LIBRARY TRENDS

Geographic Information Systems and Libraries

Jaime Stoltenberg and Abraham Parrish, Issue Editors

Volume 55
Number 2
Fall 2006

Library Trends
Geographic Information Systems and Libraries
Volume 55 | Number 2 | Fall 2006

Published by The Johns Hopkins University Press

LIBRARY TRENDS

Library Trends, a quarterly thematic journal, focuses on current trends in all areas of library practice. Each issue addresses a single theme in depth, exploring topics of interest primarily to practicing librarians and information scientists and secondarily to educators and students.

Editors: Boyd Rayward and John Unsworth
Publications Committee: Kate McDowell, Boyd Rayward, John Unsworth, Terry Weech, Marlo Welshons

Library Trends (ISSN 0024-2594) is published quarterly—in Summer, Fall, Winter, and Spring—for the University of Illinois Graduate School of Library and Information Science (GSLIS) by The Johns Hopkins University Press, 2715 N. Charles Street, Baltimore, MD 21218-4363.

Subscriptions. Institutional rate per year is $120; individual rate is $75.00. Students with ID may subscribe for $45.00. *For all subscriptions mailed outside the U.S.A., please add foreign shipping:* $7.50 (Canada or Mexico); $15.80 (outside North America). Single copy rate (current or back issues) is $36.00 for institutions; $23.00 for individuals. If applicable, please add state taxes or GST. *Prepayment is required for shipment.* All orders, address changes, and other business correspondence should be addressed to The Johns Hopkins University Press, Journals Division, P.O. Box 19966, Baltimore, MD 21211-0966 (USA); telephone (410) 516-6987; fax (410) 516-6968; email jrnlcirc@press.jhu.edu; or toll free 1-800-548-1784 (US and Canada only). Claims for replacement of missing issues must be received within three months of mailing (six months for foreign addresses). All notices of change of address should provide both the old and new address.

Periodicals postage paid at Baltimore, Maryland, and at additional mailing offices. POSTMASTER: Send address changes to *Library Trends*, The Johns Hopkins University Press, Journals Division, 2715 N. Charles Street, Baltimore, MD 21218-4363.

Library Trends is indexed or abstracted in *Current Contents, Library Literature, and PAIS.* The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences–Permanence of Paper for Printed Library Materials. ANSI Z39.48-1984.

Procedures for Proposing and Guest Editing an Issue of *Library Trends*

We encourage our readers to submit ideas for future *Library Trends* themes; issue topics are developed through recommendations from members of the Publications Committee and from reader suggestions. We also encourage readers to volunteer to be issue editors or to suggest others who may be willing to be issue editors.

The style and tone of the journal is formal rather than journalistic or popular. *Library Trends* reviews the literature, summarizes current practice and thinking, and evaluates new directions in library practice. Papers must represent original work. Extensive updates of previously published papers are acceptable, but revisions or adaptations of published work are not sought. Although *Library Trends* is not formally peer-reviewed, guest editors invite articles for submission, which are then critically reviewed by both the guest editor and journal editor. **Unsolicited articles are not accepted**.

An issue editor proposes the theme and scope of a new issue, draws up a list of prospective authors and article topics, and provides short annotations of each article's scope or else gives a statement of philosophy guiding the issue's development. Please send your ideas, inquiries, or prospectus to Library Trends, GSLIS Publications Office, 501 E. Daniel Street, Champaign, IL 61820-6211.

LIBRARY TRENDS

Geographic Information Systems and Libraries

Jaime Stoltenberg and Abraham Parrish
Issue Editors

Volume 55
Number 2
Fall 2006

Geographic Information Systems and Libraries

Contents

Introduction
Jaime Stoltenberg 217

GIS Collection Development within an Academic Library
Patrick Florance 222

Legal Considerations in the Dissemination of Licensed Digital Spatial Data
Patti Day and Chieko Maene 236

Building a System to Disseminate Digital Map and Geospatial Data Online
Tsering Wangyal Shawa 254

Libraries as Distributors of Geospatial Data: Data Management Policies as Tools for Managing Partnerships
Gail Steinhart 264

Geospatial Web Services and Geoarchiving: New Opportunities and Challenges in Geographic Information Services
Steven P. Morris 285

Digital Preservation of Geospatial Data
Julie Sweetkind-Singer, Mary Lynette Larsgaard, and Tracey Erwin 304

Building a Library GIS Service from the Ground Up
Rhonda Houser 315

Improving GIS Consultations: A Case Study at Yale University Library
Abraham Parrish 327

Centralized vs. Distributed Systems: Academic Library Models for GIS and Remote Sensing Activities on Campus
Joe Aufmuth 340

GIS Mentoring
Kim M. Ricker 349

Introduction

Jaime Stoltenberg

Geographic Information Systems (GIS) data collections and services within the library environment are not new concepts. For years libraries have been active in collecting geospatial data and handling GIS-related reference questions. In the early 1990s librarians often contributed to the literature while researching and learning the best methods for developing geospatial collections and incorporating GIS services in libraries. But during the last ten years, while there have been enormous advancements in the field, the literature related to GIS in libraries has been somewhat scarce. In a time when librarians are busy perfecting their technical GIS skills, creating and defining service models, working closely with data producers, discovering new ways of building collections, and researching ways to archive and preserve digital geospatial data, it is only appropriate that the current issues and topics within the field of GIS librarianship be highlighted in the literature.

In this first-ever issue of *Library Trends* dedicated to GIS and libraries, we present thirteen authors contributing ten profoundly informative articles that address a variety of current issues and trends in the field. These authors share their experience, expertise, and opinions regarding the most prominent concerns within the field of GIS librarianship today. In an effort to create a comprehensive issue that spans the most current and relevant topics, careful consideration was made in the selection of articles included in this issue.

This introduction provides a brief overview of the topics covered in this issue and gives readers a taste of the diversity of these topics. Readers will see emerging trends among the topics in several articles, particularly those related to geospatial data acquisitions, distribution, and preservation. They will also gain exposure to new ideas related to GIS mentoring programs,

LIBRARY TRENDS, Vol. 55, No. 2, Fall 2006 ("Geographic Information Systems and Libraries," edited by Jaime Stoltenberg and Abraham Parrish), pp. 217–221

analysis of GIS reference statistics, the application of centralized and distributed GIS service models, and geoarchiving.

The issue of how libraries acquire geospatial data is a prevalent one in this collection of articles. Many of the articles discuss the subject but focus on different aspects of building data collections. One author has devoted his entire article to geospatial data collection development. Patrick Florance from Tufts University presents some of the current issues related to building geospatial data collections within libraries. He begins with a fundamental discussion of the nature of geospatial data and stresses the importance of understanding data types, formats, and scales. Florance provides examples of some important considerations essential to successful geospatial data collection development: cost, availability, licensing, distribution policies, documentation and metadata, software, and hardware. Along with these examples, he offers suggestions for success based on his experience building the geospatial data collection at Harvard University.

The process by which data are distributed to library users depends on many factors including size, format complexity, and potential restrictions applied to the data as a result of copyright or license agreements created by data producers. Libraries striving to build collections of geospatial data are increasingly faced with a variety of legal issues related to these agreements. Patti Day and Chieko Maene, former Digital Spatial Data Librarians at the University of Wisconsin–Milwaukee, write about their experiences dealing with copyright restrictions and license agreements applied to several sets of geospatial data acquired by the American Geographical Society Library (AGSL). Day and Maene's article includes a well-researched overview of relevant legal issues and a discussion of freedom of information laws at the federal level and open records laws at the state level to help guide other libraries facing challenges in providing access to licensed geospatial data.

Expanding on the subject of geospatial data dissemination, Tsering Wangyal Shawa from Princeton University writes about his experience developing a system to provide library users with the ability to access many different forms of geographic information via the Internet. Shawa explains the processes by which paper map collections are digitized, metadata records are created, and imagery is made available for preview and direct download from Princeton's Digital Map and Geospatial Information Center. He provides detailed examples of various software packages and development technologies utilized in the project.

Building relationships with data producers is an essential role a librarian must play in order to ensure library users have the ability to access the wealth of geospatial data being produced today. In a well-researched article, Gail Steinhart of Cornell University writes on the issue of geospatial data collections in libraries, but she focuses more on library–data producer relationships. She discusses the importance of building these relationships and talks about how the formulation of an actual data management and

distribution policy can help define critical parameters within a partnership. Issues taken into consideration on both sides of an agreement include intellectual property rights, liability issues, distribution methods, data management practices, and security risks posed by geospatial data. Steinhart draws on knowledge gained from her experiences formulating a policy for the Cornell University Geospatial Information Repository (CUGIR).

Charging into the frontier of geospatial Web services, Steve Morris from North Carolina State University (NCSU) tackles the opportunities and challenges related to streaming geospatial data and geoarchiving in libraries. Past trends show many libraries working to build physical or "in-house" collections of geospatial data on disk. Today, data producers are making geospatial data available to the public through geospatial Web services published via the Internet. These geospatial Web services are prompting changes in the way libraries develop their data collections. In cases where a streaming Web service is the only mechanism available to obtain the data, libraries are faced with challenges in managing, maintaining, and archiving this information. Morris addresses important issues related to the role libraries should play in the development and utilization of emerging geospatial Web services and the long-term preservation challenges of these types of data distribution systems. He cites examples from the work being done at NCSU as the result of an award from the Library of Congress's National Digital Information Infrastructure and Preservation Program (NDIIPP).

While Morris introduces the concern of long-term preservation of geospatial data, Julie Sweetkind-Singer and Tracey Erwin of Stanford University, along with Mary Larsgaard of the University of California at Santa Barbara (UCSB), write specifically on this subject. They define and discuss issues such as data versioning, copyright, the complexity of geospatial file formats, and how these issues pose unique challenges when thinking about the ways libraries can preserve the information contained within geospatial files. Sweetkind-Singer, Erwin, and Larsgaard lay the foundation for their article by discussing the research and development currently taking place with a grant from the NDIIPP, awarded to both Stanford and UCSB. As part of this project, the development of two prototype archives for housing data and the creation of a format registry to describe the data being stored will assist in answering some very important geospatial data preservation questions.

Over the years many libraries have established collections of geospatial data, but how many have developed and implemented GIS services? In a comprehensive account of how the University of Kansas developed a GIS service model, Rhonda Houser describes how libraries can begin to offer services aimed at assisting library users with using GIS software as well as locating and analyzing geospatial data. Creating a service model for GIS in the library is an important activity aimed at meeting the needs of a rapidly growing user community that can span many disciplines. Houser describes

how GIS services including instruction, software training, and GIS project consultation can be implemented in a library. She stresses the importance of utilizing mechanisms like outreach and publicity to ensure that these services are successful and properly meet the needs of users.

Reference services related to GIS can run the gamut in variations of applications, complexity and depth of research problems, technical abilities required to perform analytical operations within GIS software, and challenges in locating the proper data sources. Keeping well-documented statistics on the types of reference questions and how they are answered is a way to create a database of knowledge that could be turned into instructional guides aimed at providing users with quick access to answers to their questions. Until now, research related to the analysis of GIS reference statistics and the quality of GIS consultations in academic libraries is virtually absent in the literature. Abraham Parrish from Yale University describes four years of accumulated reference statistics. He relates the analysis of these statistics to the librarian's ability to provide effective consultations. Based on his database of over 5,700 records, Parrish provides examples of a wide variety of questions, the average amount of reference time spent with each library user, different types of library users utilizing GIS reference services, and total average megabytes of data disseminated to a library user at a given time. He compares GIS reference techniques and processes with other traditional library reference techniques and processes to show how GIS requires more of a "consulting" approach.

GIS services in academic libraries attract users from many departments or centers on campus. To achieve success, services developed within the library should be directed toward the needs of a broad user community and should fit into the larger service model architecture of the university. In an interesting discussion of two different models, Joe Aufmuth from the University of Florida defines both the distributed and centralized models of GIS services and data delivery methods from the enterprise GIS perspective. Viewing the academic library as an enterprise, he compares the advantages and disadvantages of these service models. Following a descriptive explanation of both models, he suggests that for some, the best method for providing GIS services may be the result of a hybrid of the two models. Aufmuth cites specific examples of both service models at the University of Florida and Florida International University.

Reference and instruction services are successful when there are knowledgeable staff members available to assist library users with complex research questions or problems. Kim Ricker of the University of Maryland writes about establishing an effective GIS mentoring program to help train a staff whose goal is to meet the needs of the GIS community on campus. While literature on the general topic of mentoring in libraries is seemingly abundant, Ricker found that articles dealing specifically with GIS mentoring were nonexistent. Her innovative ideas offer a fresh perspective on develop-

ing a mentoring program designed to train a staff capable of handling the highly specialized and technical aspects of GIS reference and instruction services. She includes an informative ten-point framework for GIS mentoring largely based on her experiences implementing such a program with graduate assistants at the University of Maryland.

A shift in how libraries are developing and delivering GIS data collections to library users represents the most definitive trend in the field today. Geospatial Web services, library–data producer relationships, and license or copyright issues are just a few trends that illustrate the different ways libraries now deal with GIS collections and services in contrast to methods of the past.

The purpose of this special "GIS in Libraries" *Library Trends* issue is not only to discuss the current issues within the field but also to provide an opportunity for readers to learn from what other libraries have accomplished with regard to GIS collections and services in the last decade. Through the experiences documented by these authors, readers will learn about new ways to develop and share their geospatial collections, how to create effective service models and mentor staff, why data-sharing relationships are important, and why it is absolutely essential to think about best practices for archiving and preserving geospatial data.

Jaime Stoltenberg is the Map and Geographic Information Systems (GIS) Librarian at the University of Wisconsin–Madison's Arthur H. Robinson Map Library where she manages a collection of over 500,000 items and provides reference services in all subject areas related to maps, geography, cartography, geospatial data, and GIS. Jaime has experience in the acquisition and management of geospatial data collections as well as GIS software and project management within the academic library environment.

GIS Collection Development within an Academic Library

Patrick Florance

Abstract

Locating usable spatial data is essential for the application and use of Geographic Information Systems (GIS). GIS data collection development constitutes a core element of GIS services within academic libraries. Managers of geospatial resources require a fundamental understanding of the nature and use of GIS data. In the creation of a GIS collection development policy, library professionals should consider the established collection development policy, needs of the GIS user community, campus GIS services, and library infrastructure. Library professionals also need to employ a variety of online resource guides and spatial search engines and navigate a network of government agencies, academic institutions, commercial enterprises, and GIS professionals to locate, select, and acquire spatial datasets. When making decisions regarding GIS data acquisition, the selector should consider cost, availability, license agreements and distribution policies, documentation, data structures, and software and hardware.

Introduction

Since the late 1990s much has changed in the world of Geographic Information Systems (GIS): computer memory has become more accessible, the fields of geographic information science and spatial analysis have spread across disciplines, government agencies and commercial enterprises have developed massive spatial databases, high-resolution satellite imagery has become publicly available, a suite of software has been developed to meet the specialized needs of industry, and the Internet has emerged as a tool for data dissemination and visualization. There has also been a significant increase in new GIS positions within academic libraries as they struggle to

LIBRARY TRENDS, Vol. 55, No. 2, Fall 2006 ("Geographic Information Systems and Libraries," edited by Jaime Stoltenberg and Abraham Parrish), pp. 222–235

develop, maintain, and expand their GIS services. These positions include GIS specialists, GIS/data librarians, GIS/map librarians, digital cartographers, spatial data specialists, and GIS coordinators. Nevertheless, the principles of GIS have not changed all that much over the past few years. Tomlin defines GIS as “a configuration of computer hardware and software [and personnel] specifically designed for the acquisition, maintenance, and use of geographically referenced data” (1990, p. xi). When developing GIS services, three core components must be addressed: computer hardware and software, personnel, and data management (Longstreth, 1995). While all GIS service elements are equally important, a particular emphasis exists regarding the GIS data (Jablonski, 2004; Lamont, 1997, van Loenen & Onsrud, 2004), especially since data development or conversion can be extremely labor intensive (Goodchild & Longley, 1999). As a result, the availability of preexisting data often determines the feasibility and geographic area of a research project. This article examines the development of a spatial data collection within an academic setting and addresses the selection, acquisition, and source of spatial data.

The Nature of GIS Data

A fundamental understanding of the nature of GIS data is required before one can locate and use spatial data. The terms spatial data, geospatial data, and GIS data—that is, digital, geographically referenced data—will be used interchangeably in this article. GIS data are generally used to represent or model both physical and administrative geography. Physical features encompass both anthropogenic and natural features on or below the surface of the earth. Anthropogenic features typically include cultural phenomena, such as roads, railways, trails, buildings, and bridges. Natural features include rivers, lakes, shorelines, soils, elevations, etc. Abstract or administrative features are generally cultural divisions or boundaries created and used by organizations and agencies to administer their affairs and resources. These typically include national, state, county, election district, school district, municipal, zoning, zip code, neighborhood, census tract, and parcel or property boundaries. The Committee on Licensing Geographic Data and Services provides a detailed synthesis of geographic data types available in the United States (2004, Appendix C).

Two basic methods exist for representing geographic features within a GIS (DeMers, 1997, pp. 97–101). The vector data structure is composed of an ordered list of points and represented by points, lines, and polygons. Vector graphics model discrete geographic features such as administrative boundaries, roads, buildings, and rivers. A graphic vector object is usually combined or linked with attribute information stored in a separate spreadsheet or database. The raster data structure is composed of a grid of cells or pixels used to model continuous data. The resolution is a measure of the dimension on the ground represented by each pixel. Typical raster datasets

include digital elevation models (DEMs), satellite imagery, digital orthophotography, land use/cover, and georeferenced digital images of maps.

GIS data are scaled models or abstractions of reality (Goodchild & Longley, 1999). Understanding the scale and precision of spatial data is essential for both locating and using GIS data. The scale of data is described as a representative fraction such as 1:100,000 (Chrisman, 2002, p. 98; Clarke, 2003, p. 120). The representative fraction is a ratio of units measured on the map to units measured on the surface of the earth. In the example above, one inch on the map equals 100,000 inches on the surface of the earth. The smaller the ratio is, the larger the scale is. For example, a scale of 1:1,200 is considerably larger than a scale of 1:24,000. Datasets of larger scale usually possess more detail and a higher level of accuracy than those of smaller scale (Decker, 2001, pp. 16–19). The capture and generalization of features and attributes will likely vary from scale to scale. The map in Figure 1 shows the differences in representation of the Cape Cod shoreline using different scale data. In application, researchers must consider whether the scale of the data will yield needed results. For instance, if a researcher wants to conduct a site selection project to locate a new optimal location for assisted living in Boston based on municipal transit, a 1:1,000,000 scale transportation dataset will not be appropriate for the study. The scale is too small for the level of detailed information required. Hence, for the purposes of collection development, selectors should take into consideration the scale most appropriate to the needs of their patrons.

Scale requirements also impact where a selector must go to locate and acquire data. In the example above, a selector would probably need to go to Massachusetts state agencies as well as the City of Boston itself to acquire the necessary data because these agencies are more likely to maintain that level of detail in their GIS data.

Collection Development Policy

Much of the literature regarding GIS collection development suggests following an organization's current collection development policy (Lamont, 1997; Larsgaard, 1998, pp. 5–6; Stone, 1999). While this is often the case, there are times when the GIS collection development policy does not coincide with the organization's traditional collection policy. For example, at the Harvard Map Collection, the collection is global in nature with special emphasis on the United States, New England, and the greater Boston area. Our GIS collection certainly parallels the printed collection in this respect. However, while this is a good place to start, we must be careful: we have learned that the users of GIS are not necessarily part of the same user community as users of printed geographic information. These two groups of patrons often have different backgrounds and research needs. A typical example is an economist interested in the spatial econometric modeling of a given area. The economist wants to calculate the distances between

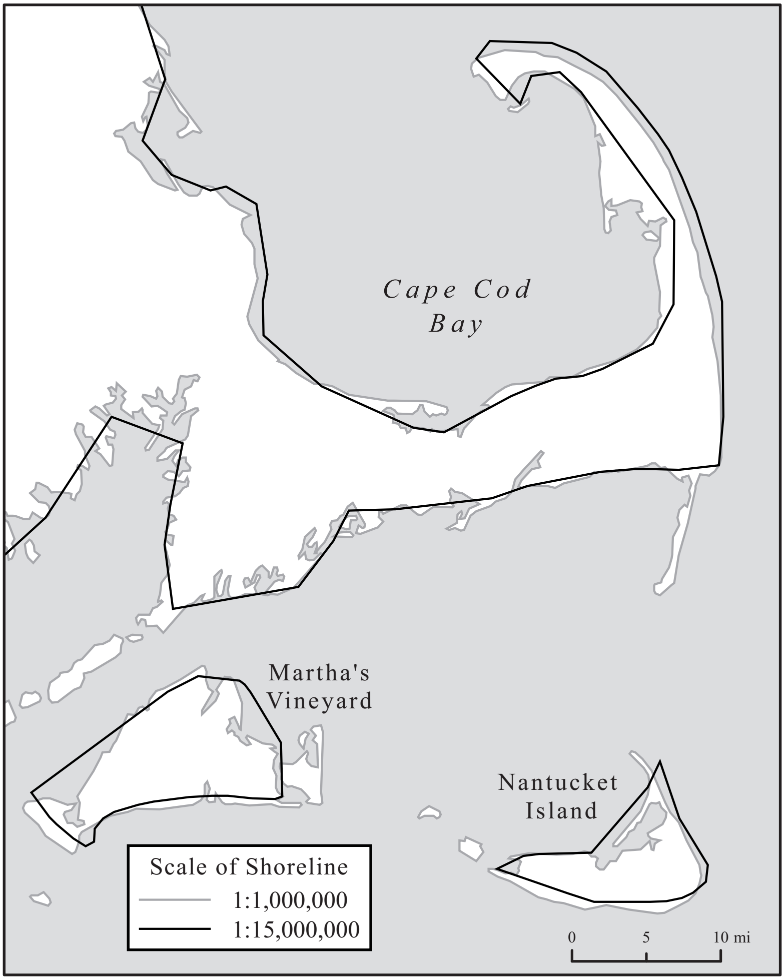


Figure 1. Cape Cod shoreline shown using different scale datasets.

thousands of features, such as cities or businesses or census block groups, resulting in robust distance matrices. Normally their studies would be too computationally intensive to use printed information. Another difference we have discovered at Harvard is that while current events influence the use of printed maps, they have had little impact on the use of GIS data. We purchased robust datasets of North Korea, Iraq, and Afghanistan, and they have received little or no use. The similarities or dissimilarities between printed map users and GIS users will vary from institution to institution. Understanding the needs of both communities of users will lead to better collection development decisions.

The content of GIS data also may span traditionally separate areas of expertise in library collection development, but it often makes sense to consolidate GIS services and data, despite varying subject matter. For example, the Harvard Map Collection does not collect printed geologic maps, but it does collect geological spatial data and applies the same approach to census data. GIS personnel are capable of helping users with this data, despite the fact that these areas do not fall within the Map Collection's areas of print-based concentration.

In addition, the differences between maintaining printed geographic information and GIS data may also impact collection development decisions. For instance, the Harvard Map Collection collects large-scale urban datasets of selected cities across the United States that it would not ordinarily collect in printed form because of the storage issues. In essence, although the established collection development policy provides a good foundation, the GIS collection development policy should also be influenced by the GIS user community, campus GIS services, and the library's infrastructure.

In formulating a GIS collection development policy, it is best to observe users' needs and requests for a few months to a year before making significant changes (Larsgaard, 1998, pp. 1–3; Stone, 1999). In an academic environment, collection development policies need to support teaching, research, and applications (Longstreth, 1995). A needs assessment is the best approach (Martindale, 2004). To begin a needs assessment, keep a database or spreadsheet of spatial datasets requested, noting the area of interest, type of data, contact information, and department. This information is useful in identifying and evaluating GIS data usage and trends. The information and statistics gathered can also be used to justify decisions regarding spatial data acquisition and services. Conducting outreach to departments using GIS can also be very useful. Ask what their most frequently used datasets or types of GIS applications are. These departments may also contribute their own data to the library's collection or contribute additional funds for data acquisition.

Data Selection and Acquisition

User demands, budgets, license restrictions, availability, data formats, and staffing resources influence decisions about selecting spatial data for acquisition. Focusing on the needs of the user community as a whole, rather than the special purpose or special project datasets, is critical (Longstreth, 1995). At Harvard, for instance, we generally only purchase and acquire datasets that we feel will be used somewhat frequently by students, faculty, and staff for teaching and research purposes. For that reason, we rarely purchase satellite imagery. To begin with, it is expensive. Also, when someone is conducting a remote sensing project, they usually need a very specific area at a very specific time period or periods. In most cases, it is unlikely that anyone else will use that dataset in the future. However, we

have purchased some high-resolution satellite imagery of foreign cities to help supplement our collection, generally when we are unable to acquire their urban-scale GIS data.

Financial considerations and licensing are certainly a key factor. Commercial and government datasets can be expensive, ranging into the thousands of dollars. Federal depository libraries may receive some U.S. government datasets for free (Lamont, 1997). Many public agencies and commercial vendors offer educational discounts. Mention that the data are for academic/noncommercial use and try to pay a one time charge, avoiding recurring subscription fees. The price of GIS data often relates to the licensing or usage restrictions of the data (Stone, 1999). When the library selector purchases GIS data, the selector is usually purchasing a licensed copy of the data. The licensing of a spatial dataset or product means "a transaction or arrangement . . . in which the acquiring party . . . obtains information with restrictions on the licensee's rights to use or transfer the information" (Committee on Licensing Geographic Data and Service, 2004, p. 25). At the Harvard Map Collection we generally encounter three types of license agreements:

- A free-use license, in which the data can be freely distributed to the general public
- A Harvard-wide site license, in which the data can be disseminated via a server or the Web to Harvard users (students, faculty, and staff) provided they have a Harvard identification or personal identification number (PIN)
- A single-use Harvard license, in which the data can only be used on a single computer at a time, but Harvard users can subset the data and take it away with them to work in a computing lab or at home

Read and negotiate the license agreement carefully. Always try to get a site license. A site license is less restrictive than a single-use license and makes the data much easier to disseminate. Some vendors provide a site license agreement at no additional cost. Some vendors charge five to ten times the amount of a single-use license for a site license, and some vendors refuse a site license altogether. Be honest and upfront with the respective agency or seller. Tell them that at the very least students and faculty need to be able to work at a single computer and subset the data in order to take some of it away to work on on their own. If an academic library cannot at least get that basic license, the product is usually not worth the money. Shop around and compare prices from different dealers. Many of them offer the same datasets at competitive prices. I have had good success with companies such as East View Cartographic, Map Mart, and LeadDog regarding licensing (see Table 1). East View Cartographic in particular has a long tradition of working with libraries. Other companies such as GfK Macon and Bartholomew generally offer more restricted-use data. Finally, partner

Table 1. Frequently Used Commercial Data Providers

Company Name	Web Site
ESRI	http://www.esri.com
East View Cartographic	http://www.cartographic.com
Map Mart	http://www.mapmart.com
GfK Macon	http://www.gfk-macon.com
GIS Data Depot	http://data.geocomm.com
LAND INFO	http://www.landinfo.com
LeadDog Consulting	http://www.goleaddog.com
Collins-Bartholomew	http://www.bartholomewmaps.com
ACASIAN	http://www.asian.gu.edu.au
Digital Globe	http://www.digitalglobe.com
GeoEye	http://www.geoeye.com
MapInfo	http://www.mapinfo.com

with other departments or libraries to combine funds to purchase the more expensive datasets. The Harvard Map Collection regularly partners with Harvard's Government Documents Library to purchase GIS data related to both collections.

Another important factor regarding the collection of GIS data is the online availability of the data. Is it better to download or order them on disk from the agency? Several issues influence this decision. Are the data frequently used? Are the data only available temporarily? Is the site unstable or unreliable? Are the files too large to download over the library's current bandwidth? Is there a cost involved? At Harvard we acquire all of the GIS data from the State (MassGIS) and the City of Boston on disk even though much of that data is available online. The data receive such heavy use and some of the files are so large that it just does not make sense to keep downloading them over and over again.

When acquiring GIS data, it is important to get the data in an easily accessible format and media. Will the data be easy to use in their delivered format, or will staff time have to be spent converting the data to make them useable? Converting data from one format to another can be time consuming, and datasets can be enormous. This has become less of an issue in recent years as data providers frequently offer a variety of deliverable formats. I strongly suggest products or formats that reduce GIS staff time. It is well documented in the literature that GIS services can easily double your staff's workload (Larsgaard, 1998, p. 8; Longstreth, 1995). At Harvard we have had great success with foreign and domestic census products that bundle the joined census attribute data with their corresponding boundaries. Although these products tend to be expensive, they greatly reduce patron and staff time spent joining the two disparate datasets. One example is the suite of census products produced by Geolytics (http://www.censuscd.com/). Their products make accessing U.S. census data within a GIS significantly easier and greatly reduce personnel time (Florance, 2004). I have also had

success with bundled census GIS products of China (All China Marketing Research Co. & China Data Center, 2004), India (ML Infomap, 2003), and Australia (Australian Bureau of Statistics, 2003).

Data is usually either downloaded from the Internet or put onto CD, DVD, or more recently portable hard drives. Raster data files tend to be very large and often come in compressed tiles or sections. In addition to acquiring the uncompressed tiles, try to get a single mosaicked image in compressed form such as JPEG2000, MrSID, or ECW. The single compressed image makes the data much easier to disseminate and work with. When ordering digital orthophotography from a commercial vendor, try to get an additional complete mosaicked image resampled at a much lower resolution. Users often want to print an image of an entire city or large area. Sending several high-resolution TIFFs or a giant SID to a printer often crashes the printer; creating pyramid layers can help as well. For very large datasets, those over a couple gigabytes, I have found portable external hard drives more stable and easier to use than CDs or DVDs. Some datasets will occupy several CDs or DVDs and switching from one CD to another is taxing and speeds are slower. In short, select formats and media that work best for the library's environment and that minimize staff time.

Acquiring local, large-scale GIS data requires establishing contacts with state, county, and municipal agencies. Many states have developed mature geographic information systems and widely disseminate their data. Policies regarding availability, pricing, and licensing vary widely among county and municipal agencies. Some local agencies provide all of their data for free, some agencies charge tens of dollars, some agencies charge hundreds of dollars, some agencies charge thousands of dollars, and some agencies refuse to provide data altogether. Try to attain the data for free or at least nearly free. I mention that the data will only be used for academic/noncommercial purposes and that we will act as an archive for the data. I usually mention that I manage and serve as the contact for GIS data for Harvard University, and I find this a more effective approach than sending all of the students directly to the agency. If the agency does not comply, I provide its contact information to all of the interested students, and students can be very resourceful and determined. Be persistent but patient: e-mail and call, but give agencies time to respond and keep in mind that their primary goal is to service and manage their town or county GIS, not disseminate their data to the world.

The final deliverable product should include documentation. The documentation or metadata is critical, as metadata provide valuable information about the data. Government agencies or private vendors usually do not provide Federal Geographic Data Committee–compliant (FGDC; http://www.fgdc.gov) metadata (van Loenen & Onsrud, 2004). However, many data providers do not provide any metadata at all! At the very least, try to get information concerning the author/creator of the data, date for

which the data are relevant, basic explanation of the attributes, source of the data, scale, projection, coordinate system/datum, and units of measure. Metadata are crucial for the use, management, and dissemination of spatial data. Without this information, the data are of little value.

Evaluation of the products before selecting them for acquisition is essential. Read the documentation carefully. What scale is the data? What sort of attributes come with the data? What is the source? What is the licensing? Download samples, open them up, and take a look at them using GIS software. Read reviews if they exist, talk to other GIS or library professionals, and post questions to listservs about the data. One of the better listservs is GIS4Lib, administered by the University of Washington (http://mailman1.u.washington.edu/mailman/listinfo/gis4lib). Be watchful of vendors who "improve" and repackage Digital Chart of the World data (VMAP0), a spatial database of the world at 1:1,000,000 developed by the National Geospatial-Intelligence Agency (NGA). Quite often the "improvements" are minimal, and they are just selling you something you can get for free.

You might find it useful to draft a formal document outlining the GIS datasets selected for potential acquisition. In this document include the dataset or product, description, cost, license, and contact information. Review the items with the appropriate staff and select products for acquisition based on user needs, budget, licensing, quality, online availability, deliverable format, and staffing resources. Although a large amount of publicly available data can be found online, institutions should own a few inexpensive core datasets that meet most users' needs (see Table 2). Locating and evaluating the datasets and negotiating price and licensing require a significant amount of time. I generally prepare a small collection development proposal for more immediate demands in the fall and a more in-depth lengthy proposal in the spring.

Finding Data

Undoubtedly the most critical part of developing a GIS collection is locating data. Navigating the labyrinth of GIS data sources is not easy. Currently there are many sources for GIS data: U.S. and foreign governments, state and local governments, academic institutions, commercial data providers, utility companies, and others (Committee on Licensing Geographical Data and Services, 2004, chap. 3). In order to find data, one must utilize electronic and print resource guides, online data dissemination engines or portals, relationships with GIS professionals, commercial data providers, and printed map resources.

Attempting to locate free or low-cost, publicly available data is essential, since most academic research in the United States relies on this form of geographic information (van Loenen & Onsrud, 2004). Over the past few years, GIS professionals and librarians have developed guides to locating geospatial data. Many academic institutions have created virtual collec-

Table 2. GIS Datasets Recommended for Acquisition

Product	Description
ESRI Data & Maps	Contains a variety of data for the world, Canada, and Mexico, as well as general and detailed data of the United States. The product should meet many users' needs. It ships with ArcGIS software. Contact your software license administrator or ESRI (http://www.esri.com) for a copy.
Global GIS—Global Coverage DVD	Contains a wealth of USGS and other public domain data, including global coverages of elevation, landcover, seismicity, and resources of minerals and energy at a nominal scale of 1:1 million. It is available at http://www.agiweb.org/pubs/globalgis/.
TIGER/Line	Extracts of selected geographic and cartographic information from the Census Bureau's TIGER database. Free for federal depository libraries. Available from the U.S. Census Bureau (http://www.census.gov/).
Landview	A desktop mapping system for Environmental Protection Agency (EPA), Census Bureau, and USGS data. Free for federal depository libraries. Available from the U.S. Census Bureau (http://www.census.gov/).
World Vector Shoreline Plus	Originally developed by the National Geospatial-Intelligence Agency (NGA), it contains worldwide coverage of political boundary lines and shorelines at a scale of 1:250,000. It is available from the USGS (http://www.usgs.gov/) as well as several commercial vendors in a variety of formats.
National Transportation Atlas Database	A set of national geographic databases of transportation facilities for the United States. Available free from the Bureau of Transportation Statistics (https://www.bts.gov/pdc/index.xml).
Statewide Data	Statewide GIS data. Contact local state agency.
Local County or Citywide Data	Large-scale GIS data. Contact local municipal or county agency.

tions of Web links to frequently used sources of geographic information. These online resource catalogs are great places to start the hunt. Three good examples are the University of Arkansas Libraries (http://libinfo.uark.edu/GIS/us.asp); Stanford University Libraries (http://www-sul.stanford.edu/depts/gis/web.html); and the Harvard College Libraries (http://hcl.harvard.edu/research/guides/cartography/resources/online.html).

Printed guides to geospatial data are another good resource. Decker's (2001) *GIS Data Sources* and Ralston's (2004) *GIS and Public Data* are helpful for understanding, finding, and using geospatial data. Decker's work provides a basic introduction to GIS data and collection development as well as useful appendices to state and federal sources of geographic information. Decker's book is a must for spatial data librarians. Ralston's guide

Table 3. Frequently Used Geospatial Clearinghouses and Data Portals

Name	Web Site
Geospatial One-Stop	http://gos2.geodata.gov/wps/portal/gos
National Spatial Data Clearinghouse	http://clearinghouse1.fgdc.gov/
GIS Data Depot	http://data.geocomm.com/
Geography Network	http://www.geographynetwork.com/data/
USGS EROS Data Center	http://edcwww.cr.usgs.gov/
The National Map	http://nationalmap.gov/
NGA Geospatial Engine	http://geoengine.nga.mil/
The Harvard Geospatial Library	http://hgl.harvard.edu
Alexandria Digital Library	http://alexandria.sdc.ucsb.edu
Global Land Cover Facility	http://glcf.umiacs.umd.edu/index.shtml

provides information for some of the most useful publicly available data in the United States and includes information regarding formats, uses, and sources of unrestricted data.

Seekers of spatial data must also make use of geospatial clearinghouses or data portals (Tang & Selwood, 2005). Geospatial clearinghouses are Internet sites devoted to disseminating spatial data online following the FGDC guidelines for organization and metadata (Decker, 2001, p. 57). Some useful data portal sites are provided in Table 3. Although the sites take a while to learn and their stability is sometimes unreliable, they offer a wealth of free data and meet much of one's data needs.

For large-scale data, such as building- or property-level information, a selector must often establish relationships with the local county or municipality itself (Cobb, 1995). GIS resides in a variety of departments within local agencies such as planning, engineering, information technology, GIS, and the assessor's office. A good place to start is the county or municipal Web site to get contact information. Some counties and municipalities disseminate their data online, but many will require an e-mail or phone call to access the data. Getting to know your local GIS professionals will greatly help you in acquiring localized datasets.

Developing GIS relationships is essential for finding GIS data in general (Cobb, 1995). Join regional professional associations and attend regional conferences and workshops. For instance, the Northeast Arc Users Group (NEARC) is a regional organization for GIS and mapping professionals in the northeast. It provides an opportunity to meet local GIS professionals and learn about regional GIS activities. New York City and Boston both have formal and informal GIS user groups that get together and discuss GIS projects, jobs, and new data acquisitions, among other topics. Building relationships and establishing a network of GIS colleagues extends nationally as well. Join national organizations and attend their conferences, such as the North American Cartographic Information Society (http://www.nacis.org) or the Environmental Systems Research Institute (ESRI) User's Confer-

ence (http://www.esri.com/events/uc/). The key is to find an organization that meets your needs. Most of these local and national organizations also maintain listservs. Listservs provide a great opportunity to post questions about data sources. GIS4Lib, mentioned previously, is particularly useful for locating GIS data. *Directions Magazine* provides an additional list of GIS listservs (http://lists.directionsmag.com/discussion/). Mapping professionals frequently e-mail or phone each other while hunting for data.

If you cannot locate data that are publicly available, look to commercial data providers (see Table 1). There is an increasing number of commercial vendors that lease and sell GIS data. Commercial vendors offer a variety of raster and vector data such as topographic and administrative boundary datasets, digital elevation models, digital orthophotography, satellite imagery, and digital georeferenced maps for a wide array of GIS applications.

When GIS data are not available or accessible for a given area, researchers must often derive the GIS data from map sources. This is most common when users need historical data or international data at a scale greater than 1:250,000. Maps must go through a process of conversion before they can be used within a GIS (Hohl, 1998; Lee & Pun, 2001). They must be converted to a useable GIS form either by using digital imaging technology or a digitizing tablet. Maps are digitally imaged either by overhead photography or by a large-format scanner. The digital map image is then georeferenced, which converts the digital image from a nonreal-world coordinate system (image space) to a real-world coordinate system (Verbyla & Chang, 1997). Next, the user digitizes or traces (vectorizes) the necessary features on-screen (heads-up digitizing) and encodes the features with the appropriate information. Using a digitizing tablet, the user tacks the map or a copy of the map on the digitizing tablet, registers the map to the tablet, and then traces or extracts features using the digitizer puck. Depending on the number of features, digitizing can be a severely labor intensive process. This process of data conversion or data development is widely used in the mapping industry. If a library does not have access to the necessary equipment for digitization, many maps can be ordered in digital form from government agencies and commercial data providers. It is important to note that just as GIS data can supplement a printed map collection, printed maps can supplement a GIS collection. Therefore, the acquisition or inclusion of maps in print and digital form can be a valuable part of a GIS collection development strategy.

In short, be prepared to spend time looking for data. However, the more one does it, the easier it gets. The Web sites become easier to find and navigate, and the data portals become easier to use. After a while, public agencies and commercial dealers occasionally contact the GIS librarian when they have new GIS data available for distribution.

A GIS collection is not built over a month or a year but matures over time. The collection development process requires a fundamental under-

standing of the characteristics and uses of spatial data. It also requires the use of online resource guides and spatial search engines as well as the development of relationships with the user community, data providers, GIS professionals, and other librarians. In the formation of a GIS collection development strategy, library professionals should incorporate the established collection development policy, needs of the GIS user community, campus GIS services, and library infrastructure. When making decisions regarding GIS data acquisition, the library professional must also consider cost, availability, license agreements and distribution policies, documentation, data structures, and software and hardware. Utilizing each of these resources and incorporating each of these issues should help library personnel build a valuable GIS collection.

References

All China Marketing Research Co. & China Data Center, University of Michigan. (2004). 2000 China county population census data with county maps (version 3) [Data file]. Beijing, People's Republic of China: All China Marketing Research Co., & China Data Center.

Australian Bureau of Statistics. (2003). *CDATA 2001: Full GIS* (2nd release) [Computer software & data file]. Canberra, Australia: Australia Bureau of Statistics.

Chrisman, N. (2002). *Exploring Geographic Information Systems* (2nd ed.). New York: John Wiley & Sons.

Clarke, K. C. (2003). *Getting started with Geographic Information Systems* (4th ed.). Upper Saddle River, NJ: Pearson Education.

Cobb, D. A. (1995). Developing GIS relationships. *Journal of Academic Librarianship, 21*(4), 275–277.

Committee on Licensing Geographical Data and Services. (2004). *Licensing geographic data and services.* Washington, DC: National Academies Press.

Decker, D. (2001). *GIS data sources.* New York: John Wiley & Sons.

DeMers, M. N. (1997). *Fundamentals of Geographic Information Systems.* New York: John Wiley & Sons.

Florance, P. (2004). Review of *CensusCD 1990 Long Form in 2000 Boundaries. Library Journal 129*(3), 172.

Goodchild, M. F., & Longley, P. A. (1999). The future of GIS and spatial analysis. In Paul A. Longley, Michael F. Goodchild, David J. Maguire, & David W. Rhind (Eds.), *Geographical Information Systems vol. 1* (2nd ed., pp. 567–580). New York: John Wiley & Sons.

Hohl, P. (Ed.). (1998). *GIS data conversion.* Santa Fe, NM: Onword Press.

Jablonski, J. (2004). Information literacy for GIS curricula: An instructional model for faculty. *Journal of Map & Geography Libraries, 1*(1), 41–58.

Lamont, M. (1997). Managing geospatial data and services. *Journal of Academic Librarianship, 23*(6), 469–473.

Larsgaard, M. L. (1998). *Map librarianship: An introduction* (3rd ed.). Englewood, CO: Libraries Unlimited.

Lee, Y. C. & Pun, L. (2001). Geographical data from analogue maps. In Yong-Qi Chen and Yuk-Cheung Lee (Eds.), *Geographical data acquisition* (pp. 65–84). New York: Springer.

Longstreth, K. (1995). GIS collection development, staffing, and training. *Journal of Academic Librarianship, 21*(4), 267–274.

Martindale, J. (2004). Geographic Information Systems librarianship: Suggestions for entry-level academic professionals. *Journal of Academic Librarianship, 30*(1), 67–72.

ML Infomap. (2003). *Standard digital map of India* [Data file]. New Delhi, India: ML Infomap.

Ralston, B. (2004). *GIS and public data.* Clifton Park, NY: Delmar Learning.

Stone, J. (1999). Stocking your GIS data library. *Issues in Science & Technology Librarianship, 21*(Winter). Retrieved October 20, 2005, from http://www.istl.org/99-winter/article1.html.

Tang, W., & Selwood, J. (2005). *Spatial portals: Gateways to geographic information*. Redlands, CA: ESRI Press.
Tomlin, D. C. (1990). *Geographic Information Systems and cartographic modeling*. Englewood Cliffs, NJ: Prentice Hall.
van Loenen, B., & Onsrud, H. J. (2004). Geographic data for academic research: Assessing access policies. *Cartography & Geographic Information Science, 31*(1), 3–17.
Verbyla, D. L., & Chang, K. T. (1997). *Processing digital images in GIS*. Santa Fe, NM: Onword Press.

Patrick Florance is the Senior GIS Specialist at Tufts University. Formerly Patrick was the Digital Cartography Specialist at the Harvard Map Collection, Harvard University. Patrick has worked in a variety of public, private, and academic settings.

Legal Considerations in the Dissemination of Licensed Digital Spatial Data

Patti Day and Chieko Maene

Abstract

State and local governments increasingly license digital spatial data, the dissemination of which by academic libraries requires specific legal and operational considerations to reconcile license conditions with public access. We examined this in the context of the American Geographical Society Library (AGSL) at the University of Wisconsin–Milwaukee during 2000-05. Wisconsin open records law protects the right of access to public records, and geographic data is intended to be in the public domain. Despite this, Wisconsin counties have dramatically increased their use of licenses for geographic data, and the use of these licenses has never been challenged under Wisconsin open records law. The AGSL negotiates existing licenses, conveying to users the licensing conditions and reassuring the data producers. We developed user sublicenses including copyright statements, original licensor's names, and signed user agreements to the terms of the original licenses. Each user agreed that failure to comply with these terms would result in disciplinary action. For security reasons, all licensed data were delivered on CD-ROMs, which incorporated the licensing information, forced users to sign the sublicense, and insured discussion of the licensing issues. To insure consistency, we developed policies and procedures to be followed for each type of data request. We also provided to faculty members and students instruction sessions dealing with data availability and acquisition.

Introduction

Digital geographic information is among the most rapidly growing components of many academic libraries (Kinikin & Hench, 2005). One

LIBRARY TRENDS, Vol. 55, No. 2, Fall 2006 ("Geographic Information Systems and Libraries," edited by Jaime Stoltenberg and Abraham Parrish), pp. 236–253

sector of this information—licensed digital spatial data—presents specific opportunities and problems for librarians. Although the situation in each individual library will reflect its size, the nature of its collections, and its mission, certain issues are universal, including legal considerations. Freedom of information laws at the federal level and open records laws at the state level influence access to digital spatial data. Here, we examine these issues in the context of the American Geographical Society Library (AGSL) at the University of Wisconsin–Milwaukee (UWM) during the period 2000-05, when we held positions there as digital spatial data librarians.

The AGS Library

The AGSL is a unit within the UWM Libraries and is one of the largest geographical collections in the world. The library contains over one million items, including maps, charts, atlases, globes, photographs, monographs, periodicals, and digital spatial data (AGSL, 2006). The last of these, which includes both electronic statistical and geographic data, is of growing importance to and is increasingly used by UWM faculty and students. The number of electronic files distributed by the AGSL increased by 3,601 percent from 2000-01 (1,026) to 2004-05 (37,974). Over the same period, the number of data CD-ROMs and DVDs burned increased 364 percent, from 119 to 552 (AGSL, 2005).

UWM has a conspicuous Geographic Information Systems (GIS) community that supports the library in its role as a campus data center. A campus-wide interdisciplinary GIS council was formed in 1990 in the early stages of GIS development at UWM, and anyone interested in GIS is encouraged to participate. The council includes representatives of the UWM information technology (IT) department, librarians, research scientists, and faculty and students from academic units including architecture, urban planning, geography, civil engineering, anthropology, urban studies, business, and economics. The AGSL is the main campus unit that actively collects and archives digital spatial data files on behalf of the UWM GIS community.

Serving a wide range of disciplines and user groups requires that the AGSL collect a wide range of digital spatial data. These data are inherently diverse in terms of origin, format, and geographic and temporal coverage. Data formats are raster (for example, digital orthophotography, satellite imagery, and Digital Elevation Models [DEMs]), vector (such as Computer Aided Design [CAD] drawing files or GIS vectors) and tabular (statistics and attribute files). Geographic coverage ranges from global to county or city level, even as localized as a quarter-quarter-section. Time ranges may be continuous (for example, the Milwaukee real property master files since 1975), irregular series, or one-time snapshots. Data producers include federal, state, and local governments, educational institutions, nonprofit organizations, and commercial enterprises.

Data Types

The AGSL's digital spatial data collection is dominated by three categories of information: research data, public registry and administrative data, and commercial data. These are the data types most frequently requested by patrons.

The research data is primarily information collected by federal government agencies for their own purposes. In general, primary users of research data are government agencies that use the data in policy making and administration. Secondary users include academics, the general public, and commercial interests, which may repackage and market the data (Eechoud, 2004). Research data is attractive to GIS users because it is widely available, in the public domain, and useful in a broad range of applications (Eechoud, 2004). U.S. Census Bureau and U.S. Geological Survey datasets typify this category and serve general or scientific research purposes. They contain no private information, and the data is presented at a relatively small scale.

The AGSL also holds state and local government research datasets and directs patrons to relevant Internet sources. For example, current Wisconsin Department of Natural Resources (WDNR) digital spatial data are available at the WDNR Web site (www.dnr.state.wi.us/maps/gis/geolibrary.html) and archived data files are available in the AGSL.

Public registry and administrative data constitutes the foremost category of data requested in the AGSL. Public registry and administrative data is that collected by governments for specific legal and regulatory purposes, such as monitoring or regulating public and private activities like collecting taxes or regulating discharge of hazardous substances (Eechoud, 2004). This information includes land registry or cadastral data, law enforcement data, zoning permissions, and derivative land information such as street center lines with complete addresses.

The advantages of public registry and administrative datasets are numerous. They are geographically accurate (that is, data are created at large scale with high precision) and are updated frequently. Moreover, longitudinal (time-series) data may be archived for the entire area of interest. However, public registry and administrative datasets are not always accessible, or access may be regulated by legislation such as privacy laws.

Commercial data are acquired for reference use, for specific purposes or projects, or as a last resort when other data are unavailable. Commercial products may be relatively expensive but may be convenient, accurate, and scale appropriate. Commercial data acquired by the AGSL include high-resolution aerial photographs, satellite imagery, CensusCD+Maps, business location data (as a part of the Environmental Research Systems Institute [ESRI] Business Analyst) and commercial street data. These were acquired mainly for faculty research when the information was not available from other sources.

Legal Issues

Legal issues affect access to geographic data in general (Onsrud, 1995a, 1995b, 2000). Issues of particular relevance to digital spatial data include public access, intellectual property rights, professional ethics, and licensing. The essential issue is that of reconciling data producers' restrictions with public access.

Public Access

The raison d'etre of public access to government information is to allow public evaluation of public officials' conduct, to make available information about public policy, to protect against secret laws and decisions, and to encourage informed participation in public affairs (Solove, 2004; Cate, Fields, & McBain, 1994; Friedley & Colbert, 1991; Braverman & Heppler, 1981). Prior to 1966 there were no federal laws concerning access to government information, but the prevailing opinion was that the U.S. Constitution implied such rights (Henrick, 1977; Board of Education, Island Trees Union Free School District No. 26 v. Pico 457 U.S. 853, 1982). During the Watergate crisis of 1974, Congress rewrote the federal "Government in the Sunshine" laws strengthening the right of access to government information (Solove, 2004; Henrick, 1977). Freedom of Information (FOI) laws had been enacted by 1983 in all fifty states and the District of Columbia (Solove, 2004).

Concerning geographic data in particular, the National Research Council states that "Government accountability and transparency require agencies to ensure that the ability to control scarce geographic data never becomes 'outcome determinative' for any political or judicial process . . . Transparency is important to agency adjudications and rulemaking, to petitions to Congress for new legislation, and to mount court challenges to illegal government acts" (2004, p. 161).

Copyright, Geographic Information, and Compositions

Copyright holders obtain exclusive rights to copy, display, distribute, adapt, and perform a protected work (Minow & Lipinski, 2003). These rights are extended as soon as an original idea, which shows a minimal level of creativity, becomes fixed in a tangible medium (Minow & Lipinski, 2003). With very few exceptions, federally produced government information is not placed under copyright protection (Dansby, 1994; Cho, 1998). Some states allow copyright of public information, but others do not (Fishman, 2004). In terms of geographic data or databases, it is important to remember that copyright protects originality, not hard work ("sweat of the brow").

Traditionally cartographers and producers of geographic data have relied upon copyright to protect the intellectual property of their works. The Supreme Court ruled in *Feist Publications Inc. v. Rural Telephone Service Co.* (499 U.S. 340, 1991) that facts per se are not copyrightable, but a slight

amount of originality, including the selection and arrangement of facts, is protected (Dando, 1991, 1993b). On these grounds, many cartographers and producers of geographic data believe that geographic data arranged within a database has copyright protection, even if the facts themselves do not. It is unclear, however, exactly what degree of originality in geographic databases is required to warrant protection. "Maps and photographic images, for example, often have been found to be copyrightable" (National Research Council, 2004, pp. 106–107). Others may extract, copy and use the factual information contained in the work as long as the creative expression is not copied These works, like factual databases, are said to have "thin" copyright (Karjala, 1995).

Section 107 of the Copyright Revision Act (1976) contains the statutory expression of "fair use" rights to use copyrighted materials. Under certain conditions, use is allowed for purposes such as criticism, comment, news reporting, teaching, scholarship, and research (Minow & Lipinski, 2003). Four factors are considered in determining if a use is "fair": (1) the purpose and character of the use (whether commercial, nonprofit, or educational), (2) the nature of the work (factual or otherwise), (3) the amount and substantiality of the portion used in relation to the whole, and (4) the effect of the use upon the potential market for or value of the copyrighted work (Minow & Lipinski, 2003). These provisions are particularly relevant because many nonfederal public sector geographic data producers are concerned about liability, proper attribution, control of third-party redistribution, and inappropriate derivative reproduction of "their" data. The general consensus is that copyright protection is not sufficient. "Fair use and the misuse doctrine represent significant limits on the copyright owner's rights. The scope of their application is sufficiently uncertain, however, that, where possible, parties should contract [license] for anticipated uses rather than rely on fair use doctrine or other uncertain legal doctrines to sanction the licensee's activities" (National Research Council, 2004, p. 110).

It is obvious that some uses of geographic data constitute fair use, for example, using a factual geographic data database for teaching purposes. Here the data producer would most likely be concerned about redistribution of the data beyond the confines of the educational institution.

Geographic Data as Public Domain Information

Federal Office of Management and Budget (OMB) Circular A-16 (1994) deals more specifically with geographic data as public domain information and includes provisions for "improvements in coordination and use of spatial data" (OMB Circular, 1994). The OMB circular incorporates Executive Order 12906 (Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure, NSDI), which require agencies to "adopt a plan . . . establishing procedures to make geospatial data available to the public, to the extent permitted by law, current policies, and relevant

OMB circulars" (National Research Council, 2004, p. 125). Similar to many federal laws, A-16 strongly advocates public availability and dissemination of geographic data acquired by the government (National Research Council, 2004). The NSDI is a vision for a nationally shared catalog of geographic data from all levels of government. Federal agency participation is mandated, and programs have been instituted to encourage participation by state and local agencies. These programs include Geospatial OneStop (Phillips, 2005), the National Map, and earlier less successful ventures. Some state and local governments are reluctant to cooperate in these efforts for several reasons, including the federal requirement that their licensed geographic data be placed in the public domain.

State government approaches to geographic data distribution vary widely on the basis of different justifications (Cho, 2005). "Some provide access rights on the basis of an exception to open records law, others depend on the nature of the request that is made" (Cho, 2005, p. 73). Some agencies distinguish between "services" and "sales" (Wells & Tsui, 2005). Some make no distinction between geographic data and other types of digital databases (Cho, 2005), while others have enacted specific legislation concerning distribution of geographic data (National Research Council, 2004). "Federal law permits state and local governments to assert copyright in works containing geographic data (if they otherwise meet the requirements for copyright protection). When consistent with local law, state and local governments may also maintain geographic data as secret, or restrict their use and redistribution" (National Research Council, 2004, p. 134). As a result, each state or local government agency may create policies that either impose prohibitive use conditions or provide open access to geographic data. Prohibitive conditions are place specific and localized; the underlying assumption, based on democratic principles as demonstrated in federal law and policies, is that the public has the "right to know."

Licensing of Geographic Information

A license is a legal contract between two parties by means of which the licensor allows the licensee to use a data collection (Cho, 2005; Wells & Tsui, 2005). Licenses are typically governed by state contract laws. "Contract law is about relationship building rather than simply attempting to either drive a hard bargain or to get out of a dispute" (Cho, 2005, p. 292). The licensee accepts certain restrictions on the use of the data, such as agreeing that no copying or further dissemination will occur. Parties can usually negotiate terms to come to a mutually agreeable arrangement. Until the mid-1990s it was uncommon for government agencies to license geographic data, but, since then, nonfederal public agencies have become more inclined to do so in order to limit the use of their data, limit their liability, or raise revenue (National Research Council, 2004; Wells & Tsui, 2005). Typically, licenses contain a statement of ownership and copyright, a product description

and statement of quality, warranties, disclaimers and indemnification, any restrictions on use or resale, specification of the length of the agreement and terms of renewal, cancellation terms, fees or in-kind exchange, and responsibilities for updates and error notification (Wells & Tsui, 2005).

State and local governments, operating under different laws and policies, provide many reasons for electing to license geographic data. These typically include cost recovery, liability concerns, as a vehicle of proper attribution, and to prevent third-party redistribution and derivative production (Dando, 1992, 1993a; Dansby, 1992; Holland, 1997; Onsrud, 1999; National Research Council, 2004; GITA, 2005). The specific goal of cost recovery has rarely been realized (Humphrey, 1995; Sears, 2001; Joffe, 2003; National Research Council, 2004). In 2003 the U.S. Geological Survey funded the Open Data Consortium (ODC) to develop a model data distribution policy for local governments in the United States. According to Joffe (2003, 2005), most local agencies that sell or license public data operate at a loss, with very few earning even modest revenues.

Open Records in Wisconsin

The Wisconsin legal system provides general guidance to both data producers and users within the state. In Wisconsin, as in most states, open records law protects the right of access to public records. Wisconsin Statutes 19.31 thru 39, subchapter II, Public Records and Property (State of Wisconsin, 2004), begins with a declaration of policy (19.31) that presents the overriding principles governing the subsequent laws that deal with public records within the state. It reads in part:

> In recognition of the fact that a representative government is dependent upon an informed electorate, it is declared to be the public policy of this state that all persons are entitled to the greatest possible information regarding the affairs of government and the official acts of those officers and employees who represent them. Further, providing persons with such information is declared to be an essential function of a representative government and an integral part of the routine duties of officers and employees whose responsibility it is to provide such information. To that end, ss. 19.32 to 19.37 shall be construed in every instance with a presumption of complete public access, consistent with the conduct of governmental business. The denial of public access generally is contrary to the public interest, and only in an exceptional case may access be denied.

In every instance, complete public access to governmental business records, except in exceptional circumstances, is the policy of the state. This policy reflects federal FOI laws and policies.

Section 19.32 provides definitions of "authority," "local governmental unit," and "record." A record is "any material on which written, drawn, printed, spoken, visual or electromagnetic information is recorded or preserved,

regardless of physical form or characteristics, which has been created or is being kept by an authority. 'Record' includes, but is not limited to, handwritten, typed or printed pages, maps, charts, . . . tapes (including computer tapes), computer printouts and optical disks" (State of Wisconsin, 2004).

Land Information in Wisconsin

Historically, Wisconsin has been at the forefront of efforts to modernize land records in the United States (Koch et al., 2001). Public agencies, cities, universities, and private-sector groups have worked individually and collectively to institute a progressive system that was formalized in 1989 through the creation of the Wisconsin Land Information Board (WLIB). Wisconsin Acts 31 and 339 (1989) assigned the board responsibility for implementing the Wisconsin Land Information Program (WLIP) (Holland, 1994). The intent of the WLIP was to develop a "decentralized confederation of systems where those with existing land records responsibilities would continue to collect, maintain and keep custody of land information. . . . Through integration, this confederation of systems will be tied by formal and/or informal data sharing agreements" (WLIB, 1994, p. 1). To emphasize the objective of providing open access to geographic data, language was added to Act 339 specifically empowering the WLIB to utilize program revenue for "Systems Integration" (WLIB, 1994, p. 2). Definition of this term was requested by the legislature, and Systems Integration was subsequently defined as "the coordination of land records modernization at all levels of government to ensure that the information can be shared, distributed and used by all participants, including state and local government, the private sector and taxpayers" (Sec. 20.505(4) Wis. Act 339, 1989). According to the WLIB, the interpretation is meant to be literal and contextual in light of legislative and gubernatorial intent, and the objective of developing systems with shared data is "clear and unambiguous" (1994, p. 3). The policy supports the assumption that geographic data is intended to be in the public domain by statutory authority.

Despite this clear intent, between 1999 and 2002 Wisconsin counties increased the use of licenses for geographic data by over 100 percent and increased the use of copyright by 108 percent (Day, 2004). Use of these licenses for geographic data has never been challenged under Wisconsin open records law, so there are no legal judgments upon which to decide whether or not these licenses are legally binding. This leaves interpretation of the law open and leaves local authorities free to impose licenses, notwithstanding the "clear and unambiguous" nature of Wisconsin Act 339 (WLIB, 1994; Shanley, 2005). Despite the ongoing ethical and legal debate, the AGSL and other Wisconsin libraries dealing with local government geographic data have only two choices: to negotiate the existing licenses or forgo access to the data.

Accessibility Issues

Accessing research data at an academic library such as the AGSL is relatively straightforward. Much federal, state, and some local government geographic information is available for download or purchase without cost or at the cost of reproduction. Libraries participating in the federal depository library program automatically receive selected geographic data produced by the federal government. By contrast, purchasing commercial data may or may not be relatively straightforward, depending on the company and their experience with libraries. Having generated the necessary funds to purchase the identified data, negotiating the license agreement terms can be challenging. A librarian who is experienced in such matters may be able to handle the negotiation independently, but consultation with university legal counsel may be necessary in certain instances.

In terms of accessibility, the most difficult data to obtain is locally produced public registry and administrative data. Data producers may hesitate to distribute the information because it may contain personal or private information about citizens. They may also fear loss of control over the information (that is, property ownership) once the data is removed from their supervision. Additionally, the organization may want (or need) to recover the cost of data production and maintenance by charging not only for reproduction but also for data creation. Agencies also may not want to supply data because the effort to extract and package data is time -consuming and is not their primary function (Cho, 2005; Wells & Tsui, 2005).

Another consideration is the value to the organization of the information itself. Locally produced large-scale geographic data is often regarded as a commodity and is considered too valuable to disseminate at no cost or at the cost of reproduction only. Therefore, individual agencies may decide to implement geographic data distribution policies that contradict the federal open records law. Federal law permits state and local governments to employ copyright protection over their geographic data if certain requirements are met.[1] Also, state and local governments are allowed to restrict access, usage, and redistribution of geographic data when it is consistent with local law (National Research Council, 2004). The result is that there is no uniform policy governing access to geographic data produced by state and local governments. "Public policy that promotes the use of and access to automated geographic information differs widely among the states from the use of open records laws through to the public records and FOI laws. There seems no model that adequately addresses the power and commercial utility of GIS databases" (Cho, 1998, p. 141). There has been discussion of how local governments should provide access to geographic data since the early 1990s, but consensus among local data producers has not been reached yet.[2]

Other issues concerning access to geographic data include privacy and confidentiality (Cho, 1998, 2005; Dillehay, 1993), liability (Cho, 1998, 2005;

Onsrud, 1999), and national security (Baker et al., 2004; Tombs, 2005). Data producers may not have clear guidelines about what, how, and to what extent personal information can be distributed in their geographic data. For example, some land information contains personal details, such as landowners' names and contact information. Uncertainty about how open records laws apply to land information published on the Internet may delay decisions about data distribution to the public (WLIA, 2003).

Liability in the use of geographic information has long been a subject of interest in the geographic information community (Onsrud, 1999). The use of warranties and disclaimers is becoming the norm among data producers seeking to minimize liability exposure, although this does not protect them entirely (National Research Council, 2004).

Invoking national security as a reason to restrict access to local geographic data is relatively new and may impact future policy developments (Zellmer, 2004). For example, a township in New Jersey blocked a resident's open records request by submitting utility geographic data that was later incorporated into the Department of Homeland Security's Critical Infrastructure Information program, hence preventing any public access (Lozar, 2005; Tombs, 2005).

Managing Licensed Data at the AGSL

The logistics of handling licensed data at the AGSL during our tenure were complex and time-consuming. Each license required that data be used only by UWM students and faculty (occasionally staff), and each placed different restrictions on use and reproduction of the data. Conveying this information to users was a major concern of the data producers, and maintaining their trust in this regard was imperative in order to obtain updates or new data (Harvey, 2003).

To address the concerns of the data producers, we developed a system of sublicensing each dataset. Submitting a copy of the sublicense to data producers usually convinced them that we were committed to complying with their original licenses and that we were taking appropriate steps to prevent misappropriation of the data beyond the university.

The system of sublicensing each dataset was developed with the assistance of UWM legal counsel. After completing negotiations for any licensed data, a user sublicense was created with the restrictions of the original license presented in nonlegal terminology (see Appendix A). Each sublicense included a copyright statement, the original licensor's name, the date, the user's name (printed) and signature, and a statement that the user agreed to the terms of the original license. Each sublicense also included an agreement that failure to comply with these terms would result in academic or nonacademic disciplinary action. Users also agreed in writing to return or destroy the data at the end of the semester in which it was requested. Each user was also informed of the restrictions verbally, and a blank copy of the

sublicense, additional to the copy completed by the user, was burned onto each CD-ROM distributed to protect the AGSL legally from any claim that a user was unaware of the restrictions.

To minimize paperwork and to track what geographical data had been distributed to each individual user, the sublicense form also served as the internal processing form. Users were made aware that the form they were signing was a legal document that the AGSL was required by law to keep for seven years (the statute of limitations in Wisconsin). Initially, we had argued that these forms should be considered circulation forms and therefore could be destroyed after the information was processed or at least at the end of each semester. This argument was overruled by UWM legal counsel, hence the requirement that the sublicense forms be retained for seven years.

Although public domain data were delivered via FTP, e-mail, CD-ROM, or the Internet, all licensed data were delivered on CD-ROM. In part this was for security reasons, in that the data could possibly be hacked off the Internet, even from a "secure" site. Using CD-ROMs also allowed us to include the licensing information, and it forced potential users to contact us in person, facilitating signing of the sublicense and insuring discussion of the licensing issues. We regarded discussion of the license as a teaching tool, conveying to users that data is licensed, it costs money, and there are consequences for its misuse. We charged a minimal processing fee ($2–3) to cover the cost of the CD and the staff time involved in repackaging each geographic area individually.

The majority of data requests were for Milwaukee County and surrounding counties in southeastern Wisconsin. The AGSL collected spatial data for as many Wisconsin counties as possible and other areas as requested. Upon receipt of the data, all available documentation (read me files, disclaimers, warrantees, metadata) was reviewed to determine the legal status of and restrictions upon each data file. Datasets were classified into four groups: public domain, copyrighted, licensed, and restricted to in-library use only. Data producers were contacted if the legal status of any data was unclear. Some datasets, such as ESRI Data and Maps and the USGS/AGI Global GIS database series, are complex, with different restrictions applying to individual data files.

Hard copy documentation relating to datasets was scanned and stored electronically with the digital files so that it was available for distribution to users. To streamline processing and minimize uncertainty, hardcopy binders were created with the original license and the sublicenses arranged alphabetically by county. The digital files were arranged on the network in the same order. Since the AGSL holds both nonlicensed and licensed data, color-coded stick-on dots were placed on the CD-ROM cases in the file cabinets to differentiate between them. Only the most frequently requested data were located on the server.

Student workers were repeatedly instructed always to check with the digital spatial data librarian if they were in any way uncertain about the conditions pertaining to data dissemination. An intern created flow charts (for an example, see Appendix B) for the most frequently requested datasets indicating the appropriate procedures to be followed for each request. These flow charts proved very valuable during student training and in day-to-day operations and were posted prominently for student workers to consult.

Instruction Sessions

One of the drawbacks in collecting and providing to university users a variety of digital spatial data is that such users become reliant upon the service and have little incentive to learn where and how to obtain the data themselves. Although users were encouraged to obtain licensed data via the AGSL, for pedagogic reasons and because of increasing demand they were urged to acquire public domain data themselves. Discussing data availability and acquisition in person proved useful in instructing users about procurement methods, but we could reach relatively few individuals in this way.

To address this issue, we encouraged faculty members to invite us to classes utilizing digital geographic data. This allowed us to discuss various issues related to geographic data information, including how to find and access information on the Web and in the library, the spectrum of public and private data producers, copyright and licensing restrictions, and appropriate acknowledgment and citation procedures. This instructional service was provided in both introductory and advanced courses in geography, urban planning, architecture, civil engineering, and other disciplines. The well-established GIS community on campus played a vital role in connecting us with faculty who use digital spatial data in their research and teaching.

Conclusion

The supply of and demand for licensed digital spatial data is increasing rapidly. Although users potentially can obtain such data themselves, academic libraries will increasingly be expected to obtain and disseminate these resources. Understanding the legal issues pertaining to such data is paramount, and the AGSL provides a model for reconciling data producers' restrictions with academic access.

The AGSL experience suggests that the greatest demand is for locally produced geographic data. In Wisconsin, such data is regulated by various and potentially conflicting statutes. Despite the "clear and unambiguous" assertion of Wisconsin Act 339 that geographic data is intended to be in the public domain (WLIB, 1994, p. 3), local government agencies remain at liberty to impose licenses on their geographic data because the use of such licenses has never been challenged under Wisconsin open records law. Between 1999 and 2002 Wisconsin counties increased the use of licenses for

geographic data by over 100 percent, and the AGSL and other Wisconsin libraries dealing with local government geographic data must either negotiate these existing licenses or forgo access to the data.

Managing licensed data is complex and time-consuming. Licenses restrict use and reproduction of the data, and maintaining the producers' trust in this regard is imperative. At the AGSL we accomplished this by developing a system of user sublicenses that were in compliance with the original licenses. For security reasons, all licensed data were delivered on CD-ROMs, which incorporated the licensing information, forced users to sign the sublicense, and insured discussion of the licensing issues. To insure consistency, we developed policies and procedures to be followed for each type of data request. We also provided to faculty members and students instruction sessions dealing with data availability and acquisition.

Appendix A: User Sublicense Agreement

UWM Libraries

Request for Electronic MCAMLIS Data Sets

Electronic data sets from the Milwaukee County Automated Mapping and Land Information System are only available to students, faculty, and staff of the University of Wisconsin-Milwaukee for educational purposes.

These electronic files, and any electronic files derived from them cannot be used for commercial application, sold**,** **redistributed**, or published **in any ELECTRONIC format** (including on the World Wide Web). **Paper** copies of the original maps created with the data or published in reports may be produced. The library reserves the right to verify the educational purpose of each request. **Violation of the foregoing provisions shall automatically terminate all access by you to the digital base maps and any material derived therefrom** because such acts jeopardize the continued availability of the data sets to the University of Wisconsin-Milwaukee.

Requestors must enter into the agreement attached below prior to being granted access to the MCAMLIS Data Sets.

Request Date (M/D/Y): / / 200	Time:	Pick-up:	Charge:
Name (Please Print) :			
Email address (UWM email address required):			
UWM ID Number:	Department/Major:		
Purpose of Request (name of class/nature of assignment): Course Number: ( ) Instructor Name: ( )			

Request (description of requested material/publisher/number of files) (**completed by staff**):

MCAMLIS preferred format

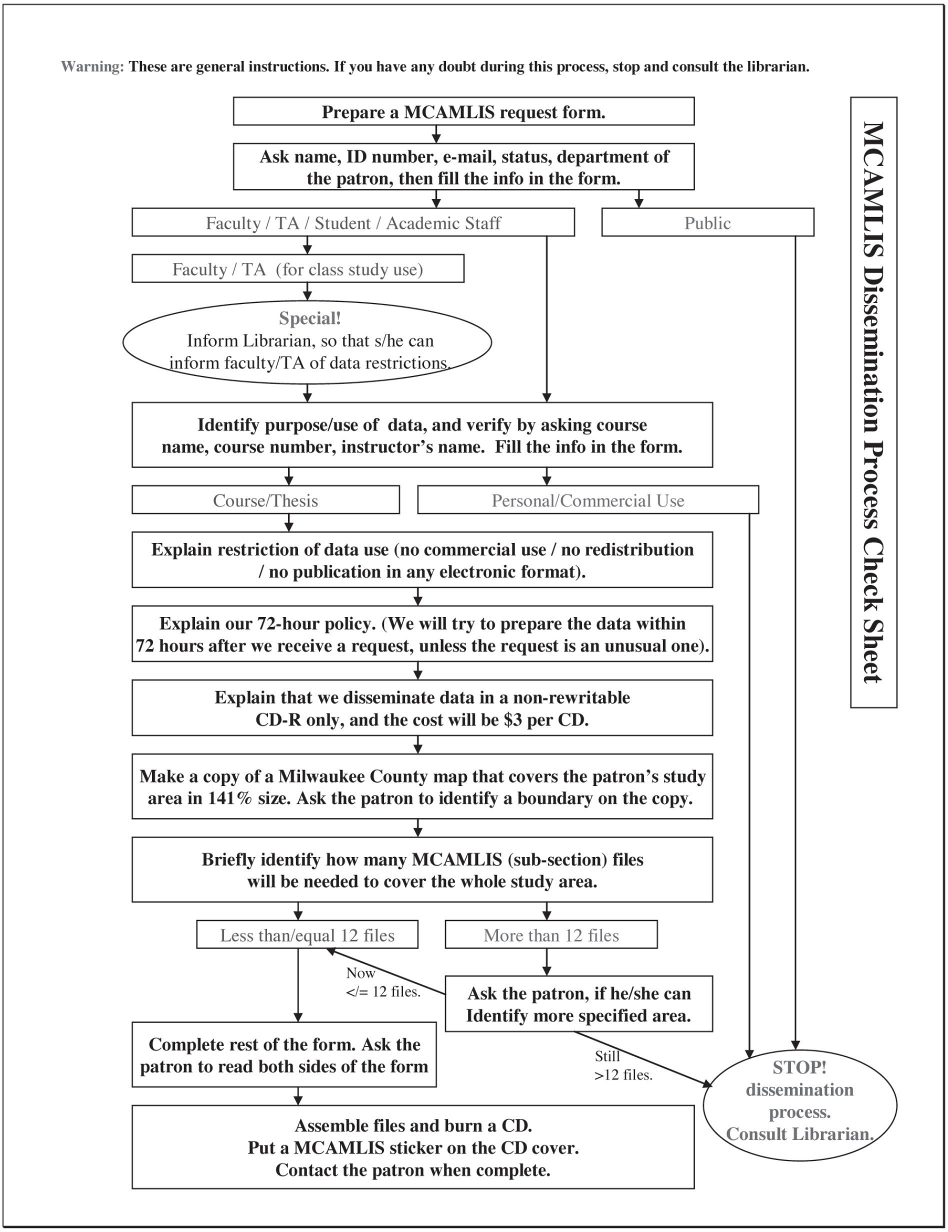
- ☐ AutoCAD DXF
- ☐ MicroStation DGN
- ☐ Other ( )

Aerial Photo preferred format

- ☐ 2000 Black & White
- ☐ 2002 Color --- AirPhotoUSA Request Form REQUIRED
  - ☐ JPEG
  - ☐ TIF
  - ☐ MrSID
  - ☐ Other ( )

Description of format distribution CD-R (copyrighted data) **or** CDRW, Zip, FTP, etc. (ONLY for public domain data). (**completed by staff**):

Appendix B: Student Flowchart



Notes

1. See more arguments on geographic data copyright protection and licensing issues in Cho (1998, 2005), Dando (1992), Holland (1997), and Petersen (1994).
2. Discussion about geographical data access issues and licensing can be found in legal periodicals and local government related periodicals, such as *GIS Law* and various URISA publications. Examples of current attempts to set up data access and distribution policy are available from organizations that deal with geospatial data issues, such as the National Research Council, Committee on Licensing Geographic Data and Services (2004), Open Data Consortium (Joffe, 2003), and Geospatial Information and Technology Association (GITA, 2005).

References

American Geographical Society Library (AGSL). (2005). *Twenty-seventh annual report. July 1, 2004-June 30, 2005.* Milwaukee: University of Wisconsin–Milwaukee Libraries.

American Geographical Society Library (AGSL). (2006). *American Geographical Society Library.* Retrieved February 12, 2006, from http://www.uwm.edu/Libraries/AGSL.

Baker, J. C., Lachman, B., Frelinger, D., O'Connell, K. M., Hou, A. C., Tseng, M. S., et al. (2004). *Mapping the risks: Assessing the homeland security implications of publicly available geospatial information.* Santa Monica, CA: Rand Corp.

Braverman, B. A., & Heppler W. (1981). A practical review of state open records laws. *George Washington Law Review, 49*(3), 720–760.

Cate, F. H., Fields, D. A., & McBain, J. K. (1994). The right to privacy and the public's right to know: The "central purpose" of the Freedom of Information Act. *Administrative Law Review, 46,* 41–47.

Cho, G. (1998). *Geographical Information Systems and the law: Mapping the legal frontiers.* New York: John Wiley & Sons.

Cho, G. (2005). *Geographic information science: Mastering the legal issues.* New York: John Wiley & Sons.

Dando, L. P. (1991). Open records law, GIS, and copyright protection: Life after FEIST. In *Urban and Regional Information Systems Association, Annual Conference Proceedings, 4*(1), 1–17.

Dando, L. P. (1992). A survey of state open records laws in relation to recovery of data base development costs: An ends in search of a means. In *Urban and Regional Information Systems Association, Annual Conference Proceedings, 4,* 85–111.

Dando, L. P. (1993a). A case for the commercialization of public information. In W. Bamberger & N. S. Bryan (Eds.), *Marketing government geographic information: Issues and guidelines* (pp. 105–108). Washington, DC: URISA.

Dando, L. P. (1993b). GIS and copyright protection. In W. Bamberger & N. S. Bryan (Eds.), *Marketing government geographic information: Issues and guidelines* (pp. 23–30). Washington DC: URISA.

Dansby, H. B. (1992). Public records and government liability Part II. *GIS Law, 1*(3), 10–14.

Dansby, H. B. (1994). Freedom of information at the federal level. *GIS Law, 2*(1), 6–11.

Day, P. A. (2004). *Copyright, license agreements and GI data access in Wisconsin.* Paper presented at the 2004 PPGIS annual conference, Madison, WI, July 19.

Dillehay, B. (1993). Access, privacy and electronic information—Time for new laws? *GIS Law, 1*(4), 1–7.

Eechoud, M. V. (2004). *The commercialization of public sector information delineating the issues.* Retrieved February 2, 2006, from http://www.ivir.nl/agenda/iter/PapersCommodification/EechoudCommodification.pdf.

Fishman, S. (2004). *The public domain: How to find & use copyright-free writings, music, art & more.* Berkeley, CA: Nolo.

Friedley, D., & Colbert, L. (1991). Reaching for consensus in public information access policies. In *Urban and Regional Information Systems Association, Annual Conference Proceedings, 4,* 50–63.

Geospatial Information & Technology Association (GITA). (2005). *Free of fee: The governmental data ownership debate.* Retrieved February 2, 2006, from http://www.gita.org/resources/whitepapers/Free_or_fee.pdf.

Harvey, F. (2003). Developing geographic information infrastructures for local government: The role of trust. *Canadian Geographer, 47*(1), 28–36.

Henrick, W. (1977). Public inspection of state and municipal executive documents: "Everybody, practically anything, anytime, except..." *Fordham Law Review, 45*(5), 1105–1153.

Holland, W. S. (1994). Historical background to the Wisconsin land information program. In *Modernizing Wisconsin's land records through decentralized and integrated land information systems.* Madison: Wisconsin Land Information Board.

Holland, W. S. (1997). *Copyright, licensing and cost recovery for geographic and land information system.* Madison: GeoAnalytics.

Humphrey, M. (1995). Pricing electronic data and data access. *Public Management, 77*(11), 16–18. Retrieved February 2, 2006, from http://www.geoanalytics.com/publications/copyrightAndLicensing.pdf.

Joffe, B. (2003). *Report from the Open Data Consortium.* Retrieved February 2, 2006, from http://www.opendataconsortium.org/documents/Data_Policy-4b.pdf.

Joffe, B. (2005). Ten ways to support GIS without selling data. *Urban & Regional Information Systems Association Journal, 16*(2), 27–33.

Karjala, D. (1995). Copyright in electronic maps. *Jurimetrics Journal, 35,* 395–415.

Kinikin, J., and Hench, K. (2005). Survey of GIS implementation and use within smaller academic libraries. *Issues in Science & Technology Librarianship,* Spring. Retrieved February 2, 2006, from http://www.istl.org/05-spring/refereed-1.html.

Koch, T. W., Hart, D., Moyer, D. D., & Niemann, B. J, Jr. (2001). *Land information modernization activity in Wisconsin: Impacts, status and future tasks, 1990–2000.* Retrieved February 2, 2006, from http://www.lic.wisc.edu/pubs/ModReport.pdf.

Lozar, R. (2005). A records roadmap: Journalists often struggle to gain access to GIS data. *The News Media & the Law, 29*(3), 25–27.

Minow, M., & Lipinski, T. (2003). *The library's legal answer book.* Chicago: American Library Association.

National Research Council, Committee on Licensing Geographic Data and Services. (2004). *Licensing geographic data and services.* Washington, DC: National Academies Press. Retrieved May 11, 2006, from http://www.nap.edu/books/0309092671/html.

OMB Circular A-16. (1994). Retrieved July 25, 2006, from http//www.whitehouse.gov/omb/circulars/a016/a016_rev.html.

Onsrud, H. J. (1995a). The role of law in impeding and facilitating the sharing of geographic information. In H. J. Onsrud & G. Rushton (Eds.), *Sharing geographic information* (pp. 292–306). New Brunswick: Center for Urban Policy Research

Onsrud, H. J. (1995b). The use of geographic information systems in local government. *Public Administration Review, 55*(5), 461–467.

Onsrud, H. J. (1999). Liability in the use of geographic information systems and geographic datasets. In P. A. Longley, M. Goodchild, D. Maguire, & D. Rhind (Eds.), *Geographical information systems: Management issues and applications* (2nd ed., Vol. 2, pp. 643–652). New York: Wiley.

Onsrud, H. J. (2000). Legal access to geographic information: Measuring losses or developing responses? In D. G. Janelle & D. C. Hodge (Eds.), *Information, place, and cyberspace: Issues in accessibility* (pp. 303–316). Berlin: Springer.

Petersen, B. A. (1994). Access to electronic records. *GIS Law, 2*(2), 1–6.

Phillips, A. (2005). A portal to reliable real estate data or a door to nowhere? A look at how state and local dissemination policies have impacted the development of the national spatial data infrastructure and geospatial one-stop portal. *Real Estate Law Journal, 34*(1), 9–42.

Sears, G. (2001). *Executive summary: Geospatial data policy study.* Ottawa: KPMG Consulting Inc. Retrieved February 2, 2006, from http://www.geoconnections.org/programsCommittees/proCom_policy/keyDocs/KPMG/KPMG_E.pdf.

Shanley, L. A. (2005). *Licensing of geographic data and services: The Wisconsin experience.* Retrieved February 2, 2006, from http://www.wlia.org/standards/Part2_Licensing.pdf.

Solove, D. J. (2004). *The digital person: Technology and privacy in the information age.* New York: New York University Press.

State of Wisconsin. (2004). *Unofficial text from 03-04 Wisconsin Statutes & Annotations.* Retrieved February 10, 2006, from http://folio.legis.state.wi.us/cgi-bin/om_isapi.dll?clientID=33068415&infobase=stats.nfo&j1=19.31&jump=19.31&softpage=Browse_Frame_Pg.

Tombs, R. B. (2005). Policy review: Blocking public geospatial data access is not only a homeland security risk. *Journal of the Urban & Regional Information Systems Association, 16*(2), 49–51.

Ventura, S. J. (1995). Overarchiving bodies for coordinating geographic data sharing at three levels of government. In H. J. Onsrud & G. Rushton (Eds.), *Sharing geographic information* (pp. 172–192). New Brunswick: Center for Urban Policy Research.

Wells, E., & Tsui, M. (2005). *Public data access, privacy, and security: U.S. law and policy.* (Urban and Regional Information Systems Association Conference Workshop, Kansas City, KS, October, 2005). Park Ridge, IL: Urban and Regional Information Systems Association.

Wisconsin Land Information Association (WLIA). (2003). *Internet privacy & open records: Final draft.* Madison: Wisconsin Land Information Association. Retrieved February 2, 2006, from http://www.wlia.org/standards/PrivacyOpenRecordsHandbook.htm.

Wisconsin Land Information Board (WLIB). (1994). *Modernizing Wisconsin's land records through decentralized and integrated land information systems.* Madison: Wisconsin Land Information Board.

Zellmer, L. (2004). How homeland security affects spatial information. *Computers in Libraries, 24*(4), 6–40.

Patti Day was responsible for the development of the digital spatial data clearinghouse at the American Geographical Society Library, where she served as the digital spatial data librarian for seven years. It was during her tenure there that she developed an interest in the legal, ethical, and policy aspects of access to geographic information. She is now a multidisciplinary doctoral student in geography and information studies at the University of Wisconsin–Milwaukee.

Chieko Maene is a map librarian in the Government Documents, Maps and Microforms Department at the Richard J. Daley Library of the University of Illinois at Chicago. She worked as an intern at the American Geographical Society (AGS) Library between 2002 and 2004 and then as an interim GIS librarian at the AGS Library between 2004 and 2005. She holds an M.S. in urban studies and M.L.I.S. in information studies, both from the University of Wisconsin–Milwaukee.

Building a System to Disseminate Digital Map and Geospatial Data Online

Tsering Wangyal Shawa

Abstract

The expectation of library patrons to get all of the information they need, including geographic information, accessible on their desktops has created challenges to map and Geographic Information System (GIS) libraries. This new expectation has forced libraries to think about how to design a system that will allow diverse geographical information to be available over the Internet. Some libraries have built a site to distribute localized data, others have developed a system to make only maps accessible online. Princeton University Library's Digital Map and Geospatial Information Center started a pilot map scanning project in early 2004 to build a system, to develop specifications for scanning maps and compressing TIFF images to JPEG2000 file format, and to establish workflows. The system was built using many off-the-shelf commercial software packages. This article discusses challenges of building a system and explains how Princeton developed a scanning process and standards, workflows, and what lessons were learned in building such a system.

Introduction

Libraries purchase and receive geospatial data and paper maps free of charge through the Federal Depository Library Program (FDLP). One of the requirements of the FDLP is to make all the materials distributed through it freely accessible to the public. Because of this requirement and demands from library users to make all the materials accessible on their desktops, many libraries scan their paper maps and make them accessible online. However, one major problem libraries face is how to design a system that will allow the user to search, view, and download diverse geospatial data

LIBRARY TRENDS, Vol. 55, No. 2, Fall 2006 ("Geographic Information Systems and Libraries," edited by Jaime Stoltenberg and Abraham Parrish), pp. 254–263.

and digital maps. This article examines the challenges of creating such a system and explains how Princeton University Library's Digital Map and Geospatial Information Center has designed a system that will allow the library to integrate various forms of geographic information and make them accessible online from one interface.

Challenges

There are numerous challenges in making geospatial data and digital maps accessible over the Internet. Many libraries have used ESRI's ArcIMS and ArcSDE, and relational databases such as Micosoft's SQL Server, Oracle, etc., but they were not very successful in making diverse collections of digital maps and geospatial data accessible online from one interface. This was due to the following reasons:

- Disseminating digital maps and geospatial data via ArcIMS technology is not practical for libraries when they have a great quantity of material covering different parts of the world at different scales and in different formats.
- There is no simple way to view and download vector geospatial data stored in ArcSDE without creating ArcIMS image or feature services. Using ArcIMS to build image and feature services to view and download vector data is not only time consuming but also uses a lot of processing power on a server.
- Many libraries are scanning large historical maps and aerial photographs. Some of them are georeferenced but many are not. Disseminating these types of materials with vector geospatial data is a real challenge.
- The file sizes of scanned maps and geospatial data could vary from a few megabytes to a gigabyte. Making a large file accessible over the Internet is a challenge.
- Designing a system that has easy workflows and ease of maintenance is difficult.

Because of these reasons, I spent a few years testing different server side technologies to build a system that will not only allow our library to organize and manage digital maps and geospatial data with easy workflows but will also allow users to search, browse, view, and download different formats of geographic information. Some of these formats include scanned historical/present maps, aerial photographs, satellite images, and vector geospatial data. The advantage to building such a system is that all kinds of geographic information can be integrated, managed, searched, and accessed from one interface. Geographic information can range from maps and geospatial data to photographs of places, etc. Many libraries have designed systems to disseminate maps and geographic data online, but the focus is either regional or item specific. In order to build an integrated system to disseminate diverse geographic information, I started a pilot map

scanning project in early 2004. The goal of the project was to design systems and specifications for scanning maps and to establish workflows.

System Design

Before designing a system I had to research what kinds of software packages were available. The Environmental Systems Research Institute (ESRI) server software packages were some of the most sophisticated software packages on the market and some of the most easily available to academic institutions because of ESRI educational licenses. The ESRI server software packages could handle most of the things that I wanted to accomplish. For instance, storing data in ArcSDE provides the flexibility to make data accessible to ArcMap users over the Internet and to store data in a relational database management system (RDBMS). However, there are some limitations to the software. The ESRI server software packages assume that all the data will be made accessible online via ArcIMS and will be georeferenced. That leaves out all the scanned maps or aerial photographs that have no georeferenced information. Another limitation with the ESRI software is that if data are stored in ArcSDE, the only way for a non–ESRI software user to access these data over the Internet is to build some sort of ArcIMS service and make it viewable and downloadable in shapefile format. This server design forced me to look for different software packages that offer the ability to disseminate non-georeferenced scanned maps and aerial photographs online and provide users with the option to view and download vector data straight from ArcSDE.

After understanding the pros and cons of using ESRI server packages, I built a system using ESRI server software packages such as ArcIMS MetadataServer, ArcSDE, Micosoft's SQL Server database, and ArcCatalog. I also used off-the-shelf commercial software packages such as Safe Company's SpatialDirect/FME and Mapping Science's GeoJP2 Encoder and Decoder and Image Server. I used ArcCatalog to create metadata; ArcIMS MetadataServer, ArcSDE, and SQL Server to publish and store all the metadata and geospatial vector data; and SpatialDirect and FME to access data from ArcSDE and convert ArcSDE data into more than thirty different file formats. I used GeoJP2 Encoder to convert and compress TIFF files to JPEG2000 (JP2) and Image Server to serve JP2 images over the Internet without plug-ins. I was able to create five databases (Metadata, Gazetteer, GISdata, SpatialDirect, and PUMapData) in the SQL server to store various components of our data. The Metadata database stores all the metadata records, the Gazetteer stores gazetteer information to help search a place name more easily, GISdata stores all the vector data, SpatialDirect stores all the vector records to interact with FME software, and PUMapData stores basic information of scanned maps and creates unique image file names. In addition to these databases, I also created two folders in our server to store JP2 images. One is for holding public domain materials, and the other is

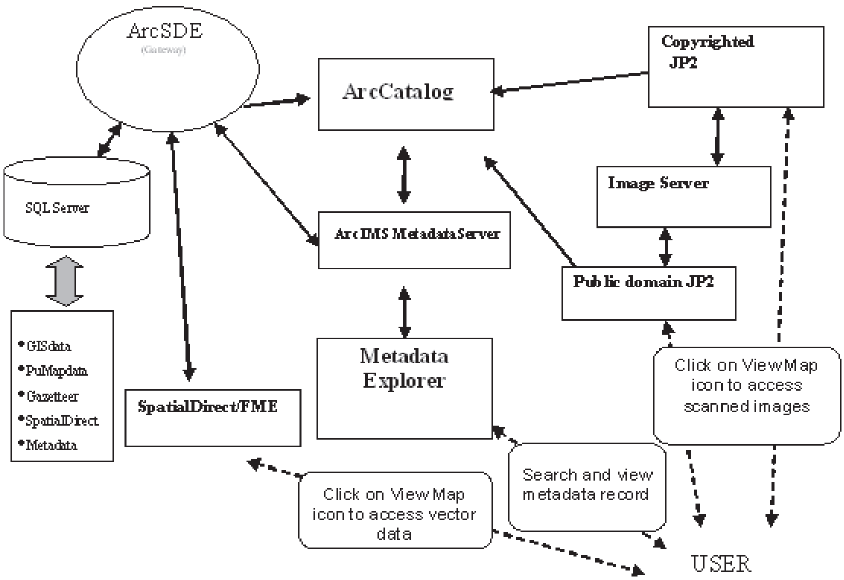


Figure 1. System Design

for storing copyrighted maps. Both of the folders are linked to JP2 Image Server. See Figure 1 for a diagram of this system.

Scanning Process

Before the scanning work was started, I researched how other institutions were scanning maps and why specific resolutions were used. The Library of Congress scans cartographic materials at 300 dots per inch (dpi) with tonal resolution of 24-bit color and saves files in TIFF non-compressed file format. The British Ordnance Survey (OS) scans maps between 254 dpi and 400 dpi in a non-compressed TIFF file with 256 colors. The United States Geological Survey (USGS) has done a lot of map scanning work. The main goal of the OS and USGS scanning work is to convert paper map information into digital geospatial data. The USGS has scanned differently scaled USGS maps, extracted map information, and created geospatial data such as digital elevation models (DEMs), digital line graphs (DLGs), and U.S. Census's TIGER street data. In researching what resolutions were used, I found that the USGS Digital Raster Graphics (DRGs) made before October 2001 were scanned at the resolution of 250 dpi. However, most DRGs made after October 2001 have scan resolutions of 500 dpi. The colors of the scanned maps were reduced to a standard color map of 13 colors. The goal of most map scanning projects is to preserve map information and extract data from the map for geospatial analysis.

An in-house test proved that scanning a paper map (USGS 1:24,000 topographic map) at 400 dpi with 256 colors versus 500 dpi with 24-bit color shows very little difference. In fact, most of the large-format sheet-fed scanners that are currently on the market have around 400 dpi as actual/optical scanning resolutions. Scanning a map higher than the scanner's optical resolution is basically interpolating actual optical resolution, which means the number of pixels and file size increase but better map information is not necessarily captured. After reading about and testing different scanning options, I came to the conclusion that a minor visual quality improvement hardly justifies the larger file sizes (500 dpi with 24-bit color: file size 441MB; 400 dpi with 24-bit color: file size 278MB; 400 dpi with 256 color: file size 96.2MB). Nor does it justify the extra time it takes to scan and save the image. Therefore, I decided to scan paper maps at 400 dpi optical resolution with 256 colors, since scanning a map to preserve map information for later Geographic Information Systems (GIS) use and scanning a map as artwork are two different things. The objective of this scanning project was to preserve map information, so it was not important to capture all the subtle color differences or color "noise" generated by the condition of the paper and the printer. Maps published by the USGS usually use less than 13 colors, and storing a scanned map as 256 colors is more than enough to preserve map information.

After making the decision on what resolution to scan the maps, I also needed to research what was the best compression ratio to encode the TIFF file into JP2 file format. By performing different compression tests I found that 10:1 was the best compression ratio in terms of visual result and file size. The maps were scanned at 400 dpi with 256 colors and were saved in a non-compressed TIFF file format for archival purposes. The TIFF images were then compressed using GeoJP2 software into JP2 files with 10:1 compression ratio for online access.

Once scanning resolution and compression ratio standards were established, the maps were scanned without making much effort in color balancing, image cleaning, or other changes in image processing software. One exception to this was that the images were cropped to delete white space that was not part of the map. Any pencil marks on a map were erased before it was scanned. In the initial stage, our library scanned maps covering different parts of the world to organize them in different geographical regions and to test how browsing options worked on the Metadata Explorer's page.

Workflow

The maps scanned as part of this project were cataloged in the GEOMAP database (our local map cataloging database). Before a map was scanned, the catalog record was located in the GEOMAP database and used to enter brief information in the PUMapData database. A simple Microsoft Access

interface was used to connect to the PUMapData database, which is located in the SQL server. Once a connection was made, a staff member entered brief information about the scanned map, such as the title, publication date, and description of how the map was scanned and encoded, etc., in the PUMapData database. After entering the basic information, the database allowed us to generate a sample text file consisting of the information entered in the database along with a unique ID and the time and date the map was scanned. This was used as a brief metadata record and was encapsulated with the scanned map when it was encoded into the JP2 file. The unique ID was also used as a file name for the scanned map. The scanned map was saved as a non-compressed TIFF file. Afterwards, all the scanned maps were compressed (encoded) with text generated from the PUMapData database, using Mapping Science's GeoJP2 Encoder software. Once the maps were compressed, they were moved to JP2 folders in our server. The public domain maps were moved to a normal JP2 folder. If the scanned map was copyrighted, it was moved to another folder called "Copyrighted." The maps from this folder are accessible only at one computer in the Map Library. The non-compressed TIFF files were moved to a specially designated hard drive space for archiving.

Once maps were in the JP2 Image Server folders, metadata records were created with ArcCatalog software. All the scanned maps were individually cataloged using the *International Organization for Standardization (*ISO) 19115 metadata standards. At this stage, the GEOMAP database was accessed in order to pull the compressed map catalog record using a GN number (all the scanned maps that were cataloged in GEOMAP database have this unique number). Most of the GEOMAP catalog record is used for creating metadata for scanned maps in the ArcCatalog. Once a metadata record is created, it is published to the ArcIMS MetadataServer. As soon as metadata is published, a scanned map is immediately accessible to our users. Before publishing metadata, we created different folders in the MetadataServer that are based on some geographical hierarchy such as continents, regions, etc. (for example, North America ⑧ United States ⑧ New Jersey ⑧ Mercer County ⑧ Princeton). These folders are used for publishing our metadata records and will help our users to browse and select a map based on some well-known geography hierarchy.

After publishing the metadata, the scanned map ID and name were entered in the Excel spreadsheet with a note stating the metadata record was created. If somehow a metadata record could not be created or there was a problem with a compressed image, that information was entered in an Excel spreadsheet for a substitute record.

Vector data workflow processes are slightly different. First the data were uploaded in the ArcSDE using ArcCatalog, and SpatialDirect's Spatial Assistant connected ArcSDE tables (this connection allows SpatialDirect to read the data directly from ArcSDE without creating ArcIMS services). After

making the connection between ArcSDE and SpatialDirect, we opened SpatialDirect's Administration Interface Web page, created a map image, generated a unique URL, and entered the necessary information such as file name and size in the database called SpatialDirect. This database is located on the SQL server. We then opened ArcCatalog and made a connection to the ArcSDE database. We selected data and created a metadata record for that data, and while creating the record we inserted a unique URL that was generated in SpatialDirect in the Online Linkage space. Next we saved the metadata record and published it in the ArcIMS metadata server. The published metadata and data were then ready to search, view, browse, and download from Metadata Explorer immediately. Figure 2 shows a snapshot of a Metadata Explorer page.

How the System Works

This system helped the library develop an easy workflow and also helped patrons search and browse geographical information including geospatial data, maps, and aerial photographs from one interface without searching different databases. The system has also allowed our library to scan copyrighted maps in addition to those in the public domain. Copyrighted maps are scanned for two purposes: for archival reasons and to give a general picture of how a map looks. This is possible because the scanned materials that have metadata records also have thumbnail images of the map. This thumbnail image of the map will give our user some idea of whether the map in our library will be useful for his/her research. If we did not provide this option, users would need to come to the Map Library to look at the maps.

This system design has given our patrons the option of accessing our materials on their desktops, either by searching or by browsing. Once the material is found, a user can click on the *View Map* Icon to view a map as a digital image or vector data. If it is a public domain map, the user can view and download the map in either JPEG or TIFF. If the map is georeferenced, the user can not only view the map but can also download it in JPEG and TIFF with a world file. This allows patrons to use a downloaded map in GIS software. If it is a copyrighted map, then another window will pop up with a message stating it is copyrighted material and that the map is not accessible over the Internet but can be viewed at the Map Library.

If the user is accessing vector data, the system will force the user to type his/her user name and password. User names and passwords are necessary to protect misuse of SpatialDirect/FME software. These software packages are free for academic institutions for educational use, but the Safe Company does not allow use of the software by the general public. Once the proper information is provided, a general coverage of the map will be shown, which allows the patron to download the file in more than thirty different

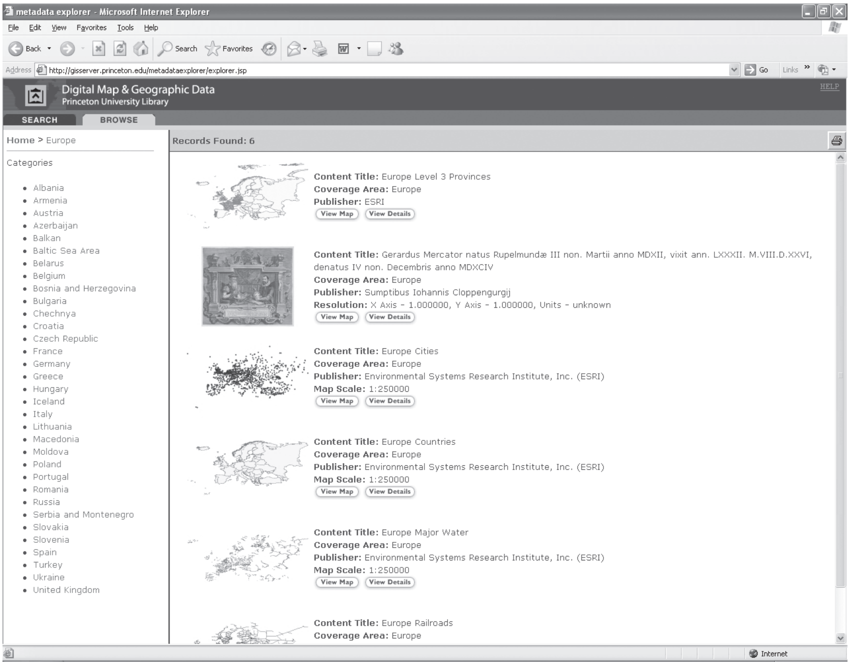


Figure 2. Sample Metadata Explorer Page

file formats. This has allowed our users, who may not use ESRI software, to access and download data in their preferred software file formats.

Lessons Learned

Building a digital data infrastructure has helped me to understand what resources are needed and how to build such a system. I found that it is crucial to get support from the systems department, a database specialist, and a programmer to design a system; without their help it would be very difficult to build and maintain a system. To continue with scanning, creation of metadata, and uploading of vector data in ArcSDE, having a dedicated support staff is essential. Based on these experiences, we have found that hiring student workers may not be the best option. The high turnover among student workers every semester demands too much time and resources for training. This high turnover can also lead to inconsistent quality of work.

Throughout this project, we found it was important to make the library administration understand what size of disk space we needed for our work. After I was initially given a server with roughly 300 GB space, I informed

our administrator that this was not enough. I suggested a minimum of a few terabytes of server space to continue with our map scanning project and making geospatial data accessible online. Unlike other digital projects, scanned maps and geospatial data take up a lot of disk space, and therefore it is important for the library administration to understand the need for the larger amounts of disk space to continue with the work. In addition to disk space, I also learned from my experience the importance of building a redundancy system on our server so that if anything unexpected happens, our services will remain accessible to our users. Because of this, we decided to move our system to a new server that is based on a cluster server. This server has two nodes, both of which will be running the same application but data will be stored in another drive. This server design will help us to build a redundancy system. The final lesson that I learned was the need to create an alias name for the server. This way, when we move the project to another server, we can keep the same alias server name and will not have to change the Web page address/name.

Conclusion

The pilot map scanning project was very helpful to our library. It helped us build a system that will allow our library the flexibility of disseminating diverse geographic information over the Internet. Before the system was built we did not have the tools to make maps, aerial photographs, and geospatial data accessible online from one interface. The project allowed us to use a new file format called JP2 and to develop our map scanning and file compression standards, which we continue to use. It helped us to estimate the size of disk space we need to continue making our diverse geographic information available to our library users online. It also helped me make our administrator aware of what supports and resources were needed to integrate diverse collections of geographic information and make them accessible online. One of the goals in designing this system was to encourage other libraries to build similar systems for their own use. In addition, the project led me to ask the president of ESRI to develop a similar system for the map and GIS library community. If ESRI does design such a system, my hope is that it will minimize the complexity I found in integrating different software packages. Whether libraries manage to build their own systems or are able to use a new package from ESRI (if they do design such as system), I hope that more libraries will be encouraged to make their diverse geographic data accessible online from one interface.

Appendix: Suggested Reading

British Ordnance Survey. (n.d.). *1:25 000 Scale Colour Raster: technical information*. Retrieved November 12, 2003, from http://www.ordnancesurvey.co.uk/oswebsite/products/25kraster/techinfo.html.

British Ordnance Survey. (n.d.). *1:10 000 Scale Raster, technical information.* Retrieved November 12, 2003, from http://www.ordnancesurvey.co.uk/oswebsite/products/10kraster/techinfo.html#gr.

GPO. (2005). *About the Federal Depository Library Program (FDLP).* Retrieved February 8, 2006, from http://www.gpoaccess.gov/fdlp.html.

Library of Congress. (n.d.). *Scanning cartographic materials.* Retrieved November 12, 2003, from http://memory.loc.gov/ammem/gmdhtml/gmddigit.html.

Shawa, T. W. (2003). Review of JPEG2000 and GeoJP2 Compression Software. *Baseline: A Newsletter of the Map and Geography Round Table, 24*(3), 8–10.

Shawa, T. W. (2003). What is the best resolution to scan a map? *Baseline: A Newsletter of the Map and Geography Round Table, 24*(6), 6.

Shawa, T. W. (2005) From the chair. *Baseline: A Newsletter of the Map and Geography Round Table, 26*(5), 4–5.

USGS. (2001). *National mapping program technical instructions, standards for digital raster graphics.* Retrieved November 12, 2003, from http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/drg_temp/Pdrg0401.pdf.

USGS. (2001). *National mapping program technical instructions, part 1, general standards for digital raster graphics.* Retrieved November 12, 2003, from http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/drg_temp/1drg0401.pdf.

USGS. (2001). *National mapping program technical instructions, part 2, specifications, standards for digital raster graphics.* Retrieved November 12, 2003, from http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/drg_temp/2drg0401.pdf.

Tsering Wangyal Shawa is a Geographic Information Systems Librarian at Princeton University. He has widespread experience in geospatial data selection, software, and hardware and holds degrees in the areas of library science, education, geography, and cartography. He is the current chair of the American Library Association Map and Geography Round Table (2005–2006) and is the chair of the Geographic Technologies Committee. He was selected by the National Research Council and the Federal Geographic Data Committee's Homeland Security Working Group to study and publish reports on "Licensing Geographic Data and Services" and "Guidelines for Providing Appropriate Access to Geospatial Data in Response to Security Concerns." He was a consultant for the Tibetan and Himalayan Digital Library (THDL) based at the University of Virginia and was a Cartographic Users Advisory Council (CUAC) member from 2002 to 2005. He was born in Tibet and has lived and taught geography and cartography to high school and college students in India, Nepal, Kenya, and Sudan.

Libraries as Distributors of Geospatial Data: Data Management Policies as Tools for Managing Partnerships

Gail Steinhart

Abstract

Libraries can bring substantial expertise to bear on the collection, curation, and distribution of digital geospatial information, making them trusted and competent partners for organizations that wish to distribute geospatial data. By developing a well-thought-out data management and distribution policy, libraries can define the parameters of a data distribution partnership and reinforce a data provider's confidence in the library's role as a data custodian and distributor. In developing a policy, data distributors are advised to consider such issues as intellectual property rights, liability issues, distribution methods and services, data and metadata management practices, security risks posed by geospatial data, and user limitations. This article describes the most common elements of data sharing and distribution agreements and describes the development of a data management policy for the Cornell University Geospatial Information Repository (CUGIR).

Introduction

Although libraries are generally not producers of geospatial data, they are effective institutions to serve as distributors of geospatial data within larger spatial data infrastructures (SDIs). The process of managing distribution partnerships with data providers touches on virtually every aspect of managing and distributing digital data. This article will present a brief overview of some of the issues influencing organizations' decisions to share data and distribute data, the strengths libraries bring to data distribution, and an overview of issues that a library, acting as a data distributor, should

LIBRARY TRENDS, Vol. 55, No. 2, Fall 2006 ("Geographic Information Systems and Libraries," edited by Jaime Stoltenberg and Abraham Parrish), pp. 264–284

consider when formulating data management policies or agreements. The article concludes with a description of the process of developing a data management policy for the Cornell University Geospatial Information Repository (CUGIR).

Evolution of Attitudes Toward Data Sharing and Distribution

Born digital, geospatial data lends itself to distribution via the Internet. It is easily reused, well-developed standards for metadata exist, and while there are multiple proprietary formats for geospatial data, some are cross-platform and many applications are capable of reading or importing multiple formats. Initiatives at local, state, and national levels and beyond encourage, or at times require, producers of geospatial data to share or distribute data publicly. Systems such as the National Spatial Data Infrastructure gateways and Geospatial One-Stop (in the United States) exist to facilitate discovery of and access to geospatial data from multiple providers.

The benefits of sharing for providers and users of geospatial data are generally well recognized. Specific benefits to a data provider depend on its mission and mandates, data needs, and the type of sharing or distribution arrangements the organization enters into. Some of the benefits of sharing or distributing data may include

- enhancing interorganization activities by sharing information
- enabling the reuse of geospatial data by other organizations and resulting cost savings
- improving and correcting errors in data in response to feedback from users
- fulfilling public data distribution requirements
- developing competencies in and promoting data and metadata standards.

When a data provider enters into a partnership with a data distributor, additional benefits may accrue: the data provider may receive support or consulting services for metadata development; the distributor's services may make the data discoverable by new or additional means; and the distributor may take responsibility for being the first point of contact for data users.

Early development of data-sharing arrangements and SDIs was sometimes characterized by reluctance on the part of data producers to share data. Where the direction and management of the relationship was perceived as top-down and remote, there may have been resistance to participation. Issues related to the potential loss of local control were the main reason for resistance to data sharing; and some of these issues included meeting local requirements for data management and access, standards requirements (particularly for metadata), time requirements, management of data updates, and cost (Meredith, 1995).

There has been substantial progress in sharing data and developing SDIs over the last several years, but in some cases these concerns persist. Harvey (2003) asserts that trust is fundamental in establishing partnerships and sharing data. A survey of local government agency contacts in Kentucky showed that while local governments share data in a variety of ways, these relationships are based on trust rather than formal agreements. Nearly half of Harvey's survey respondents had no data-sharing agreements. What formal agreements Harvey did encounter were largely post-hoc agreements, formalizations of informal and preexisting arrangements. In a survey of agencies whose activities affect transportation systems, where most of the responding agencies recognized that sharing data can enhance interagency coordination, Zimmerman (2002) also found that about half the agencies she surveyed had a formal data-sharing policy. These agencies report sharing data with other agencies as well as distributing information on travel conditions to the public. Respondents reported protecting their interests in the data they shared by a variety of means, although most of these were relatively unrestrictive and the most common practice was a requirement to acknowledge the source agency.

On a national level, in the United States federal laws and regulations have influenced the data-sharing and distribution policies of federal agencies. One of the most important of these is OMB Circular A-130 (Office of Management and Budget, 1996), which governs the management of federal information resources, pursuant to the Paperwork Reduction Act. Its most salient provisions are that federal agencies should actively disseminate public information without restrictions or conditions and that data should be provided at not more than the cost of dissemination. States also often have policies in place mandating or encouraging the sharing of information among agencies or with the public; Cho (2005) reports that every state has a statute or policy related to Geographic Information Systems (GIS) data distribution. In New York State, Technology Policy 96-7 establishes the New York State GIS Data Sharing Cooperative and encourages data sharing among state and local agencies (Governor's Task Force on Information Resources Management Technology, 1997).

In spite of some apparent lingering concerns regarding loss of local control over data, there has been an evolution of thought with respect to data sharing with SDI participation. Masser (2005) describes several such trends in SDI development. One is the movement from a product-focused model—that is, the development of datasets and databases—to a process-focused model—the ongoing management, updating, creation, and distribution of data. Architectures have evolved as well, from centralized, top-down structures to more distributed models. Finally, management functions are maturing from formulation to implementation and are becoming sufficiently flexible to accommodate multiple levels of participation and new organizational structures. If these trends hold true, it would seem

many of the early objections to data sharing and SDI participation are less important than they once were, that the nature of SDIs has evolved in such a way that some of these concerns have been effectively addressed, or that various mandates have simply removed these concerns as significant barriers to data sharing and distribution.

Why Partner with Libraries for Data Distribution

Libraries can be effective participants in SDI development and data distribution and have a proven track record as partners in data distribution, evidenced by their role in the Federal Depository Library Program (McGlamery, 1995). Libraries also possess well-developed expertise in several related areas, including collection development, archival practices, cataloging and indexing, development of platforms for discovery and distribution, and education and user support. In a paper on the creation of the New York State GIS Clearinghouse, Dawes and Oskam (1999) described an important additional characteristic that made the New York State Library, the original operator of the clearinghouse, an effective partner in a statewide effort to distribute GIS data: the library was perceived as a neutral party. Making a New York State agency the primary distributor may have given the appearance that a particular agency was the leader with respect to GIS operations, but the library was not perceived as a rival by other New York State agencies. This characteristic neutrality of libraries can be important for establishing trust with prospective data providers. Finally, many libraries, either by virtue of their participation in the Association of Research Libraries' (ARL) GIS literacy project, or through their own deliberate development of expertise in GIS technology and services, have acquired the more specialized knowledge of GIS and geospatial data that is required to support a distribution system (Herold, 1997; McGlamery, 1995).

Managing Partnerships

Libraries are generally recognized as trusted custodians of information, and one of a library's core responsibilities is to manage information in such a way that both safeguards the integrity of the information and facilitates access. Libraries acting as partners in the distribution of geospatial information must both meet these core responsibilities and ensure that the requirements of the cooperating data providers are met. Creating a data management and distribution policy can serve to clarify and make explicit both participants' expectations and lend predictability and stability to data distribution arrangements.

Three types of participants are involved in the distribution of geospatial data: data providers (the creators of geospatial data), data distributors (who may be the same as the data provider or may be a third-party distributor), and data users. The channels of communication between the participants may be unidirectional or bidirectional and are illustrated in Figure 1. Com-

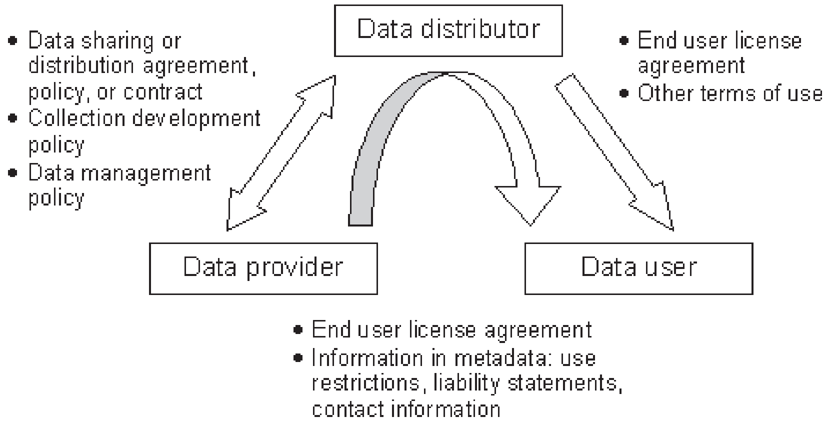


Figure 1. Communication between Participants in Geospatial Data Distribution

munication between data providers and distributors may be bidirectional, with both parties having the opportunity to specify their own policies and terms and accepting the other party's, or they may be entirely dictated by the data distributor. Both the data distributor and the data provider may wish to impose certain restrictions on the use of data, to exempt themselves from liability, and to communicate other information to the user. This information is usually communicated unilaterally, by means of end-user license agreements, use constraints or other information included in a dataset's accompanying metadata, or other terms of use, such as those posted on a Web site. In the case of terms imposed on the end-user by the data provider, while the information to be conveyed may be determined by the provider, the communication is usually accomplished by the distributor.

Distribution partnerships may range from very open to fairly specific and restrictive in terms of the degree of oversight and control exercised by either the data provider or data distributor. As evidenced by the lack of universal creation and adoption of data-sharing and distribution agreements, management of various aspects of such partnerships may be formal or informal. More formal arrangements may take the form of legal contracts or nonbinding agreements or policies. One drawback to legal contracts is the obligation to negotiate terms with each partner, and in some cases, a nonbinding agreement or policy may be the preferred approach (Longhorn et al., 2002). Existing models of formal statements of data-sharing practices include agreements and contracts published by various governmental agencies, data repositories, and archives, both for geospatial data specifically and for other types of data more generally. Among GIS practitioners and creators of geospatial data, many agreements are bilateral, governing the

exchange of data between two organizations, rather than distribution arrangements between a data provider and a data distributor. Nevertheless, many of the same issues and principles apply whether the communication is intended to facilitate sharing or exchange of data between two parties or it is intended to facilitate distribution of data more broadly (Dangermond, 1995).

Elements of Data-Sharing Agreements

To identify the most common elements of data-sharing agreements, policies, and contracts, sixteen actual and sample or model agreements were reviewed (see Table 1). These were found by searching the Internet, visiting individual data repositories and locating relevant documentation, and reviewing literature on best practices for data sharing and distribution. The most common elements were identified and summarized in Table 2.

There is no single approach to articulating data management and distribution practices, data-sharing agreements, or the terms of these types of partnerships. Some agreements include information both on the details of managing the relationship between two parties as well as information on actual operations, including data management practices. Other agreements focus primarily on the former, with data management practices outlined separately. A complete treatment of all the potential elements of a data-sharing policy or agreement is beyond the scope of this article; hence, following a brief overview of the elements listed in Table 2, this discussion will focus on those topics in which libraries have particular strengths and where CUGIR has significant experience: data management and collection development policies, including some issues related to the management of security concerns with respect to geospatial data.

Definitions and Procedural Information

Definition of terms and procedural information is fairly standard and straightforward material in contracts. This information serves to identify the participating organizations and, in the case of contracts, to outline the rules of engagement for executing, amending, and terminating agreements, as well as dispute resolution.

General Legal Issues

Applicable law, or jurisdiction, is commonly declared in contracts. It is of little relevance in agreements that are nonbinding. Intellectual property rights in geospatial data are likely to be a matter of copyright, but copyright law with respect to geospatial information is not clear-cut. Facts are not copyrightable, but compilations of facts or databases may be if they entail sufficient creative expression. Some argue that the representation of geographic features leaves no room for creative expression in the context of geographic information systems without adversely impacting the accuracy of the information or greatly diminishing its value by depicting

Table 1. Data-Sharing Agreements, Policies, and Contracts Reviewed for This article

Organization	Type of Data	Type of Agreement	Reference
Charlevoix County GIS Program	Geospatial	Cooperative	Charlevoix County GIS Program, 2004
County of Hunterdon, New Jersey, Division of Geographic Information Systems	Geospatial	Usage	County of Hunterdon, New Jersey, Division of Geographic Information Systems, n.d.
Geography Network	Geospatial	Distribution	Environmental Systems Research Institute, Inc. (ESRI), n.d.
GeoNOVA Geographic Gateway to Nova Scotia	Geospatial	Cooperative, distribution, usage	Barrington Consulting Group, 2005
Geospatial One-Stop	Geospatial	Distribution	Geospatial One-Stop, n.d.
Global Biodiversity Information Facility (GBIF)	Various (biodiversity)	Distribution	Global Biodiversity InformationFacility) (GBIF), n.d.a; Global Biodiversity Information Facility (GBIF), n.d.c
Global Biodiversity Information Facility (GBIF)	Various (biodiversity)	Usage	Global Biodiversity nformation Facility (GBIF), n.d.b
Macomb County (MI) GIS Services Division	Geospatial	Cooperative	Macomb County (MI) GIS Services Division, 2002
MetroGIS	Geospatial	Cooperative, distribution	MetroGIS, 2004
New York State Office of Cyber Security and Critical Infrastructure Coordination	Geospatial	Cooperative, distribution ,	New York State Office of Cyber Security and Critical Infrastructure Coordination, 2005
North Carolina and State of North Carolina Centerfor Geographic Information and Analysis (CGIA)	Geospatial	Cooperative	North Carolina Center for Geographic Information and Analysis (CGIA), n.d.
Open Data Consortium Project	Geospatial	Distribution	Joffe, 2003
Somerset County, New Jersey	Geospatial	Cooperative	Somerset County, New Jersey, n.d.
U.S Global Change Research Program	Various (global change research)	General policy	USCGRP Data and Information Working Group, 2002
University of Michigan School of Natural Resources and Environment	Geospatial	Distribution	University of Michigan School of Natural Resources and Environment, 2003
Wyoming Geographic Information Advisory Council (WGIAC)	Geospatial	General policy	Wyoming Geographic Information Advisory Council (WGIAC), 2000

Note: This table include actual agreements and policies, as well as recommended or model agreements and policies. Cooperative agreements refer to agreements made between two or more parties that govern the sharing or use of data by one or more of the parties. Distribution agreements are agreements between a data provider and a data distributor. Usage agreements are agreements or conditions posted on a Web site or otherwise specified by a data distributor. General policies describe the goals and policies of organizations that coordinate data-sharing activities and may lack specific information on the responsibilities of participants.

Table 2. Common Components of Data-Sharing and Distribution Policies

Component	Issues to Consider
Definitions	Definitions of terms and acronyms
Procedural Information	Primary points of contact Duration of contract or agreement Applicable fees Procedures for amendment Procedures for notification Procedures for dispute resolution Procedures for termination
General Legal Issues	Applicable law Intellectual property rights, including distribution permissions and limitations Liability statements
Distribution Methods and Services	Modes of distribution (media, Internet, direct database connection, Web services) Distributor-provided services such as data extraction and reformatting
Data Management Practices	Verification of provider's authority to make data available for public distribution Distributor's collection development practices Data requirements and standards Metadata requirements and standards Maintenance and improvement of data Archival policies and practices Limitations on access to data Policies and procedures for accepting and distributing sensitive data Privacy and confidentiality policies
End-User License Agreement Terms	Statement of copyright Limits to warranty Liability statements Attribution requirements Use restrictions Redistribution limitations Delivery of derivative works to data provider Rights in value-added datasets

or transmitting it in a nonstandard way (Onsrud & Lopez, 1998). Others argue that there is substantial latitude for creative expression, especially cartographic expression, even in digital form (Cho, 2005). Contract law and licensing agreements present alternatives to copyright protection when a data provider or distributor must retain a proprietary interest in data (Onsrud & Lopez, 1998). Regardless, the law is not entirely settled on this issue, so agreements should clearly state whether the data provider claims copyright, what rights are transferred to the distributor, and applicable distribution permissions and limitations (Committee on Licensing Geographic Data and Services, 2004). In addition, derived or value-added datasets and products may present complex intellectual property rights issues (Longhorn et al., 2002).

Liability in the use of geospatial data generally arises because the data are used to make decisions, and errors in the data that result in inappropriate decisions or actions are at the root of liability cases. The issues are usually ones of contract law and warranty (Onsrud, 1999). An additional liability risk posed by the distribution of geospatial data is infringement upon intellectual property rights (Cho, 2005). In either case, strategies to manage liability risks might include disclaimer statements and management practices that explicitly track and document data quality. Such practices include evaluating and documenting data currency, accuracy, and lineage. Much of this information can be expressed in geospatial metadata (Cho, 2005).

Distribution Methods and Services

Geospatial data may be distributed by a variety of means, on- or offline. Modes of online distribution for geospatial datasets may include data repositories, data clearinghouses, direct connections to databases, and Web mapping applications.

Data-related services that might be provided by a distributor could include extraction of parts of a dataset or reprojection of a dataset, either manually upon request or by providing users with Web-based tools. Some data distributors may add value to datasets by supplying additional attribute data.

Data Management Practices

Data Provider's Authority to Make Data Available for Public Distribution To guard against infringement of copyright or other applicable laws, it is essential that the data provider have the authority or permission to allow the public distribution of the data in question.

Distributor's Collection Development Practices Some aspects of collection development policies and issues related specifically to geospatial data are listed in Table 3. Elements of a collection management policy may influence, or be influenced by, general decisions related to data and metadata management. A policy can ensure consistency in collection development and can help guide decisions when resources for acquiring items are limited. For some GIS data, there may be no cost to acquiring data, but a significant amount of staff time may be required to process new datasets, create or edit metadata, and maintain and support the distribution system. Criteria that might be considered in any collection development policy also apply to geospatial data, such as subject area and geographic scope and data format, but even these raise specific questions with respect to geospatial datasets.

Data Requirements and Standards Data distributors should give some thought to several characteristics of data they might distribute. File format is one important consideration. There are many geospatial data formats; some are proprietary and not all are equally accessible in all GIS software applications. Whether data must be georeferenced and projected, and whether there is a preferred coordinate system, are also important con-

Table 3. Elements of Collection Development Policies

Policy Element	Issues to Consider
Subject Scope	What is the subject scope of the collection?
Geographic Scope	What is the geographic scope of the collection? If the geographic scope is defined by political boundaries, how should datasets that are distributed by nonconforming or overlapping boundaries (such as watersheds or 7.5 minute quad sheets) be treated?
Data Quality	Are there minimum standards for data quality? Does the responsibility for maintaining standards of data quality rest with the original data provider or with the repository?
Distribution Constraints	What distribution constraints apply to the library or repository? Is the repository to be the sole distributor of the data or may the data be distributed by other channels? What distribution constraints apply to end users of data in the repository?
Security Issues	Do the datasets under consideration pose security risks? Does the repository accept for distribution datasets that may pose a security risk, and if so, does the repository restrict access in any way to such datasets?
Metadata Availability	Is metadata required for the datasets? Does the responsibility for creating metadata rest with the original data provider or with the repository?
Metadata Standards	Is adherence to a specific metadata standard required? Is adherence to a specific metadata standard the responsibility of the original data provider or the data repository? Does the repository provide support to data providers for creating standards-compliant metadata?
File Format	Are specific file formats supported or not supported? Are proprietary or open (platform- and application-independent) formats favored for distribution? Will the same data be provided in more than one format?
Unit of Distribution	Is it preferable to distribute data files individually or as packages? What are the preferred geographic units for distribution?

siderations. Finally, distributors should consider their preferred units of distribution. This can apply to geographic units (should files be distributed by the largest or smallest possible areas?), and also to whether it is preferable to distribute packages of related files or if data should be distributed in single layers.

Metadata Requirements and Standards Metadata are essential for providing the means to discover geospatial data, for users to evaluate a dataset's fitness for use for their particular application, and for documenting important information about a dataset. The Content Standard for Geospatial Metadata (CSDGM) (Federal Geographic Data Committee, 2000), promulgated by the Federal Geographic Data Committee (FGDC), is currently the most widely used standard in the United States. The International Standards Organization (ISO) has published an international standard for geographic metadata (International Organization for

Standardization, 2003) that defines the schema required for describing geographic information and services, and various groups are working to harmonize the CSDGM and ISO standards. If they have the resources to do so, data distributors may offer data providers some guidance in creating standards-compliant metadata. Finally, distributors may want to add supplementary information to a data provider's metadata. Such additions might include additional contact or liability information pertaining to the distributor and enhancements or improvements to metadata. Maintenance and Improvement of Data Currency and accuracy are two critical aspects of geospatial data. Data providers may need to provide updated or corrected datasets for distribution. Whether a new version of a dataset represents an update or a correction and the disposition of superseded datasets should be considered.

Archival Policies and Practices When geospatial data are to be distributed by a party other than the creator of the dataset, both groups should be clear as to whether preservation or archival services are to be provided and by whom. RLG's report on trusted digital repositories (RLG-OCLC Working Group on Digital Archive Attributes, 2002) and audit checklist for certifying trusted repositories (RLG-NARA Task Force on Digital Repository Certification, 2005), and the Open Archival Information System (OAIS) reference model (Consultative Committee for Space Data Systems, 2002) provide useful guidance with respect to digital preservation in general. Others have considered the special challenges presented in preserving geospatial data (Brown, Welch, & Cullingworth, 2005; Center for International Earth Science Information Network, 2005). Even if preservation services are not provided by the distributor, some geospatial datasets are updated frequently, and the distributor will need to distinguish between updates and new versions (Hyland, 2002).

Limitations on Access to Data Limitations on who may access data may take the form of written statements, such as end-user license agreements, or technological controls, such as user authentication. Levels of access for different users may take the form of read- or view-only access controls or methods of distribution.

Policies and Procedures for Accepting and Distributing Sensitive Data A distributor is well advised to consider whether it wants to take responsibility for distributing data that may pose a security risk and what procedures must be in place to ensure the security of the data in its collection. For a thorough review of these issues, as well as a framework for assessing the risks associated with geospatial datasets, see the Rand Corporation report on the topic (Baker, 2004). The Rand report framework takes into account three main characteristics of geospatial information: usefulness to would-be attackers, uniqueness of the information, and the potential costs and benefits associated with restricting access.

Privacy and Confidentiality Policies The high degree of geographic specificity that exists in some geospatial datasets makes it imperative that data providers and distributors consider the protection of the privacy of personal information (VanWey et al., 2005). Both should ensure that their practices are in compliance with the privacy policies of their institutions and any applicable laws. The Federal Geographic Data Committee's (1998) policy on personal information privacy also serves as a general guide to protecting the information privacy of individuals while promoting public access to geospatial data.

End-User License Agreement Terms

End-user license agreements (EULAs) serve to communicate a data provider or distributor's terms to an end-user. These terms may include statements of copyright, limits to warranty and liability, attribution requirements, and user and redistribution limitations. In addition, it is useful to recognize two types of end users—consumers and "value-added" users, who may improve or integrate datasets and redistribute them as new products (de Sherbinin & Chen, 2005). Additional requirements may apply to value-added users, such as requirements to deliver derivative works to the original data provider and statements of rights in value-added or derivative datasets.

Developing a Data Management Policy for the Cornell University Geospatial Information Repository

About CUGIR

Created in 1998, the Cornell University Geospatial Information Repository (http://cugir.mannlib.cornell.edu/) is an online repository providing access to digital geospatial data and metadata for New York State. As a service of Albert R. Mann Library, the library serving the College of Agriculture and Life Sciences and the College of Human Ecology at Cornell University, the focus of the collection is on features and data relevant to agriculture, ecology, natural resources, and human-environment interactions. The CUGIR workgroup is responsible for the development and maintenance of the repository and usually consists of four to five staff from public services, information technology services, technical services, and collection development.

At its inception, a grant from the FGDC's cooperative agreements program made possible the conversion of TIGER/LINE files to GIS format, and the CUGIR collection consisted entirely of data from the U.S. Census Bureau. Soon after, the New York State Department of Environmental Conservation (NYSDEC) and the Soil Information Systems Laboratory (SISL) at Cornell University began distributing their data via CUGIR. There are now more than a dozen CUGIR data providers, which include national, state,

and local agencies, as well as members of the academic community and the private sector. Currently, the repository has more than 7,500 datasets, has supported more than 350,000 downloads since 2001, and provides Web mapping for selected datasets. All data files are cataloged in accordance with the FDGC CSDGM and made available in widely used geospatial data formats. CUGIR is a participating node of the National Spatial Data Infrastructure (NSDI) and registered publisher with Geospatial One-Stop. CUGIR is one of two statewide clearinghouses for GIS data in New York State and coordinates its efforts with the New York State GIS Clearinghouse.

Implementing the CUGIR Data Management and Distribution Policy

The CUGIR work group recently implemented a data management and distribution policy. A primary motivation in developing the policy was to communicate our data management and distribution practices to our data providers. While all of our data providers were probably already aware of how we manage and distribute their data and metadata, because our practices sometimes include modifications to data or metadata and distribution or publication beyond CUGIR itself, we thought we should document our practices and share this information with our data providers. A secondary purpose in creating the policy was to formalize a security review process that was initiated following a request to disable the entire repository some time after the terrorist attacks of September 11, 2001.

The process began with a review of the literature and data-sharing agreements and policies described in the first part of this article. We identified the main elements that should be included and drafted a policy. We considered the possibility of creating a legal contract rather than a policy, but after consulting with Cornell University legal counsel, we decided against this for two reasons. First, because much of the geospatial data distributed via CUGIR are in the public domain or are available with no or minimal restrictions, issues of intellectual property are simple or nonexistent. Second, we could not discern significant enough benefits to having a legal contract that would justify the burden or risk of negotiating agreements with the legal representatives of numerous organizations, including state and federal agencies. CUGIR may be considered unique compared to government-based repositories because participation by providers is voluntary rather than legally mandated. The Governor's Task Force on Information Resources Management Technology (1997) Policy 97-6 on GIS Data Sharing directs all New York State public agencies to "share in the creation, use, and maintenance of GIS datasets" and to deposit their data with the New York State Clearinghouse. No such mandate exists for CUGIR. Nevertheless, some issues related to data management and distribution seemed to warrant a formal expression of CUGIR's data management and distribution practices, if not a legal contract. The probability of our data providers approving an informal policy seemed much greater than if we required a

legally binding agreement. We asked Cornell University legal counsel to review the final draft policy, and then sent it to three of our data providers for preliminary review. Two had no comments, and one had comments that resulted in minor revisions. We then sent the policy to all of our data providers, along with a data inventory for each provider. We asked for their approval of the policy, as well as updates to the information on the inventory. No data providers had any objections to the policy, and as of this writing we are awaiting approval or information from only two data providers.

Elements of the CUGIR Data Management and Distribution Policy

Our policy addresses three main areas: data and metadata management; security; and use, distribution, and rights (CUGIR Work Group, 2005a). CUGIR also has a separate collection development policy (CUGIR Work Group, 2003).

Data and metadata management Our concerns with respect to data and metadata management have to do with issues of file format, geographic projection, updates to data, metadata management and harvesting, and Web mapping. Our guiding principles for establishing guidelines with respect to format and projection were to maximize the utility of CUGIR data. This meant promoting the use of commonly used file formats and projections appropriate to the extent and location covered by the data. CUGIR does, on occasion, request permission from the data provider to distribute the dataset in a format or projection other than the original.

We also wanted to be explicit about the disposition of superseded datasets. There is significant interest in being able to track change over time in a particular location, and if possible, we prefer to make older versions of data available. However, under some circumstances an update to a dataset may represent a change in legal boundaries, and the data provider may prefer to have only the most current data available. The data inventories we sent to our providers included what information we had on whether older versions of their datasets should remain publicly available. In some cases, we had no information, and the process clarified for us how we should handle updated datasets. We should also note that while CUGIR attempts to maintain copies of superseded datasets or other datasets even if they are no longer available for public use, it does not serve as a preservation repository for geospatial data. A possibility for future work in this area is to assess our collection to identify datasets that are good candidates for preservation and to develop the capacity to preserve geospatial data.

Finally, we wanted to convey information about our metadata management and harvesting practices. Because CUGIR participates in various geospatial data clearinghouse initiatives, all data available in CUGIR must have FGDC CSDGM metadata. In some cases, CUGIR metadata librarians will work extensively with a data provider to create or improve metadata.

As the data distributor, we also add information to and enhance the original metadata, replacing the provider's metadata with our version. Additions include Library of Congress place names and keywords, as well as distributor contact and liability information for Mann Library. In addition to clearinghouse initiatives, CUGIR converts metadata records to MARC format for inclusion in Cornell's library catalog, as well as online union catalogs such as OCLC's WorldCat and the Research Libraries Information Network (RLIN).

Security The terrorist attacks of September 11, 2001, substantially increased awareness of and concern about the security risks posed by freely accessible geospatial information. In February of 2002 the New York State Director of Public Security issued a memo to agency heads in New York State, directing them to immediately conduct a review of all sensitive information in the agencies' possession and made available to the public by any means (OMB Watch, 2003). CUGIR was not one of the original recipients of the memo but learned from user inquiries at that time that the New York State GIS Clearinghouse was offline. After CUGIR staff contacted the clearinghouse, Mann Library received a copy of the security memo by fax and was asked to disable access to the site pending a full content review (Hyland, 2002; Martindale, 2002). The library and CUGIR staff, in consultation with Cornell University legal counsel, decided not to disable the site because the directive was intended for state agencies, which CUGIR and Mann Library are not. Instead, we decided to conduct the content review as requested, inform the data providers of the results, and act accordingly. Before the review was completed, one data provider requested that access to all of their data be disabled while they conducted their own content review. Although an operating principle of CUGIR is that access to the collection is free and unrestricted, the CUGIR work group honored this request. We felt it was important to do so in order to maintain trust in the data distribution partnership. Eventually, access to all but three datasets was restored.

This experience led us to consider permanently formalizing the security review of datasets at the point of addition to the repository so we would have that information at hand in the event of any similar requests in the future. We reasoned that it would be easier and faster to defend a decision to keep the repository online if we could provide documentation on the security risks (or lack thereof) posed by the data in the collection. It is worth reiterating that the focus of the collection is largely on geospatial data related to the environment and natural resources. There is little information on critical infrastructure, but the collection does contain, for example, digital raster graphics, which do depict facilities such as power plants and dams. On the other hand, digital raster graphics are widely available from other sources and as paper maps.

The initial security review of CUGIR data was based on two factors (Martindale, 2002): inherent risk (utility of the information to potential

attackers) and distribution level (availability of information from other sources). Each dataset was assigned a numeric score for these risks and for distribution level. The scoring scheme was loosely based on a preservation risk assessment model used by Mann Library for numeric data the library makes available online in cooperation with the United States Department of Agriculture (Hyland, 2002). These two factors correspond nearly perfectly to two of the three factors identified in a report published by the Rand Corporation (Baker, 2004); they were adopted to update the security assessment of all CUGIR datasets in 2005 and to establish a procedure for security assessment. The Rand report framework also takes into consideration the costs and benefits of restricting access to geospatial information. Because a fundamental principle of CUGIR is that the information in the collection is freely available, we did not incorporate the third factor—the costs and benefits of restricting access to geospatial information—into our assessment procedure. This revised CUGIR data security assessment procedure (CUGIR Work Group, 2005b) guided our updated review and was sent to all active CUGIR data providers for their input. Upon completing the review, active data providers were asked to approve or suggest changes. Only minor changes were requested (adjusting a score up or down one point, at most).

Use, Distribution, and Rights CUGIR provides unrestricted access to geospatial data. The one exception we make with respect to this policy is to honor security-related requests made by our data providers. We permit data providers to impose use constraints, as long as they are not in conflict with the rest of our data management policy.

As noted earlier, intellectual property issues with respect to data distributed via CUGIR are simplified by the fact that much of it is in the public domain or otherwise free of copyright and other distribution restriction.

Collection Development Policy

CUGIR's collection development policy was developed about two years before the rest of the data management policy. Some elements of the data management policy are briefly addressed in the collection development policy, but in general the collection development policy is more narrow in scope. The policy describes the overall nature and purpose of the repository, acknowledges CUGIR's data providers as the owners of the data in the repository, and provides guidelines for the scope of the collection. The policy also includes some suggested requirements of data and metadata, although the data and metadata guidelines have already been discussed in more detail in the context of the newer data management policy.

In terms of collection scope, the policy addresses both subject and geographic scope. Generally, most New York State data related to natural resources, the environment, and human-environment interactions are appropriate for inclusion in CUGIR. Examples of such data include topography, soils, hydrology and water resources, environmental hazards, agricultural

activities, wildlife, and natural resource management. We have included datasets from immediately adjacent areas when those data may provide some benefit to CUGIR users. To date, that practice has been limited to some digital raster graphics in neighboring states along the New York State border. The policy also stipulates that CUGIR's distribution policy is an open one and that there is no requirement that CUGIR be the sole distributor of any datasets.

Lessons Learned

Developing a data management policy forced us to consider all aspects of our data management and distribution practices. Because we already had a collection development policy in place that addressed several important issues related to data management, our most significant motivations for developing the policy had to do with communicating our practices that result in modifications to a provider's data or metadata and collecting additional information from our data providers to help us better manage their data.

We have not operated with our data management policy in place long enough to evaluate the results, but we are encouraged by the fact that none of our data providers had any objections to the policy and pleased that the process helped us update our records about how certain datasets should be managed. Some of our providers were surprised by the question of what to do with superseded datasets and had to give the issue some thought before responding. For data providers with whom we have infrequent contact, the process provided us with an opportunity to "check in" with them and provide them with some assurance that we are attentive and responsive to their needs with respect to data management. We are also pleased to have complete security risk information at hand, which would permit us to respond and make decisions quickly in the event of any future requests to restrict access to data in the repository.

Conclusion

Libraries can bring substantial expertise to bear on the collection, curation, and distribution of digital geospatial information. This expertise makes libraries trusted and competent partners for organizations that wish to distribute geospatial data. Managing and distributing geospatial data raises some unique concerns, including information privacy, security issues, complex and unsettled legal issues related to intellectual property rights, and preservation challenges. In formulating data management and distribution policies, libraries or other organizations entering into data distribution arrangements with data providers are well advised to consider the main components of data-sharing and distribution policies described here and to identify those that are most important and relevant to them. This should be

done with an eye toward the library's level of commitment to maintaining the various components of a data distribution system. CUGIR, for example, provides a fairly high level of service in the area of metadata preparation and consulting. Data distributors who choose not to commit that much staff time to metadata development may elect to have strict requirements that all data providers supply the distributor with standards-compliant metadata and provide no additional enhancements or processing. In general, whether in the form of a legal contract or a less formal policy, a well-thought-out data management policy can clarify the expectations of participants, guard against future misunderstandings, and provide stability and predictability in transactions between participants.

Acknowledgements

Special thanks to Kathy Chiang, Jon Corson-Rikert, Anne Kenney, Jeff Piestrak, and Kornelia Tancheva for providing helpful comments on an earlier version of this article, and to the CUGIR working group: Jon Corson-Rikert, Keith Jenkins, Jeff Piestrak, and Elaine Westbrooks.

References

Baker, J. C. (2004). *Mapping the risks: Assessing homeland security implications of publicly available geospatial information.* Santa Monica, CA: Rand Corp. Retrieved October 23, 2005, from http://www.rand.org/publications/MG/MG142/MG142.pdf.

Barrington Consulting Group. (2005). *GeoNOVA exchange agreement template.* Retrieved November 20, 2005, from http://gov.ns.ca/GeoNova/pdf/GeoNOVA_Exchange_Agreement_template.pdf.

Brown, D. L., Welch, G., & Cullingworth, C. (2005). *Archiving, management and preservation of geospatial data: Summary report and recommendations.* Retrieved November 20, 2005, from http://www.geoconnections.org/programsCommittees/proCom_policy/keyDocs/geospatial_data_mgt_summary_report_20050208_E.pdf.

Center for International Earth Science Information Network (CIESIN). (2005). *Guide to managing geospatial electronic records.* New York: Columbia University.

Charlevoix County GIS Program. (2004). *Charlevoix County intergovernmental digital geographic data sharing agreement.* Retrieved November 20, 2005, from http://www.charlevoixcounty.org/downloads/chxcounty_data_sharing_agreement.pdf.

Cho, G. (2005). *Geographic information science: Mastering the legal issues.* Hoboken, NJ: Wiley & Sons.

Committee on Licensing Geographic Data and Services. (2004). *Licensing geographic data and services.* Washington, DC: National Research Council, Committee on Licensing Geographic Data and Services.

Consultative Committee for Space Data Systems. (2002). *Reference model for an Open Archival Information System (OAIS).* Washington, DC: CCSDS Secretariat. Retrieved November 20, 2005, from http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf.

County of Hunterdon, New Jersey, Division of Geographic Information Systems. (n.d.). *Spatial data distribution agreement.* Retrieved November 9, 2005, from http://gis.co.hunterdon.nj.us/hunterdon/Agreement.htm.

CUGIR Work Group. (2003). *Collection policy: Cornell University Geospatial Information Repository (CUGIR).* Retrieved November 29, 2005, from http://cugir.mannlib.cornell.edu/CUGIRCollectionPolicy.20030423.pdf.

CUGIR Work Group. (2005a). *CUGIR data management and distribution policy.* Retrieved November 29, 2005, from http://cugir.mannlib.cornell.edu/CUGIRpolicy.pdf.
CUGIR Work Group. (2005b). *Security assessment procedure.* Retrieved November 29, 2005, from http://cugir.mannlib.cornell.edu/CUGIRSecurityAssessment.pdf.
Dangermond, J. (1995). Public data access: Another side of GIS data sharing. In H. J. Onsrud & G. Rushton (Eds.), *Sharing geographic information* (pp. 331–339). New Brunswick, NJ: Center for Urban Policy Research.
Dawes, S. S., & Oskam, S. (1999). The Internet, the state library and the implementation of statewide information policy: The case of the NYS GIS Clearinghouse. *Journal of Global Information Management, 7*(4), 27–33.
de Sherbinin, A., & Chen, R. S. (2005). *Global spatial data and information user workshop: Report of a workshop.* Retrieved November 27, 2005, from http://sedac.ciesin.columbia.edu/GSDworkshop/GlobalDataWorkshop_report_web.pdf.
Environmental Systems Research Institute, Inc. (ESRI). (n.d.). *Geography network participant agreement.* Retrieved November 26, 2005, from http://www.geographynetwork.com/publishing/index.html.
Federal Geographic Data Committee (FGDC). (1998). *FGDC policy on access to public information and the protection of personal information privacy in federal geospatial databases.* Retrieved November 26, 2005, from http://www.fgdc.gov/policyandplanning/privacy%20policy.
Federal Geographic Data Committee (FGDC). (2000). *Content standard for digital geospatial metadata workbook* (Version 2.0). Retrieved November 28, 2005, from http://www.fgdc.gov/metadata/documents/workbook_0501_bmk.pdf.
Geospatial One-Stop. (n.d.). *Responsibilities of a publisher.* Retrieved November 20, 2005, from http://gos2.geodata.gov.
Global Biodiversity Information Facility (GBIF). (n.d.a). *Data sharing agreement.* Retrieved November 20, 2005, from http://www.gbif.org/DataProviders/Agreements/DSA.
Global Biodiversity Information Facility (GBIF). (n.d.b). *Data use agreement.* Retrieved November 20, 2005, from http://www.gbif.org/DataProviders/Agreements/DUA.
Global Biodiversity Information Facility (GBIF). (n.d.c). *Guiding principles regarding intellectual property rights.* Retrieved November 19, 2005, from http://www.gbif.org/DataProviders/Agreements/GBIFdataIPRprinciples.html.
Governor's Task Force on Information Resources Management Technology. (1997). *Governor's Task Force on Information Resources Management Technology policy 97-6.* Retrieved November 20, 2005, from http://www.oft.state.ny.us/policy/tp_976.htm.
Harvey, F. (2003). Developing geographic information infrastructures for local government: The role of trust. *Canadian Geographer-Geographe Canadien, 47*(1), 28–36.
Herold, P. (1997). *Maps and legends: Plotting a course for geographic information systems.* Retrieved November 28, 2005, from http://www.ala.org/ala/acrlbucket/nashville1997pap/herold.htm.
Hyland, N. C. (2002). GIS and data sharing in libraries: Considerations for digital libraries. *INSPEL, 36*(3), 207–215.
International Organization for Standardization. (2003). *Geographic information, metadata* (1st ed.). Genève, Switzerland: Iso.
Joffe, B. A. (2003). *Model data distribution policy.* Retrieved November 25, 2005, from http://www.opendataconsortium.org/documents/Data_Policy-4b.pdf.
Longhorn, R. A., Henson Apollonio, V., White, J. W., & International Maize and Wheat Improvement Center. (2002). *Legal issues in the use of geospatial data and tools for agriculture and natural resource management: A primer.* Mexico: CIMMYT.
Macomb County (MI) GIS Services Division. (2002). *Intergovernmental data sharing agreement for Macomb County digital geographic data sets.* Retrieved November 19, 2005, from http://macombcountymi.gov/gis/Documents/intergovernmental_license.pdf.
Martindale, J. (2002). National security and access to GIS data via the Internet: The Cornell University Geospatial Information Repository (CUGIR). *Proceedings of the Annual ESRI Education User Conference, San Diego, CA.* Retrieved October 23, 2005, from http://gis.esri.com/library/userconf/educ02/pap5165/p5165.htm.
Masser, I. (2005). *GIS worlds: Creating spatial data infrastructures.* Redlands, CA: ESRI Press.
McGlamery, P. (1995). Libraries as institutions for sharing. In H. J. Onsrud & G. Rushton (Eds.), *Sharing geographic information* (pp. 319–330). New Brunswick, NJ: Center for Urban Policy Research.

Meredith, P. H. (1995). Distributed GIS: If its time is now, why is it resisted? In H. J. Onsrud & G. Rushton (Eds.), *Sharing geographic information* (pp. 7–21). New Brunswick, NJ: Center for Urban Policy Research.

MetroGIS. (2004). *Regional parcel data sharing and distribution agreement for public parties between the Metropolitan Council and the counties of Anoka, Carver, Dakota, Ramsey, Hennepin, Scott, and Washington.* Retrieved November 20, 2005, from http://www.metrogis.org/about/history/agreement_3rd.pdf.

New York State Office of Cyber Security and Critical Infrastructure Coordination. (2005). *The New York State Geographic Information Systems (GIS) cooperative data sharing agreement for use with local governments of New York State and not-for-profit entities.* Retrieved November 19, 2005, from http://www.nysgis.state.ny.us/coordinationprogram/cooperative/agreement.cfm.

North Carolina Center for Geographic Information and Analysis (CGIA). (n.d.). *Memorandum of agreement between <Community>, North Carolina and State of North Carolina Center for Geographic Information and Analysis (CGIA) to enable and advance the sharing of strategic geospatial data resources and associated documentation between the agencies and among their data users.* Retrieved November 19, 2005, from http://cgia.cgia.state.nc.us/gicc/cdsa/moa.pdf.

Office of Management and Budget. (1996). *Circular no. A-130—Transmittal Memorandum no. 4.* Retrieved November 1, 2005, from http://www.whitehouse.gov/omb/circulars/a130/a130trans4.html.

OMB Watch. (2003). *NY State confidential memorandum re: agency sensitive information January 17, 2002.* Retrieved October 23, 2005, from http://www.ombwatch.org/info/2001/NYSinventory.html.

Onsrud, H. J. (1999). Liability in the use of GIS and geographical datasets. In P. Longley, M. Goodchild, D. Maguire, & D. Rhind (Eds.), *Geographical Information Systems: Management issues and applications* (pp. 643–652). New York: John Wiley & Sons.

Onsrud, H. J., & Lopez, X. R. (1998). Intellectual property rights in disseminating digital geographic data, products, and services: Conflicts and commonalities among European Union and United States approaches. In P. A. Burrough & I. Masser (Eds.), *European geographic information infrastructures: Opportunities and pitfalls* (pp. 153–167). London: Taylor & Francis.

RLG-NARA Task Force on Digital Repository Certification. (2005). *An audit checklist for the certification of trusted digital repositories.* Mountain View, CA: Research Libraries Group (RLG). Retrieved December 21, 2005, from http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf.

RLG-OCLC Working Group on Digital Archive Attributes. (2002). *Trusted digital repositories: Attributes and responsibilities.* Mountain View, CA: Research Libraries Group (RLG). Retrieved November 26, 2005, from www.rlg.org/legacy/longterm/repositories.pdf.

Somerset County, New Jersey. (n.d.). *Digital data sharing agreement.* Retrieved November 19, 2005, from http://www.opendataconsortium.org/documents/SomersetCo_NJ_DSA_Form.pdf.

University of Michigan School of Natural Resources and Environment. (2003). *MRI data sharing agreement.* Retrieved November 19, 2005, from http://rivers.snre.umich.edu/mri/mrishare.htm.

USCGRP Data and Information Working Group. (2002). *USGCRP DIWG data guidelines.* Retrieved November 25, 2005, from http://globalchange.gov/policies/diwg/diwg-guidelines.html.

VanWey, L. K., Rindfuss, R. R., Gutmann, M. P., Entwisle, B., & Balk, D. L. (2005). Confidentiality and spatially explicit data: Concerns and challenges. *Proceedings of the National Academy of Sciences of the United States of America, 102*(43), 15337–15342.

Wyoming Geographic Information Advisory Council (WGIAC). (2000). *Spatial technology and Geographic Information System policy (Draft).* Retrieved November 20, 2005, from http://wgiac2.state.wy.us/files/other/revised.pdf.

Zimmerman, C. A. (2002). *Sharing data for public information: Practices and policies of public agencies.* Retrieved November 13, 2005, from http://www.ops.fhwa.dot.gov/Travel/DatShare.htm.

Gail Steinhart is Environmental Sciences and GIS Librarian at Albert R. Mann Library, Cornell University. She is the coordinator of the Cornell University Geospatial Information Repository (CUGIR) work group and provides instruction and support for GIS users. Prior to joining the staff of Mann Library, she worked in environmental research for fourteen years.

Geospatial Web Services and Geoarchiving: New Opportunities and Challenges in Geographic Information Services

Steven P. Morris

Abstract

Over the course of the past fifteen years the role of Geographic Information Systems (GIS) has changed significantly. Initially the role of the map library was confined to that of building and providing access to collections of hard copy maps and imagery. Later, digital data, whether on CD-ROMs or network based, was added as a new type of resource within that collection and service model. By the late 1990s some academic libraries began to take on a Web map server role, providing interactive Web mapping access to collections of digital geospatial data. In the new era of distributed, interoperable map services, libraries will have an opportunity to explore new roles as portals to streaming content available in the form of geospatial Web services. At the same time, the increasingly ephemeral nature of digital geospatial content will make even more critical the need to address the long-term digital preservation challenges that are facing geospatial content.

This article focuses on two major geographic information issues facing academic libraries as well as libraries in general. First, what role should libraries play in the development and utilization of emerging geospatial Web services? Second, how should libraries address the challenge of long-term preservation of digital geospatial data in light of a shift to distribution methods that make the content ever more ephemeral?

Introduction

Over the course of the past fifteen years the role of Geographic Information Systems (GIS) has changed significantly. Initially the role of the map library was confined to that of building and providing access to collections

LIBRARY TRENDS, Vol. 55, No. 2, Fall 2006 ("Geographic Information Systems and Libraries," edited by Jaime Stoltenberg and Abraham Parrish), pp. 285–303

of hard copy maps and imagery. Later, digital data, whether on CD-ROMs or network based, was added as a new type of resource within that collection and service model (Journal of Academic Librarianship, 1995, 1997). By the late 1990s some academic libraries began to take on a Web map server role, providing interactive Web mapping access to collections of digital geospatial data. In the new era of distributed, interoperable map services, libraries will have an opportunity to explore new roles as portals to streaming content available in the form of geospatial Web services. At the same time, the increasingly ephemeral nature of digital geospatial content will make even more critical the need to address the long-term digital preservation challenges that are facing geospatial content.

This article will focus on two major geographic information issues facing academic libraries as well as libraries in general. First, what role should libraries play in the development and utilization of emerging geospatial Web services? Second, how should libraries address the challenge of long-term preservation of digital geospatial data in light of a shift to distribution methods that make the content ever more ephemeral? Specific experiences with engaging geospatial Web services and with instituting preservation-focused action responses will be drawn from the North Carolina State University (NCSU) Libraries data services program and the North Carolina Geospatial Data Archiving Project, a cooperative effort with the Library of Congress and the NC OneMap Initiative.

Brief Overview of Digital Geospatial Data Services in Academic Libraries

There are many components of academic library GIS services. At the core is the data collection, but accompanying the data is a mix of services that vary from campus to campus. A brief summary of typical service components follows.

Data Collections

Libraries acquire, license, catalog, make discoverable, archive, and carry out value-added processing on digital geospatial data. While, in the United States at least, much data is available in the public domain, the data is not always organized or readily accessible in such a way as to allow the user to easily sort through the wide range of data options available, and effort is required to make such freely available data discoverable. Furthermore, in order to improve data availability it is sometimes necessary to acquire and license additional commercial or fee-based government data for use. In some cases libraries also engage in large-scale value-added work—retiling, projecting, or otherwise converting and reorganizing data resources into a more convenient form for the libraries' target audience.

Data Discovery Tools and Support

Libraries support the discovery, selection, and use of geospatial data. While the most common form of promoting access to data collections has been the development of Web documentation for data collections, in some cases searchable databases of geospatial metadata are also made available. Data resources may also be included in the library's catalog, but the catalog is not usually the most effective vehicle for exposing or searching for digital geospatial data.

Technical Support

The line between providing reference support for finding and selecting data and providing actual technical support for using the data is a blurry one, and it has become more common for academic libraries to play a prominent role in providing technical support to GIS users. At NCSU, for example, the library holds one of four "right to call" spots for the campus Environmental Systems Research Institute (ESRI) site license and provides technical support as needed to campus users. Libraries also play varying roles in supporting campus software licenses, facilitating distribution of software, and troubleshooting installations.

Workshops and Training

As an extension of reference and technical support many libraries offer workshops on a variety of topics such as introductory GIS, data discovery, or use of specific software tools. The mix of workshops offered often reflects the sort of reference and technical support demands placed on the library. Increasingly, in-library workshops have now been complemented by and even supplanted by online training resources. At NCSU, for example, the library supports over 600 registrations per year for the ESRI Virtual Campus online courses.

Marketing and Outreach

Another academic library function, which goes hand-in-hand with workshops, is marketing and outreach—promoting geospatial resources and services to the campus community. GIS activity typically initially takes root in one or just a few core departments where there is a high level of activity and support. Meanwhile, latent demand exists in a broad range of academic disciplines where awareness of geospatial tools and resources is lacking, or where there is a perceived barrier to entry in terms of lack of access to tools, data, training, and support. Libraries, as a neutral space focused on customer service, are well positioned to cultivate new GIS users by promoting the use of geospatial tools and content and by providing ready access to software, data, training, and support. At NCSU the number of academic departments engaged in GIS grew from fewer than ten to thirty-five within just a few years as a result of combined campus and library efforts to develop a campus GIS infrastructure.

Evolution of Technical Approaches to Delivering Geospatial Data

The manner in which libraries have provided access to geospatial information has changed significantly in recent years, with analog map and image offerings increasingly being supplemented by or replaced by digital resources. At NCSU campus-wide networked access to data was initiated in 1993, with data made available both for download and for use online from GIS workstations in a networked environment. By 2000 one began to see more libraries offering Web mapping services, making the GIS content available to a much broader audience, including those who otherwise lacked the skills, software, and data access ordinarily needed to utilize GIS content.

The Early Library Experience with Web Mapping

While the Web mapping approach was initially fruitful—and still is in some contexts—these library-based map servers have increasingly risked becoming liabilities to the extent that volatile state and local content is included. State and local agency data producers are typically better positioned to manage data updates, and the number of available state and local map servers has risen steadily since 2000. In North Carolina, for example, the number of county map servers increased from 15 in 2000 to 77 out of 100 counties in 2005 (NCSU Libraries, 2006a). User demand for county and city data is high because it is larger scale, more detailed, more current, and more accurate than state and federal alternatives. Furthermore, many resources, such as cadastral data, zoning, and building footprints, tend to be available only at the local level. Meeting real user demand for data has increasingly required that local content be made available, yet the rate of update of that data has made it increasingly unfeasible to integrate and successfully update such content within library-based Web mapping services. The existence of stale data hosted on library servers, coupled with concern some data producers have about liability issues, have made the library Web map servers an increasingly untenable option. At NCSU Web map services, which began in 1997, were ceased in 2001 in deference to emerging state and local map services.

Data Interoperability and Emerging Geospatial Web Services

By the year 2000 producer-operated map servers were proliferating, but these emerging federal, state, and local map servers remained data islands that could be viewed only in isolation from one another. There was no way to zoom in and see federal, state, and local content together for a particular location. There was also no easy way to view adjacent county or municipal services in a side-by-side manner. Social, environmental, and economic processes did not stop at county borders, but local map services did.

Around the same time, however, the various initiatives of the Open Geospatial Consortium (OGC) began to bear fruit and some key initial

steps toward data interoperability were made. Data interoperability is necessary to integrate disparate data resources; allow sharing of content; allow interoperability between resources in different formats, commercial software environments, and coordinate systems; and facilitate service chaining (Reichardt, 2005). A key initial OGC specification, the Web Map Server (WMS) 1.0 specification, was adopted in 2000 (OGC, 2004), and activities related to the Web Mapping Testbed led to a subsequent explosion in the development of WMS services (Doyle, 2000). Initiatives such as the National Map, at the national level, and NC OneMap, at the state level, helped to further the integration of federal, state, and—increasingly—local map services in a flexible interoperable environment. In North Carolina, for example, by virtue of extensive outreach carried out by the NC Center for Geographic Information and Analysis and the U.S. Geological Survey with their partners, the number of state and local WMS services in the state grew from two in 2002 to seventy-four in February 2006 (NCSU Libraries, 2006a; NCGICC, 2006a). As a standalone system as well as a component of the National Map, NC OneMap provides services in the context of statewide needs while also feeding content directly into the National Map system (NCGICC, 2003).

The rapid growth in availability of geospatial Web services has been followed by the development of new services focused on geospatial Web service discovery and integration. Initial work in ESRI's Geography Network, available from 2000, was followed by development of the National Map Catalog and later Geospatial One-Stop. At the same time, commercial geospatial Web services also began to proliferate, with offerings such as ArcWeb Services from ESRI and other commercial services from firms such as TopoZone. While such commercial mapping services initially took the form of noninteroperable Web mapping services, it has increasingly been the case that these offerings are interoperable services based on OGC interoperability specifications, Simple Object Access Protocol (SOAP), or Application Programming Interfaces (API) that support application integration.

The Attraction of Geospatial Web Services

Geospatial Web services are potentially attractive to libraries and their users for a number of reasons:

- The services are available in a time and location independent manner.
- Access to extremely large datasets is possible even over low bandwidths.
- The most current data is readily available and data update does not require local maintenance action by libraries or other intermediate information providers.

- Differences in native formats and coordinate systems can become less of a barrier to use.
- Access to data can be more efficiently offered for regions where demand does not merit static data purchase.

Map services can be used in a broad range of situations, from complex projects involving application fusion to rather basic one-off uses. For example, one of the common uses of paper maps seen in the map library is that of tracking down the locations and coordinates of specific places on USGS topographic quad sheets. Since the late 1990s there have been a variety of commercial and public domain servers that allow users to examine topographic maps online, identify coordinates, and make annotations. There is no question that examining a topographic map by holding the large-format analog copy is preferable, from an ergonomic or aesthetic perspective, to looking at a smaller map area on the computer screen, or that many map analysis tasks can be more effectively carried out using large-format analog maps and images. Yet when one factors in issues such as convenience of access, travel time to the library, expanded resource availability, and other factors, the Web-based option becomes an attractive alternative for many map uses.

Drawbacks of Geospatial Web Services

There are of course many drawbacks to utilizing geospatial Web services or relying on them as a core information resource within geographic information services, including the following:

- Application performance when using Web services will frequently not match that which can be achieved using locally loaded data.
- Uptime reliability can be a problem, lead to service chain failures, and threaten project work.
- Some services are of a demonstration nature and can disappear without notice.
- While the content underlying Web services might be updated with some frequency, some applications may have a need to rely upon static, snapshot content for consistency in results and analysis.
- Screen-generated maps are aesthetically and ergonomically no match for large-format analog maps and images.

Geospatial Web services are clearly more useful in some situations than others, depending upon application and user requirements. These services are probably most useful when

- the user needs the most current data;
- the data is subject to frequent change;
- the user needs to make use of extremely large datasets, perhaps over lower bandwidth connections;

- the user wishes to preview the data prior to acquiring it;
- the user just needs the data for background use;
- the data needs to be integrated into remote or portable devices;
- the data is not otherwise available or cannot be efficiently acquired and stored for local use.

Integrating Geospatial Web Services into the Library Environment

Awareness of and promotion of these emerging Web services are still rather low both on the part of end-users and on the part of academic libraries. Integrating and managing access to services presents some problems that are very different from those associated with locally hosted content, including the following:

- Geospatial Web services have been difficult to discover and select from.
- In the case of commercial services, sustainable licensing models that work on a campus scale have yet to be worked out to satisfaction (problems include allowing for the volume of requests related to simple operations such as pan zoom, the ability to restrict access to authorized users, and anticipating an unknown volume of requests).
- Linking data resources with services that act upon them has been a sticky issue, with metadata standards and practices not adequately addressing the linkage of data resources with services that act upon them.
- Rights issues and approved use are in many cases ambiguous, with Web services in something of a “Wild West stage” (for example, it is not clear whether it is acceptable to extract data from ArcIMS services through ArcGIS connections; this is technically possible but not typically an intended use of the service).
- Integrating Web services into the physical browsing environment of the map collection in order to stimulate awareness of these new resources is tricky given the transient nature of such services.

With regard to the issue of physical browsing, while libraries have become increasingly if not overwhelmingly digital, the map room still provides a rich and effective browsing environment. While volatile resources such as Web services do not lend themselves easily to hard-coded representation on shelving or in map cases, emerging mobile device technologies might, in time, make it more feasible to integrate discovery and use of these resources within the context of the physical browsing environment.

Possible Library Roles vis-à-vis Geospatial Web Services

So what might academic libraries do to promote and facilitate access to geospatial Web services? Some possible roles might include facilitating discovery of services; producing new map services to fill the gaps in service availability; building new map portal services on top of existing map ser-

vices; licensing commercial Web services for use; and utilizing Web services consumption data to inform collection development planning.

Facilitating Discovery and Selection Libraries can support user discovery and selection of resources by incorporating such services into catalogs, GIS data collections, and the physical map room browsing environment. Just as libraries provide support in user selection of maps or datasets, support can also be offered in selecting from among competing service options. The notion of the reference interview, as it applies to geospatial data, can be extended to geospatial Web services.

The more traditional geospatial data-focused reference interview will tend to focus on content issues, a *subset* of which might include the following questions:

- Data extent: Does the data cover the study area as required?
- Thematic content: Does that street dataset have street centerlines or curbs and gutters?
- Attribute availability: Are there street addresses? Are they complete across the entire dataset? Is the format friendly to geocoding processes?
- Currency: How recently was the data produced? What real world time period does it represent? How concurrent is it with other data to be used in the project?
- Format: Is the data in a vector format that the project's software can support or at least convert without unacceptable data loss? In the case of imagery, has a level of compression been used that entails unacceptable data loss?
- Openness of licensing: Can the data be copied off of the CD-ROM or server? Can maps created from the data be used in publication? Can the data be used in a Web mapping application? Can a value-added derivative of the data be redistributed?
- Ease of access: Can the data be downloaded right now? In the case of very large datasets, is it possible to connect directly to the resources and use the data across the network? Is it possible to extract data for extremely large areas, or must one make numerous much smaller extractions to assemble data for the study area?
- Coordinate system, datums, etc.: Will it be necessary to re-project the data? Will a datum conversion be necessary? Is this information even recorded in the metadata?

In the Web service context, some of the content facets, such as format, can become less important, while some additional service or "functional" metadata come into play. These facets might include the following:

- Type of service: Image service, feature service, geocoding service, etc.
- Access protocol: ArcIMS image service, ArcIMS feature service, WMS, WFS (Web Feature Service), SOAP, and other methods such as the

Google Maps API. Is the service exposed through a protocol that is compatible with the user's technical environment?

- Reliability and uptime: Will downtime impact project work or service chaining? Is this a demonstration service that is liable to disappear at an inconvenient moment?
- Licensing or pricing scheme: How will trivial transactions such as pan and zoom count against overall service consumption costs? Can licensing effectively be extended to multiple, concurrent users within a constrained domain of authorized and authenticated users?
- In the case of image services, what image formats are offered (GIF, JPEG, PNG, etc.)?

Service discovery is available through the National Map catalog, Geospatial One-Stop (GOS), and regional services such as NC OneMap, but exhaustive, comprehensive access is still not available. The National Map Catalog covers a subset of the services available in GOS, and GOS covers a subset of all available services. The National Map Catalog exposes an API for application developers, raising the possibility of drawing from these service metadata collections to develop specific local catalogs (USGS, 2005). Other more extensive service catalogs are being developed, including the Naval Research Laboratory (NRL) GIDB Portal, which lists nearly 1,400 map servers and over 300,000 individual data layers (Naval Research Laboratory, 2006a), and Mapdex, which lists over 1,700 servers (Mapdex, 2006). The NRL is working on a searchable catalog system that will be compliant with the OGC Catalog Services Specification and will provide the capability to browse, search, and query using any OGC Catalog client application (Naval Research Laboratory, 2006b).

Providing Map Services Another possible library role lies in the area of helping to fill the holes in map service availability by, for example, serving up WMS layers that are not otherwise available. Rather than risk providing stale data that are better provided by the data producers, libraries might focus on serving out specific strategic content that users and other services could choose to consume. NCSU Libraries, for example, is deploying census data map services that will be integrated with the NC OneMap environment, helping to plug a hole in data availability within the statewide framework. Libraries, by virtue of their mission, might be more predisposed than other organizations to serve out lower-demand older or archival content that is not served up by data producers, who may tend to focus on the highest-demand, most current data.

Map Portals and Cascading Map Services Libraries may also have a role to play in deployment of the next-generation version of the old map server: setting up map servers that draw from and build on top of multiple existing map services, thereby creating single map interfaces. The USGS National Map at the national level and NC OneMap at the state level are two promi-

nent examples of cascading map services. In general, one of the things libraries try to do is build windows to the world of information where the window is orientated in a way that best suits the library's client base, often resulting in a particular geographic focus. In the case of map services, this notion might be translated into building specialized views that integrate existing map services.

In practice there are many complicated issues involved with setting up cascading map services: services adhere to different versions of OGC standards, use different symbolization, apply different scale restrictions, and name their data in different ways (for example, land parcels versus cadastral or property boundaries). Also, metadata that is needed to properly integrate resources may be missing, and rights issues concerning services are often ambiguous.

Building an effective cascading map service often becomes an exercise in community building that the technical interoperability specifications do not themselves address. Service builders must work closely with data providers to standardize service characteristics such as symbolization, classification schemes, scale thresholds, and layer naming. The relevant community must agree to and promote a set of practices that go beyond whatever requirements the actual standards or specifications might impose, as has been illustrated in the NC OneMap experience of developing a statewide integrated set of services (NCGICC, 2006c). The reality is that federal and state agencies participating in spatial data infrastructure are usually better positioned to carry out the community-building process.

Licensing Commercial Web Services Another opportunity for libraries lies in the area of licensing fee-based services for use by patrons. Such services may offer more than just content, with functions such as geocoding and routing being offered by emerging commercial services. Key challenges lie in the area of working out effective licensing models and in integrating campus identification and authorization schemes with these commercial products.

Using Web Services Consumption as a Measure Demand Another possible use of geospatial Web services is in the measurement of data demand associated with a library's user market. Development of digital geospatial data collections that fit the spatial demand footprint of the library's audience can be a challenging task. Funds are limited, and only so much data covering so much territory can be acquired and managed. To the extent that content exists, user demand can be measured based on data downloads by region, but if data holdings do not exist for given areas then demand cannot easily be assessed. It might be possible to carry out more rigorous market analysis if, for example, libraries were able to obtain zoom-in density maps from aggregated data reflecting their institution's traffic on national portals such as Geospatial One-Stop and the National Map.

Web Mashups, Geo-Hacking, and the New Geospatial Frontier

Geospatial Web services in the form of Web map servers, OGC services, and SOAP services have grown rapidly in the past five years, but these services have for the most part only been exposed to a confined market of geospatial data users. Geospatial Web services are entering a new phase of wider use with the availability of new, more mainstream services such as Google Maps, Google Earth, MSN Virtual Earth, and Yahoo Maps. APIs have made it possible for third-party developers and the general public to build applications on top of these offerings, which are more accessible than traditional geospatial industry offerings. These new services have experienced rapid growth in use since their inception in early 2005, with a vast new audience of "geo hackers" without traditional GIS backgrounds beginning to work with geospatial content and creating "web mashups" or "map mashups," which integrate content from multiple, distributed environments using AJAX (Asynchronous Javascript and XML) and other technologies ("Mashing the Web," 2005). The explosion during 2005 of creative activity on top of these services is likely to be just the beginning of a revolution in how geospatial content is used and republished.

The geospatial content available in these environments is still limited, with only a very, very small slice of all available geospatial content exposed for use with these systems; however, holes are being poked and then widened through the walls that separate the new commercial Web mapping realm from the much more content-rich traditional GIS realm as developers create tools that integrate WMS or WFS services with Google Maps (Flood, 2005; Mulka, 2005) or convert traditional geospatial data to Keyhole Markup language (KML) for integration with Google Earth (Martin, 2005). The new mainstream mapping space has a very large audience, and yet only a relatively small proportion of available data is exposed to these environments. Meanwhile, in the traditional geospatial industry space there is a relatively small audience and a very large amount of data available. As these two information spaces begin to connect and merge, a number of new opportunities are likely to emerge for libraries.

One very immediate impact of Google Maps, Google Earth, and the like will be the creation of a much larger audience and market for geospatial information resources. While those doing Web mashups are often commercial information technology developers, they are also often members of the general public developing maps for their churches, schools, or community groups. Mainstream developers crossing over to geospatial systems, while initially naïve on the topic of data quality, are developing a more sophisticated understanding of the qualitative differences between data alternatives and are seeking guidance from others in the selection of data sources for integration. One opportunity for libraries will be in the area

of exposing archived content to the Web mashup environment for before-and-after and time-related uses, as the emerging services currently focus, for the most part, on delivery of only the most current data.

Digital Preservation Challenges in the Web Services Era

While the emergence of geospatial Web services has opened up a number of opportunities for libraries, a significant threat is also posed. History has shown that it is quite often secondary archives that preserve content over long periods of time rather than the original content producers. For example, libraries typically preserve books rather than publishers. Until recently, in order to provide efficient access to content it has been necessary to physically acquire the data in order to make it available to users. As a result, data archives have often evolved as a somewhat accidental by-product of the process of providing access. With the emergence of Web services it will be much easier to point to the data source and avoid handling data and storage media altogether. This is convenient for the user and eases the burden on the library, but who then archives and preserves the data? If preservation of digitally born resources was already a problem before the advent of Web services, the shift to new distribution efforts will require an even more focused and intentional effort on the part of libraries to preserve data.

Many GIS professionals will readily admit that retention of older content is often very low on the list of priorities. "Kill and fill" is often the operating archival strategy. In the early years of GIS it may have made more sense to ignore the temporal component of geospatial data resources: there was little older content so time series analysis was out of the question anyway, barring massive and expensive vector digitizing of old maps. Most GIS projects are focused on problems that require the use of the most current data. Issues of convenience also undermine demand for older content: the fact that students, during their formative training, will tend to build class projects around available data perhaps reinforces the inclination to focus on more current content and topics.

Yet there is increasing evidence of a rise in demand for older content and of interest in doing associated temporal analysis. GIS has been in use for decades now, so users—especially younger users—are starting to expect that older content will exist. More projects are focusing on time series components—looking backwards at land use change and looking forward at business trends, for example. As GIS becomes more of a core enterprise resource at the local levels, the stakes are raised vis-à-vis accountability for the disposition of taxpayer-funded data development work.

Early Geoarchiving Efforts at NCSU

Geoarchiving is one term that has been used to describe the problem of preserving digital geospatial content (Maine GeoArchives, 2004). In 2000

the coincidence of emerging local agency data, rising user demand for that data, and a growing sense of long-term risk to data sparked an NCSU project targeting county and city data for acquisition and archiving. One learning outcome of that project was a deeper understanding of the complexity of the process of identifying data resource availability across many counties and municipalities. Another learning outcome was an awareness that more efficient and effective data management processes were needed. It was surprisingly easy to turn on the "fire hose" that sent torrents of data into the library collection, but the "plumbing" to deal with all of the content that could be acquired needed to be developed.

The Need for an Infrastructure-Based Approach to Preservation

This early archiving effort made it clear that a statewide infrastructure-based approach was required, one that would build from existing geospatial data infrastructures that were evolving under the auspices of the National Spatial Data Infrastructure (NSDI), Federal Geographic Data Committee, and Geospatial One-Stop. Two key developments in 2003 helped push NCSU Libraries' preservation effort to the next level: the NC OneMap Initiative and the National Digital Information Infrastructure and Preservation Program (NDIIPP), with its Cooperative Partnership Program.

NC OneMap Initiative In February 2003 the NC OneMap initiative was announced (NCGICC, 2003). NC OneMap is a combined state, federal, and local initiative that is focused on allowing users to view geographic data seamlessly across North Carolina; search for and download data for use on their own GIS; view and query metadata; and determine who has what data through an online data inventory (NCGICC, 2006b). Included in the NC OneMap vision statement was the assertion that "Historic and temporal data will be maintained and available" (NCGICC, 2003). While primarily focused on access and content standardization, NC OneMap has offered a scalable framework by which the 100 counties and many municipalities in the state might be engaged in the problem of preservation.

NDIIPP In August 2003 the Library of Congress put out a call for proposals in connection with a new congressionally funded initiative focused on preservation of digitally born content: the National Digital Information Infrastructure and Preservation Program. In this first funding round of the program, entitled "Building a Network of Partners: Collaborative Collection Development," the Library of Congress sought to engage with a diverse set of partners in a "dual effort to identify, get, and sustain significant material while also collaborating with the Library and the other partners to advance digital preservation methods and best practices" (Library of Congress, 2003). The eight selected projects address a range of content types, including Web pages, numeric social sciences data, business records, and cultural heritage resources (Library of Congress, 2006). One of the NDIIPP cooperative projects is the NC Geospatial Data Archiving Project

(NCGDAP), a partnership between NCSU Libraries, the NC Center for Geographic Information and Analysis, and NC OneMap (NCSU Libraries, 2006b).

North Carolina Geospatial Data Archiving Project

NCGDAP is focused on collection and preservation of digital geospatial data resources from state and local government agencies in North Carolina. The objectives of NCGDAP include

- identification of available resources through the NC OneMap data inventory;
- acquisition of at-risk geospatial data, including static data such as digital orthophotos as well time series data such as local land records and assessment data;
- development of a digital repository architecture for geospatial data using open source software tools such as Dspace;
- enhancement of existing geospatial metadata with additional preservation metadata using Metadata Encoding and Transmission Standard (METS) records as wrappers;
- investigation of automated identification and capture of data resources using emerging Open Geospatial Consortium specifications for client interaction with data on remote servers;
- development of a model for data archiving and time series development.

The project is operating under a three-year timeline from late 2004 to late 2007. Since the project is set within the context of an emerging Web services framework—NC OneMap and the National Map—the project is especially focused on responding to evolving data distribution methods and engaging emerging geospatial Web services in the archive development process.

Geoarchiving Challenges

Although the Web services aspects of the preservation problem are the focus of discussion here, a few salient issues related to the challenge of long-term preservation of digital geospatial data should be highlighted.

Geospatial Data Formats The absence of reliable, open vector formats is a stumbling block to preservation. SDTS (Spatial Data Transfer Standard), while open, has proven problematic and is not in wide use. The initial plan of the NCGDAP project involves retention of the data objects in the format received, while also exporting the content into a safer commercial vector format and buying time until a reliable, open alternative emerges. It is considered preferable to retain the content in a widely understood and supported commercial format rather than to rely solely on a migration of the content to an open format that may not be widely supported and con-

version to which may involve subjecting the content to some unfortunate transformations and data loss.

One thread of investigation involves the use of Geography Markup Language (GML) in an archival capacity. The challenge with this approach is that GML is not really a format per se but rather a means to define something akin to formats in the form of GML profiles and GML application schemas (Lake, 2005). The emerging Simple Features Profile for GML provides a potential solution in the form of a widely supported GML profile that is more sustainable over time, though quality and functionality tradeoffs against industry-specific GML application schemas will be a consideration (OGC, 2005a). NCGDAP will be participating in a broader effort by the National Archives and Records Administration (NARA) and the Federal Geographic Data Committee (FGDC) Historical Data Working Group to investigate the role of GML in preservation (FGDC, 2006).

Another area of investigation concerns mining of data inventories, possibly using the emerging RAMONA system being developed by the National States Geographic Information Council (Indiana GIC, 2006), to detect format "doppler signals" using a format's loss of market share as a possible indicator of format risk.

Geospatial Databases Another problem is the widespread emergence of complex spatial databases, either of the commercial variety or in more open varieties such as PostGIS-based systems. A spatial database stores geographic features and attributes as objects hosted inside a relational database management system. Multiple data layers may be stored in a single database, which may also host elements such as topology, relationships, behaviors, and annotations that are not exportable to conventional vector file formats. Within the project domain, the ESRI Geodatabase format is a prominent example of this approach to data management. Until recently spatial databases were relatively rare in the project domain, but local agencies—especially municipalities—are increasingly turning to the ESRI Geodatabase format in particular to manage geospatial data (NCGICC, 2004).

Preserving Cartographic Representation The true counterpart to the old, preserved map is not the current GIS dataset but rather the cartographic representation that builds on that data. The representation is the result of a collection of intellectual choices and application of current methods with regard to symbolization, classification, data modeling, and annotation. One goal of capturing cartographic representation will be to preserve data in the form that decision makers and others encountered and interpreted it. Another goal, in the case of image capture approaches, would be to provide a stable, preservation-friendly—though "dumbed down"—alternative in the case of long-term failure in the vector data preservation process. The derived image might also serve as a content preview, helping future researchers decide whether to commit time and resources to do whatever "digital archeology" (Ross & Gow, 1999) might be necessary to resurrect

the underlying content. Any preservation of cartographic representation should, ideally, occur in addition to preserving the underlying data.

In the Web services context, one issue to consider is that decisions will increasingly be made on the basis of ephemeral maps created online, making it difficult to document the basis for decisions. The OGC Web Map Context specification addresses the issue of saving the application state in order to re-create maps but does not address the issue of saving data state (OGC, 2005b).

Time-Versioned Content Many of the vector data layers to be acquired are subject to frequent update. County cadastral (land parcel) datasets, for example, are typically updated on a daily or weekly basis. Such time-versioned content, if preserved, can form the basis of time series analyses such as land use change analysis.

Version-handling over time, however, can be quite difficult to manage within the archive. And experience in the content domain has shown that some resources of only a few years of age have already lived in two or three repository environments, so any single repository cannot be expected to have all of the versions.

Content Packaging One of the points of frustration in working with geospatial content in a library context has been the absence of a packaging or bundling scheme for data. Geospatial data is characterized by complex multifile formats that need to be tied together, bundling data with associated metadata and ancillary documentation. Content packaging mechanisms may be used to bundle different versions of the dataset (by format, coordinate system, tiling scheme, etc.), to attach rights information and licensing, to supplement FGDC metadata with additional technical and administrative metadata, and to link objects with services that act upon them. The NCGDAP project will experiment with the use of the Metadata Encoding and Transmission Standard, a technology that has emerged in the library community, as a data bundling scheme. Other packaging schemes, such as the MPEG 21 Digital Item Declaration Language, are being considered in connection with the OGC Geo Digital Rights Management (GeoDRM) initiative (OGC, 2006).

Other Geoarchiving Challenges Other preservation challenges include securing and adequately defining archival and use rights for content; preserving semantic information associated with datasets; providing long-term support of coordinate systems and datums; and maintaining the independence of the preserved content from any particular repository software environment.

Putting Web Services to Work in Geoarchiving

While the shift toward Web services–based distribution of geospatial data may pose a threat to long-term preservation of content, it is also possible that those same geospatial Web services might in the future aid in the

onerous process of developing archives on the basis of widely distributed sources. Taking the example of North Carolina alone, there are 100 counties and over 140 municipalities. Nearly all North Carolina counties have GIS systems, as do many municipalities. Keeping track of data availability across this many agencies is not a trivial problem. Even more problematic is the task of routinely harvesting content from such a diversity of agencies.

In this context Web services become interesting from the point of view of automating inventory creation and automating extract and transfer of content. One of the difficult selling points for digital preservation has been the level of effort that must be applied to solve a problem that is very low on the list of priorities of data producers. If the process of archive development can be automated using Web services, then the barrier to participation in the preservation process might be lowered considerably.

Unfortunately, currently deployed services based on OGC specifications are not really fashioned to the needs of archive development processes. In terms of data transfer, WMS involves transfer of "dumb" images with the data intelligence removed. Web Feature Service (WFS), which involves transfer of the actual data as GML, is perhaps not really optimized for full-scale transfer of entire datasets or databases (OGC, 2005c). Furthermore, WFS is not yet widely deployed. What is lacking, so far, is a sort of rsync-like layer in the spatial data infrastructure that allows for efficient, full-scale replication of data resources while also being informed by data update processes, rights arrangements, and metadata. In cases where delta files—or change files—are used as a means of transferring database changes across the network, archival processes will need to handle conflation of the delta files with the archived database and certify that no delta files have been missed.

Conclusion

Geospatial Web services, which may be image services, feature services, geocoding services, or offer other functionality, are clearly on the rise. These new, dynamic resources are more useful for some applications than others, where access to static resources will continue to be more suitable. These services are notably difficult to discover, creating opportunities for libraries in the area of facilitating discovery of and access to them. The rise of more mainstream map services such as Google Maps, through its API, appears to be leading toward a rapid growth in the use of geospatial data and services by a broader audience.

At the same time, digital geospatial data is becoming increasingly ephemeral. The challenges in preserving static geospatial data are already daunting as we face the issue of preserving proprietary formats and spatial databases, capturing time series snapshots, and preserving cartographic representation. The advent of geospatial Web services raises additional challenges to data preservation, as static files are replaced by dynamic, changing services. At the same time, new Web services technologies may offer some possibility

of making the process of archive development more efficient through the use of automated approaches.

References

Doyle, A. (2000). *Web mapping testbed.* Retrieved February 24, 2006, from http://www.intl-interfaces.com/wmt2/wms.html.

Federal Geographic Data Committee (FGDC). (2006). *Federal Geographic Data Committee (FGDC) Historical Data Working Group.* Retrieved February 24, 2006, from http://www.fgdc.gov/participation/working-groups-subcommittees/working-groups.

Flood, B. (2005) *Google Maps API and WMS servers.* Retrieved February 24, 2006, from http://www.spatialdatalogic.com/cs/blogs/brian_flood/archive/2005/07/11/39.aspx.

Indiana Geographic Information Council (GIC). (2006). *Beta RAMONA GIS inventory.* Retrieved February 24, 2006, from http://in.gisinventory.net.

Journal of Academic Librarianship. (1995). *21*(4).

Journal of Academic Librarianship. (1997). *23*(6).

Lake, R. (2005). *Application schemas help build the Geo-Web.* GeoWorld. Retrieved February 24, 2006, from http://www.geoplace.com/uploads/FeatureArticle/0502tt.asp.

Library of Congress. (2003). *Program announcement to support building a network of partners: Collaborative collection development.* Retrieved February 24, 2006, from http://www.digitalpreservation.gov/partners/pa_081203.pdf.

Library of Congress. (2006). *Digital preservation partnerships.* Retrieved February 24, 2006, from http://www.digitalpreservation.gov/partners/project.html.

Maine GeoArchives. (2004). *GeoArchives project agenda.* Retrieved February 24, 2006 from http://www.maine.gov/geoarch/agendaminutes/11_15_04_Agenda_files/11_15_04_Agenda.htm.

Mapdex. (2006). *Mapdex.* Retrieved February 24, 2006, from http://www.mapdex.org/search.

Martin, K. (2005). *Export to KML 2.1.* Retrieved February 24, 2006, from http://arcscripts.esri.com/details.asp?dbid=14273.

"Mashing the Web." (2005). *Economist 376*(8444), Special section, p. 4.

Mulka, K. (2005). *WMS in Google maps.* Retrieved February 24, 2006, from http://blog.kylemulka.com/.

Naval Research Laboratory. (2006a). *GIDB OpenGIS Web services.* February 24, 2006, from http://columbo.nrlssc.navy.mil/ogcwms/servlet/WMSServlet.

Naval Research Laboratory. (2006b). *Digital mapping, charting and geodesy analysis program team.* Retrieved February 24, 2006, from http://columbo.nrlssc.navy.mil/dmap/idx.jsp.

North Carolina Geographic Information Coordinating Council (NCGICC). (2003). *North Carolina OneMap: Geographic data serving a statewide community.* Retrieved February 24, 2006, from http://www.nconemap.net/documents/visiondoc.pdf.

North Carolina Geographic Information Coordinating Council (NCGICC). (2004). *2003 statewide local government GIS data inventory.* Retrieved February 24, 2006, from http://cgia.cgia.state.nc.us/nconemap/inventory.html.

North Carolina Geographic Information Coordinating Council (NCGICC). (2006a). *Current NC OneMap participants.* Retrieved February 24, 2006, from http://www.nconemap.net.

North Carolina Geographic Information Coordinating Council (NCGICC). (2006b). *NC OneMap.* Retrieved February 24, 2006, from http://www.nconemap.net.

North Carolina Geographic Information Coordinating Council (NCGICC). (2006c). *NC OneMap: Map service, layer naming and symbology standards.* Retrieved February 24, 2006, from http://204.211.135.110/onemap_standards/.

North Carolina State University (NCSU) Libraries. (2006a). *North Carolina county GIS data.* Retrieved February 24, 2006, from http://www.lib.ncsu.edu/gis/counties.html.

North Carolina State University (NCSU) Libraries. (2006b). *North Carolina geospatial data archiving project.* Retrieved February 24, 2006, from http://www.lib.ncsu.edu/ncgdap/.

Open Geospatial Consortium (OGC). (2004). *OpenGIS Web Map Service (WMS) implementation specification 1.3.* Retrieved February 24, 2006, from http://portal.opengeospatial.org/files/?artifact_id=5316.

Open Geospatial Consortium (OGC). (2005a). *GML simple features profile: Request for comments*. Retrieved February 24, 2006, from http://www.opengeospatial.org/specs/?page=requests&request=rfc22.

Open Geospatial Consortium (OGC). (2005b). *OpenGIS Web map context implementation specification 1.1*. Retrieved February 24, 2006, from https://portal.opengeospatial.org/files/?artifact_id=8618.

Open Geospatial Consortium (OGC). (2005c). *Open Web Feature Service (WFS) implementation specification 1.1*. Retrieved February 24, 2006, from https://portal.opengeospatial.org/files/?artifact_id=8339.

Open Geospatial Consortium (OGC). (2006). *Geo Digital Rights Management (GeoDRM) working group*. Retrieved February 24, 2006, from http://www.opengeospatial.org/groups/?iid=129.

Reichardt, M. (2005). *The havoc of non-interoperability*. Retrieved February 24, 2006, from http://portal.opengeospatial.org/files/?artifact_id=5097.

Ross, S., and Gow, A. (1999). *Digital archaeology: Rescuing neglected and damaged data resources. A JISC/NPO study within the Electronic Libraries (eLib) Programme on the Preservation of Electronic Materials*. Retrieved February 24, 2006, from http://eprints.erpanet.org/47/.

U.S. Geological Survey (USGS). (2005). *The National Map catalog service: A guide for application developers, document version 0.3.3, for catalog service versions 1.1.0 and 2.0.0*. Retrieved February 24, 2006, from http://mcmcweb.er.usgs.gov/catalog/docs/catalog_api.pdf.

Steve Morris is Head of Digital Library Initiatives as North Carolina State University (NCSU) Libraries and has worked for the past dozen years in the area of facilitating access to geospatial data. He is principal investigator in a partnership with the Library of Congress focusing on preservation of digital geospatial data. Steve has a master's degrees in both geography and library and information studies.

Digital Preservation of Geospatial Data

Julie Sweetkind-Singer, Mary Lynette Larsgaard,
and Tracey Erwin

Abstract

The selection, acquisition, and management of digital data are now part and parcel of the work librarians handle on a day-to-day basis. While much thought goes into this work, little consideration may be given to the long-term preservation of the collected data. Digital data cannot be retained for the future in the same way paper-based materials have traditionally been handled. Specific issues arise when archiving digital data and especially geospatial data. This article will discuss some of those issues, including data versioning, file size, proprietary data formats, copyright, and the complexity of file formats. Collection development topics, including what to collect and why, will also be explored. The work underlying this article is being done as part of an award from the Library of Congress's National Digital Information Infrastructure and Preservation Program (NDIIPP).

Introduction

Digital geospatial data is now routinely found in libraries that carry cartographic data, geologic information, social science datasets, and other materials in support of disciplines using Geographic Information Systems (GIS) in their research and work. Over the course of years, the data have been received on floppy disks, CD-ROMS, DVDs, and hard drives or are available for free or for a fee over the Internet. In the paper world, ensuring longevity of items means creating ideal conditions in which to store collections. Materials will last longer if kept in a cool space without much light and correct humidity and handled as seldom as possible.

LIBRARY TRENDS, Vol. 55, No. 2, Fall 2006 ("Geographic Information Systems and Libraries," edited by Jaime Stoltenberg and Abraham Parrish), pp. 304–314

The same is not true for digital data. As Clay Shirky (of New York University's Interactive Telecommunications Program) pointed out in July 2005 at the bi-annual meeting of the National Digital Information Infrastructure and Preservation Program (NDIIPP), digital materials must be touched and manipulated on a regular basis if they are to survive. Leaving digital data alone will certainly cause it to be lost, and the time frame may be surprisingly short. Technology is changing at such a rapid pace that it can now be a challenge to find a machine that will read floppy discs, much less the obsolete program on which the data was supposed to run. Web sites can be and are removed at a moment's notice. This is especially frustrating for the federal depository libraries that formerly received paper copies of government information now available only in digital formats. Clearly, librarians must begin thinking about long-term preservation of their digital collection, from what to collect to ensuring that it is preserved with the same thoughtfulness and care that is given to hardcopy materials.

The Library of Congress and the NDIIPP Awards

In December 2000 Congress appropriated nearly $100 million dollars in funds to underwrite the cost of studying the issues related to the long-term preservation of digital data. The program was to be administered by the Library of Congress and was named the National Digital Information and Infrastructure Preservation Program (Library of Congress, 2006a). Conference Report H. Rept. 106–1033 stated that

> The overall plan should set forth a strategy for the Library of Congress, in collaboration with other Federal and non-Federal entities, to identify a national network of libraries and other organizations with responsibilities for collecting digital materials that will provide access to and maintain those materials. . . . In addition to developing this strategy, the plan shall set forth, in concert with the Copyright Office, the policies, protocols, and strategies for the long-term preservation of such materials, including the technological infrastructure required at the Library of Congress. (Library of Congress, 2006b)

The goal of the program was to create a network of committed partners willing to work on the policies, protocols, and architectures needed to build a series of archives to house digital materials.

The first round of major funding was announced in September 2004 with eight projects receiving a total of $13.8 million dollars in funding over a three-year period. Two of these projects focused specifically on geospatial data. The North Carolina State University Libraries partnered with the North Carolina Center for Geographic Information and Analysis to create a model for archiving the local and state government output of digital geospatial resources, including digitized maps. The project is designed to be a demonstration project for other states. The second contract was given jointly to the University of California at Santa Barbara

(UCSB) and Stanford University to underwrite the creation of the National Geospatial Digital Archive (NGDA). The NGDA's goal is to design repository infrastructures at each university and to collect materials across a broad spectrum of geographic formats. The team will work to expand the network of organizations committed to preserving geospatial content (Library of Congress, 2004).

The NGDA Project

The NGDA project has both research and development components. Research topics include considerations for long-term preservation; collection development, including prioritization and scope; architectural and economic models; rights issues; and best practices. The two libraries are developing prototype archives for housing the data and jointly creating a geospatial format registry to describe the data being stored. During the second year of the grant the two archives will be federated using the Alexandria Digital Library (ADL) software interface (see Figure 1).

Technical Architectures

The two repositories are being built using similar technologies while at the same time meeting the specific needs of each institution. Both architectures contain standards-based interfaces, clearly defined metadata formats, an underlying format registry, a goal of end-to-end automation of the systems, and exploration into open source front ends. UCSB has developed a repository specifically to house geospatial information, with tools and templates designed around common data structures. Stanford is building a repository to hold all of its digital content no matter what its nature; the goal is to determine if a general digital repository can adequately handle the complexities of geospatial data formats using standard metadata and a content transfer manifest, which include provisions for geospatial information. As of the end of December 2005, both repositories were complete through their first stages and had ingested geospatial data.

Format Registries

Technically, geospatial data is more complex than standard digital formats. This must be accounted for when archiving the data. In order to preserve a data format, information about that format must be known. The archive has to have an automated way to understand the file it has received and to verify that it is what it purports to be. This format information is typically stored in a registry, which records detailed metadata about the types of files. For example, format information for a GeoTIFF would include specifications for the correct TIFF standard and explanations of any accompanying files, such as those containing projection information. The format registry can be as complex as a custom-made database or as simple as a Web page or text document.

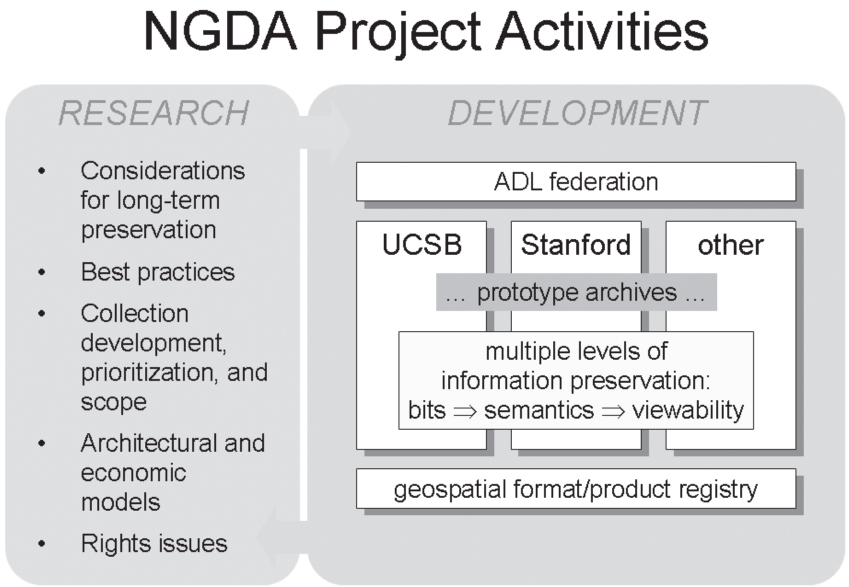


Figure 1. The overarching project activities for the three-year life of the NGDA contract with the Library of Congress

The Library of Congress, along with many other organizations, has spent a great deal of time describing formats and housing that information in format registries. Caroline Arms and Carl Fleischhauer compiled the format registry for the Library of Congress in order to help determine the sustainability of any given format throughout its content life cycle, to gain an understanding about which formats are more sustainable than others, and to develop strategies for sustaining the content they receive. The content categories studied include still image, sound, text, and moving image. They did not populate a format registry for geospatial data formats (Arms & Fleischhauer, 2005).

After searching unsuccessfully for other groups that had created a format registry for geospatial data, the NGDA team decided to build its own. Work is ongoing to describe the data elements necessary for preservation on four formats: digital orthographic quarter quadrangles (DOQQs), digital raster graphics (DRGs), Environmental Systems Research Institute (ESRI) "shapefiles," and Landsat imagery. (Other document formats will be analyzed as content selection for the archives progresses.) These file formats have differing levels of complexity. All of them contain multiple files that must travel together in order to make the format usable now and in the

future. For example, ESRI shapefiles are in a proprietary data format and are used in proprietary software. According to the specifications published by ESRI, only the .shp, .dbf, and .shx files are mentioned as part of the shapefile itself. But, with each shapefile there may also exist numerous other files, such as .sbn, .ain., and .prj files. Public documentation to reference these files and their role in the playability of the file itself is not available. Correspondence with ESRI staff was necessary to ascertain whether or not they considered the last three files necessary to preserve in the archive along with the published specification files. Building the format registry is labor intensive as it is necessary to trace the dependencies of files (a GeoTIFF must also include the correct TIFF specification, for example), and one must locally collect as much documentation as possible about each format. However, the set of format specifications should have to be created only once and then updated as necessary.

Rights Management and Contracts

Information regarding the rights governing the ownership, use, and copyright status of the data is associated with each file included in the repositories. A great deal of domestic geospatial data is produced by the U.S. government, which allows for wide use of its output due to the fact that most of it is in the public domain. But even government data may have copyright stipulations attached to it if it has been distributed through a third-party vendor contracted through a Cooperative Research and Development Agreement (CRADA) or has had value added to it by a commercial firm.

Important datasets, such as California's SPOT Image coverage and the base data on ESRI's Data & Maps CD-ROMS, are governed by strict licensing and use agreements. These datasets provide high-quality base map layers for GIS work and, especially with the yearly release of ESRI Data & Maps, provide longitudinal data that allow for the study of change over time. As the NGDA project moves forward, these agreements may make it impossible to federate data if the other potential repositories in the NGDA federation do not also have the legal right to hold the data. During the second year of the grant, the NGDA staff will begin a dialog with commercial data and imagery producers to assess their preservation strategies, awareness, and willingness to work with preservation archives.

In order to codify the rights and responsibilities of the repositories, each depositor will sign a contract licensing their content for preservation in the NGDA repositories. The goal is to create a single contract that can be used, and modified if necessary, by both Stanford and UCSB. The contract (in draft form as of this writing) governs the use, display, delivery, and preservation of materials in the NGDA. It clearly states who owns the copyright to the materials and ensures that those depositing materials in the archive have a right to do so. It further clarifies that the copyright stays with the original depositor and that the archives are not responsible in cases

of copyright infringement. It explains what may be distributed from the archives—the metadata, the data, or both—and to whom. It details how the repositories' rights and responsibilities will be carried out, including the need to use best practices and standards for preservation. The archives agree to take measures to prevent unauthorized access to the data, to permit only authorized users to access the content, to credit the copyright holders, and to use the utmost care in the preservation of the content. The contract explicitly allows the archives to manage the data to maximize its chances of survival over the long term.

In addition to legal protection for both parties afforded by the contract, a well-thought-out contract explicating the roles of each party builds an important element of trust that will encourage content creators to deposit their content in our repositories. The contract embodies one of the aspects of the trust-building activities recommended in the Research Libraries Group/Online Computer Library Center report, "Trusted Digital Repositories: Attributes and Responsibilities" (Research Libraries Group, 2002).

In order to further investigate how copyright law affects archiving of digital data, the Library of Congress has convened the Section 108 Study Group. Section 108 of the Copyright Act, created in 1976 and amended in 1998, governs the use of copyrighted materials held in libraries and archives. It is believed that even with the 108 revisions, the law is designed to meet the needs of the analog world, not the complex issues and needs of the digital one. This group has been charged with reviewing existing copyright laws as they pertain to libraries and archives, and specifically as they apply to digital media. The group will advise the Librarian of Congress in May 2006 on their findings and make recommendations based upon the needs of the content producers as well as those wishing to archive and access their output (Library of Congress, n.d.).

Collection Development

When the Library of Congress announced the Digital Preservation Program in August 2003, they enunciated the following three goals:

> "The continuing selection, collection, and organization of the most historically significant cultural materials and of important information resources, regardless of evolving formats,
>
> The long-term storage, preservation, and authenticity of those collections, and
>
> Persistent, rights-protected access for the public to the digital heritage of the American people." (Library of Congress, 2003)

Nature of Risk A required outcome of the project is to focus on materials that are deemed to be "at-risk" of disappearing or have no analog counterpart. While the Library of Congress did note that they considered historical and cultural materials or information "that document[s] key social and political developments necessary to understand contemporary

events" (Library of Congress, 2003) to be preservation worthy, they did not specifically define what it meant to have materials be "at risk." This is not surprising given the broad range of information across all disciplines in need of preservation.

Digital geospatial data may be deemed to be at risk because of many factors. The sheer magnitude of geospatial data being created and in existence makes it nearly impossible to collect it all for the future without significant efforts toward collaborative collecting models. MODIS data, used to study global dynamics and processes on the Earth, are being captured in thirty-six spectral bands from the MODIS satellites at a rate of a terabyte a day; over two petabytes of MODIS data are now stored at NASA. It is highly unlikely that a university, even the largest, would want to archive the whole MODIS output. On a more localized level, the problem of data storage is still significant. The state of California as represented in the Digital Orthophoto Quarter-Quadrangles includes approximately 13,200 scenes requiring roughly 670 gigabytes of storage space. In order to ensure viability of this dataset into the future, geographic redundancy is necessary in addition to the information being stored on different types of storage systems to lessen the chance of loss or corruption of data. This means the large datasets cannot be stored in a single location, creating the need for numerous, large, robust preservation environments.

In addition to the volume of data being produced, geospatial data are often updated and changed, creating the need to save different versions of the same information. How often the versions are collected will have to be decided on a case-by-case basis. For example, the National Elevation Dataset is updated on a bi-monthly basis by the United States Geological Survey (USGS) as higher-resolution or higher-quality data become available. Even a single data layer of a city GIS that is used by many different departments may be updated as often as several times a day. The different versions may be considered to be at risk because of the possibility that each iteration may need to be preserved (for example, for legal reasons, such as to prove when a change in a city's infrastructure was made). A strong argument can be made that each version need not be preserved in order to get a valid snapshot of the data environment.

Government geospatial data may well be considered at risk given the sensitive nature of some of the information, the decentralization of the computing environment, the lack of distribution of digital content that used to come to libraries as part of the Federal Depository Library Program, and the ease with which content can be removed from a government Web site. According to OMB Watch, the Bureau of Transportation Statistics (BTS) removed all GIS data, maps, and resources from their Web site after September 11, 2001. These data were later restored after the decision was made that their release did not pose a threat to national security (OMB Watch, 2005). Pipeline mapping data was removed from the Department of

Transportation's (DOT) Web site around the same time and has not been released to the public again. The DOT notes on their Web site that the data is now restricted "to pipeline operators and Local, State, and Federal government officials ONLY" (PHMSA, 2005).

Geospatial data is also potentially at risk for long-term preservation when it is produced by a small group or a single person. The ease with which content is now created and displayed has caused an explosion of small producers of high-quality geospatial content. Digital preservation requires a good deal of planning and expertise. It may also be prohibitively expensive to undertake. Simply making a backup copy of these data does not ensure that sufficient metadata has been captured to understand the environment in which the data was created, guard against failure of the storage mechanism, allow for geographic distribution, or solve the problem of file format migration over time. It is hoped that through the work done by this group and others the ability for small groups and individuals to archive safely their data for the long term will increase.

Collection Development Policies Collection development policies play a critical role in map libraries and have been important for many decades. The University of California/Stanford Map Libraries Group (UCSMLG) is still using the Research Library Group (RLG) conspectus portion for maps and geospatial data. The cooperative agreement is updated every five years and clearly spells out policies related to collaborative purchasing, collecting commitment levels for cartographic types of data and regions, and interlibrary loan. This agreement and the list of collecting responsibilities assigned to each university by call number have proved to be useful to this day (UC/Stanford Map Libraries Group, 2006).

Collection development policies typically do not include directives for long-term archiving of the collection itself. It has proven useful for us to review the work being done in the archiving community. While research libraries do, in general, keep their materials for a long period of time, they also weed with impunity for reasons of cost, space, and lack of use. An archive has made a commitment to keeping the material with the idea of turning it over to another trusted archive when they cannot or do not want to steward it any longer. In archives provenance is an integral component of responsible stewardship. Provenance details who or what group created and/or managed the records and traces the history of ownership of the records. This is critical information for a geospatial archive as well, and it must be included in the metadata. A good primer on archival practices is available from the University of Albany's M. E. Grenader Department of Special Collections and Archives (Parker et al., 2005).

Another area where archival practices influence long-term preservation is multiple file dependencies. Archival practices have codified the process of accessioning items in a specific order. This is necessary to preserve the contents as they were originally received and/or arranged. This is impor-

tant for a digital archive as well. Preserving such dependencies becomes critical when one thinks about long-term preservation of geospatial data incorporated on a Web site. The Arizona Model, being developed by the Arizona State Library, Archives, and Public Records, is using the framework of archival records management for the curation of collections of Web documents. They note that archiving Web documents by order translates into the correct management of the directories and subdirectories, which are called *series* and *subseries* in archival parlance. They argue that only through judicious use of archiving practices can large amounts of data be captured with a relatively small amount of human input. The system created must be able to scale and cannot do so if curators must select items one by one (Pearce-Moses & Kaczmarek, 2005).

The collection development policy for the geospatial archives being built by the NGDA will be a hybrid between a library policy and an archival one. It will include standard sections of a collection development statement that outline the user community; the geographic scope; the methods, scales, and frequency by which the materials are collected and updated; and the types of materials included. In addition, the policy statement will include descriptions of the type and quality of metadata that need to be included for ingestion into the repositories. Widely used file formats and types will be explained on a general level with the expectation that these will need to be updated over time. The Cornell University Geospatial Information Repository (CUGIR) has posted its collection development policy on its Web site, and it is a good example of this hybrid format (CUGIR WorkGroup of Mann Library, 2006).

The NGDA librarians will also produce specific collection development guidelines for their respective institutions. We hope that over time there will be many partner repositories in the federated network with broad collecting responsibilities. The individual nodes will focus on the needs of their primary audience, revising the policy to reflect individual institutional priorities. It is expected that areas of collecting interest will fall roughly along the same lines that were used when accessioning print materials. For example, the UCSB Map and Imagery Laboratory has a long history of collecting aerial and satellite imagery, while the Stanford Map Collections has focused heavily on geologic mapping and data. It is imperative that multiple collecting bodies be engaged in the process of selection and retention. There is just too much geospatial information being produced for a few libraries or institutions to preserve it all.

Conclusion

The National Geospatial Digital Archive team has completed the first of three years of their contract with the Library of Congress. Much more will be learned over the next two years from our research, the research of

other NDIIPP grants, as well as the work being done by others around the world in this field.

Year two goals include investigating to what degree, if any, commercial geospatial data producers are concerned with archiving and whether there is an interest in partnering with academic institutions; gaining a better understanding of existing mandates for archiving government-produced geospatial data; continuing to grapple with complex legal issues surrounding archiving; and ongoing technical development of the repositories themselves.

References

Arms, C., & Fleischhauer, C. (2005). *Sustainability of digital formats planning for the Library of Congress collections.* Retrieved January 24, 2006, from http://www.digitalpreservation.gov/formats/.

CUGIR WorkGroup of Mann Library. (2006). *About CUGIR: CUGIR collection development policy.* Retrieved January 31, 2006, from http://cugir.mannlib.cornell.edu/about.jsp.

Library of Congress. (n.d.). *The Section 108 working group.* Retrieved January 24, 2006, from http://www.loc.gov/section108/.

Library of Congress. (2003). *Program announcement to support building a network of partners: Collaborative collection development.* Retrieved January 24, 2006 from http://www.digitalpreservation.gov/partners/pa_081203.pdf.

Library of Congress. (2004). *Library of Congress announces awards of $13.9 million to begin building a network of partners for digital preservation.* Retrieved January 24, 2006, from http://www.loc.gov/today/pr/2004/04-171.html.

Library of Congress. (2006a). *Digital preservation: National Digital Information Infrastructure and Preservation Program.* Retrieved January 31, 2006, from http://www.digitalpreservation.gov.

Library of Congress. (2006b). *Funding legislation for NDIIPP.* Retrieved January 24, 2006, from http://www.digitalpreservation.gov/about/fund.html.

OMB Watch. (2005). *Access to government information post September 11th.* Retrieved January 31, 2006, from http://www.ombwatch.org/article/articleview/213/1/DOT.

Parker, J., Williams, G., Keough, B., & Schnidler, A. (2005). *M. E. Grenander Department of Special Collections and Archives: Accessioning and processing manual.* Retrieved January 31, 2006, from http://library.albany.edu/speccoll/processing.htm.

Pearce-Moses, R., & Kaczmarek, J. (2005). *An Arizona Model for preservation and access of Web documents.* Retrieved January 31, 2006, from http://www.lib.az.us/DigGovt/azmodel/AzModel.pdf.

PHMSA. (2005). *National Pipeline Mapping System: Data dissemination.* Retrieved January 31, 2006, from http://www.npms.rspa.dot.gov/data/data_dissem.htm.

Research Libraries Group, Inc. (2002). *Trusted digital repositories: Attributes and responsibilities, an RLG-OCLC report.* Retrieved July 11, 2006, from http://www.rlg.org/legacy/longterm/repositories.pdf.

UC/Stanford Map Libraries Group. (2006). *Cooperative agreements.* Retrieved January 31, 2006, from http://library.ucsc.edu/maps/ucsmg/agreemts.html.

Julie Sweetkind-Singer is the Head Librarian at the Branner Earth Sciences Library and Map Collections at Stanford University. Her subject specialization is maps and GIS. She is currently Stanford's project lead on an NDIIPP grant from the Library of Congress. She has worked at Stanford since May 2000. In 1999 she worked jointly with David Rumsey on the Rumsey Map Collection Web site, which displays over 12,000 maps from the eighteenth and nineteenth centuries. She was the President of the Western Association of Map Libraries from June 2004 to July 2005. She was the Vice-President of the California Map Society, Northern Chapter, in 2001 and 2002. She received her M.L.I.S. from San Jose State University and M.B.A. from the University of Colorado at Boulder.

Mary Lynette Larsgaard has been Assistant Head of the Map and Imagery Laboratory, Davidson Library, University of California at Santa Barbara, since 1988. The Map and Imagery Lab has a collection of remote-sensing imagery and maps of approximately 5.5 million items and is the largest of its kind in any university library in North America. Larsgaard has published extensively in the field of geospatial data in libraries, most notably with a widely used text, *Map Librarianship: An Introduction* (now in a third edition, published in 1998 by Libraries Unlimited); she is also a coeditor (with Paige Andrew) of the *Journal of Map and Geography Libraries/Geoscapes*. Her specialties are cataloging/metadata creation and twentieth-century and more recent topographic and geologic maps. In the year 2000 she was promoted to Librarian, Distinguished Step, a promotion given only to librarians who have demonstrated superior competence and are internationally recognized as an authority in an area of library science.

Tracey Erwin is a Geospatial Librarian for the National Geospatial Digital Archive (NGDA), a grant-funded project of the University of California at Santa Barbara and Stanford University. The NGDA is a collaborative initiative funded by the Library of Congress to identify, collect, and preserve geospatial digital materials within a nationwide digital preservation infrastructure. Tracey is a recent graduate of San Jose State University's Library and Information Science Program.

Building a Library GIS Service from the Ground Up

Rhonda Houser

Abstract

Geographic Information Systems (GIS) services in academic libraries tend to differ, based on availability of GIS data, software, hardware, and staff expertise. GIS services at the University of Kansas are closely aligned with support for government information, data, maps, and statistics. Thus, our responses to users' needs are often naturally collaborative, optimizing the expertise of multiple staff members and various types of resources. The GIS and Data Specialist assists campus researchers with spatial data and software, as well as facilitating access to GIS data. Lab space for research and coursework involving spatial data is a core component of GIS services. In addition, various levels and types of GIS workshops are offered each semester, and custom training sessions are also available. "Word of mouth" and hands-on workshops are some of the most effective methods of outreach.

Introduction

This article describes the development, current profile, and future directions of Geographic Information Systems (GIS) services at the University of Kansas (KU) Libraries. The intended audience of the article has at least a basic understanding of GIS and related terminology. Academic libraries considering implementing new GIS services or evaluating existing services may find this article helpful. KU Libraries supports 25,000 students and 1,300 faculty with academic and research resources. Library collections consist of approximately 3.7 million volumes and 33,000 current serial titles, housed in six facilities in Lawrence and one in Kansas City.

LIBRARY TRENDS, Vol. 55, No. 2, Fall 2006 ("Geographic Information Systems and Libraries," edited by Jaime Stoltenberg and Abraham Parrish), pp. 315–326

KU Libraries, Information Technology, and the Policy Research Institute founded a new committee several years ago called the Academic Data Research Services Alliance (ADRSA).[1] Members assist the campus research community with data identification, access, and analysis. ADRSA was instrumental in establishing the Library Geographic Information Systems (GIS) and Data Lab and in creating the GIS and Data Specialist position (hereafter referred to as GIS Specialist). GIS and Data Services is a transparent network of ADRSA partners and associated staff, which pulls from the expertise of staff in maps, data, GIS, government information, and statistics.

GIS and Data Lab

The GIS Specialist supports GIS-related coursework, research, and teaching at KU. Prior to such assistance being formally available, the T. R. Smith Map Collection (henceforth referred to as the Map Collection) provided basic assistance with GIS. The GIS and Data Lab (or Lab) provides a quiet working environment for individuals or groups and it includes space for use of GIS and statistical workstations, laptop use, and research design and discussion.[2] The Lab was initially located in the rear of a meeting room on a lower floor of the library, with two statistical workstations set up for social science data users. Use was low, perhaps due to minimal resources and staff support and the limitations of appointment-only use. In 2002 the Lab was moved to a room next to the Map Collection, which also contained offices for the Map Librarian and new GIS Specialist. Users had relatively easy access to staff, and staff could easily monitor the Lab.

These staff offices were moved to a space behind the Lab in 2004, coinciding with initial integration of the campus statistics lab with the GIS and Data Lab. These moves added workspace and resources in the lab and grouped staff offices, but they also removed staff from the immediate Lab area. A new doorway was created to directly connect and improve communication between Lab users and support staff. Workstations now have log-ons for user authentication that are consistent with other public machines in the libraries and standardized images, such that all computers are loaded with the same basic set of programs. This profile is more secure, easier to maintain, and gives users uniformity. Users have dedicated network space for storage of data and related files. Due to the merger of these two labs, users now benefit from extensive spatial and statistical resources and a wider range of staff expertise available in the same location.

The GIS and Data Lab is generally open sixty hours per week for walk-in use and for scheduled consultations with on-site staff. Full-time staff that support users in various capacities include a GIS Specialist, Map Librarian, Data and Government Information Librarian, and Statistical Consultant. Seven powerful workstations are loaded with major GIS and statistical applications, such as ArcGIS and ArcView, SAS, SPSS, and other programs used in conjunction with GIS software, such as Excel and PhotoShop.[3] Lab

users also have access to small scanners and printers, as well as oversize scanning and printing services, formally offered for the last several years. Additional resources include GIS-related reference books and software manuals, datasets, tutorials, and current publications.

Lab workstations are also loaded with the datasets and digital manuals that accompany ArcGIS software, or “Environmental Systems Research Institute’s (ESRI) Data and Maps” and “Digital Books.” These datasets and digital manuals are commonly requested and provide comprehensive information on using the software and a wide array of vector and raster data for the world (Andronache et al., 2005). However, users are often unaware that these resources are available or otherwise do not install them on their own workstations. The GIS Specialist worked with other Information Services staff to make these resources available online to KU affiliates.

In addition, Lab machines include a customized ArcGIS application with current, detailed data for Lawrence obtained from the city GIS program. Aerial photos, streets, contour lines, and city parcels are some of the more commonly used datasets. This tool is particularly helpful to novice and non-GIS users, such as Urban Planning and Architecture students working on site or building design projects. Students often generate a “map” of their project area and use other programs to manipulate the resulting image. Members of the general public also use this application for local mapping information, such as a “picture” or “map” of their neighborhood.

The Lab is located in one of the two larger campus libraries that holds most of the science-oriented materials. The central location of this facility, and its proximity to one of the main campus computer labs, is convenient to most users. Furthermore, this library is near buildings that house the Architecture, Urban Planning, Geography, and Geology Departments, some of our most active patrons. A potential drawback is that we are on the ground floor of a building for which the third floor is the main entrance, and this lower visibility limits accidental discovery by new users. However, being “off the beaten path” does result in relatively quieter and cleaner work areas, and those who need to find us can do so.

A nearby campus lab has approximately twenty GIS workstations available for walk-in use. This lab serves a different purpose than the GIS and Data Lab, as it does not have GIS support staff; it is primarily used for technical workshops and so is often unavailable on weekdays. However, its evening and weekend hours complement those of the GIS and Data Lab. These labs are the only two on campus that support GIS and that are open to anyone regardless of departmental or other affiliation.

The GIS and Data Lab is next door to the Map Collection, and cooperation among map, GIS, data, and government information resources extends beyond sharing physical space. GIS often acts as a bridge between geographical and tabular information since it is comprised of both. Such service groupings have been successfully implemented at various universi-

ties (Czarnocki and Khouri, 2004). Combined data, GIS, and maps services seemed to work well at Queen's University in Kingston, Ontario, Canada, which provides resources related to both "merging *and* emerging technologies" (Moon, 1993, p. 33). Soete (1997) also noted the benefits of GIS services that are closely tied to related services.

GIS Assistance

The GIS Specialist provides assistance via in-person consultations, email, and telephone through the GIS and Data Lab with frequent referrals from the Map Collection. Some common requests for assistance include the following issues and associated questions.

- Finding Specific Datasets:
  I'm looking for administrative boundaries for the Soviet Union.
  Is there GIS data for New Orleans (post-hurricane), such as transportation networks, census data, and debris site locations?
- Generating a Map or Image from Spatial Data:
  I want to create a map (aerial photo) of downtown Kansas City, near the river market.
  I need a map (aerial photo) of downtown Lawrence, showing streets and contour lines.
- Converting Data Among Various Formats, Such as Open Source Proprietary, Vector, Raster, and Compressed:
  How do I convert elevation data, downloaded in DEM (Digital Elevation Model) format, into something (grid) I can use in ArcMap?
  How do I convert this file for hiking trails in Yellowstone (trails.e00) into something ArcGIS will read (ArcInfo coverage)?
- Subsetting Data or Clipping Data Layers:
  From country boundaries for the world data, I need Africa as a separate (vector) layer.
  I only need tract boundaries and income data for Douglas County, Kansas, not the entire state.
- Creating Data, Such as Point, Line, or Polygon Features and Editing Associated (Attribute) Information:
  I need to identify and digitize forested areas on top of this aerial photograph, with a topo map (DRG) as another base map.
  I want to calculate the percent of population change in Montana counties for the last 100 years.
- Mapping Tabular Data Using Geographic Coordinates for Point Locations, or Collected with a Global Positioning System (GPS):
  How can I map my data on protest events in North Korea? I have XY coordinates in Excel for each location where a protest occurred.
  I need to get my field data out of this GPS unit and map it (that is, sunflower populations, ice core samples, wind farms)

- Mapping Tabular Data by Linking (Joining) to a GIS Layer with the Same Geographic Variables:
  I want to map child welfare data by zip code from this Excel file. How can I display the data based on a particular variable?
  I have readership data for the *Saturday Evening Post* back to the 1920s by U.S. county and state. How can I put these data on a map?
- Integrating an Image, Such as a Scanned Map, into a GIS Using Geographic Coordinates (Georeferencing):
  I need city footprints in digital format for Costa Rica.
  I want to compare modern maps to historical maps created by an explorer and plot the plant populations he documented.
- Spatial Data Analysis:
  What types of land cover are associated with these bird nesting areas?
  How many elementary schools are located within Richmond city limits?

The source of information that fills a user need may be either paper or digital, regardless of the final format needed. For instance, information originating in digital format can be manipulated using GIS or other software, remain in digital format, or be printed via desktop printers or plotters. Likewise, information beginning in paper format can be converted to digital format via an oversize scanner and printed to a plotter for what is essentially a copy. Alternatively, a scanned item can be manipulated in GIS or other software and retained in digital format or printed to hard copy. Thus, the Map Collection may bring new users to GIS and vice versa (Cobb, 1999).

For example, a graduate student in ecology working with distribution and protection of rare bird species walks into the Map Collection looking for digital data, constituting base map features for part of Ecuador. No sufficiently detailed GIS data are freely available, but a map in the Joint Operations Graphics series is found at an appropriate scale for the desired area. As the student wants digital data to incorporate into a GIS, we scan the map and the student later georeferences the image, feeding latitude and longitude values from the map into ArcGIS.

The number of data consultations nearly doubled from 2003 (eighty-two) to 2004, and this year's monthly average was 16 percent higher than it was in 2004. The majority of users assisted in 2005 were KU-affiliated and most commonly were from the Departments of Urban Planning, Architecture, Ecology and Evolutionary Biology, and Journalism. In 2005 meetings were held most often with graduate students, followed by faculty and then students who were either undergraduates or graduates. The profile for walk-in, unassisted GIS users of the Lab is similar to the consultation user base.

GIS consultations are typically longer and more complex and technical than traditional reference questions, as is consistent with GIS services at

other institutions (Abbott and Argentati, 1995; Deckelbaum, 1999; Suh & Lee, 1999). To balance the increasing demand for GIS services and the extensive time required, the GIS Specialist at KU continually works to provide self-help tools, such as online datasets, software manuals, and tutorials. This does not undermine the ongoing need for in-person assistance but rather supports and may make such meetings more productive (Argentati, 1997).

Consultations are carried out such that users are guided in using spatial data and software but the work is not done for them; these meetings are learning opportunities for all parties involved (Soete, 1997). At one end of user experience and informational needs are those wanting simple facts or statistics that may be generated by stand-alone or Web-based mapping applications. These patrons may never use GIS software, which involves a fairly steep learning curve (Suh & Lee, 1999; Deckelbaum, 1999; Soete, 1997). At the opposite end are those who want to use raw spatial data on a GIS workstation. These user profiles are similar to what Kowal (2002) defined as low- and high-level users and are reflected in the complexity of data and software appropriate for such users (Abbott & Argentati, 1995).

For others who enjoy digging for data, some requests inevitably remain unanswered by freely available data. Some recent examples from GIS and Data Services include

- data for building footprints and heights in London
- city limits for cities in Central America
- air photos for a wine-growing region in Tuscany
- zoning data for New Orleans
- street and contour data for Kristiansend, Norway.

Data Management

In developing KU Libraries GIS services, it was important to survey library-owned spatial data, including federal depository items, commercial data, and internally generated data. Library holdings, mainly from federal agencies such as the Census Bureau and Geological Survey, are underused and very difficult to find through traditional library search mechanisms (Soete, 1997). Initial inventory attempts were unsuccessful, as library cataloging tools and staff were ineffective in generating even a rough inventory. Instead, inventory is ongoing through data requests, coincidental discovery, collection development, and monitoring of new resources. Considering the strengths of KU Libraries sheet map and data holdings, and freely available data, it is challenging to find large-scale data for places outside the United States and Canada, such as topography, political boundaries, and roads. Sweetkind-Singer and Williams (2001) also found such data onerous to obtain for free.

Library GIS staff at KU and other institutions rarely purchase datasets, as a wealth of spatial and tabular data is freely available online and through government agencies (Abbott & Argentati, 1995; Lamont, 1997; Kowal, 2002; Soete, 1997). Many small-scale commercial datasets are already freely available online in raw or alternative formats. Moreover, the availability of library staff with GIS expertise to assist with data manipulation reduces the need for acquiring most commercial datasets.

At KU GIS datasets and related resources are not allotted a separate collection development budget, so they are purchased through subject liaisons. KU Libraries did purchase data to complete the Global GIS database set, which provides a variety of small-scale data layers for the world (Hearn et al., 2003; American Geological Institute, 2003). GIS data generated by campus researchers and through internal and joint projects may be made available through such archival and access tools as the institutional repository and the state GIS data clearinghouse.[4]

Instruction

Teaching GIS

The GIS Specialist teaches GIS workshops throughout the semester with assistance from colleagues. Workshops offered on an ongoing basis are designed to teach basic to advanced skills, and they cover specialized topics in ArcGIS, such as the Spatial Analyst extension.[5] Lessons focus on applying geographic concepts and questions related to spatial data while developing familiarity with software and terminology. The GIS Specialist leads these hands-on workshops, which are usually two hours long. Workshops are scheduled in cycles, as participants tend to sign up for an entire cycle instead of one at a time. Attendance is generally higher during the first half of a semester and when new workshops are introduced, but it has been steady for most workshops over the course of three years. Figures 1 and 2 describe the status and affiliation of workshop attendees.

The GIS Specialist also provides custom training sessions, such as "Moving from ArcView 3.x to ArcGIS" and "Mapping Tabular Data in a GIS." In addition, instructors may request a GIS session for a particular class. Departments that have requested GIS sessions are Architecture, Ecology and Evolutionary Biology, Environmental Studies, Geography, Journalism, Political Science, Public Administration, and Urban Planning. Teaching a workshop often initiates dialogue among participants and instructors, and many attendees use this time to discuss their GIS uses and needs or to make appointments for in-depth assistance. The GIS Specialist and related staff also provide informal training sessions, presentations, and tours to educate library staff about GIS services and help them make referrals.

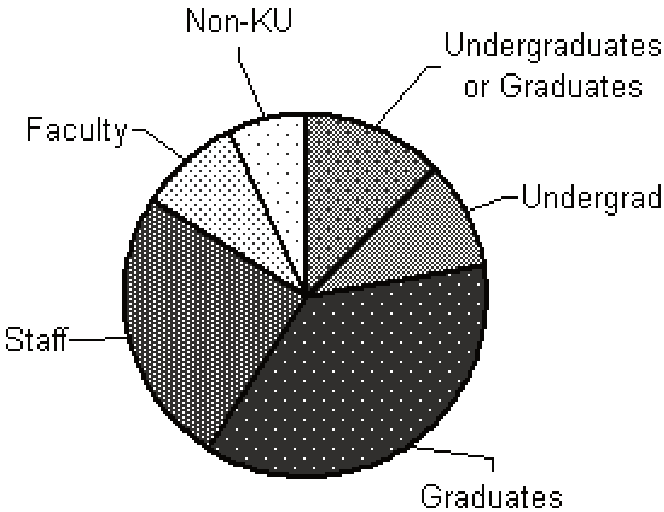


Figure 1. KU Libraries GIS Workshop Attendees, 2002–2005

GIS Training

The GIS Specialist facilitates access to educational resources such as software tools and applications, how-to guides for data and software, email list services, and tutorials. The Environmental Systems Research Institute (ESRI) online tutorials, free to KU affiliates through the campus site license, are frequently used by individuals seeking self-paced, focused training and by instructors to augment a semester course. The GIS Specialist also advises users on selecting and acquiring GIS software.

Outreach

Of perhaps more importance than developing the nuts and bolts of a GIS service is building relationships with library staff and with faculty, graduate students, and other GIS users. These relationships are the key to an effective and sustainable service and can help it evolve to meet changing user needs, as was also found by Atkins (1999) and Soete (1997). Referrals and projects arise from such relationships, both personal and professional, as patrons often come in to use the Lab by recommendation from instructors, other students, or friends. A dual method of outreach is the hands-on workshops, which help identify user needs as well as provide GIS training (Atkins, 1999).

In addition, the GIS Specialist and related staff participate in various campus events, such as GIS Day. GIS Day is a symposium sponsored primarily by the Geography Department and aimed at increasing knowledge of local GIS applications and of geographic information science. Library staff

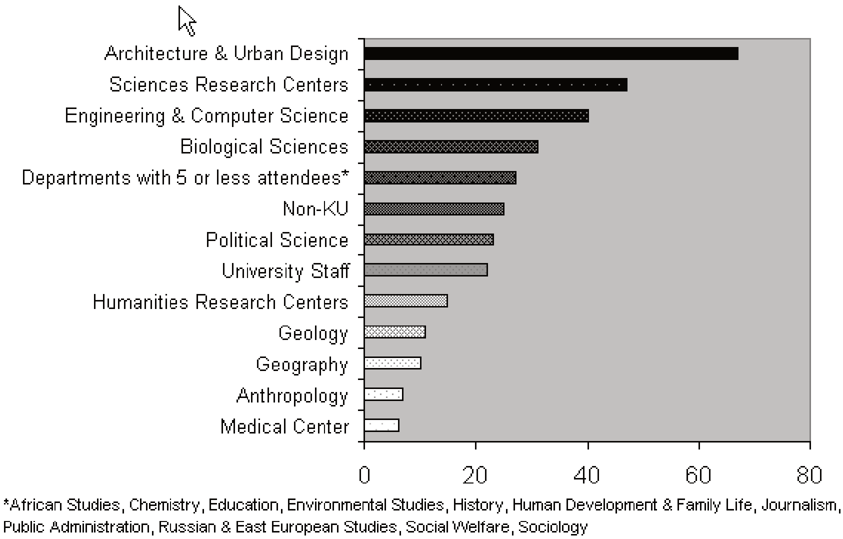


Figure 2. KU Libraries GIS Workshop Attendees, 2002–2005

have displayed GIS research posters as part of this event, judged student paper presentations, and promoted the event through GIS workshops. Events such as GIS Day are excellent opportunities to make and maintain contacts and raise awareness of library GIS services. Ongoing modes of outreach are also useful and accessible, such as email list services, direct emails, presentations, Web pages, and various publications. The GIS Specialist also works with library subject specialists in communicating with academic departments regarding GIS issues.

Future Directions

Interest in GIS is spreading steadily across many campuses (Badurek, 2000; Andronache et al., 2005; Sweetkind-Singer & Williams, 2001). At KU interest has sparked lately in the Political Science, Geology, Environmental Studies, Public Administration, and Engineering Departments. Many of these departments have neither staff nor facilities to support GIS, and users may only consist of one or two faculty and a few graduate students. These GIS users need access to software, data, and expertise from a centralized, accessible source of research assistance, such as the library.

GIS training is also an ongoing need at KU, particularly that which is inexpensive and convenient. The campus seems hungry for GIS training of any kind. However, hands-on training options outside the university are often cost-prohibitive. Workshops, online tutorials, and other educational

resources help address but do not seem to fill this need. Workshop evaluations indicate that attendees would like a wider variety and higher frequency of workshops. However, developing new workshops requires a substantial amount of time; existing workshops must be revised with software upgrades and changes in data format. The library cannot and probably should not satisfy all campus GIS training needs (Argentati, 1997).

GIS and data staff hope to fully merge the GIS and Data Lab and the campus statistics lab in the coming year. Challenges include maintaining security, providing quiet work space, and giving users a sense of ownership, while remaining readily available to assist users. The GIS Specialist also seeks to increase awareness of the Lab, library spatial data holdings, and related resources such as free software training and online tutorials. Additional signage for both public space and staff offices is one way to heighten visibility of GIS services. Finally, it is demanding but essential to continually engage library staff in discussions and training related to GIS services.

Conclusions

Researchers have access to increasingly, and seemingly user-friendly, spatial applications, which offer powerful display and analysis options. Such applications can deceptively appear overly simplified because users can obtain "glossy" output from complicated datasets with the click of a button. The user may be overwhelmed or fail to appropriately interpret this surfeit of results. This can lead to the common misconception that GIS is easy to use and learn. Realistically, finding or processing spatial data usually takes twice as long as initially expected, an experience shared by Sweetkind-Singer and Williams (2001). However, visualization and analysis capabilities, and skills gained through working with the data, are worth the effort. Graduate students seem the most willing to recognize whether a new application such as GIS software is a useful research tool for their discipline, and subsequently they dive in and learn to use the program proficiently.

Lab users also often see GIS as merely a map-making tool. Compared to professional cartographic software, most GIS programs can produce basic maps with minimal effort. However, novice GIS users often generate rudimentary maps lacking a title, projection, legend, or identification of data sources. Committing more time and effort can result in fairly detailed and accurate maps, but users rarely spend this much time. Basic paper maps generated in ArcGIS should not be used as a final medium for data but rather to examine preliminary data patterns or trends, to contribute to further digital processing, or for collection of additional data (Deckelbaum, 1999). GIS is about visualization and analysis of digital spatial data, seeking answers to questions of geography, and examining relationships between geographic features within or between layers; GIS is not about creating pretty maps (Cox & Gifford, 1997; Deckelbaum, 1999).

Others providing campus GIS services agree that every university has different GIS needs. What is the library's niche in supporting these needs (Abbott & Argentati, 1995; Adler, 1997; Argentati, 1997; Badurek, 2000; Lamont, 1997)? One must try to understand the GIS environment by consulting with stakeholders, such as other libraries, computer labs, research centers, and academic departments interested in GIS. For instance, GIS and maps staff at KU have developed solid working relationships with the Geography Department and the Kansas Biological and Geological Surveys, which are significant GIS users. One may also obtain useful information about campus GIS users from the GIS site license administrator. Some issues to consider are the following:

- Do GIS users have access to necessary data, software, and hardware?
- Is there a GIS lab open to anyone, regardless of department or affiliation?
- Are data storage and distribution needs met?
- Can users obtain GIS assistance from campus staff?
- Do users have access to needed training and related resources?
- How is GIS software purchased? Is a site license available, and are users taking advantage of the benefits?
- Are there commonly used datasets to which the library could facilitate access?
- Is grant funding available to support GIS services?
- Is there a means for communication among campus GIS users?

Support from library administration and related campus departments is fundamental to a successful GIS service. In addition, staff involved must possess a strong knowledge of major GIS software and computer applications in general and be committed to ongoing education on GIS data, software, and applications (Argentati, 1997; Cox & Gifford, 1997; Kowal, 2002). Many libraries offer GIS services. As Adler also noted (1997), the question for larger academic institutions may no longer be whether to offer GIS services but what level of service to offer.

Notes

1. KU Academic Data Research Services Alliance: http://www.ku.edu/adrsa/.
2. KU Libraries GIS and Data Lab: http://www.lib.ku.edu/gis/.
3. Adobe (1990–2002), Photoshop 7.0 Software; Apache Software (1989–2003), SPSS 12.0 Software; Autodesk (2005), AutoCad; Environmental Systems Research Institute (1992–2002), ArcView 3.3 Software; Environmental Systems Research Institute (1999–2004), ArcGIS 9.1 Software; Microsoft (1985–1999), Excel 2000 Software; SAS Institute (2002–2003), SAS 9.1 Software.
4. KU Institutional Repository: https://kuscholarworks.ku.edu/dspace/.
5. GIS workshop descriptions and handouts available from "GIS & statistics workshops" link: http://www.lib.ku.edu/gis/.

References

Abbott, L., & Argentati, C. (1995). GIS: A new component of public service. *Journal of Academic Librarianship 21*(4), 251–256.

Adler, P. (1997). An introduction—Where are we heading? Geographic Information Systems (GISs) Revisited: A Symposium [Special issue]. *Journal of Academic Librarianship, 23*(6), 447–448.

American Geological Institute. (2003). *Global GIS database: Digital atlas of North Eurasia.* Alexandria, VA: U.S. Geological Survey.

Andronache, C., Mento, B., Gaiser, T., Hon, R., Dalgin, R., & Schreiner, A. (2005). Increasing GIS resources for teaching and research. In *ESRI GIS Education Users Conference,* ESRI, July 2005, San Diego, CA. Redlands, CA: ESRI.

Argentati, C. (1997). Expanding horizons for GIS services in academic libraries. *Journal of Academic Librarianship, 23*(6), 463–468.

Atkins, A. (1999). Library-based GIS initiative at the Illinois Institute of Technology. *Illinois Libraries, 81*(2), 67–74.

Badurak, C. (2000). Managing GIS in academic libraries. *WAML Information Bulletin, 31*(2), 110–114.

Cobb, D. (1999). GIS: Its impact on library service. *Meridian, 16,* 5–8.

Cox, A., & Gifford, F. (1997). An overview to geographic information systems. *Journal of Academic Librarianship, 23,* 449–461.

Czarnocki, S., & Khouri, A. (2004). A library service model for digital data support. *IASSIST Quarterly, Summer/Fall,* 30–34.

Deckelbaum, D. (1999). GIS in libraries: An overview of concepts and concerns. *Issues in Science & Technology Librarianship, 21.*

Hearn, P., et al. (2003). *Global GIS databases: Digital atlases of Africa, Central and South American, Europe, North America, North Eurasia, South Asia, and South Pacific.* Alexandria, VA: U.S. Geological Survey.

Kowal, K. (2002). Tapping the Web for GIS and mapping technologies: For all levels of libraries and users. *Information Technology & Libraries, 21*(3), 109–114.

Lamont, M. (1997). Managing geospatial data and services. *Journal of Academic Librarianship, 23*(6), 469–473.

Moon, J. (1993). Integration of "traditional" data centre services with GIS technology. *IASSIST Quarterly, 17,* 30–34.

Soete, G. (1997). *Issues and innovations in geographic information systems.* Washington, DC: Association of Research Libraries, Office of Management Services.

Suh, H., & Lee, A. (1999). Embracing GIS services in libraries: The Washington State University experience. *Reference Librarian, 64,* 125–137.

Sweetkind-Singer, J., & Williams, M. (2001). Supporting the information needs of Geographic Information Systems (GIS) users in an academic library. *Information and the Professional Scientist and Engineer, 16,* 175–191.

Rhonda Houser is the GIS and Data Specialist at the University of Kansas Libraries in Lawrence, Kansas. She holds a bachelor's and master's degree in Environmental Studies. Her thesis research focused on using Geographic Information Systems (GIS) to examine Eastern Cougar habitat in the Northern Blue Ridge of Virginia. She did GIS and mapping-related work for the Virginia Department of Conservation and Recreation before completing her master's degree in 2002.

Improving GIS Consultations: A Case Study at Yale University Library

Abraham Parrish

Abstract

For the last decade and a half, Geographic Information Systems (GIS) services in academic libraries have been developing, and GIS librarians have been experimenting with different ways to provide these services. However, there has been virtually nothing in the literature with respect to GIS consultation statistics. One goal of this article is to discuss a four-year case study on the use of GIS consultation statistics to give a better understanding of what GIS librarians might typically expect as far as number of patrons, their characteristics, amount of time spent with them, and the amount of data distributed to them while running a GIS service at an academic library. Techniques for reducing the amount of time spent with patrons while developing a higher degree of efficiency and effectiveness in conducting GIS consultations will also be explored. Finally, a juxtaposition of GIS consultations with other types of library reference services will reveal significant differences between them.

Introduction

What kind of workload can a Geographic Information Systems (GIS) librarian expect from GIS consultation services? How much time will he or she be spending on this activity, or what average amount of time will they spend on a single consultation? How much data will they be distributing to patrons on average? What can be done to make consultations less time-consuming and more efficient and effective? Answers to these questions will be explored through the use of GIS consultation statistics that have been collected for a four-year period at the Yale University Library Map

LIBRARY TRENDS, Vol. 55, No. 2, Fall 2006 ("Geographic Information Systems and Libraries," edited by Jaime Stoltenberg and Abraham Parrish), pp. 327–339

Collection GIS Service. However, a brief review of the current literature on reference statistics will prepare us for this exploration.

With the progression of digital technology over the last decade and a half, librarians have been able to provide patrons with access to a larger array of digital data and information more rapidly and have used electronic sources more often for reference (Cardina & Wicks, 2004). This progression has allowed an increase in the capability of what can be provided through reference, but it has also created an increase in the demand for reference information via these new technologies (Tenopir, 1998). Even though the number of reference interviews has decreased during this time, they now take longer (Mayfield, 2000; Warner, 2001).

Even with an evident trend toward providing reference for digital services and content in academic libraries, there is not much in the library literature on detailed statistics for reference services. Spencer and Dorsey (1998) identify total and average times for reference interviews broken down by affiliated and nonaffiliated patrons for an Arizona State University West study over the span of a year with one week a month selected randomly for data collection. They cite an overall mean of five minutes per patron and identify reference exceeding eleven minutes having occurred only 27.2 percent of the time. Most other studies had smaller samples and even less distinction among types of patrons.

The library literature for GIS reference in particular was even sparser. Kinikin and Hench (2005) present a weekly GIS service utilization table based on a survey of eleven libraries; the survey identifies one GIS user per week for four libraries, one to two users per week for another four libraries, no libraries with three to four users a week, and five or more users for one library. This fails to provide a clear picture of different types of GIS users and the actual time spent on consultations. This article will reveal more detailed statistics for GIS consultations based on a case study at Yale University Library. First, however, a preliminary review of the differences between GIS consultations and other types of library reference is warranted to set the stage for understanding these statistics.

It can be argued that the growth of digital technology has had a greater impact on GIS services in libraries compared to other emerging services that incorporate electronic resources since GIS depends completely on this type of technology. It follows that GIS reference or consultations can be much more involved and time-consuming not only because they completely depend on computer technology but also because of the more complex structure and variety of GIS software and geospatial data.

Robust GIS software is not as simple to use as a software program that may display a textual source of information such as a digitized journal article or a raster image such as a digital photograph. Patrons must learn how to interpret spatial data and create information from this data via manipulation through GIS software. This software can contain hundreds

of tools, extensions, and additional scripts that can potentially be used to manipulate GIS data. This means that a GIS reference librarian must take into account a much broader range of service issues when conducting consultations for patrons, which add significant challenges that exceed those of general library reference.

These challenges include first and foremost providing training for patrons to enable them to use the GIS data they acquire. GIS software has a steep learning curve and takes an individual with a variety of abilities to successfully employ the technology (Deckelbaum, 1999). For robust GIS software such as ArcGIS, it can take a minimum of fifteen hours just to learn the basics. There are also many different applications of the software that span many different disciplines. Simply knowing what specific tools can do in the software does not necessarily or easily translate into knowing what spatial methodology to apply for specific disciplines. As a result, patrons can easily spend thirty or more hours just learning enough of the software to tackle a significant research project. Thus, the GIS librarian is required to possess a substantial skill set in order to be an effective reference librarian or consultant. This skill set includes familiarity with GIS software in many disciplines, available training courses or tutorials, sources of additional or extensible GIS software, and hundreds of software and application books.

Another challenge is training patrons on how to manipulate GIS data through processing tasks such as file format conversion, re-projection, and geo-processing. Information is often created and developed by these processes, unlike a published book or print map that has already been turned into information. So simply having the data in hand—or in this case, on digital media—often does not provide a patron with enough to glean any substantial information. Having more data layers that are spatially synchronized will better enable patrons to garner more information and perform deeper analysis. For the GIS librarian, this again requires a substantial skill set—in this case, skill in data manipulation. It is often the case that most of the time patrons spend utilizing GIS as a tool for their research project is spent manipulating GIS data.

The third and final significant challenge is providing the patron with training in information management. This challenge may well be a significant part of general library reference as well, particularly with those who handle digital data or information. However, GIS patrons often collect tens of layers of data, which can be manipulated several times. For some GIS data formats, one layer can be composed of up to eighteen different files organized in multiple folders. This can result in hundreds of files that need to be managed and whose structure must be understood, even if a patron is only using as little as ten layers. The patron must also understand at least the basics of geodesy (coordinate systems and projections) and its relationship to the organization and display of their layers. This can be a

very difficult topic to teach to patrons who have no background in geography. If patrons do not acquire data management skills, particularly for large projects, they could lose track of layers they had manipulated, end up with layers that are not spatially synchronized, and potentially waste hours of time trying to locate specific layers.

These additional significant challenges that are part of GIS consultations affected the resulting statistics that were collected at Yale University Library, mostly by adding a significantly longer amount of time to the average GIS consultation.

Description of Yale University

Before analyzing the GIS consultation statistics that have been collected, a description of Yale University and its library will provide a better understanding of the results. The environment in which the following statistics were collected certainly does not fit the mold of every university, as different universities may have different models of service, which may vary according to the expertise of staff and level of service it provides (Deckelbaum, 1999). However, this sample study may at least offer librarians providing GIS services a general guide on what can be expected for GIS consultations.

In the four-year period of the study, which spans from July of 2001 to June of 2005, Yale University had an enrollment of approximately 11,000 students, of which about half were undergrads and the other half graduate students. There were approximately 3,000 faculty and 7,000 staff at the university. The university contained no geography department and only offered GIS courses in two departments—Forestry and Environmental Studies, and Epidemiology and Public Health. There were a few service centers on campus that provided some level of GIS service but not at the comprehensive level provided by the Yale University Map Collection GIS Service. Many of the service centers simply offered lab computers with GIS software or had a collection of a particular type of geospatial data such as Federal Depository data or satellite photos.

Description of GIS Service

The Yale University Library Map Collection GIS Service is responsible for serving all Yale-affiliated patrons, which include students, faculty, and staff. Public patrons are not eligible for GIS Service. The staff of the GIS Service is made up of a permanent GIS Specialist along with an array of constantly changing staff. During the four-year period of the study, the GIS Service has included up to as many as four concurrent student employees, a casual employee, and two clerical and technical employees, all of which were part-time employees. Currently, there is one GIS Specialist, one full-time managerial and professional GIS Assistant, and one student worker.

Due to its location in the Yale Map Collection, the GIS Service had some degree of overlap with the print map collection. This overlap most

often manifested itself in the form of digital scans of paper maps in the collection. However, the Map Collection and the GIS Service were still considered separate services and the statistics collected were applicable to GIS patrons only. One exception was patrons who requested digital scans of maps and who may or may not have had them georeferenced for use in a GIS. Since the GIS Service ran the scanning service, these patrons were added to the statistics.

Statistics Collection Method

Martindale (2004) points out that it is important to keep statistics on GIS consultations because they are useful as a measure of successful service. The Yale Map Collection GIS Service maintains a Microsoft Access database called the Daily Log that is used to track patron consultations as well as other tasks that are performed on a daily basis. It has been in operation since the first day the GIS Service was started on July 2, 2001. It was created for many reasons, but the two most important are to keep track of the patrons being helped (the GIS Service can be involved with as many as thirty consultation projects at a given time), and to formulate statistics to discover trends and make adjustments in the service.

The statistics that have been collected include the time spent not only finding GIS data for patrons but also assisting patrons in the manipulation of this data as well as providing software and information management training. Group instruction sessions were counted as one consultation. For example, if a GIS workshop was given to a class of fifty students, the Daily Log would record it as one consultation. Also, there were cases where a single patron consulted with the GIS Service for more than one GIS project, or several patrons working in a team consulted the GIS Service. Therefore, consultations rather than patrons are described as single entities in the statistics.

Data on consultations were entered daily into the Daily Log on a continual basis. The disadvantage of this was that it took a long time to enter the data, and employees had to be constantly aware of how they spent every minute. However, it has been well worth the time compared to the time that would be spent trying to keep up with and remember the progress of several simultaneous consultations. The ability to be able to profile a patron consultation project at any time on the database and get up to speed with a patron about the current progress of their project makes a GIS service very efficient. Another advantage is being able to have multiple staff members handle a single consultation.

Evaluation of Statistics

Table 1 shows the total and average hours spent on consultations broken down by patron type. It shows that master's students took up the most amount of time with about 30 percent of all consultations. All graduate

Table 1. Time Spent on GIS Consultations by Patron Type, 2001–2005

Patron Type	Total Minutes	Consultations	Total Hours	Percentage of Total Time	Average Mean Time (Hours) per Consultation Project
Faculty	30,070	61	501.17	19.81	8.22
Total Graduate Students	74,413	346	1240.22	49.03	3.58
Masters Students	45,751	291	762.52	30.14	2.62
Doctoral Students	28,662	55	477.70	18.88	8.69
Undergrads	15,524	54	258.73	10.23	4.79
Staff	21,159	137	352.65	13.94	2.57
All Others	10,608	20	176.80	6.99	1.84
Grand Total	151,774	618	2,529.56	100	4.09

Table 2. GIS Consultation by Contact Type, 2001–2005

Contact Type	Total Contacts	Average Mean Number of Contacts per Consultation Project
In Person	1,839	2.98
Email	500	0.8
Phone	164	0.26
All Contact Types	2,503	4.05

students accounted for about half of total consultation time. Given that about half of Yale students are graduates, this appears to explain the statistic. However, while this fact may explain most of this number, the total number of undergraduate consultations (54) was low because undergrads typically do not have time to incorporate GIS in their coursework. Even though they are required to write a senior thesis, which most of the undergrads consulted were working on, there were not as many undergrads working on long-term projects as graduates or faculty. These statistics show that GIS consultations are typically geared toward patrons who have time to work on long-term projects.

Table 1 also reveals that the average amount of time spent per consultation project is about four hours. Faculty and doctoral students typically average about twice as much time per consultation project as undergrads or graduates. This is useful information when starting a consultation project and trying to determine how much time you will spend consulting. You can develop a prepared plan for dealing with each of the different types of patrons.

Table 2 shows what you might expect by way of contact when conducting GIS consultations. The GIS Service averaged about four different visits by patrons overall, most visits being in person. This makes sense, as it is difficult to conduct a consultation over the phone or email unless it is for a specific and quickly solved problem. The highly visual nature of GIS usually requires an in-person demonstration to explain a GIS problem to a patron. This also

Table 3. Time Spent on GIS Consultations by Department, 2001–2005

Department	Hours	Percentage of Total Time
Forestry & Environmental Studies	659.68	26.08
Anthropology/Archaeology	361.33	14.28
Art & Architecture	310.75	12.28
Epidemiology & Public Health	108.73	4.30
Political Science	99.83	3.95
Economics	88.08	3.48
History	138.42	5.47
American Studies	67.80	2.68
International & Area Studies	61.05	2.41
School of Management	34.42	1.36
All Other Departments	599.46	23.70
Grand Total	2,529.56	100

Table 4. GIS Consultation Statistics by Year

	2001–2002	2002–2003	2003–2004	2004–2005	All Years
Total Consultations	114	187	211	162	618
Total Time Spent on all Consultations (hours)	613.71	709.68	844.71	361.46	2,529.56
Average Mean Time Spent per Consultation (hours)	5.38	3.79	4	2.23	4.09

Note: Total consultations for all years do not add up to 674 because there were consultations that overlapped from one year to another and were counted twice in more than one year.

shows that GIS consultations rarely can be completed in one session. The GIS Service has consulted with some patrons in as many as thirty different sessions spanning several months or even several years.

Table 3 shows consultation projects by department and the percentage of time out of the total hours spent on consultations by each department. Clearly, Forestry and Environmental Science, Archaeology, and Art and Architecture have far more users of GIS than any other department. These statistics can be useful in determining for which departments to focus workshops, targeting departments with much potential but little use, or determining the discipline-specific types of GIS resources to collect. It can also be useful in determining the cost share of a GIS software site license for which a department should be responsible.

Table 4 shows the trend over time of the number of consultations along with total and average time spent on them for each of the four years in the study period. It can be seen that the GIS Service saw a steady increase in the number of and total time spent on consultation projects, but the average amount of time spent on each consultation decreased. This can be explained by the initial creation of the GIS Service in 2001. There was no GIS Service in the library prior to July 2001 and it was subsequently built from scratch. As more patrons learned of the service, word spread and the number of consultations increased until reaching a peak in the 2003–2004

period. Around this time, the GIS Service had applied enough efficient consultation techniques and made enough faculty contacts to reach more patrons with fewer consultations by providing more workshops and class demonstrations (as stated above, group consultations were counted as one consultation). This explains the drop in the number of consultations and total time spent on them. Besides these reasons, the experience of the GIS Service for the first three years of the study had a significant effect on increasing the efficiency and effectiveness of consultation projects.

Suggestions for Improving GIS Consultations

One of the best ways to improve the efficiency and effectiveness of GIS consultations is to recognize different types of patrons and have a plan to deal with their particular situations in the most effective manner. The following is a list of the ten most distinct types of patrons that have been encountered at the Yale Map Collection GIS Service in the four-year consultation period, along with suggestions on effective consultation techniques.

The Sleeper

These patrons will initially come in for a consultation in which you spend a significant amount of time consulting with them on how to incorporate GIS into their research. They then show up six months later, never having made one bit of progress. They have not completed any self-paced training you have given them or even looked at or lost any data you distributed to them; they want to start all over again with the GIS consultation as if you were meeting them for an initial visit.

To avoid wasting reference time on the Sleeper, try to gauge how much work patrons are willing to put into utilizing GIS as a tool in their research. Tell patrons that learning to use GIS software will not only take a significant period of time, but data processing, analysis, and cartography can be very time-consuming as well. Try to start off novice GIS users slowly so you do not invest too much consultation time unnecessarily. If they show progress on something small, such as completing an online training course or processing a small set of data, then you can provide continued assistance with more confidence that they will follow through for the rest of the GIS project.

The Data Collector

These patrons want to collect every little bit of GIS data they can get their hands on even if it is not necessary for their research. These are the patrons who want over a thousand census attributes when you are creating a census layer for them or want data that over-expands their study area by an unnecessary amount "just in case" they might need it. Try to get these patrons to focus on their research questions and the specific datasets they will need to answer those questions. It always helps to establish a geographic study area with bounding coordinates with the patrons before rushing into

a data search. Explain or demonstrate to them how long it takes to process or analyze GIS data with a small sample so they have an idea what it will take for larger datasets. You could also just distribute the data they need and show them how to acquire the "just in case" data themselves.

Seeking a Professional Cartographer

Although cartography is an important part of using GIS in research or coursework, it can be a tricky issue when it comes to GIS consultations. Yale University has no geography department or any other department that teaches a cartography course. Therefore, there are not many students or faculty who are familiar with cartographic techniques, particularly with GIS software. However, often the best way to share research that utilizes GIS is through a map.

The Map Collection GIS Service has been dealing with this issue for many years and has developed a policy of assisting patrons with cartography but only to a certain point. You want to avoid patrons with no cartographic background requesting a map be made for them. A professor who wants you to make a map of X, Y, and Z and have it ready by next week as if you were running a professional cartography business, or a student hovering over your shoulder telling you to move a label a few millimeters to the right then a few millimeters back to the left, is not the most efficient use of your time as a GIS librarian. Unless your library has the resources for it, it is best to limit cartographic assistance to specific cartographic techniques. Patrons can consult with you just for advanced cartographic techniques and save your consultation time by learning basic techniques on their own. There are several short online tutorials available for this type of training, such as the Environmental Systems Research Institute's (ESRI) "The 15-Minute Map."

The Enigma

This type of patron is probably common among many types of library reference services. They are the type of patrons who want your assistance but do not want to tell you much about their research, as if it were classified Top Secret. The goal with these patrons is to gain their trust and explain to them that there are certain things you need to know in order to help them. However, there may be cases where they just will not reveal certain types of information. In these cases, it still may be possible to help the patrons with certain consultation techniques. For example, if an archeologist is doing a dig at a sensitive site for which she does not want to reveal the location, but she wants to plot Global Positioning System (GPS) locations collected from the site in a GIS, you could show her how to plot the points using a different set of GPS points or XY data so she can repeat the technique in private.

Don Quixote

These patrons do not realize the limitations of using GIS or acquiring data for their research. Don Quixote will ask for a GIS layer of all the streets in Connecticut in 1930 so he can geocode addresses from the 1930 U.S. Census (which is only available on paper) by next week. Don Quixote will ask for 1-meter color satellite photos for the whole country of Zimbabwe. For this type of patron, it is best to explain the limitations of acquiring or developing GIS data due to time and budget. Make it clear how much time you are willing to spend acquiring or processing GIS data for them and determine if they are willing or even able to spend the rest of the time needed to reach the goals of their project.

The Lounge Lizard

These patrons will try to utilize every possible second of your time to help them with their GIS project. They will call or email you several times a day asking what button to click next in the GIS software. They will pop up in your office without an appointment, even when you are in an appointment with another patron, to ask you a burning GIS question that just can not wait. They hang out in the GIS lab constantly on a computer to be near you in case they have a question (even though there are many labs with GIS software on campus that can be used). They will even ask for help from other advanced users in the GIS lab trying to concentrate on their own projects.

Get these patrons to invest time into learning at least the basic GIS software tools and analysis techniques early in their project. Suggest online training courses such as ESRI's Virtual Campus (if your library uses ArcGIS software) or GIS courses that your university offers. Build a substantial GIS reference collection that includes books on GSI software and applications to provide patrons with other sources of information they can consult besides you. Lastly, build your own GIS tutorials on short, task-based techniques such as georeferencing or geocoding so you can hand patrons a sheet of instructions or point them to a Web site instead of spending many minutes or even hours explaining techniques to them.

Indiana Jones

Lamont and Marley (1998) point out that digital map collections can be modeled on print map collections that are focused on the geographic region in which the institution is located. While it holds true that most patrons will most likely request data for a local area such as New Haven or the state of Connecticut, I have found that a substantial number of patrons at Yale University conduct research all over the world and therefore need data that spans this geographic extent—often at a large scale. Indiana Jones is a patron who is studying yak herds in central Asia or rain forests in Costa Rica, or conducting an archaeological excavation in Syria. It can be difficult to deal with these patrons due to the lack of data available

for their geographic areas of study, which are often third world countries with sparse mapping. You must often resort to supplying satellite images or scanned and georeferenced maps of their area if little or no GIS data can be obtained.

However, these patrons can often still be helped and can also be an advantage for collection development. Although you may not be able to acquire data through normal channels, the patron can often serve as his own resource with a little guidance. These patrons often spend a significant amount of time in their geographic study area, speak the local language, and have made local contacts. You can steer them in the right direction to find the data they need, which may be available from a national mapping agency or a local company only if you are actually physically present in that country and know the right people. Sometimes deals can be worked out so that the patron can acquire a standard dataset for the whole country or region for your GIS collection.

Also, these patrons are often doctoral students who may spend much of their time at their study site collecting their own data via GPS or surveying devices. They can be helped by teaching them how to take their own raw data and make it into GIS layers with which they can perform more analysis. For example, a patron who collects XYZ locations can be shown how to interpolate the data into a high-resolution digital elevation model, which can be further developed into slope or aspect surface layers.

The Sponge

These are patrons who can learn GIS very quickly and have a strong desire to become proficient in using GIS software. They often start out as non-GIS or novice users and quickly soak up anything you expose them to. They finish sixteen-hour online GIS courses in one weekend, familiarize themselves with enough data sources to acquire most or all of the data they need, and consult with you only after unsuccessfully tackling a GIS problem themselves for two hours.

These are highly desired patrons that make your consultations with them less involved. However, after discovering a patron is a Sponge, try to convey to them that they need not waste too much time trying to tackle a GIS problem themselves. A two-hour problem may be answered in a one-minute consultation. It may also be beneficial to consult with them a little longer than normal. The fact that they learn quickly will make them into advanced users in a short period of time. As a recognized advanced user, they may often help colleagues in their department with minor GIS problems or may even end up working for the library GIS service as a student employee.

The Ninja

The Ninja is another desirable type of GIS patron. This is a patron who is already a highly skilled and deft user of GIS. These patrons may have

already taken several advanced GIS courses or may have had several years of real-world project experience with GIS. They often use more than just standard GIS desktop software, utilizing and disseminating GIS data on interactive mapping sites or relational databases. They may have mastered additional software for spatial analysis such as remote sensing, statistical, or Computer Aided Design and Drafting (CADD) software.

During consultations, they usually just need help finding a particular dataset or just need you to purchase one they have already found. However, they may sometimes ask complex questions that strain the limits of your own GIS experience and expertise. Be prepared for the Ninja by having a plan for questions you may not be able to answer immediately. One plan may be to refer the patron to another librarian or faculty member in the area of expertise about which they are asking. For example, if they are asking questions about spatial statistics and this is not your strongest area, you could refer them to a statistics/data librarian or mathematics professor. Another plan could be contacting GIS colleagues or your GIS software vendor, or posting to a GIS listserv. A final solution could be spending the time to figure out a solution to a GIS problem on your own. Try to avoid spending extended periods of time researching an answer to a tough question while the patron is consulting with you in person. This may end up wasting both the patron's and your time. Help the patron with whatever you are immediately able to, and then contact him later to set up another appointment once you have had time to research the question.

The Philanthropist

This patron can be a wonderful resource to a GIS service and can also be identified as a virtual employee. These patrons may be developing Sponges or GIS Ninjas who take it upon themselves to share their collected data or expertise. They may offer to provide a copy of a significant or valuable dataset that they acquired to the library so other users may access it for research. They will get involved with you in projects that involve building geodatabases or interactive mapping sites, lending their time and technical skill to advance the GIS infrastructure in your library or university. It can be very productive to accept the generosity of the Philanthropist, as it can make your job as a GIS librarian easier by having additional GIS expertise that can be tapped. However, be sure not to be too much of a drain on this type of patron and try to work with them in a way that makes it just as beneficial to them as it is for you. A GIS project that satisfies part of their research requirements and builds needed GIS infrastructure in the library would be a prime example of this balance.

Conclusion

It is apparent the GIS reference or consultations are more involved and require more time and expertise from librarians compared to other types

of library reference. The four-year statistical study shows that GIS librarians can expect to spend an average of four hours for a single consultation that can span months of time interspersed with several meetings. It also shows that about 155 consultations a year can be expected from an academic university similar to Yale. Statistics, such as those from the four-year study, can be useful for organizing and determining the future direction of a GIS service. And finally, recognizing different types of patrons and utilizing techniques to deal with them can lead to more effective and efficient consultations.

References

Cardina, C., & Wicks, D. (2004). The changing roles of academic reference librarians over a ten-year period. *Reference & User Services Quarterly, 44*(2), 133–142.

Deckelbaum, D. (1999). GIS in libraries: An overview of concepts and concerns. *Issues in Science and Technology Librarianship, 21,* 1.

Kinikin, J., & Hench, K. (2005). Follow-up survey of GIS at smaller academic libraries. *Issues in Science & Technology Librarianship, 43,* 1.

Lamont, M., & Marley, C. (1998). Spatial data and the digital library. *Cartography & Geographic Information Systems, 25*(3), 143–149.

Martindale, J. (2004). Geographic Information Systems librarianship: Suggestions for entry-level academic professionals. *Journal of Academic Librarianship, 30*(1), 67–72.

Mayfield, K. (2000). Ask a librarian, not Jeeves. *Wired News.* Retrieved February 16, 2006, from http://www.wired.com/news/culture/0,1284,40308,00.html.

Spencer, J., & Dorsey, L. (1998). Assessing time spent on reference questions at an urban university library. *Journal of Academic Librarianship, 24*(4), 290–294.

Tenopir, C. (1998). Plagued by our own successes. *Library Journal, 123,* 39–40.

Warner, D. G. (2001). A new classification for reference statistics. *Reference & User Services Quarterly, 41*(1), 51–55.

Abraham Parrish has over twelve years of GIS and information management experience. He has served as the Geographic Information Systems Specialist at the Yale University Library Map Collection for the last five years. He received a master's degree in Library and Information Science as well as a master's degree in Geography in 2001 from the University of Wisconsin, Milwaukee, where he also worked as a GIS Librarian intern at the American Geographical Society Library for three years. He also served in the U.S. Army as an intelligence analyst for four years.

Centralized vs. Distributed Systems: Academic Library Models for GIS and Remote Sensing Activities on Campus

Joe Aufmuth

Abstract

Academic libraries are a prime example of an enterprise whose mission is to support the information needs of its institution. Geographic Information Systems (GIS) and remote sensing (RS) are popular topics for academic research and are used globally. Two major enterprise information service and data delivery models, centralized and distributed, describe how enterprises approach information sharing. Simply stated, centralized systems provide services and data through a single individual or departmental unit. Distributed systems rely on many interconnected individuals or units to supply services and data. There are advantages and disadvantages to both, which may lead to a hybrid model of combined elements or a movement away from one and toward the other. This article discusses centralized and distributed enterprise information service and data delivery models and how two Florida university libraries deploy these models to deliver enterprise GIS services and data to their institutions' user communities.

Introduction

Information services and data products such as those associated with Geographic Information Systems (GIS) and remote sensing (RS) are provided by many public and private organizations. By definition, "Any organization [public or private] that needs to support multiple concurrent users accessing a shared information resource" (Rich, Das, & Kroot, 2001) can be called an enterprise. The enterprise's concurrent users can be both internal and external to the organization. Academic libraries are a prime

LIBRARY TRENDS, Vol. 55, No. 2, Fall 2006 ("Geographic Information Systems and Libraries," edited by Jaime Stoltenberg and Abraham Parrish), pp. 340–348

example of an enterprise. If viewed in the context of an academic institution, one might even view libraries as an enterprise embedded in the larger institutional enterprise. At the academic institution level, administration, teaching, research, and technology transfer are intended to meet the information needs of concurrent users in society. Simultaneously, the people, information services, and materials physically owned or licensed virtually by the library directly support the mission and information needs of the academic institution, that is, faculty, staff, students, and others outside the institution. For the purposes of this article, GIS will be used to refer to both GIS and RS as combined spatial services and data.

Enterprise GIS

An enterprise GIS is typically viewed as the infrastructure (hardware, software, and personnel), spatial data, and applications used to inventory, manage, and analyze an institution's own spatial resources. In "Developing Enterprise GIS for University Administration: Organizational and Strategic Considerations," McCormick discusses enterprise GIS for university administration (McCormick, 2003). He points out three GIS usage areas: technical reference (planning and facilities), public reference (cartographic visitor maps), and decision support (student recruiting, locating facilities, etc.). He also describes three scales of enterprise GIS usage: a single departmental GIS, which is created for a specific purpose; a loose confederation of departments, which might share spatial software and data; and a fully integrated enterprise system, which may even be used for detailed records management. The scale of enterprise usage may change due to success or failure at any one particular scale.

The literature reveals that McCormick's public reference category exists in academic libraries. Some articles have proposed GIS systems that provided patrons with collection location information (Xia, 2005), developed GIS interfaces to digital historic collections (Haas et al., 2005), and investigated geographical interfaces to catalog records (Haas, Aufmuth, Coleman, & Uhlinger, 2002). While GIS has been used to locate new public library facilities (Koontz & Jue, 2001), nothing in the literature suggests GIS is internally used in locating new academic library facilities, library facilities management (McCormick's technical reference), or library decision-making support.

Service and Data Delivery Models

Two significant models for delivering an enterprise's information services and data are the centralized system and the distributed system. Simply stated, centralized systems provide services and data through a single individual or departmental unit. Distributed systems rely on many interconnected individuals or units to supply services and data. There are advantages and disadvantages to both that may lead to a hybrid model of combined

elements or a movement away from one and toward the other. For instance, the single point of access in a centralized system requires users to have a high level of trust in the provider's ability to meet user needs. A lack of trust may lead a department to develop its own GIS. In distributed systems some of the services and data delivery points may be redundant and lead to consolidation to reduce costs. In a hybrid system, some features of a distributed system may become centralized, such as software licensing for multiple users.

An academic library's choice of GIS service model is influenced by current campus-wide GIS activities—or "enterprise scale"—as defined by McCormick, as well as the library's willingness to compete with other centers on campus. Libraries at institutions without GIS research centers or GIS in academic departments might initially choose the centralized system and become the hub of campus GIS activity. The centralized model may start out in a collaborative mode, but as campus departments develop GIS expertise, libraries may find themselves competing for funding and space. If an established network of campus GIS and remote sensing research centers and programs already exists, libraries might choose to provide GIS services to faculty, staff, and students in those distributed departments rather than create its own center. In distributed systems the library can also serve as a neutral place for highly competitive research and educational centers to come together. Whether centralized or distributed, library GIS services and data delivery may include data acquisition (creation, download, or purchase), data processing and analysis, data distribution, data maintenance, data archiving (institutional repositories), GIS applications (development and/or implementation), software (purchasing, licensing, distribution, and installation), hardware support, teaching, research, and consulting.

A Tale of Two Libraries

The State University System (SUS) of Florida has eleven member institutions. Many of the institutions have prominent GIS and RS teaching and research programs. Four of the institutions' academic libraries—University of Florida (UF), Florida International University (FIU), University of South Florida (USF), and University of Central Florida (UCF)—have hired or identified library faculty responsible for spatial data and services. All four libraries exhibit some traits of either the centralized or distributed models for enterprise GIS information services and data delivery. Although the traits are exhibited, the libraries did not plan to follow a specific model. Instead several factors have influenced the direction each has taken. Each university's history, breadth and depth of academic programs (McCormick's "scale of usage": single department, loose collection, or large-scale enterprise), and institutional culture helped shape library services to meet users' needs. Two of the SUS libraries that typify the centralized and distributed models are FIU and UF respectively.

Centralized Model: Florida International University

Florida International University, which celebrated its thirtieth anniversary in 2002, has been successfully growing academic GIS teaching and research programs. Library GIS and remote sensing services and data delivery at FIU have evolved from a one-person government documents department initiative in 1995 (formerly known as the Geographic Information Systems Remote Sensing Applications Laboratory, GISRSAL) to an autonomous full-service GIS and remote sensing library department and campus center, known as the GIS-RS Center. In the years between 1995 and 2000 GISRSAL was the only GIS and RS research and teaching laboratory on the FIU campus. All of the other departments on campus utilized the library's facilities. Over the ten-year period since 1995, the number of GIS-RS Center personnel has grown from a couple of individuals to six full-time individuals, including a GIS-RS Center Head who reports to the Director of Libraries, a GIS Research Manager, a Web Developer/GIS Programmer, two GIS Research Associates, an IT Administrator/Developer, and an IT Associate/Web Designer. The GIS-RS Center has an advisory committee consisting of thirteen individuals representing a total of nine departments and other campus centers with a vested interest in GIS and remote sensing. The committee advises the library administration and the Center Head on user community service needs and direction. Hardware listed on the center's Web site includes twelve staff computer workstations, forty-five teaching and research workstations, three Internet servers for applications and data, a fiber optic network, a large format scanner, a digitizing tablet, a color printer, and a large format plotter. Center software includes the Microsoft Office Suite; SAS for statistical analysis; and ArcGIS, ArcView, and Leica's ERDAS Imagine for GIS and remote sensing. The center also distributes and maintains campus-wide GIS software and licenses. In order to fiscally support the center's activities, a fee-based approach to services has been adopted. Fees charged for mapping and data analysis services are listed at $50/hour for FIU students and $75/hour for faculty and non-FIU affiliated patrons. Plotting charges vary by type and size of print. If print files need to be restructured to fit printer dimensions, an additional $50/hour fee is charged. Faculty and other campus labs using the GIS and RS software are charged an annual software user fee, which includes the license, installation, and troubleshooting support. Software fees vary by application. Besides the typical GIS and RS services, the center actively pursues related grants, collaborates on research, participates in teaching GIS classes and labs, and coordinates a graduate GIS certificate.

The FIU GIS-RS mission statement summarizes campus GIS efforts and the center's centralized role:

> GIS-RS Center at the FIU Green Library primarily supports all the computing and research needs in the areas of Geographic Information Systems, Remote Sensing, geo-statistical analysis and Computer-Aided

> Design (CAD) for the FIU community. Consulting services include assistance with geographic information systems, data analysis and extraction, statistical software, scientific visualization and remote sensing software. We also provide training in the uses of desktop GIS programs and organize seminars and workshops in GIS. The GIS Center has two physical divisions, the Teaching Lab and the Research Lab. The High Performance Database Research Center, International Hurricane Center, South Florida Environment Research Program, FIU Department of Geology, Environmental Studies, and Landscape Architecture and Design all help to equip the lab with hardware, software and expertise enabling the center to fulfill its goals and objectives effectively. (FIU, 2006)

Distributed Model: University of Florida

The University of Florida, which recently celebrated its 150th anniversary, has an established history of conducting GIS and remote sensing activities on campus. In 1984 the Geo-Facilities Planning and Information Research Center (GeoPlan) was established. The College of Design, Construction, and Planning (formerly the College of Architecture), in response to UF's growing GIS research and teaching needs on campus, created GeoPlan in the Department of Urban and Regional Planning. The center is still very active today and serves as a spatial data node on the National Spatial Data Infrastructure (NSDI) network. GeoPlan is widely acknowledged throughout the state of Florida for the creation and maintenance of the Florida Geographic Data Library's (FGDL) 350+ GIS layers. FGDL grew out of a joint Florida Department of Environmental Protection (FDEP) grant project between the UF Map and Imagery Library and GeoPlan Center. Since 1989 various departments on campus have been conducting GIS and remote sensing activities. GIS and RS activities at UF span several colleges (Engineering, Architecture, Liberal Arts, and Agriculture), and research labs are distributed over many departments (Civil Engineering, Environmental Engineering, Computer Science, Electrical Engineering, Urban and Regional Planning, Landscape Architecture, Geology, Geography, Forestry, Botany, Wildlife, Soil Science, Food and Resource Economics, Entomology, Anthropology, and others). In the early 1990s UF became one of the first universities to have a university-wide Environmental Systems Research Institute (ESRI) ARC/INFO site license, and initially software license costs were shared by individual departments. In the late 1990s the UF provost's budget began paying for the license and GIS became freely available to all departments.

Prior to the late 1980s the approach to higher education at UF was to eliminate redundancy in academic courses and programs. If one program taught a subject or had an area of expertise, other programs on campus were discouraged from pursuing those same subjects. Between the late 1980s and mid- to late 1990s, a former university president implemented a business model approach to higher education. Colleges and departments

were rewarded for the number of class seats filled and the amount of research dollars in their laboratories.

The cultural shift, combined with freely available GIS software, resulted in a surge of GIS and RS courses and research efforts on campus as well as fierce competition between colleges and departments. GIS evolved as a significant focus in many disciplines, and some departments sought recognition as UF's expert. The constant competition caused a former provost to create UF's Interdisciplinary Concentration in GIS (ICGIS) Committee and graduate-level certificate program. Because the Head Map and Imagery Librarian and the GIS Librarian of the UF Library actively participated as ICGIS committee members and because the library had no GIS or RS center of its own, the library became a neutral committee meeting site. Eleven academic departments with GIS and RS components now participate in the certificate program.

In 2000 UF's George A. Smathers Libraries Government Documents Department hired its first faculty GIS Librarian. Establishing a GIS Librarian position was a five-year administrative process. The major driving force for the position was numerous research and student requests for spatial data and processing, mostly related to census data. The first release of FGDL in 1998 and the Map Library's large digital spatial collection also contributed to the need for GIS expertise. Additionally, there was a need on campus for an "in-house" GIS and RS consultant for faculty, staff, and students not associated with a particular research or teaching lab on campus. Besides consulting with faculty, students, and staff, the GIS Librarian teaches courses in the Geomatics Department. Due to the increased demand for GIS consultations, in 2005 a vacant government documents position was converted to a faculty spatial and numeric data librarian, who will actively collect and distribute data.

GIS and RS service and data delivery at UF libraries now includes six public computer workstations, two 11" x 14" scanning stations, four research and consulting workstations, a large format grey-scale scanner, a color laser printer, and a large format plotter. Software includes the Microsoft Office Suite, sixty ArcGIS license seats, and three seats of Leica's ERDAS Imagine software. The GIS Librarian coordinated a shared campus-wide purchase of Leica ERDAS Imagine image processing software. All services and equipment use are provided free to patrons. However, consultations for large projects or involvement in data creation for grants are subject to negotiation. The goal of the GIS Librarian is to provide patron guidance in GIS education in order to enable patrons to produce their own products.

GIS and RS projects at UF libraries have evolved into a collaborative and distributed team of individuals from multiple library departments. The Government Documents GIS Librarian coordinates acquisition, development, and distribution of spatial data in addition to maintaining an Internet Map Server (IMS). The Map and Imagery Library houses paper

and digital collection materials. The Digital Library Center scans, archives, and distributes images that can be used in GIS. Library Systems assists in server maintenance, database development, and Web site programming. Cataloging works to incorporate digital collection records into the library catalog. Lastly, the Florida Center for Library Automation (FCLA) hosts imagery.

The UF Library's mission statement summarizes a commitment to a distributed set of services:

> The mission of Smathers Libraries is to support the university community in its pursuit of knowledge, thus contributing to the advancement of the University of Florida to a ranking among the top 10 public universities. Our strategic goals aim at providing excellent information resources and tools to faculty, students, and staff when and where they are needed. We are committed to a service model that imbeds library resources in academic programs, reaches out and collaborates with scholars and others within and beyond the university and adapts quickly to take advantage of technology developments. (University of Florida, 2005)

While the mission of a university library may be to support the academic community in its pursuit of knowledge, namely research and teaching, strategic goals help shape development of services provided and how they are delivered. In this instance goals that deliver "information resources and tools . . . when and where they are needed" and a service model that "imbeds library resources in academic programs, reaches out and collaborates with scholars and others within and beyond the university" are consistent with distributed models of service.

Conclusions and Trends

Academic GIS and remote sensing teaching and research efforts are growing on large and small campuses across the country. Increasingly, academic libraries regardless of size are developing spatial data services to meet user needs (Kinikin & Hench, 2005). Many library GIS and RS efforts begin as part of another library department, typically government documents or a map library; however, specialized science-oriented libraries are also typical incubators for these efforts. UF libraries' GIS efforts are organized under the Government Documents Department. At FIU, GIS and remote sensing started in Government Documents, but spatial data services have become a unique library department. While university libraries often consider the placement and scope of GIS services, the service and data delivery model are not developed in relation to existing campus enterprise GIS infrastructure.

This article has examined two spatial service and data distribution models, centralized and distributed, that can be applied to any size academic library. By comparison, campus GIS and RS activities and expertise at FIU

began in the library, and as a result FIU libraries developed a GIS-RS Center similar to UF's Urban and Regional Planning GeoPlan Center. In the case of FIU and the centralized model, the cost of maintaining a library center is partially shared by other departments. However, when other college departments already have distributed research centers and teaching labs, such as at UF, they are not motivated to contribute to a library center. Creating a library GIS center in a distributed environment of existing research and teaching labs may force a library to compete against other centers for research grants and budget line item funding. By focusing on the user community and adding coordinate information, or spatial value, to library collections, libraries may be able to avoid competition with other campus centers. Consequently, UF has evolved into a distributed model for GIS service and data delivery that relies on many library departments.

Although FIU and UF libraries embody different service models, GIS efforts of both libraries are centralized in a single library. At UF this raises the administrative question of how to best meet user needs for nine separate satellite library locations, eight different specialized library collections, and over eleven academic units. One current consideration consistent with a distributed network of services is basic GIS and RS education for a select number of reference librarians throughout the libraries. A concern with this approach is the significant amount of time needed to achieve a sufficient level of technological knowledge in hardware, software, and spatial data to meet a user's particular need.

Providing spatial technology services is often seen as prestigious and may encourage some academic libraries to invest in GIS and RS centers; or libraries may feel compelled to keep pace with other university departments or academic institutions. Prior to investing in GIS centers and services, evaluating library enterprise GIS service models in the context of the broader university enterprise will help libraries plan and implement spatial information services that meet a patron's needs.

References

Florida International University (FIU). (2006). *Mission statement.* Retrieved March 23, 2006, from http://gislab.fiu.edu/portal.

Haas, S. C., Kesse, E. J., Aufmuth, J. A., Sullivan, M. V., & Renner, R. D. (2005). From the air: The photographic record of Florida's lands. *OSS: International Digital Library Perspectives, 21*(2), 131–139.

Haas, S. C., Aufmuth, J. A., Coleman, D., & Uhlinger, E. (2002). MARC and ARC: Geospatially enabling bibliographic records. *Proceedings from the 28th annual IAMSLIC conference, Mazatlan Mexico* (pp. 119–126). Retrieved January 29, 2006, from http://www.iamslic.org/proceedings_papers/proc02119.pdf.

Kinikin, J., & Hench, K. (2005). Survey of GIS implementation and use within smaller academic libraries. *Issues in Science & Technology Librarianship, 42.* Retrieved February 6, 2006, from http://www.istl.org/05-spring/refereed-1.html.

Koontz, C. M., & Jue, D. K. (2001). The location of your library building: Why it is important, and how to do it, using GIS (Geographic Information System software). In M. F. Bisbrouck, *Library Buildings in a Changing Environment,* IFLA Publications 94 (pp. 141–153).

McCormick, B. G. (2003). Developing enterprise GIS for university administration: Organizational and strategic considerations. *New Directions for Institutional Research: Using Geographic Information Systems in Institutional Research, 120,* 63–76.

Rich, S., Das, A., & Kroot, C. (2001). Spatial data management in an enterprise GIS. *Proceedings of the 2001 ESRI User Conference.* Retrieved January 29, 2006, from http://gis.esri.com/library/userconf/proc01/professional/papers/pap742/p742.htm.

University of Florida. (2005). *UF Library mission statement.* Retrieved May 19, 2006, from http://www.uflib.ufl.edu/staff-site/budget/ProgramReview05-06.doc.

Xia, J. (2005). Locating library items by GIS technology. *Collection Management 30*(1), 65–74.

Joe Aufmuth is a faculty GIS Librarian and the interim Head of the University of Florida's George A. Smathers Map and Imagery Library. He currently serves as one of ALA MAGERT's Cartographic Users Advisory Council (CUAC) representatives and is also a member of the American Society for Photogrammetry and Remote Sensing (ASPRS). He earned an M.S. degree from UF's geomatics program in civil engineering and a B.S. in ecology, ethology, and evolution from the University of Illinois at Urbana-Champaign. He has seventeen years of experience working with GIS and remote sensing in academic research and environmental consulting.

GIS Mentoring

Kim M. Ricker

Abstract

The implementation and development of effective mentoring is crucial to the growth and success of Geographic Information Systems (GIS) librarianship and staffing. Mentoring is necessary to fill the gaps for all staff members, especially students. I propose that mentoring can assuage many of the staffing obstacles to successful GIS programs. Effective mentoring will create a healthy and productive work environment as well as nurture future GIS librarians and staff members. Although mentoring within librarianship has been discussed in library literature, unfortunately work on the area of GIS mentoring is lacking. This article discusses the benefits of mentoring and demonstrates how the need for GIS mentoring is different from other library mentoring, specifically because of staffing. It also offers strategies for effective GIS mentoring.

Introduction

"To teach is to learn." This Chinese proverb embraces the cyclical nature of mentoring, in which knowledge is gained by everyone in the mentoring relationship—student and teacher, protégé and mentor. Unfortunately, it is not always easy to know how to teach and mentor effectively. While literature exists about mentoring in relation to libraries, this is not the case for mentoring in relation to Geographic Information Systems (GIS). As GIS is a relatively new discipline, circumstances regarding staffing, financial obligations, time obligations, and isolation set it apart from other areas of library mentoring. This article discusses these differences and offers practical strategies for implementing an effective GIS mentoring program.

LIBRARY TRENDS, Vol. 55, No. 2, Fall 2006 ("Geographic Information Systems and Libraries," edited by Jaime Stoltenberg and Abraham Parrish), pp. 349–360

Mentoring for Libraries

Simply defined, a mentor is "someone who helps someone else learn something that he or she would have learned less well, more slowly, or not at all if left alone" (Bell, 2000, p. 133). This definition can be applied to everything that many of us, as librarians, do on a daily basis with all patrons, and it needs to be tailored to working with students and other assistants. A library is a business—to the extent that it has financial concerns, competition, and employees and offers services. Therefore, it is appropriate to embrace a business-oriented definition. Richard Luecke, author of *Harvard Business Essentials: Coaching and Mentoring: How to Develop Top Talent and Achieve Stronger Performance,* redefines mentoring as "a means of developing human resources" (2004, p. xi). Libraries, although not always thought of as such, are corporate entities and should be investing in ways to develop human resources. In order to offer high-quality services, attention must be spent on recruiting, developing, and keeping staff members. With students, mentoring plays a key role in demonstrating why librarianship is a good career choice, developing skills and knowledge, and creating a positive work environment that may reduce turnover.

The benefits of mentoring are numerous and diverse for the mentor and the recipient of the mentoring (referred to as the protégé in this article). Benefits, however, apply to more than just those involved in the mentoring relationship. In his book *The Mentoring Manager,* Gareth Lewis (1996) divides the beneficiaries into three categories: protégé, mentor, and organization (see Table 1). Gail Munde (2000) confirms and adds to Lewis's list of benefits in her article, "Beyond Mentoring: Toward the Rejuvenation of Academic Libraries."

> Positive outcomes for protégés have included higher salaries, promotions, overall career satisfaction, and satisfaction with their organizations. Benefits for mentors include a renewal of professional purpose, a briefing in new or emerging aspects of a profession, a sense of satisfaction that one has helped to influence the future of the profession, the loyalty and support of the protégé, and recognition for the mentor's ability to identify and advise promising employees who will contribute to the organization. Positive outcomes to organizations have included increased employee retention, reduced turnover, faster and more efficient introduction of junior employees to organizational norms and expectations, and improved coherence of leadership through the organization. (Munde, 2000, p. 172)

The organizational benefits—although commonly overlooked when considering mentoring—are the foundations of a successful library department or program.

GIS Mentoring Versus Other Library Mentoring

GIS librarianship is specialized and highly technical. Although mentoring is mentioned in library literature, a review of the literature did not find

Table 1. Mentoring Benefits Lewis Finds for the Protégé, Mentor, and Organization

Recipient	Type
For the Protégé	Greater understanding of the total organizational perspective Personal benefits that come from relationships Learning outcomes developing his/her own learning to learn skills Career benefits Learning problem solving and problem-solving approaches Self-analysis of strengths and weaknesses
For the Mentor	Widening the perspective of business functions Increases in personal satisfaction that come from relationships Role enhancement and the expansion of repertoire of skills Personal development Career enhancement
For the Organization	Better-trained staff Development of an organizational culture Effective management development A positive orientation to learning Empowered staff

Source. Based on Lewis, 1996, pp. 13-15.

any mentoring models that fit the circumstances faced by GIS librarians and staff. For instance, the mentoring model used by the Louisiana State University Libraries, although successful, is not appropriate for GIS mentoring for two main reasons: goals and audience. The audience is librarians, and the goal is promotion and tenure. This is the case for the majority of mentoring articles that exist. When contemplating mentoring in relation to GIS, goals and audience must be taken into account.

The current literature is filled with general articles and those that address issues faced by other specializations such as minority librarians and academic librarians facing tenure. Although many GIS services are offered in academic libraries, because of the technical aspect of GIS, it was difficult to draw close comparisons between mentoring in these libraries and that which should be offered in relation to GIS. The closest relationship to GIS found was an article geared toward medical reference librarians in academic libraries. Hongjie Wang (2001), in "Academic Mentorship: An Effective Professional Development Strategy for Medical Reference Librarians," provides an excellent literature review of academic mentoring and captures the technical aspect of medical librarianship. Most relevant to GIS is mention of a 1990 national survey of 210 health sciences librarians affiliated with 70 academic medical libraries in the United States. The survey indicated that the specialized skills necessary for professional medical librarianship were acquired on the job. Wang writes, "This survey result supports the popular belief among information services professionals that, while an absolute prerequisite for the profession, a master's degree in library and information science is not equivalent to the skills of a competent medical reference librarian" (2001, p. 26). I believe that this is also the case for GIS librarianship. Although the foundation of librarianship, which includes

critical reference and database skills, is learned in library school, many skills required for a GIS specialization are not.

Unfortunately, the literature on medical librarianship does not completely apply. The most obvious difference is that medical librarianship is more established. Medical librarians are often in separate libraries that are more fully staffed. Although GIS librarians are often in map libraries, staffing numbers are not commonly equivalent. Adequate staffing of GIS services is vital. In his article "GIS Collection Development, Staffing, and Training," Karl Longstreth states, "For an academic library (indeed, any library) with spatial data from the government or other sources, the basic need for implementing a GIS will be to provide access to those data. The development of staff to provide this access is crucial" (1995, p. 270). Despite the critical need for staffing, GIS services are often given by a mélange of providers: students, nonlibrarian staff, and GIS professionals—in addition to librarians. GIS professionals are those individuals who have come from another job in which they worked in the field of Geographic Information Systems. These include managers, developers, consultants, data collectors, sensor and system developers, or academics and researchers (GIS Professional, n.d.). Staffing has been an important issue associated with GIS services since they began. In her 1995 article "Expanding Horizons for GIS Services in Academic Libraries," Carolyn D. Argentati asserts: "Development of GIS services has involved, and in some cases, transformed, the efforts of government documents librarians, map librarians, reference and subject specialist librarians, and others throughout the academic library organization. In this process, libraries have had to address many questions regarding staffing and management of these new services and the allocation of resources to support them (p. 463).

Although GIS technology and understanding has grown, issues regarding staffing have not changed significantly. At this point in time, I believe that staffing is the key difference between GIS librarianship and other technical forms of librarianship. Mentoring is a key way to develop staffing in order to provide a quality GIS program.

Issues Surrounding GIS Staffing

Demands on GIS staff are high. Argentati (1995) sets the stage for GIS staffing concerns when discussing "a series of developments [that] began to intensify the demands on GIS Team members" during the initial launch of GIS services at North Carolina State University (NCSU) Libraries.

> First, the number of GIS users on campus was increasing rapidly, partly because of campus licensing agreements with several major software vendors. Second, the software itself was continually evolving, and learning new GIS and data applications or staying up-to-date on changing functionality requires considerable time and effort. Finally, the proliferation of spatial data available to be collected, along with data-related

> issues such as network storage infrastructure for multi-platform access and WWW-based documentation and retrieval mechanisms, demanded sustained attention and activity. (p. 464)

These basic issues have not changed. In fact, they have multiplied. Issues regarding GIS staffing are complex and include a number of factors: financial responsibilities, a diverse set of service providers, time obligations, lack of training, and isolation. In this article, the phrase GIS staffing refers to all staff (librarians, nonlibrarian staff, and students) that provide assistance with GIS. The first issue associated with GIS staffing is the high financial obligation associated with running a GIS program. In order to meet the other financial costs for hardware, software, and data programs, the costs of staffing are often minimized, which results in low total staffing numbers and low numbers of librarians. In addition, staffing of GIS programs is not always full-time. Staff members often have other responsibilities that are not related to GIS. Further, GIS services are fairly new. Although library literature began to discuss GIS with some frequency in the late 1990s, not all libraries offer GIS services. In fact, many libraries are just beginning to start GIS programs. Even those institutions with established programs have small programs and are faced with financial issues that may prohibit rapid growth. Lastly, because of the low number of GIS staff, the issue of isolation is common in GIS staffing.

There are five components to a GIS: hardware, software, data, people, and methods. Each component is critical and the first four require a level of financial obligation. Computer *hardware* consists of input devices such as keyboards, scanners, and digitizers; output devices such as monitors and printers; and processors. Libraries commonly have public computers, but computers with GIS may require extra devices such as scanners, digitizers, or color printers. In addition, GIS software may require higher processing speeds for the hardware. GIS *software,* which is commonly upgraded almost every year, can also be costly. Larger institutions with other departments on campus who use GIS may benefit from sharing the cost of a site license. This, however, is not always the case. Although geospatial *data* are distributed at no cost to Federal Depository libraries, it is not comprehensive. Most geospatial data that is distributed through the Federal Depository Library Program (FDLP) were created for a specific project and are not applicable for most general use. Therefore, purchasing commercial data for foundational purposes, such as boundary files, or to fill in gaps is common. Lastly, *people* are needed to use the GIS. Because using a GIS is sometimes very technical, people with GIS knowledge are needed to assist and teach others. As GIS services are not common enough yet, in many libraries there is not enough demand to justify libraries devoting finances to hire a full-time or multiple full-time GIS librarians or staff. As a result, GIS staffing is often filled by a diverse set of individuals with different backgrounds: full-time/part-time, and those with or without a geography/GIS background or

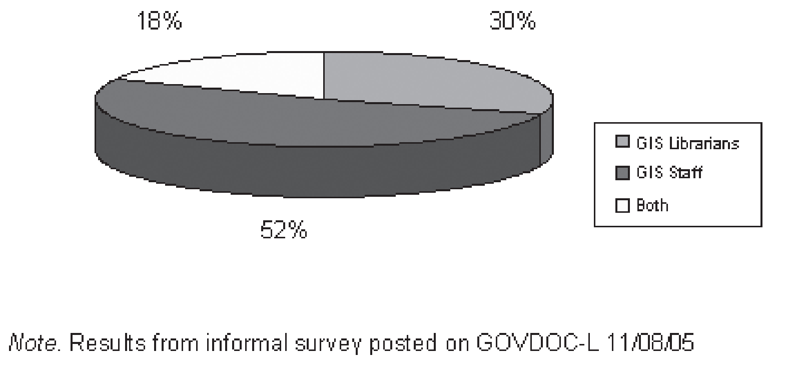
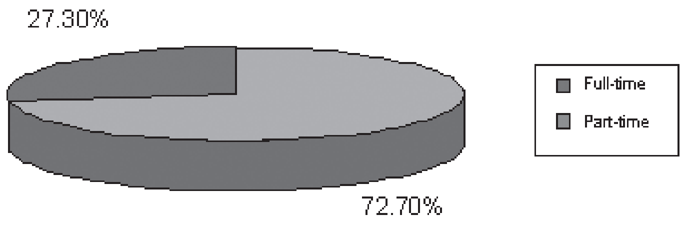


Figure 1. Breakdown of GIS Staffing in Federal Depository Libaries: Librarians and Staff

M.L.S. Many programs have a combination of all of these and rely heavily on students and volunteers.

In addition to the challenges of the financial obligations, GIS services and staffing are constrained by the concerns of sharing time with other responsibilities. Staffing issues for librarians and staff members trickle down to students who assist with GIS. In order to learn more about GIS staffing and libraries, I conducted an informal survey to depository libraries through GOVDOC-L, a "listserv-based discussion forum about government information and the Federal Depository Library program" (GOVDOC-L, n.d.). Although, as indicated above, GIS services and hence issues related to GIS staffing and mentoring pertain to libraries outside of the FDLP, this scope proved to be a good starting point.[1] The survey netted a 5 percent response rate (73 out of a possible 1,232 depositories). Of these, 33 of the 73 respondents indicated having GIS services. It is interesting to note that 6 of 43 libraries that responded they did not have GIS services mentioned that GIS services were planned for an undisclosed date in the future.

Although GIS services are staffed by librarians, nonlibrarians that include GIS professionals, and students, the survey only asked for a distinction between librarians and nonlibrarians (see Figures 1 and 2). Of the libraries indicating GIS services, 30.3 percent (10 libraries) had GIS librarians, with the responses divided equally between full-time and part-time. 51.5 percent (17 libraries) answered yes to GIS services and no to a librarian. The remaining 18.2 percent (6 libraries) indicated that they have GIS services provided by both a GIS librarian and a staff member in some capacity. For the libraries who did not have a librarian, the majority (17 libraries or 77 percent) had one part-time staff member. Many respondents specified that they are only able to designate a portion of time to GIS as they are responsible for the map collection or other duties. Although GIS services are being



Note. Results from informal survey posted on GOVDOC-L 11/08/05

Figure 2. Full-time versus Part-time Staffing of GIS Services in Federal Depository Libaries

more widely offered, 72.7 percent (24 libraries) have one librarian or staff member working only in a part-time capacity as compared with 27 percent (9 libraries) that have full-time GIS staffing. While these percentages are by no means indicative of GIS services on the national or international levels, they display a pattern of disjointed GIS staffing that I believe to be true for libraries in a broad sense. As the survey shows, many staff members responsible for GIS have other duties and obligations.

Learning and maintaining the skills necessary to manage GIS software is very time intensive. If staff resources are spread too thin there is not adequate time to manage GIS duties. Although this is changing with institutions now hiring GIS-specific technicians and librarians, many of those currently responsible for GIS services had these responsibilities added to already existing duties. Because the area of GIS is technical, time must be spent gaining personal knowledge and honing software skills. “A library should establish a set of goals specifically for GIS implementation. . . . Including a goal for training is important; as GIS software and databases are sufficiently different from other library resources, most staff and users will not effectively use GIS without instruction” (Longstreth, 1995, p. 271).

In addition to the scarcity of time, many librarians or staff may feel they are not adequately trained. Longstreth lays out these areas of knowledge and skills needed to provide GIS services:

> Library staff need to understand, and be proficient in, several areas. Training implies learning to use GIS software, but it is important to have a conceptual understanding and knowledge of real GIS applications in order to make training useful. Staff must know more than how to operate the GIS software; they need instruction in the issues of GIS theory, GIS databases, and GIS applications in a discipline. This instruction is necessary because the ability to add, manipulate, and analyze data in a GIS intelligently requires understanding; the inputs and processes

> needed to yield a meaningful result are a function of employing a GIS in an intellectually appropriate way. (1995, pp. 271–272)

Even those who have a strong background in geography or another field using GIS may feel challenged by new developments in software and resources. Because in many libraries there is only one person with knowledge of GIS, current librarians and staff may also experience isolation. Mentoring may feel like the last thing one has the resources to do successfully. There are a few things that can be used to resolve these problems. The following section discusses additional roles to be included in the definition of mentoring and suggests strategies for effective mentoring.

Suggested Strategies and Structure

The definition of mentoring must be molded to fit overall goals of GIS programs and the appropriate audience. At this time, I believe the goals of GIS staffing are to build future GIS librarians and professionals and gain assistance in providing services. The audience (protégés) for GIS mentoring is primarily students. Strategies presented here can be modified for others, however, based on goals for the relationship and amount of time invested. Taking into consideration the goals and audience for GIS staffing, the standard definition of mentoring is too vague. For GIS, coaching and training should be part of the mentor role in order to train assistants in necessary skills for the job. Luecke states, “Coaching is about your job; mentoring is about your career” (2004, p. 78). Similarly, in “Formalizing an Informal Process: Developing a Mentoring Program in a University Library,” Catherine Wojewodzki, Linda Stein, and Tommy Richardson state, “Technical, teaching, and management skills are obtained with time on-the-job and guidance from a supervisor, but professional development is a different, more personal process that can be accelerated by mentoring assistance from experienced colleagues” (1998, p. 442). These definitions of mentoring assume that the protégé is invested in a career already. For GIS, although many students/staff are devoting time to GIS, it may be on a more temporary basis. The initiative in mentoring comes from the person seeking greater understanding. “That person—the protégé—must take responsibility for his or her own growth and development” (Luecke, 2004, p. ix). Although many students, especially graduate students, take their jobs seriously, their main priority is completing their education. It is incorrect to assume that they have decided on GIS as their chosen career and that they will take the initiative to learn both GIS and library skills on their own. In the case of GIS services, which can be very technical, I believe that a combination of many roles addressing job and career development—including both mentoring and coaching—should be taken. In addition, the roles of process advisor and consultant should be added as well. A mentor should take on these roles to help the learner set objectives, manage time, monitor progress, and check skills (Lewis, 1996, pp. 90–93). While coaching is

very specific, the roles of advisor and consultant examine the larger role that the protégé fulfills.

With turnover a natural part of using student assistants, many people may question the efficacy of investing time and resources in mentoring. This is an issue for all jobs. Mentoring is most time-consuming in the beginning stages of development. Once a structure and plan are formed, time is more pleasantly spent developing a relationship with the student(s). Demonstrating that working with GIS is challenging, rewarding, and fun will reduce turnover and address staffing issues for the future. The long-term goal is for students to realize that the library is not simply a place for part-time employment but a possible, and very rewarding, career option.

A ten-point framework for mentoring is presented below. This has been developed and implemented over the past year and a half at the University of Maryland with graduate assistants, student assistants, and volunteers. I have found it particularly effective.

Set Goals

When faced with the task of offering GIS services with limited staffing, it is important to look beyond a growing list of small everyday jobs that need to be fulfilled. While this list can be helpful in designing a plan, it is crucial to look at the overall picture of what the protégé will be accomplishing both for the library and personally. For example, if it is a priority to offer GIS instruction for the campus, it is important to envision what role the protégé will play in it.

Have Regular Meetings with an Agenda

It is necessary to have regular meetings to answer questions, get feedback, and check on the progress of the protégé. In order to make meetings easier to remember, plan them for a set time and day of the week. To be productive, these meetings should have an agenda with a plan, future tasks with deadlines that are mutually agreed upon, and a rough plan for the next meeting. For instance, set an agenda for meeting A that has a plan for training or discussion for that meeting and a list of tasks outlined clearly with specific deadlines. It makes sense to have the deadlines set for meeting B or a day or two before meeting B so you have time to look over the results of tasks, for example, a list of GIS Internet resources. The agenda could also have an estimated plan for the next meeting. This allows you to look ahead. For instance, if you are using a book like Getting to Know ArcGIS (Burke et al., 2004) for training, meeting A would discuss/demonstrate components covered in chapters 1–3 as a preview, assign the task of reading chapters 1–3 with a deadline of meeting B, and plan for meeting B to preview chapters 4–6. Allow for flexibility. You may find that the agenda needs to be revised if demonstrations or questions run longer or if the protégé has been working through the instructional material faster than originally planned. After the meeting give the protégé a final agenda

for the next meeting. It helps to create a long-term adaptable plan before beginning work with the protégé. While advance planning takes a considerable amount of time, keep in mind that it will make creating productive agendas easier and cut down on time in the future when developing training for new students or protégés. When planning, remember to keep the goals you have set in mind. This will cut down on time lost to ideas outside of what you want to accomplish.

Challenge the Protégé

We often learn best and enjoy our jobs more thoroughly when challenged. Give the protégé the confidence that you believe that he/she is able to complete the tasks and challenges that you have assigned him/her. This said, however, take care not to push the protégé beyond what he or she can accomplish. Keep in mind that with students, the number one priority is for them to earn a degree. Have an open dialogue about tasks and a flexible agenda. It is ideal to incorporate tests or demonstrations into the agenda. For instance, after previewing chapters 1–3 and assigning reading chapters 1–3 in meeting A, have meeting B begin with questions about chapters 1–3 and then have the protégé demonstrate for you key items from the chapters that you have already laid out in the agenda. In addition, after a certain period of time have a more comprehensive test. This is to determine what areas need more training and give the protégé confidence in what he/she has learned. Give the protégé a list in advance of what will be covered so he/she has a way to prepare. It is helpful to have tests allotted in the plan you have created. This way the protégé will be clear of the overall goals for training and that the tests are not the result of how you feel about how he or she is progressing.

When You Learn, Teach Them

Mentoring is a learning process, for both the protégé and the mentor. As stated earlier, many GIS staff may feel as if they do not know enough to train a protégé. This is an ideal time for the mentor to learn, both in advance of training and with the protégé. Share new concepts and resources that you have learned and encourage the protégé to do the same. Working in a library, there should be no shortage of print and electronic resources for training. In addition, many academic campuses have site licenses for software, which include free training through resources like the Environmental Systems Research Institute (ESRI) Virtual Campus.

Take the Protégé to Meetings

I highly recommend attending professional meetings and conferences, especially those that are informal (and less costly) and that deal with GIS. The content is extremely useful and creating friends and contacts in the fields of GIS and libraries is invaluable. Take your graduate and undergraduate students and staff with you to appropriate meetings. In some cases this may

be difficult because of finances. If possible, work with your library to get additional funds for graduate assistants. If this is not possible, informal meetings, like ESRI user group meetings, are free. Training students creates the opportunity and structure to learn new things and hone skills. Mentoring "is a means of developing human resources . . . The mentor acts as a trusted guide, offering advice when asked and opening doors to learning opportunities when possible and appropriate" (Luecke, 2004, p. xi). In addition to learning new things, the protégé will be able to make contacts in the field, which is valuable if he/she is considering GIS librarianship as a career.

Be Patient

Mentoring is not a quick process. It involves developing a relationship. Be patient with both the protégé and with yourself. As you gain more experience as a mentor, it is possible to re-use the templates already created. Keep in mind, however, that each protégé is an individual and that his or her learning style may differ from yours or other protégés. Be flexible and modify your plan to accommodate the way that he or she learns best.

Give Protégés Room to Develop Their Own Areas

As long as it fits into the overall goals, allow the protégé to develop an area of interest. For instance, if one of the program goals is to develop GIS instruction and she is interested in history, encourage her to find literature or other resources relating to GIS and history and develop ideas for incorporating it into campus instruction.

Provide Group Training Sessions

Create a staff training session for people who are not GIS-designated staff but may have to provide some assistance with the GIS computers in your absence. This session should not be comprehensive but rather address common issues that they may need to deal with without your help. Make sure you provide a handout for their reference.

Be Available for Questions

Whether it is related to the library or to GIS, there is a plethora of new things that the protégé will be learning. It is crucial for you to be available—and make it known that you are available—to answer any questions. This will strengthen the relationship and make the learning process proceed more smoothly.

Be Actively Involved

Mentoring is a relationship. In the end, "the development of a working relationship requires the active participation of both its parties" (Portner, 2002, p. 5). While you cannot always rely on the active participation of the protégé, you can encourage it and provide a good example with your dedication to the relationship.

Conclusion

Mentoring geared toward GIS is severely needed in order to combat staffing shortages and nurture a positive learning and working environment for the department. More importantly, GIS mentoring is crucial for the survival of GIS services in the library. GIS mentoring has not been previously discussed in library literature. Although literature about mentoring in relation to health sciences or libraries in general is helpful, these examples are limited in their relevancy to GIS. The strategies provided are ones personally developed at the University of Maryland and are presented in hopes that they will begin a dialogue on effective GIS mentoring. Comments are welcome and encouraged.

Acknowledgements

Special thanks to Jeffrey DiScala and Chellammal Vaidyanathan.

Notes

1. The questions included on the survey were (1) Does your library have GIS services (yes or no?); (2) Does your library have a GIS librarian (yes or no); and (3) What is the total number of staff devoted to GIS at your library?

References

Argentati, C. D. (1995). Expanding horizons for GIS services in academic libraries. *Journal of Academic Librarianship, 23*(6), 463–469.

Bell, C. R. (2000). Mentoring as partnership. In Marshall Goldsmith, Laurence Lyons, & Alyssa Freas (Eds.), *Coaching for leadership* (pp. 131–142). San Francisco: Jossey-Bass/Pfeiffer.

Burke, R., et al. (2004). *Getting to know ArcGIS desktop*. Redlands, CA: ESRI Press.

GIS Professional. (n.d.). *GIS Professional Magazine*. Retrieved January 17, 2005, from http://www.pvpubs.com/gisprofessional_magazine.asp.

GOVDOC-L. (n.d.). *About GOVDOC-L*. Retrieved December 1, 2005, from http://www.govdoc-l.org.

Lewis, G. (1996). *The mentoring manager: Strategies for fostering talent and spreading knowledge*. London: Pitman Publishers.

Longstreth, K. (1995). GIS collection development, staffing, and training. *Journal of Academic Librarianship, 21*(4), 267–275.

Luecke, R. (2004). *Harvard business essentials: Coaching and mentoring: How to develop top talent and achieve stronger performance*. Boston, MA: Harvard Business School Press.

Munde, G. (2000). Beyond mentoring: Toward the rejuvenation of academic libraries. *Journal of Academic Librarianship, 26*(3), 171–176.

Portner, H. (2002). *Being mentored: A guide for protégés*. Thousand Oaks, CA: Corwin Press.

Wang, H. (2001). Academic mentorship: An effective professional development strategy for medical reference librarians. *Medical Reference Services Quarterly, 20*(2), 23–31.

Wojewodzki, C., Stein, L., & Richardson, T. (1998). Formalizing an informal process: Developing a mentoring program in a university library. *Technical Services Quarterly, 15*(3), 1–19.

Kim Ricker is the GIS/Data Librarian in the Government Documents and Maps Department at the University of Maryland Libraries. Before completing her M.L.S. at the College of Information Studies at the University of Maryland, she was a GIS graduate assistant (and protégé) at McKeldin Library. She has firsthand experience on both ends of the mentoring relationship.