to increasing their usefulness is interoperability. The first step is the recognition of the importance of *indirect geographic referencing* through placenames to digital earth information system applications. The second step consists of working together to establish the standards necessary to share gazetteer data across numerous distributed activities. These standards include applying techniques of controlled vocabularies (i.e., thesauri and domain lists) to bring consistency to the terms used to designate categories, and making such schemes themselves interoperable. This is a difficult but not impossible task and will be aided by automated processes of placename analysis such as those discussed here.

References

1. Hill, L.L., Frew, J., and Zheng, Q.: Geographic Names: The Implementation of a Gazetteer in a Georeferenced Digital Library. D-Lib, January (1999). http://www.dlib.org/dlib/january99/hill/01hill.html.
2. Hill, L.L.: Gazetteer and Collection-level Metadata Developments. In Kaser, R.T., Kaser, V.C. (eds.): Metadiversity. The Grand Challenge for Biodiversity Information Management through Metadata. The Call for Community. Proceedings of the Symposium sponsored by the U.S. Geological Survey Biological Resources Div. & the National Federation of Abstracting & Information Services. NFAIS (1999) 141-145.
3. Hill, L.L. and Zheng, Q.: Indirect Geospatial Referencing through Place Names in the Digital Library: Alexandria Digital Library Experience with Developing and Implementing Gazetteers. In Knowledge: Creation, Organization and Use. Proceedings of the 62nd Annual Meeting of the American Society for Information Science, Washington, D.C., Oct. 31-Nov. 4, 1999. Information Today, Medford, NJ (1999) 57-69. www.alexandria.ucsb.edu/~lhill/paper_papers/ASIS99_confpaper2_final.pdf.
4. Goodchild, M.F. The Future of the Gazetteer. Presented at the Digital Gazetteer Information Exchange Workshop, Oct 13-14, 1999. Transcribed and edited from audiotape. http://www.alexandria.ucsb.edu/~lhill/dgie/DGIE_website/Perspectives/brandt.htm.

5. U.S. Federal Geographic Data Committee. Content Standard for Digital Geospatial Metadata. [Accessed 4/21, 2000]. http://fgdc.er.usgs.gov/metadata/contstan.html.
6 Marchionini, G. Information Seeking in Electronic Environments. Cambridge Series on Human-Computer Interaction. Vol. 9. Cambridge University Press (1995).
7. Hill, L.L. Access to Geographic Concepts in Online Bibliographic Files: Practices and the Potential of a Graphic Interface. Ph.D. Dissertation. University of Pittsburgh, School of Library and Information Science (1990) 113.

Acknowledgements

The Alexandria Digital Library gazetteer development work has been financially supported by NASA Research Grant NAG5-8542 (ESDIS Prototype, Goddard Space Flight Center), by National Science Foundation Grant EIA99-78209 (a Digital Government grant for the DGIE workshop), and by the NSF Grant IR94-11330 (Digital Libraries I). UCSB graduate student Doug Tallman, Bren School of Environmental Science and Management, programmed the surname extraction programs. Ajay Kochhar, UCSB graduate student in computer science, has helped in converting sets of data and loading them into the ADL Gazetteer. Michael Goodchild contributed editing comments to a draft version.