A New Method for the Specification of Geographic Footprints in Digital Gazetteers

John P. Wilson, Christine S. Lam, and Deborah A. Holmes-Wong

ABSTRACT: This paper presents the strategy used to add neighborhood names and footprints to the Los Angeles Digital Gazetteer. The gazetteer database currently contains 4,500 features and is needed to: (1) facilitate the specification of geographic footprints in the Qualified Dublin Core metadata records that are used to describe digital assets; and (2) support the search for and retrieval of selected objects based on location, time, format, and/or keyword. The role of the digital gazetteer and a new browser which will offer the library patron a web-based query form with an interactive map is explained. The interface can be used to draw a query on a map, and it provides a series of pull down menus that can be used to specify time periods, formats, collections, and key words of interest. A new method for specifying neighborhood footprints in the digital gazetteer is described in some detail, and opportunities are highlighted for generalizing the method to help with search and retrieval using the map browser.

KEYWORDS: Geolibraries, geographic footprints, maps, gazetteers, interface design, search and retrieval, digital libraries

Introduction

Many digital library projects in recent years have utilized geographic organization and access to advance digital archives. The two fields—geography and library science—share a long-standing relationship through their work and interest in map libraries (Boxall 2003). These specialized libraries serve as both active learning centers and storehouses of cartographic materials and geospatial information (Hawkins and Bratton 1998). There are numerous examples of digital library projects that are focused on geographic information (with the Alexandria Geospatial Digital Library Project (Smith 1996) being perhaps the most notable), and some research has been conducted on the design, use, and impact of digital map libraries (e.g., Buttenfield 1998;, Lopez and Larsgaard 1998; Millea 1998).

However, the early digital map library projects were soon overwhelmed by more ambitious digital library initiatives that have led some commentators to wonder about the future of map libraries (e.g., Keller 2001; Perkins and Parry 2001). Many of these new initiatives are organized around the concept of a "geolibrary" (Mapping Science Committee 1999) which extends well beyond the traditional scope of map libraries and archives to include almost all information contained within libraries. Goodchild (1998), for example, defined the "geolibrary" as one filled with "geo-referenced information" which might include photographs, videos, music, and literature (things for which a geographic "footprint" can be specified) in addition to the maps, atlases, and satellite imagery that were the traditional focus of the map library.

The emergence of the idea of a geolibrary ran parallel with the development of national spatial infrastructures. These initiatives, which seek to facilitate the discovery, evaluation, and application of geographic data, have now evolved into the Global Spatial Data Infrastructure (GSDI) initiative (see http://www.gsdi.org for additional details). Boxall (2002; 2003) noted the connections between these concepts and how the idea of spatial information being essential to an information infrastructure—and of geography being a means to organize information and access—emerged quickly in a span of just a few years. He thought that the renewed focus and attention being paid to "things geographic or spatial" was so pervasive that it was not surprising to see the "geolibrary" concept replaced by the new and nearly all encompassing Digital Earth metaphor within just a few years of it first being proposed.

John P. Wilson and **Christine S. Lam**, Department of Geography, University of Southern California, Los Angeles, CA 90089-0255. E-mail: <jpwilson@usc.edu>. **Deborah A. Holmes-Wong**, Digital Information Management Department, Information Services Division, University of Southern California, Los Angeles, CA 90089-2571.

There are now two large geolibrary projects organized around the Digital Earth metaphor—the Alexandria Digital Earth ProtoType (ADEPT) (Smith et al. 2001) and the Digital Library for Earth Sciences Education (DLESE) (Mogk et al. 2000a; b)—and numerous examples of smaller, regionally based digital library projects such as our own which use geography as a foundation for organization and access (see Herold et al. 1999; Lim et al. 2002; and the Electronic Cultural Atlas Initiative (ECAI) at the University of California, Berkeley (http://www.ecai.org) for additional examples).

The ADEPT system builds on the Alexandria Digital Library (ADL) and seeks to support the creation of personalized digital libraries of geospatial information (I-spaces or information landscapes) and to investigate their utility in post-secondary science education. The Alexandria Digital Library is an operational digital library that provides access to collections of maps, images, and other geo-referenced materials from the University of California, Santa Barbara's Map and Imagery Laboratory. The ADL, which went online in Fall 1999 as part of the California Digital Library (http://www.cdlib.edu), is especially noteworthy because it also provides new types of library services based on gazetteers and other information access tools (see Hill 2000 for additional details). The ADEPT project, also centered at UCSB, is developing a Digital Earth metaphor to facilitate the discovery, manipulation, and display of dynamic geographic processes by instructors and students and, as such, it provides an interesting case study of the deployment of a digital library in a university setting.

Our own approach is narrower in some and broader in other aspects compared to these two Digital Earth projects. When completed, the USC Digital Library will house many of the specialized libraries and archival collections (SLAC) on campus and provide a substantial digital archive of primary research materials about what today is the Los Angeles Metropolitan Area. The system aims to integrate radically different information types rapidly through place, time, format, and keyword indexing, using the Dublin Core Metadata Element Set (DCMES) with qualifiers as the discovery metadata across all collections as well as for data sharing with other institutions.

The power of this new archive can be traced to the metadata that are used to support searches based on place, time, format, and keyword. Place here refers to the geographic footprint (i.e., the area covered by a photograph of a well-known building or landmark). Time refers to the temporal footprint (i.e., the date when a building was photographed or possibly the date when the building was erected). Format refers to the method used to represent individual items. The eight DCMI types—collection, dataset, event, image, interactive resource, service, software, sound, and text—and an additional type of geospatial datasets are currently supported. All items are part of larger collections (e.g., strip maps are part of the Automobile Club of Southern California collection) and can be retrieved by collection name as well. The architecture is designed to support flexibility in how a specific collection is described, and Qualified Dublin Core (QDC) is used as the common search and retrieval glue to enable one-stop shopping to search all collections at once. Some collections such as those in SLAC have additional information collected about them that is not covered by QDC but is covered by the Encoded Archival Descriptions (EADs) that will be used for data sharing with the Online Archive of California and may also be used to generate standard finding aids (in a future phase). This approach supports library metadata standards while also providing flexibility for collection managers whose collections require different standards and/or have different needs.

The vision for the new USC Digital Library was first laid out and implemented as a part of the "ISLA-Information System Los Angeles" prototype developed by Philip Ethington and colleagues (see Hunt and Ethington 1997 for details). The ISLA system was conceived as a web-based, distributed system that would allow separate institutions to either migrate existing data contained in other repositories into our repository or enable the direct ingest of materials into our system when an institution has a collection but no digital repository. The benefit of this is that smaller regional collections that do not have the resources or expertise to design, build, and maintain their own repository or digitize collections could still build and deliver digital archives via the web, with USC providing the repository infrastructure as well as the digitizing facilities. To date these collections have included the Automobile Club of Southern California, California Historical Society, Chinese Historical Society of Southern California, Huntington Library, as well as participation in the Green and Green Virtual Archive and Internet Missionary Photographic Archive collaborative projects featuring holdings from USC and other universities.

The system now under construction implements the ISLA vision and will provide scholars and students immediate access to a large archive of primary materials focused on southern California. The new system has an ArcIMS-based client interface and a back-end supported by Oracle 9i, Documentum,

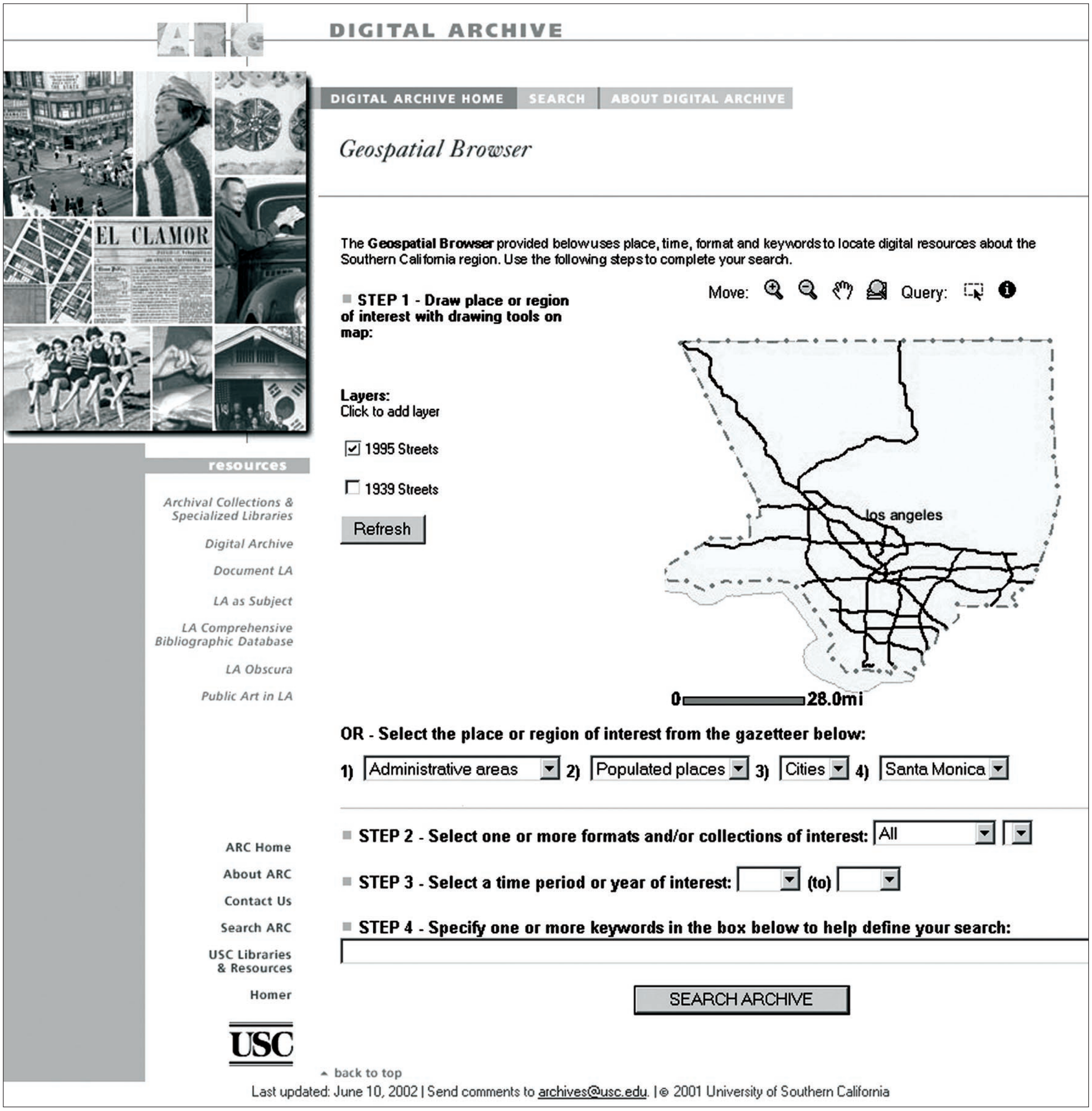

Figure 1. University of Southern California geospatial browser.

ArcSDE, and XML. The browser will offer the library patron a web-based query form with an interactive map that can be used to draw a query on a map—similar to the Alexandria Digital Library, ADEPT, Global Atlas (Lee et al. 2001) and work by Zhou et al. (2001)—and a series of pull-down menus that can be used to specify the time periods, formats, collections, and keywords of interest (Figure 1). The system will also incorporate a gazetteer or a list of geographic names that can be used in place of or with the interactive map display to specify places of interest.

Several displays will also be provided for the presentation of the search results. The first will give a brief description of the selected items and a thumbnail image if the object was saved as an image file or geospatial dataset. From there, the user will be able to click on the thumbnail image or metadata summary to obtain the full record display—containing both the full rendition of the image and the full metadata description—which they can subsequently choose to download and print or e-mail to their own personal computers or workspaces. For objects such as census information which are stored as geospatial datasets, the library patron will be able to construct on the fly and download one or more thematic maps, as well as the geospatial data sets themselves.

The USC Digital Library project is being implemented in several phases. The first phase, scheduled for completion by July 2005, has focused on what Soergel (2002) has referred to as the "computer" component of digital libraries as well as on a basic user interface. The digital library team selected Documentum for digital asset and content management and Oracle 9i for the repository to support the public interface, and we have built an ArcIMS prototype to verify that we can link the various software systems and perform the types of geographic queries envisaged in our design. We have also started work on a local gazetteer (see Lam et al. 2002 for additional details), which will be incorporated into the map browser and used in the contributor modules to facilitate the specification of geographic footprints when new objects are added to the digital archive.

The remainder of this paper focuses on the development of this gazetteer service and database for the Los Angeles region. Both will provide a bridge between the vernacular place names and other terms that we use to talk about this region and other parts of the world and the formal spatial referencing systems used by computers (Goodchild 1999). This linkage serves two purposes: (1) it will allow contributors to specify geographic names instead of geographic coordinates for the footprints that are saved as a part of the QDC metadata records; and (2) it will allow library patrons to start with a geographic name and find objects that are described with either geographic names or coordinates. The first will reduce the time, effort, and cost incurred in specifying the geographic footprints in the metadata records, whereas the second will allow the library patron to send a query to the gazetteer database to obtain the geographic location which can then be used as a spatial query to find the relevant objects and/or metadata records. We have developed a new method for rapidly generating footprints for geographic features such as neighborhoods with indeterminate boundaries (Lam et al. 2002) as a part of the gazetteer project that we think can be employed more widely to specify geographic footprints for new archive materials.

1. **Geographic Feature ID**
2. **Geographic Name**
    1. **Name *** [the primary name for feature in a particular gazetteer application]
    2. Name Source
    3. Etymology Language (default is English)
    4. Pronunciation
    5. Transliteration Scheme Used
    6. Character Set (default is ASCII)
    7. **Current / Historical Note *** (default is Current)
    8. Beginning Date
    9. Ending Date
    10. Time Period Note
    11. Source Mnemonic
    12. Entry Date
3. Variant Geographic Name (R)
4. **Type of Geographic Feature** (R)
5. Other Classification Terms (R)
6. Geographic Feature Code (R)
7. **Spatial Location (R)**
8. Street Address
9. Related Feature (R)
10. Description
11. Geographic Feature Data (R)
12. Link to Related Source of Information (R)
13. Supplemental Note
14. Metadata Information

Source Information * (R)
1. **Source Mnemonic** *
2. **Contributor Organization ***
3. Contributor Web Site
4. Contact Person
5. Email
6. Telephone Number
7. **Contributor Address ***
8. **Source Information * (R)**

* required element   (R) repeatable

Figure 2. Alexandria Digital Library Gazetteer Content Standard [Source: Modified from Hill et al. 1999; Hill 2000].

The remainder of the paper is divided into three sections. The next section describes some of the challenges and problems encountered in building digital gazetteer components, and the results of some recent work that may offer a way forward. The following section describes the methods and data sources used to add neighborhood names and footprints (i.e., shapes) to our Los Angeles Digital Gazetteer. The final section describes the strengths and shortcomings of the new method for specifying geographic footprints and some ideas on how this method might be generalized to help with

Administrative Areas		
Military areas	Postal areas	Statistical areas
Parks	Reference locations	Territorial waters
Political areas	Reserves	Tribal areas
Populated places	School districts	
Hydrographic Features		
Aquifers	Estuaries	Lakes
Bays	Floodplains	Seas
Channels	Gulfs	Streams
Deltas	Guts	Thermal features
Drainage basins	Ice masses	
Land Parcels		
Manmade Features		
Agricultural sites	Hydrographic structures	Storage structures
Buildings	Landmarks	Telecommunication features
Cemeteries	Launch facilities	Towers
Disposal sites	Mine sites	Transportation features
Fisheries	Monuments	Wells
Fortifications	Oil fields	Windmills
Historical sites	Recreational facilities	
Physiographic Features		
Alluvial fans	Cliffs	Moraines
Arroyos	Craters	Mountains
Badlands	Dunes	Natural rock formations
Banks (hydrographic)	Flats	Plains
Bars (physiographic)	Gaps	Plateaus
Basins	Isthmuses	Playas
Beaches	Karst areas	Reefs
Bights	Ledges	Seafloor features
Capes	Massifs	Tectonic features
Caves	Mesas	Valleys
Cirques	Mineral deposit areas	Volcanic features
Regions		
Biogeographic regions	Coastal zones	Land regions
Cadastral areas	Economic regions	Map regions
Climatic regions	Firebreaks	Research areas

Figure 3. ADL feature type thesaurus showing top- and second-level terms [Source: Modified from Hill and Zheng 1999; ADL Project Website].

search and retrieval in the map browser which we are building as part of the next phase of the USC Digital Library.

Building the Los Angeles Digital Gazetteer

The initial version of the Alexandria Digital Library (ADL) gazetteer was constructed by extracting the name, location, and type of selected entries from two online gazetteers maintained by the U.S. government covering the United States and its territories (the U.S. Geological Survey's Geographic Names Information System (GNIS); see http://www-nmd.usgs.gov/www/gnis for further details) and the remainder of the world (the National Geospatial-Intelligence Agency's GEOnet Names Server; see http://earth-info.nga.mil/gns/html for further details) (Hill and Zheng 1999). These sources contain over 6 million names but their use demonstrates a high level, top-down approach and means that the library patron is not likely to find large numbers of gazetteer entries for local regions and places due to the geographic extent of the coverage. However, the initial experience of building this gazetteer led to further work to develop a Gazetteer Content Standard and Feature Type Thesaurus, and these innovations are important because they can support the development and deployment of gazetteer services at multiple sites.

The Gazetteer Content Standard specifies a common set of terminology and definitions for the documentation of data to guide the development of metadata for each name (Hill and Zheng 1999). This standard provides lists of names and variant names for places and information about these names, the geographic footprints, the source or authority of the name, the language etymology, pronunciation, dates when the name was/is used, etc. Each name is assigned to one or more Feature Type Thesaurus categories, and features can be related to one another with links such as "IsPartOf" and "IsCapitalOf" (Hill 1999). Figure 2 shows the major features of this standard. The asterisks (*) and bolded names indicate required (mandatory) data elements and (R) indicates repeatable data elements. There are varying numbers of sub-elements for all of the major sections (similar to that shown for Geographic Name in Figure 2), and most of the major sections also include attribution elements for contributors and source to facilitate contributions to one geographic feature description from multiple contributors.

The ADL team has also developed a new thesaurus of place/feature categories to facilitate the consistent description of types of places and features in the ADL Gazetteer (Hill and Zheng 1999). The candidate terms were selected from the GNIS and GEOnet Names Server and used to build a hierarchical thesaurus. Six top-level terms (major categories) and 87 second-level terms were chosen to give structure to the hierarchy (see Figure 3 for additional details), and the depth of the hierarchy which sometimes extends to three or four levels but occasionally stops after one or two levels was heuristically determined based on ADL needs. The Feature Type Thesaurus, which includes broad term/narrow term relationships, synonymous terms and related terms, is an evolving document that can be modified by adding additional term variants and making changes to the structure and preferred terms as necessary (Hill and Zheng 1999). The current version of this thesaurus has close to a 5:1 ratio of lead-in vocabulary to preferred terms (209 preferred terms and 978 lead-in terms), and it bridges the various typing terminologies incorporated in six online gazetteers (see Hill 2000 for additional details).

The need to specify geographic footprints for the names is perhaps the most expensive and difficult task in building gazetteers for specific regions or places. The pre-ADL gazetteers and gazetteer-like services often represented the geographic footprints as points, whereas digital libraries require footprints that specify the geographic extent as polygons and/or bounding rectangles and not just as point locations to facilitate search and retrieval of generated content (Harpring 1997a, b; U.S. Geological Survey 1998). Two characteristics contribute to the difficulty of this task. The first is the fact that many natural and man-made features have fuzzy or indeterminate boundaries—see Burrough and Frank (1996) for an extended discussion of this phenomenon. This problem is often compounded by a second one—the tendency for the geographic footprint representing the location of a named place or feature to take on multiple representations, depending on the type of representation method used (points, bounding box, line, polygon, grid cell), source, resolution (scale, level of generalization), and time period (given that the extents of some features will change over time).

Hill (2000) also noted these problems and pointed to the multitude of sources and technologies (i.e., existing gazetteers, GIS datasets, research publications, GPS measurements, satellite imagery) which might be used to find and/or generate footprints. The challenge is two-fold: first, there is a need to select one or more appropriate footprints for each specific place and second, there needs to be a way to record multiple representations with specific entries in the Gazetteer Content Standard (realizing that the character and extents of many places in southern California and elsewhere have changed tremendously during the past 200 years). Fortunately, the second need was anticipated and an effective solution has been incorporated in the design of the ADL Gazetteer Content Standard (see Figures 2 and 4 for details).

Turning to the first need, Hill (2000) advocated using a "satisficing" criterion to select geographic footprints for digital gazetteers. Modeled after the law of diminishing returns as expressed by Simon (1979), this approach means that we do not seek out optimal solutions (i.e., the most precise and detailed footprints because the costs are too high) and we instead focus on solutions that are satisfactory given the cost. Hill (2000) used two examples—the choice of bounding boxes over detailed polygons and level of detail (i.e., number of vertices) used to describe the boundary in the latter case—to illustrate this approach. However, the ramifications extend further than both of these examples and include the choice of representation method (point, bounding box, linear, complex object) and

Spatial Location (R)
Detailed Spatial Geometry Representation (R)

[set of points; dependent on system capabilities and requirements; can represent set of non-contiguous areas]

1. **Detailed Spatial Geometry Representation*** {point, bounding box, linear, complex object}
2. **Number of Points***
3. **Points Order***
4. **(Longitude, Latitude)***
5. **Current/Historical Note*** (default is Current)
6. **Beginning Date**
7. **Ending Date**
8. **Time Period Note**
9. **Measurement Date, Beginning Date**
10. **Measurement Date, Ending Date**
11. **Method of Measurement**
12. **Accuracy of Measurement**
13. **Source Mnemonic**
14. **Entry Date**

* required element (R) repeatable

Figure 4. Specification of location in ADL gazetteer content standard [Source: Hill 2000].

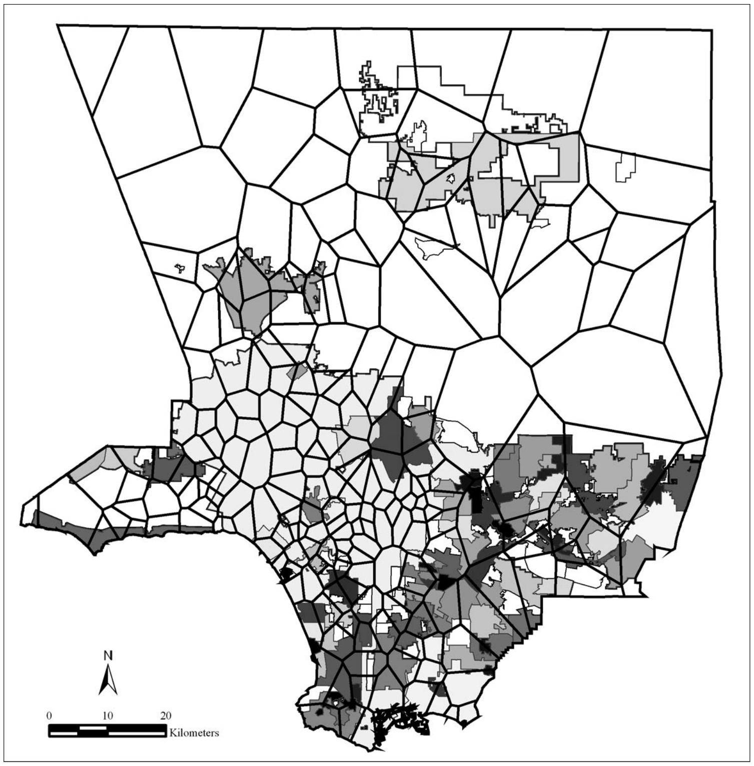

Figure 5. Map showing division of Los Angeles County into a series of Thiessen polygons using neighborhood point file and DSAM method.

the sources and/or methods used to generate the geographic footprints.

Alani et al. (2001) recently proposed the Dynamic Spatial Approximation Method (DSAM) for estimating spatial footprints from the locations of point sites that lie inside or outside some region of interest. DSAM is based on Voronoi diagrams in which the polygons represent the space closest to the associated point relative to all other points (Figure 5). This method can be used to infer spatial relations between regions in the absence of digitized boundaries; the resulting approximations provide measures of areal extent and can be used to evaluate spatial relationships such as distance, direction, and spatial adjacency.

These qualities suggest that the Thiessen polygons identified with DSAM might also be used to facilitate the search for and retrieval of geocoded objects in a digital library; however, we found that the results from using this approach were very sensitive to geographic variations in the character and density of the built environment in our work in Los Angeles County, and these same criticisms would apply to the ability of this method to show change through time. The shortcomings are illustrated by the polygon boundaries generated with the DSAM method in Figure 5, where the polygons increase in size from south to north even though most of the northern half of Los Angeles County is very sparsely populated. A series of maps showing change through time is likely to suffer similar problems since the gradual increase in the number of commonly used neighborhood names throughout the past 150 years would produce larger numbers of smaller polygons—both of the abovementioned problems occur because neighborhoods have inherently fuzzy boundaries and they seldom if ever are used to describe the entire land surface (i.e., they never fill space) as geographic entities (e.g., cities, counties) with legal standing are.

New Method for Specifying Geographic Footprints

The cities and neighborhoods in Los Angeles County were used to develop a new method for specifying geographic footprints. Many authors have described the historical evolution of Los Angeles' social landscape and highlighted its demographic transition from a small frontier town to a metropolis and world city (e.g., Hise 1997; Fulton 2001). The metropolitan area now spills over the borders of Los Angeles County into neighboring Ventura, San Bernardino, Riverside, and Orange Counties, and the built-up areas in many of the cities have coalesced to form one large contiguous urban area. Several books and articles describe the emergent spatial patterns that have produced a series of very distinctive neighborhoods and identified the historical processes that led to their production (e.g., Allen and Turner 1997; Modarres 1998). These neighborhoods are important in people's everyday lives and for navigating their way around this large metropolitan area, and there are processes now underway to use them to mobilize community participation in planning and decision-making (Talen 1999).

Given this background, we chose cities and neighborhoods to illustrate how geographic footprints can be specified rapidly. Cities are straightforward because their inception dates are well known and their boundaries are captured in numerous GIS datasets. The fuzzy and imprecise nature of neighborhoods, in contrast, causes many difficulties and pointed to the need to find a creative "satisficing" solution for this feature type. The new method described below incorporates three tasks—building a point shapefile,

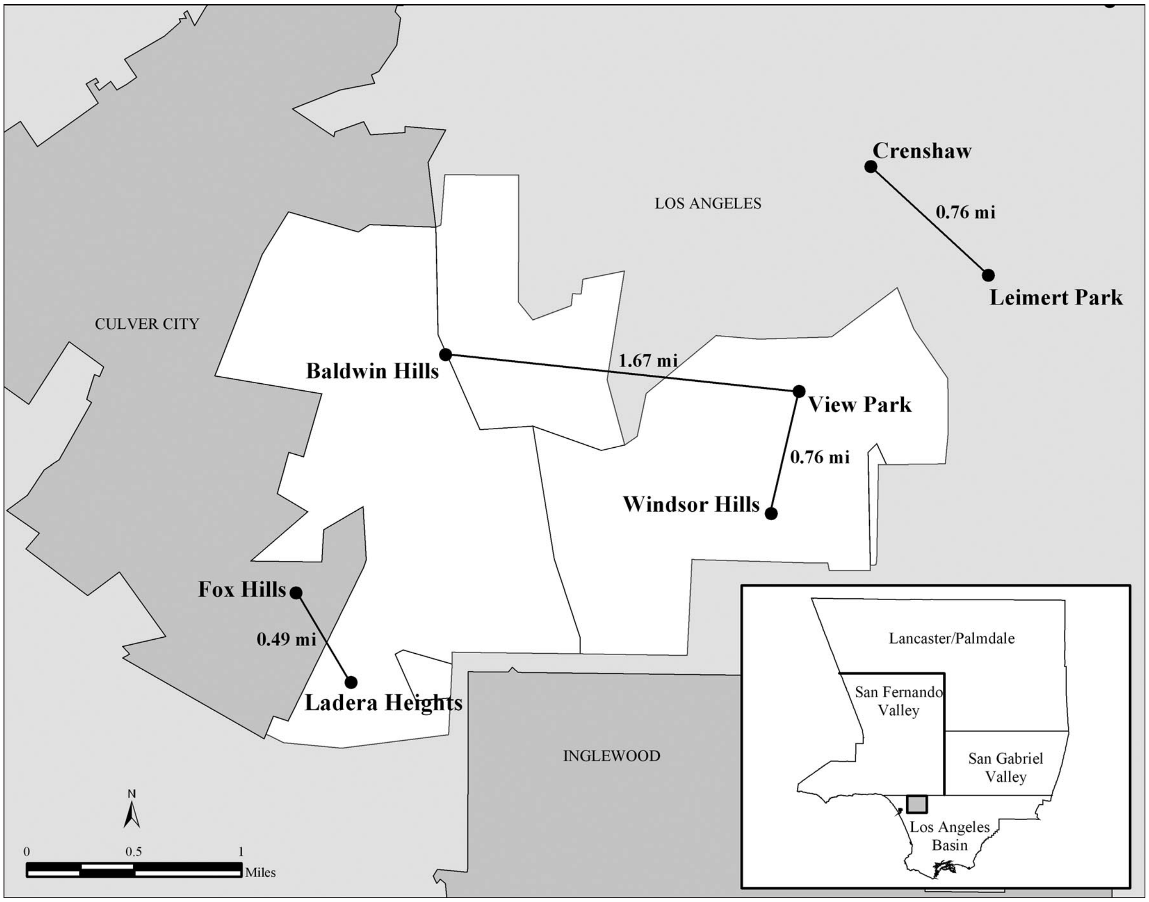

Figure 6. Map showing distance measurements for adjacent neighborhoods in a portion of Los Angeles County.

measuring the proximity between adjacent pairs of points, and generating circular-shaped footprints (i.e., polygons) for specific neighborhoods. The data sources and methods used for these tasks are described in the subsections that follow.

Building the Point Shapefile

A point shapefile was constructed using the neighborhoods listed in the 2001 Los Angeles County Street Guide and Directory (Thomas Bros 2001). These popular books, which are updated annually, show city boundaries, freeways, streets, and other major landmarks. The labels identifying the 88 cities and 218 neighborhoods recorded in the 2001 Los Angeles County Street Guide and Directory are printed in blue and a larger font size is used to distinguish city from neighborhood labels. In addition, some neighborhoods appear more than once because their names are printed on more than one page.

Street and highway layers acquired from Geographic Data Technology, Inc. were added to a new view and used along with the "heads-up" digitizing tools in ArcView™ (ESRI, Inc., Redlands, California) to add point features to a neighborhood point theme. The neighborhood labels in the Thomas Guide were treated as neighborhood centroids and recorded as point features. Mean x, y coordinates were calculated and used for neighborhoods with two or more labels in the 2001 Los Angeles County Street Guide and Directory.

Proximity Measurements

The distances between each of the digitized points and the closest neighboring points were measured next in ArcView™. The distances between these points varied substantially (Figure 6), although they were generally shorter in densely populated areas and larger in the northern half of the county. Los Angeles County was consequently divided into four regions—Lancaster/Palmdale, San Fernando Valley, San Gabriel Valley, and Los Angeles Basin—and mean distances between adjacent pairs of neighborhoods were calculated and used to build circular polygons in each of these regions (Figure 6).

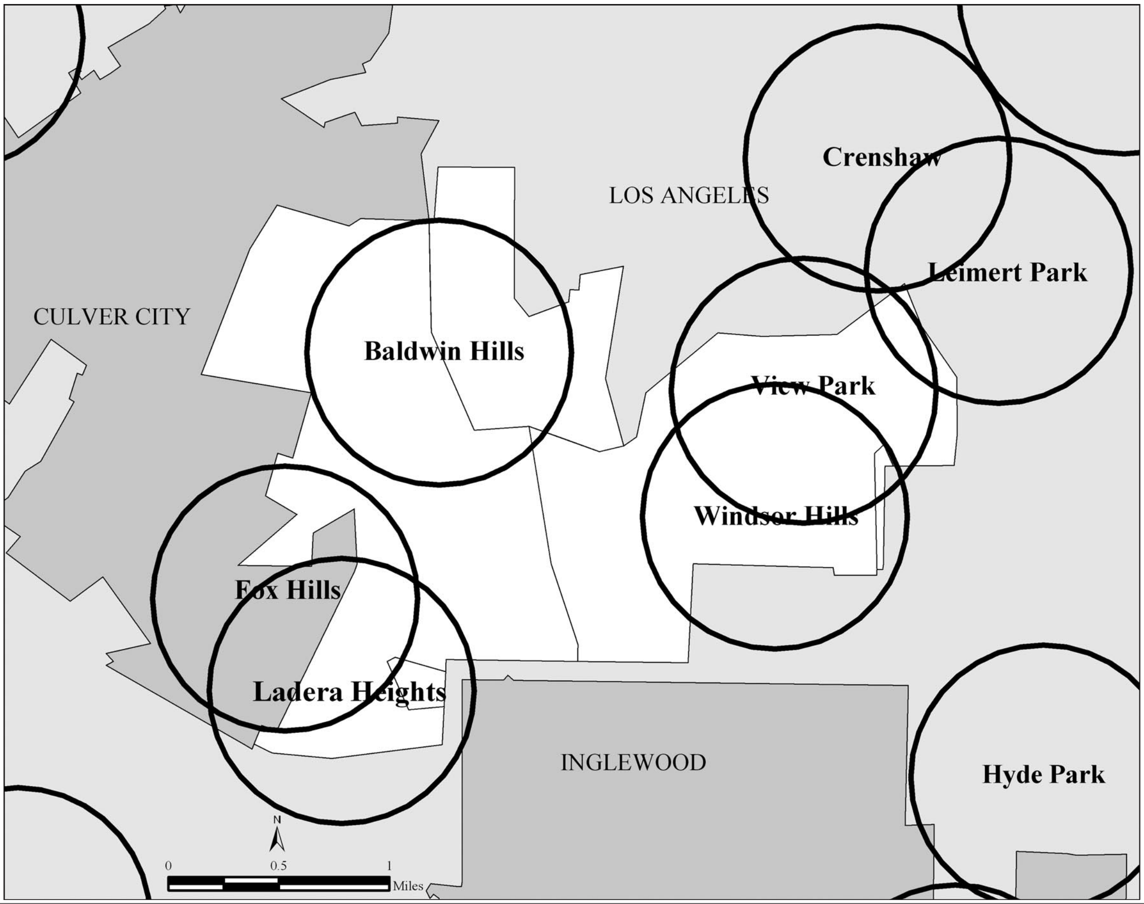

Figure 7. Map showing neighborhood footprints generated with new method for geographic area displayed in Figure 6.

Generating Circular-shaped Polygons

The four sets of points in this point theme (i.e., one set corresponding to each of the aforementioned regions) were then buffered using the four mean distances between adjacent neighborhood centroids calculated with the previous step. These buffer operations produced a series of circular-shaped polygons, and Figure 7 shows the final neighborhood footprints that were generated with this new method for the geographic area displayed in Figure 6.

Final Los Angeles County Neighborhood Footprints

Figure 8 shows the 218 neighborhood shapes (i.e., footprints) that were delineated with this new method superimposed on a series of polygons that record the boundaries of the 88 cities in Los Angeles County. Table 1, in turn, lists the four regions and summarizes the number of neighborhood pairs for which distances were measured, mean distances between adjacent neighborhood centroids, and radii used to construct the circular-shaped polygons shown in Figure 8. It is important to note that the circles vary in size, depending on the region in which they fall and taken as a whole, they share two desirable characteristics: (1) they cover approximately 11.1 percent of Los Angeles County (i.e., they are not space filling), and (2) 103 of the 218 neighborhood footprints (47 percent) overlap one or more other neighborhood footprints.

The first characteristic noted above (i.e., that the neighborhood shapes do not cover the entire county) is desirable because there are many unpopulated areas in the northern half of the county and in the Santa Monica and San Gabriel Mountains. Hence, there is no reason to extend the neighborhood boundaries to fill the entire land area in these parts of the county (as happened with the Thiessen polygons constructed with Alani et al.'s (2001) method in Figure 5, for example). In addition, the reliance on the mean distances reported in Table 1 to construct the circular-shaped polygons shown in Figure 8 meant that there was no need to delineate built-up areas throughout the county. This would constitute an especially difficult, time-consuming, and expensive task in the urban core because extensive fieldwork would probably be needed to clarify the transition zones between neighborhoods in these areas. This task might be equally difficult at the margins of urban areas, even though the rural–urban boundaries might be delineated with satellite imagery in these instances because the analysis would need to be repeated at regular intervals to capture the changes in urban boundaries over time.

The second characteristic noted above (i.e., the overlapping neighborhood footprints) is also desirable because the boundaries are inherently fuzzy and there are numerous societal actors and trends that will tend to keep them imprecise. Real estate agents and residents living close to more affluent neighborhoods and/or areas that are perceived to be more desirable may want to extend the boundaries of these adjacent neighborhoods to include their properties, for example. Figure 7 shows six areas of the county that might be assigned to two or more neighborhoods with the method described in this paper. The Fox Hills neighborhood spills over into Ladera Heights (and vice versa), Windsor Hills spills over into View Park (and vice versa), View Park spills over into Crenshaw and/or Leimert Park and vice versa. Several of these neighborhoods cover parts of two or more cities or the white areas in Figures 6 and 7 that show unincorporated parts of the county which are completely surrounded by one or more cities. Overall, there are 86 instances where some part of the county is assigned to two or more neighborhoods in Figure 8.

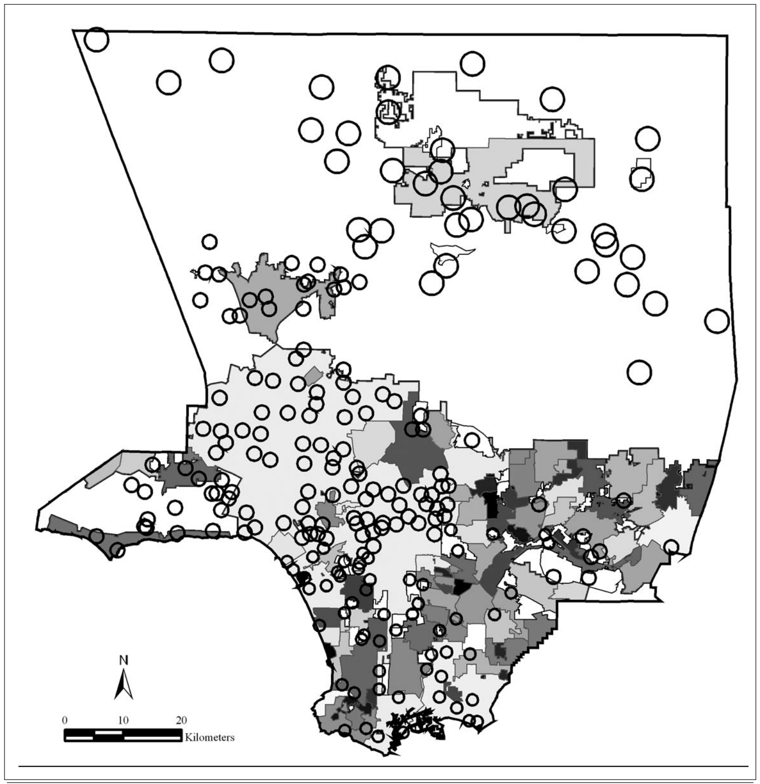

Figure 8. Neighborhood footprints currently used for Los Angeles Digital Gazetteer.

Discussion and Conclusions

The first efforts to design and build digital map libraries started nearly ten years ago, but these were soon overwhelmed by several more ambitious projects that aimed to use geography to guide the organization of and access to much more broadly conceived digital libraries. Most of these projects were organized around either the "geolibrary" and/or Digital Earth metaphors, and their ultimate success will depend on the validity of the argument that geographical organization and access to digital libraries will provide improved access to digital assets and contribute to new ways of thinking.

The USC Digital Library seeks to test the assertion that geographical organization and access will have the desired types of impacts. It complements the DLESE, ADEPT, and numerous regional projects by utilizing standard, off-the-shelf technologies and tools, avoiding the huge startup expenses incurred in other projects, and tackling a larger and more diverse group of digital assets and library patrons. The decision to build the geospatial browser and digital gazetteer components was inspired by our desire to build a digital library that provided multiple paths for accessing information and supported new ways of intellectual work (see Soergel (2002) and Crane (2004) for examples of the types of innovations needed here).

This new digital library is being constructed at a time when many in the library field are tackling the challenges and issues of how to deliver digital assets to users when there is currently no digital library standard, just many standards for component parts of a digital library. The USC Digital Library project is utilizing many of these component standards in order to build a usable, scalable, and robust digital library that can help the library field identify best practices on the road to building a model for digital libraries. We have or plan to utilize several types of standards (including DCMES, Open Archives Initiative Protocol for Metadata Harvesting (OAI-

Regions*	Number of Neighborhood Pair Measurements	Mean Distance between Adjacent Pairs (km)	Radius Used to Construct Circles (km)
Lancaster/Palmdale	14	1.997	2.013
San Fernando Valley	76	1.224	1.208
San Gabriel Valley	8	1.240	1.208
Los Angeles Basin	48	0.998	0.966

* As shown in Figure 6.

Table 1. Final distance measurements used to generate neighborhood footprints.

PMH), and LDAP with Shibboleth for authentication) to develop our system and to share our resulting digital archive over the web. We plan to use off-the-shelf tools to build and organize a large digital library that utilizes geography as one of the building blocks for organization and access. We will produce a new digital gazetteer focused on what today is the Los Angeles Metropolitan Region that can be utilized in other digital library projects. Hence, we have implemented the Gazetteer Content Standard and Feature Type Thesaurus proposed by Hill and Zheng (1999) and specified geographic footprints and the accompanying metadata for nearly 4,500 names in this gazetteer to date. The choice of names to be included is partially guided by the list of names specified in the ADL Gazetteer (Hill and Zheng 1999), the Getty Thesaurus of Geographic Names (Getty Research Institute 2000), and the Los Angeles Comprehensive Bibliographic Database (Ethington et al. 2004).

The new method presented in this paper used the distances between pairs of nearest neighborhoods to generate circular-shaped neighborhood footprints for inclusion in a Los Angeles Gazetteer. The results show that this method produces neighborhood shapes with several desirable features given the imprecision inherent in neighborhood characteristics and boundaries. In addition, this new method is reproducible and extendable. The circular-shaped footprints recorded in Figure 8 were generated using simple GIS tools and a published street map guide as input. The method is reproducible given that others could utilize this same method with the same data sources and generate the same results. The method is extendable in two ways: first, these circular-shaped footprints can be treated as the first approximation and subsequent research efforts might be used to delineate more precise shapes for one or more time periods (as discussed below); and second, others could use our method with equivalent data to generate neighborhood shapes in other metropolitan areas (e.g., Houston, Texas). This method could also be simplified further by specifying a default distance (i.e. radii for the specification of the circular-shaped polygons) and skipping the second step in the three-step method described earlier in this paper.

The circular shape was chosen for two reasons. First, circles represent the most compact shapes and we thought this was a desirable property for an initial attempt to delineate (estimate) "real-world" boundaries. Second, the reliance on circles meant that the shapes of adjacent neighborhoods were defined independently of one another. New research (i.e., a master's thesis that mapped the changes in the boundaries of a neighborhood by decade from inception to the present day) that provided more detailed boundary information for one neighborhood would not automatically alter the boundaries of one or more adjacent neighborhoods given our approach.

Another important advantage of this new method is the opportunity it provides to specify geographic footprints for new archive materials. We plan to establish a collection development working group in the next year that will be charged with prioritizing materials for ingest into the digital archive and helping to identify funding sources to help defray the cost of collection development. The specification of geographic footprints is expensive and may not be appropriate for every collection that is added – in some instances, the materials may have multiple geographic references as is the case for a book written in New York about a summer spent in Brussels or Paris for example. This exemplifies the policy issue of indexing the instantiation date or location or subject and/or the date, location, or subject of the content. This issue is further complicated by multiple instantiations. DCMES was specified to record information on the digital instantiation rather than the original format instantiation or the actual intellectual content. But in practice, its users have taken a mixed approach (see http://dublincore.org for additional details). In trying to plan for precise yet comprehensive search results for users, these policy issues need to be carefully considered and implemented. On a more positive note, we believe that the method we have developed for generating geographic footprints for gazetteer features may be applied more generally and may offer a way to reduce the costs incurred specifying footprints for many of the materials that are candidates for ingest in the next few years. These initial footprints could be automatically upgraded when the footprints incorporated in the Los Angeles Digital Gazetteer are upgraded so long as their source and lineage is tracked as part of the metadata record saved with each of the digital assets.

Moreover, it may be possible to implement the new method for specifying footprints that was outlined in this paper on the fly using the map browser illustrated in Figure 1. Geographic footprints could be specified as points and circular-shaped polygons when materials are ingested into the archive and the map browser then used to vary the size of the circular-shaped polygons that serve as geographic search windows based on distances specified by the library patrons themselves. This last approach would provide a great deal of flexibility since library patrons would be able to start with the neighborhood shapes discussed in this paper but then specify their own areas of interest based on some fuzzy space (like a neighborhood) or one or more well-known landmarks in the event this initial search window retrieved too many or too few digital assets of interest. We plan to implement these types of capabilities in the next phase of the USC Digital Library so we can test whether or not: (1) library patrons find them useful; and (2) using geography as a foundation for organization and access in a regional archive can help to promote new forms and styles of student learning.

ACKNOWLEDGEMENTS

The authors would like to thank Ms. Barbara Shepard, Information Services Division, University of Southern California, and three anonymous reviewers for their helpful comments and suggestions on an earlier draft of this manuscript.

REFERENCES

Alani, H., C.B. Jones, and D.Tudhope. 2001. Voronoi-based region approximation for geographical information retrieval with gazetteers. *International Journal of Geographical Information Science* 15: 287-306.

Allen, J.P., and E. Turner. 1997. *The ethnic quilt: Population diversity in southern California*. Center for Geographical Studies, California State University, Northridge, California.

Boxall, J. 2002. Geolibraries, the Global Spatial Data Infrastructure, and Digital Earth: A time for map librarians to reflect upon the moonshot. *INSPEL* 36: 1-21.

Boxall, J. 2003. Geolibraries: Geographers, librarians, and spatial collaboration. *The Canadian Geographer* 47: 18-27.

Buttenfield, B. P. 1998. Looking forward: Geographic information services and libraries in the future. *Cartography and Geographic Information Systems* 25: 161-71.

Burrough, P. A., and A. U. Frank (eds). 1996. *Geographic oObjects with indeterminate boundaries*. London, U.K.: Taylor and Francis,

Crane, G. 2004. The Perseus Digital Library [http://www.perseus.tufts/edu/].

Ethington, P. J., H.L. Rudd, L. Sipe, and L. McCann. 2004The Los Angeles Comprehensive Bibliographic Database [http://www.usc.edu/isd/archives/arc/lacbd/].

Fulton, W. 2001. *The reluctant metropolis*. Baltimore, Maryland: The John Hopkins University Press.

Getty Research Institute. 2000. *Getty thesaurus of geographic names*. J. Paul Getty Trust, Los Angeles, California.

Goodchild, M. F. 1998. The geolibrary. In: S. Carver (ed.), *Innovations in GIS 5*. London, U.K.: Taylor and Francis. pp. 59-68.

Goodchild, M. F. 1999. The future of the gazetteer [http://www.alexandria.ucsb.edu/~lhill/CGIE_website/perspectives/brandt.htm].

Harpring, P. 1997a. The limits of the world: Theoretical and practical issues in the construction of the *Getty Thesaurus of Geographic Names*. In: *Proceedings of the Fourth International Conference on Hypermedia and Interactivity in Museums (ICHIM '97)*, Archives and Museum Informatics, Le Louvre, Paris. pp. 237-51.

Harpring, P. 1997b. Proper words in proper places: The thesaurus of geographic names. *MDA Information* 2: 5-12.

Hawkins, B. L., and P. Bratton. 1998. *The mirage of continuity: Reconfiguring academic information resources for the 21st century*. Council on Library and Information Resources and Association of American Universities, Washington, D.C.

Herold, P., T.D. Gale, and T.P. Turner. 1999. Optimizing web access to geospatial data: The Cornell University Geospatial Information Repository (CUGIR). *Issues in Science and Technology Librarianship* 21 [http://www.library.ucsb.edu/istl/00-winter/article2.html].

Hill, L. L. 1999. Gazetteer and collection-level metadata developments. In: R. T. Kaser, and V. C. Kaser (eds), *Metadiversity: The grand challenge for biodiversity information management through metadata*. U.S. Geological Survey, Biological Resources Division, Reston, Virginia. pp. 141-5.

Hill, L. L. 2000. Core elements of digital gazetteers: Placenames, categories, and footprints. In: J. Borbinha, and T. Baker (eds), *Proceedings of the Fourth European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2000)*, Berlin, Springer. 280-90.

Hill, L.L., and Q. Zheng. 1999. Indirect geospatial referencing through place names in the digital library: Alexandria Digital Library experience with developing and implementing gazetteers. In: *Proceedings of the Sixty-second Annual Meeting of the American Society for Information Science,* Washington, D.C. pp. 57-69.

Hill, L.L., J. Frew, and Q. Zheng. 1999. Geographic names: The implementation of a gazetteer in a georeferenced digital library [http://www.dlib.org/dlib/january99/hill/01hill.html].

Hise, G. 1997. *Magnetic Los Angeles: Planning the twentieth century metropolis*. Baltimore, Maryland, The John Hopkins University Press.

Hunt, L., and P.E. Ethington. 1997. The utility of spatial and temporal organization in digital library construction. *Journal of Academic Librarianship* 23: 475-83.

Keller, C. P. 2001. The map library's future. *Cartographic Perspectives* 38: 73-7.

Lam, C. S., J.P. Wilson, and D.A. Holmes-Wong. 2002. Building a neighborhood-specific gazetteer for a digital archive. In: *Proceedings of the Twenty-second International ESRI User Conference*, San Diego, 7-11 July 2002. Environmental Systems Research Institute, Redlands, California. CD-ROM.

Lee, F., S. Bressan, and B. Ooi. 2001. Hybrid transformation for indexing and searching web documents in the cartographic paradigm. *Information Systems* 26: 75-92.

Lim, E-P., D. H-L. Goh, Z. Liu, W-K. Ng, C. S-G. Khoo, and S.E. Higgins. 2002. G-Portal: A map-based digital library for distributed geospatial and georeferenced resources. In: *Proceedings of the 2002 ACM-IEEE-JCDL Workshop on Visual Interfaces (JCDL'02),* Portland, Oregon.

Lopez, X. R., and M.L. Larsgaard. 1998. Towards a California geospatial digital library: A strategy for networked knowledge. *Cartography and Geographic Information Systems* 25: 133-41.

Mapping Science Committee. 1999. *Distributed geolibraries: Spatial information resources*. National Academy Press, Washington, D.C.

Millea, N. 1998. Delivering digital data into the library: The Digimap project and its impact on the map room—the Bodleian Library experience. *LIBER Quarterly* 9: 189-200.

Mogk, D.W., C.A. Manduca, and M.R. Marlino. 2000a. Buiding a community Digital Library for Earth System Education (DLESE). *EOS Transactions of the American Geophysical Union* 81(19): S69.

Mogk, D. W., C.A. Manduca, and M.R. Marlino. 2000b. The Digital Library for Earth System Education (DLESE): A community resource to meet community needs. *EOS Transactions of the American Geophysical Union* 81(48): F289.

Modarres, A. 1998. Putting Los Angeles in its place. *Cities* 15: 135-47.

Perkins, C. R., and R. B. Parry (eds). 2001. *The Map Library in the new millennium.* Chicago, Illinois. ALA Editions.

Simon, H. 1979. Rational decision-making in business organizations. *American Economic Review* 69: 493-513.

Smith, T. 1996. A digital library for geographically referenced materials. *IEEE Computer* 29(5): 54-60.

Smith, T., G. Janee, J. Frew, and A. Coleman. 2001. The Alexandria Digital Earth ProtoType (ADEPT) System. In *Proceedings of the First ACM+IEEE Joint Conference on Digital Libraries (JCDL'01),* Roanoke, Virginia.

Soergel, D. 2002. A framework for digital library research. *D-Lib Magazine* 8(12) [http://www.dlib.org/dlib/december02/soergel/12soergel.html].

Talen, E. 1999. Sense of community and neighborhood form: An assessment of the social doctrine of new urbanism. *Urban Studies* 36: 1361-79.

Thomas Bros. 2001. *The Los Angeles County Street Guide and Directory.* Thomas Bros Maps, Irvine, California.

U.S. Geological Survey. 1998. Geographic Names Information Systems (GNIS) [http://mapping.usgs.gov/www/gnis].

Zhou, X., J.D. Yates, and G. Chen. 2001. Using a visual search interface for WWW applications. *Information Systems* 26: 61-74.