
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

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-

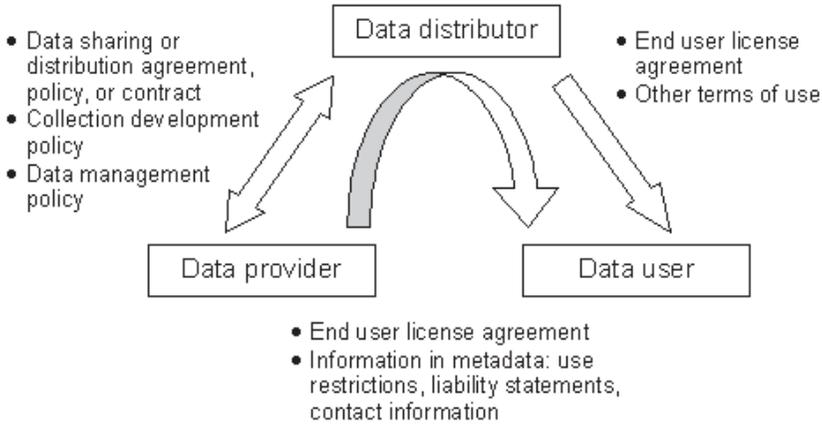


Figure 1. Communication between Participants in Geospatial Data Distribution

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 Global Biodiversity Information Facility (GBIF)	Geospatial Various (biodiversity)	Distribution Distribution	Geospatial One-Stop, n.d. Global Biodiversity Information Facility (GBIF), n.d.a; Global Biodiversity Information Facility (GBIF), n.d.c
Global Biodiversity Information Facility (GBIF)	Various (biodiversity)	Usage	Global Biodiversity Information 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 Center for 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-

STEINHART/DATA MANAGEMENT POLICIES

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/programs/Committees/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/GBIFdataIPprinciples.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>.
- USGCRP 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.