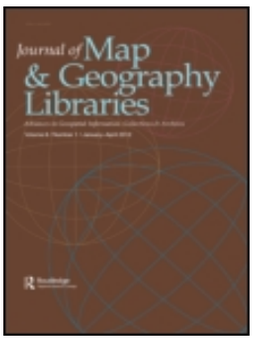

Routledge
Taylor & Francis Group

Journal of Map & Geography Libraries
Advances in Geospatial Information, Collections & Archives

ISSN: 1542-0353 (Print) 1542-0361 (Online) Journal homepage: http://www.tandfonline.com/loi/wmgl20

Placing Data in the Land of 10,000 Lakes: Navigating the History and Future of Geospatial Data Production, Stewardship, and Archiving in Minnesota

Kevin R. Dyke, Ryan Mattke, Len Kne & Shawn Rounds

To cite this article: Kevin R. Dyke, Ryan Mattke, Len Kne & Shawn Rounds (2016) Placing Data in the Land of 10,000 Lakes: Navigating the History and Future of Geospatial Data Production, Stewardship, and Archiving in Minnesota, Journal of Map & Geography Libraries, 12:1, 52-72, DOI: 10.1080/15420353.2015.1073655

To link to this article: http://dx.doi.org/10.1080/15420353.2015.1073655

Journal of Map & Geography Libraries, 12:52–72, 2016
Published with license by Taylor & Francis
ISSN: 1542-0353 print / 1542-0361 online
DOI: 10.1080/15420353.2015.1073655

Routledge
Taylor & Francis Group

Placing Data in the Land of 10,000 Lakes: Navigating the History and Future of Geospatial Data Production, Stewardship, and Archiving in Minnesota

KEVIN R. DYKE, RYAN MATTKE, and LEN KNE
University of Minnesota, Minneapolis, Minnesota, USA

SHAWN ROUNDS
Minnesota Historical Society, St. Paul, Minnesota, USA

KEYWORDS history of geographic information systems, spatial data curation, data curation, Minnesota State Archives, Center for Urban and Regional Affairs, University of Minnesota, Minnesota

INTRODUCTION

Thanks to effective collection development and preservation, when researchers come to the John R. Borchert Map Library at the University of Minnesota today, they can access original plat maps and aerial photographs nearly a century old. Historical spatial data is used for a wide range of planning, academic, and business purposes, ranging from settling land use disputes to genealogy. Yet, unless something is done, library users of the future looking for maps and other spatial data from the late twentieth and early twenty-first centuries will not be as fortunate. This is because over the last several decades governmental agencies have published fewer and fewer paper maps, and in many cases stopped publishing maps entirely. We are left with "born digital" material with no analog counterpart, material that without thoughtful attention to preservation and curation will rot into unusability.

The goal in this article is to provide background and offer a framework for developing a program to archive the state of Minnesota's spatial data at the University of Minnesota. It begins with a review of the literature on geospatial data archiving, preservation, and curation, and goes on to make

Address correspondence to Kevin R. Dyke, University of Minnesota, S76 O Meredith Wilson Library, 309 19th Avenue South, Minneapolis, MN, 55455 USA. E-mail: dykex005@umn.edu

a business case for archiving geospatial data. It next provides a brief history of Geographic Information Systems (GIS) in Minnesota, beginning in the late 1960s with the University of Minnesota's Center for Urban and Regional Affairs' (CURA) Minnesota Land Management Information System (MLMIS) project through the establishment of the Minnesota Geospatial Information Office (MnGeo) in 2009. After providing this historical context, this article builds upon existing organizational infrastructure in Minnesota to sketch a preliminary framework for archiving the state's spatial data. The ultimate goal of the project would be active curation, including format transformations, metadata enhancement, and pathways for discovery. Within the state, we explore MnGeo and the State Archives. At the regional level of the Twin Cities we examine the role of the MetroGIS collaborative. We then describe the situation at the University of Minnesota and suggest that it is an ideal site to coordinate the archiving of Minnesota's geospatial data, preserving it for use by future generations. With the launch of the state's first comprehensive data clearinghouse, the MnGeo-led Minnesota Geospatial Commons, Minnesota is now ready to undertake systematic data archiving.

PRESERVING AND ARCHIVING GEOSPATIAL DATA

Review of Literature

Digital maps and geospatial data have become "ubiquitous artifacts in the twenty-first century, even though the digital era in map making is a relatively recent phenomenon of the last 50 years," a result of decreased costs, more accessible software, and the rise of web mapping (Lauriault and Taylor 2014, 326). In spite of this ubiquity, or perhaps because of it, these sources of knowledge are frequently not being preserved in a systematic manner.

As map libraries began to grapple with digital materials, issues involving data formats, hardware, software, and user training emerged. Twenty years ago, Larsgaard (1995) predicted the ubiquity of spatial data and also persistent issues around the lack of data standards. In the years since, many different approaches to providing access have been taken, including local institutional collections, statewide clearinghouses, academic library and university consortia, and national-level discovery services. For any institution, factors such as funding (internal or external), customer needs, current data availability, technology resources, staff time, support from administrators, and preservation issues dictate the appropriate course of action (Kollen et al. 2013).

Aspects of providing long-term access to spatial data include engaging decision makers and users who value historical content, providing researchers access to that superseded content, and taking actions to preserve the usability of that content (Bethune, Lazorchak, and Nagy 2009). The issues that are complicating these efforts are the variety and proprietary nature of

spatial data formats, the difficulty of archiving a continuously updated relational database, the size of datasets (such as satellite imagery or LiDAR point clouds), and the lack of essential metadata (Erwin and Sweetkind-Singer 2009). Further, the terms "archiving" and "preservation" often mean different things to different entities, so while many entities back up their data, as Erwin, Sweetkind-Singer, and Larsgaard put it, "there is a difference between saving bits and truly preserving them" (2009, 493). Other major challenges include data identity (how to know if two files are identical or not, how to uniquely identify data), discovery, and standardization (Ramapriyan et al. 2010).

Another challenge is preserving context. Preserving the data itself is the first step. A second, equally necessary, and sometimes more difficult step is capturing enough contextual information to enable future use of the data. For some objects such as documents using the Portable Document Format (PDF), the format itself provides enough context, but geospatial data often requires more complex contextual information for future reuse (2009).

Related to issues of context is the importance of locally created geospatial data sets. Geospatial data created by state and local government agencies are frequently more detailed and current compared to federal sources (Morris 2009). Morris suggests that this specificity and currency makes archiving these data sets challenging as they are often produced by various agencies across a given state and the data are often updated on a daily or weekly basis. Curation decisions are also often difficult, as "collection development is an educated guess regarding what will be valued later" (Erwin, Sweetkind-Singer, and Larsgaard 2009, 493). The available literature yields many recommendations to address the above challenges archiving schedules, online storage for archival data, format recommendations, consistent file naming, metadata, making data readily available to the public, periodic review of retention policies, and publicizing geospatial retention schedules and archival practices (Morris 2009).

While a substantial body of knowledge exists around issues relating to "data standards, best practices, data sharing agreements, metadata production and harvesting, catalog development, and services integration," these efforts do not encompass archiving and preservation (Morris 2009, 27). That is not to say there is a total lack of work for us to draw upon. Two regional archiving projects provide a foundation for our incipient efforts in Minnesota.

The Geospatial Multistate Archive and Preservation Partnership (GeoMAPP)

The Geospatial Multistate Archive and Preservation Partnership (GeoMAPP) was a collaboration "addressing the preservation of at-risk and temporally significant digital geospatial data content" (Bethune, Lazorchak, and Nagy

2009, 46). The project ran from 2007 to 2011. GeoMAPP emerged from a series of workshops sponsored by the Library of Congress in 2005 on the preservation of digital data produced by state governments.

While GeoMAPP involved collaboration between several states, many of the strategies for coordination can be applied internally to state agencies. In particular, their efforts regarding outreach and community building and how those processes built trust between participating agencies are quite encouraging (Bethune, Lazorchak, and Nagy 2009, 54). The project is a well-documented source of best practices and strategies for archiving spatial data.

The National Geospatial Digital Archive (NGDA)

The National Geospatial Digital Archive (NGDA) is a project administered through the National Digital Information Infrastructure and Preservation Program (NDIIPP) and was proposed jointly by Stanford University and the University of California, Santa Barbara. As Sweetkind-Singer (2009) explains:

> The goals of the NGDA were to create a national federated network committed to archiving geospatial imagery and data, to investigate preservation strategies, to collect "at-risk" content across a spectrum of formats, and to develop policy agreements governing retention, rights management and obligations of the partners. (36)

The NGDA project was funded beginning in 2004. As of January 2015, the project is ongoing and counts Stanford, UCSB, Tennessee, and Vanderbilt as partners (NGDA Members).

As with GeoMAPP, the NGDA project involves multiple states, rather than agencies within a single state, but the results can still be applied to our endeavor. The outcomes so far include two functioning data repositories, the successful creation of formal legal agreements between partners, and significant work in format registry (Erwin and Sweetkind-Singer 2009). The success of the format registry work is of particular interest, as this work can be repurposed by other groups with similar goals for geospatial data preservation.

Making the Case for Minnesota

Historical geospatial data has value that can be difficult to quantify. For example, with cadastral data, a county may be primarily interested in the most recent version of the parcel data for ownership and tax purposes. The county may keep snapshots of the data through time, even if there is no policy to do so; however, if disk space becomes scarce, historical versions of the data might be purged, even if there are a great number of stakeholders

who are interested in the data. Data creators are certainly aware of the importance of the information they provide, but their first priority is to meet the business needs of their organization, especially in light of the financial constraints most face.

Discussions of the importance of archiving geospatial data have been active for some time. Common themes for why archiving practices have not been implemented include: limited resources; legal concerns; not a clear business need; and in some cases a lack of understanding by decision-makers or practitioners about the benefits of preserving geospatial data (Kne et al. 2014). The reasons for not implementing archiving differ greatly when asking government GIS personnel versus data experts as seen in the literature review above. In many cases, the people interested in using historic data are not the people responsible for creating the data. This disconnect, along with a need for clearly defining what is meant by archiving (i.e., saving data on a file server is not an archive) round out the reasons why archiving spatial data has not been implemented in Minnesota. The ultimate goal of a data archiving project in Minnesota would be active curation, which would include format transformations when necessary, metadata enhancement, and pathways for discovery.

A federated system of data providers and consumers was sketched out in the Minnesota State GIS Enterprise Conceptual Architecture Design in 2005, and became a reality in 2015 with the launch of the Minnesota Geospatial Commons (Minnesota Governor's Council on Geographic Information 2005). The Commons targets a broad audience of both traditional and non-traditional GIS users who need data for a project, services for an application, or some other web-based resource. The Commons does not provide web mapping functionality like "Google Maps" or "ArcGIS Online," but focuses solely on resource provision ("About the Minnesota Geospatial Commons"). The creation of the Minnesota Geospatial Commons demonstrates the growing commitment in Minnesota to making geospatial data freely available. The metadata requirements of the Commons lessen what is usually a major obstacle to effective data curation, while also standardizing file formats and data structures. Thus, with the launch of the Commons, Minnesota is more prepared than ever to undertake systematic data archiving. The next section provides some background on the circumstances in which GIS developed in Minnesota.

A BRIEF HISTORY OF GIS IN MINNESOTA

The state of Minnesota first began experimenting with computerizing geospatial data handling nearly half a century ago. The history of GIS in Minnesota and the lessons it offers to us as we consider the appropriate approach to data archiving is valuable. As such, this section presents a condensed

account of where GIS began and how it developed within Minnesota, first at the University of Minnesota's Center for Urban and Regional Affairs (CURA), shifting to the State Planning Agency as the Land Management Information Center, and ultimately forming MnGeo.

Minnesota Land Management Information System (1967–1977)

In 1967, CURA began the Lakeshore Development study (known as LSD) in Minnesota with the support of the Minnesota Resources Commission (MRC) and as a joint effort with the University of Minnesota's Department of Geography. The LSD sought to compile information about the current and potential development along Minnesota's lakeshores, with a stated goal "to build a predictive model of lakeshore development: to predict where development will go and the rate of that development," a goal necessitating the collection of a wide array of information on the state's lakes (Orning, Sjoberg, and Maki 1975, 1). The MRC (*MRC Report* 1969), comprised of Minnesota state representatives and senators, was charged by the legislature "to provide the Legislature with background necessary to evaluate programs to preserve, develop and maintain the natural resources of this state" (4) and crucially, to make any compiled data available to the legislature. Being the "Land of 10,000 Lakes", the project necessitated data collection, collation, and analysis on a larger scale than had previously been attempted in the state (Borchert et al. 1970). The MRC combined the lakeshore study with the Statewide Land Use Mapping project (SLUM), another CURA project, as a single item in the Commission's work plan for the 1971–1973 biennium (*MRC Report* 1969), a combination that further contributed to the transformation of the database into a true multipurpose GIS.

Even in its pre-conception, GIS in Minnesota was a problem of data discovery and reusability, namely how those tasks were nearly impossible at the time. The study authors reflect on the dispersed nature of state created data, reflections whose relevance carry through to the present:

> Much of the social, economic, and physical resource data necessary to support land use decisions is already routinely collected by various public agencies that deal with particular sectors of the state's total resource. Unfortunately, after its initial use by the collecting agency, much of the information is often tucked away in obscure files. Others that could make use of it may be forced to collect it again, sometimes at public expense, because they do not know that the data exist, because it is not readily available, or because it is not in a format that is useful. (Center for Urban and Regional Affairs and Minnesota State Planning Agency 1976, pp. 1–2)

In addition to the problem of duplicated data collection efforts, Orning, Sjoberg, and Maki (1975) detail the multiple stops needed to acquire data for the LSD:

> As a first step we determined how many lakes there were, where they were, and some of the physical characteristics of them. . . . This immediately got us into the state agency records. Here we ended up in the Department of Natural Resources in the Division of Waters, Soils and Minerals. They had a very good system to locate lakes and tell you how big they were. But they couldn't tell you how deep they were or what kind of fish were in them. We got this information from another division of the Department of Natural Resources, namely Game and Fish. (1–2)

At this stage the researchers realized that the system being built would be the only standardized source of these data and thus possessed significant administrative value. The LSD/SLUM team continued assembling data, pulling from the Soil Conservation Service, U.S. Forest Survey, State Highway Department, and county assessor's offices (Orning, Sjoberg, and Maki 1975). The core database of what became Minnesota's first GIS thus emerged as an outgrowth of two University of Minnesota-led projects.

As its base unit the MLMIS database used the forty-acre tract (the "forty"), the smallest unit of the Public Land Survey (Center for Urban and Regional Affairs and Minnesota State Planning Agency 1976). Through its use of the forty, MLMIS became one of the first raster based GISs (that is, a GIS operating on more or less a Cartesian grid). The choice of the relatively coarse size of the forty was in order to allow for the incorporation of county assessor land development records. Kozar and Schmitt (1974) explain that some underlying data structure was needed and that since existing records were oftentimes organized using the public land survey, it was decided to build MLMIS similarly (1). With a base unit chosen, other data to be included in MLMIS were generalized (i.e., made coarser) to fit the forty. The forty made sense because it was considered convenient for regional scale analysis, and also accommodated the computational capacity at the time (Center for Urban and Regional Affairs and Minnesota State Planning Agency 1976). The initial emphasis on regional scale, combined with hardware limitations of the era, would eventually necessitate some major retrofitting of the database in the 1980s when support for large scale vector data was added, straining an already strained budget even further; see below in the section on Land Management Information Center (LMIC).

Once assembled for the entire state, MLMIS contained 1.4 million forties, many of which varied considerably in both shape and size. The initial data collected for the state's forties consisted of county, township and range, section and 40-acre parcel location, government lot number, latitude and longitude, minor civil division, contiguity to lakes and watercourses, land use, and ownership (Orning and Maki 1972). Orning, Sjoberg, and Maki (1975) write that "the primary goal of the system is to improve the quality of environmental decisions made by Minnesota public officials. Another goal is to improve the efficiency and quality of research within the University of

getting people within the different disciplines to integrate research efforts" (5). This statement illustrates the tight relationship between the state and university in the coordination of the state's GIS efforts and in making the information valuable to both researchers and public officials.

The fluid relationship between the work and personnel of the university and state agencies involved is an important characteristic of the development of MLMIS. Orning, Sjoberg, and Maki (1975) make it clear when they write as follows: "We are encouraging public officials to use our new techniques such as computer mapping. We are doing this by moving personnel from the MLMIS project into state government" (10). The framework below builds upon this history of collaboration. After a decade as a part of CURA, MLMIS itself was transferred into state government, forming the technical core of the new Land Management Information Center in the State Planning Agency.

Land Management Information Center (LMIC) (1977–2009)

Founded in 1977, LMIC was the "first state agency anywhere exclusively devoted to providing GIS services within state government" ("History of MnGeo and LMIC"). As a service bureau providing GIS services for a fee to clients elsewhere in the state, management of the MLMIS database became not an end to itself, but rather the means to fulfill client contracts. As Robinette (1984) puts it:

> The Minnesota system has shifted its emphasis away from geographic information systems and toward that of a geographic information center. This shift is away from the tool and toward the application; away from differences in technology and toward the commonality of purpose; and, away from capability and toward service delivery. (155)

This transition continued throughout LMIC's existence and extended into its spiritual successor MnGeo, as well as the effort put into establishing the Minnesota Geospatial Commons.

Yet, the transition to a fee-for-service model was not entirely smooth, resulting in some unintended consequences and inefficiencies. In the late 1980s, Robinette (1988) wrote that "Minnesota now has a considerable unused capacity to produce GIS products" (141). Paul Tessar of LMIC (1988) described the situation in greater detail as part of a plenary panel at the International Geographic Information Systems Symposium, saying that originally, "charging users was seen as a way of limiting demand to only the most important projects," but that now "we often find ourselves talking with potential users who have very juicy, interesting, environmentally significant GIS projects but no cash" (154). He went on to say that "many useful, important projects have no budget, and we periodically have excess capacity, and we

find this situation somewhat troubling. We do a great deal of valuable work, but we're only scratching the surface" (154). Another serious problem with the fee-based model was the tying of MLMIS database development solely to user fees. Tessar argued that users were not willing to pay for "comprehensive database development" (153). That unwillingness combined with LMIC's reliance on user fees for survival resulted in the core MLMIS database growing increasingly dated. This aspect of LMIC's history serves as a reminder that for state agencies the focus is on having the most up to date data for the project at hand, and that expecting agencies under budgetary and time constraints to engage in formalized data archiving without support in the form of financial or human resources is unrealistic and even misguided. Rather than being the responsibility of the data producers, responsibility for archiving should rest at an institution with an established business interest in historical data being preserved, such as the University of Minnesota. We argue that such a match will result in the greater long-term success of an archiving project.

While LMIC was stifled by budgetary cuts and database stagnation, demand for and use of GIS spread across Minnesota. The diffusion of GIS was a sign of increased popularity and lower technological and financial barriers to entry, as well as dissatisfaction with the regional focus and age of the core MLMIS database. That is, because LMIC lacked the data users wanted, more and more were inclined to invest in their own GIS capability. As Craig (1989) explains, "A significant part of the explosion of new computer mapping systems resulted from the need for large scale maps of relatively small areas," a task for which the original MLMIS database was ill equipped (4).

An initiative characteristic of the challenges LMIC faced while attempting to coordinate the state's spatial data was the Minnesota Geographic Data Clearinghouse. In many ways the predecessor of the Minnesota Geospatial Commons (see below), the Minnesota Geographic Data Clearinghouse was first conceptualized in the mid-1990s, partially in response to President Clinton's 1994 executive order calling for the establishment of a National Geospatial Data Clearinghouse ("Executive Order 12906: Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure" 1994). Archiving assumed a position of some importance in early deliberations on the creation of a clearinghouse. In fact, it is one of the three main components of the clearinghouse as laid out in the Minnesota Governor's Council on Geographic Information report "Laying the Foundation for a Geographic Data Clearinghouse" (1997). This component would

> ensure the physical and intellectual integrity of data and supporting documentation over time by developing guidelines for the archive, providing storage and archiving alternatives for data producers, offering training in archive preservation processes and defining criteria for identifying data sets to be included in the archive. (1)

The report lays out several tasks necessary for the development of sound archiving practices among them:

> Develop a collection policy for the geographic data archive and build a statewide consensus on selection criteria for data valued by the geographic data community both now and in the future. (2)

It also states as follows:

> Set priorities for data acquisition and determine who will be responsible for dealing with incomplete documentation, particularly for existing data sets; Checking data quality; and determining conditions under which a data set should be improved before it is included in the clearinghouse. (2)

While the clearinghouse launched in 2000, it lacked implementation of any of the archiving components described above. The Clearinghouse was able to aggregate a substantial number of resources over the years, but LMIC's lack of legislative mandate proved to be an obstacle. Some state agencies were more enthusiastic about cooperating than others, a key component of what was essentially a voluntary effort. LMIC's lack of statutory authority had become a major obstacle hampering the evolution of GIS in Minnesota.

In 2005, the National States Geographic Information Council (NSGIC) published a set of nine criteria evaluating the effectiveness of GIS coordination within each state. The criteria focused on the presence of authority and support. NSGIC found Minnesota lacking in several of criteria (Arbeit 2007). While in many ways LMIC operated as de facto coordinator of the state's GIS activity, its mission was reflective of its past as a fee-based service center, and it lacked statutory authority, a major component of the NSGIC criteria. In 2004, a vision for statewide GIS laid out in a report published by the Minnesota Governor's Council on Geographic Information (Arbeit et al. 2004) was put into motion thanks in large part to a grant from the FGDC's Fifty States Initiative in 2006. A primary outcome of the plan was the inclusion of enterprise GIS in Minnesota's Drive to Excellence Initiative, which led to the official dissolution of the service bureau model and LMIC and the formation of the Minnesota Geospatial Information Office in 2009 (Arbeit 2008).

Minnesota Geospatial Information Office (2009–Present)

The Minnesota Geospatial Information Office (referred to almost exclusively as MnGeo) was established in 2009, and along with it a position for a chief geospatial information officer. Its creation marked the first time a state agency was given a legislative mandate for coordinating GIS

activities within Minnesota ("History of MnGeo and LMIC"). All functions of LMIC deemed necessary to MnGeo's operation were transferred to the new agency.

A FRAMEWORK FOR ARCHIVING MINNESOTA'S SPATIAL DATA

With the University of Minnesota's leadership, Minnesota is now ready to pursue a data archiving project wholeheartedly. At a meeting of Minnesota's Statewide Geospatial Advisory Council (SGAC) in January 2015, the authors gave a presentation broaching the topic of the University of Minnesota taking up archiving the state's geospatial data. With its broad coverage of the state's data producers and its role directing the course of MnGeo, the SGAC was an ideal place to begin making a pitch for geospatial data archiving in Minnesota. The reaction to the presentation was supportive and for a few participants, enthusiastic, although one longtime state official warned that data archiving had been discussed many times over the past twenty years without concrete results and that without clear criteria for determining what is worthy of preservation, new efforts seem likely to suffer a similar fate. This is certainly an important and valid point, and we argue that the framework described below addresses these concerns.

The framework for creating an infrastructure for statewide archiving is simple; create an environment fostering the collaboration and communication among all stakeholders. While there are long lists of details and policies to create this environment, the University is ready to champion the task.

We believe the University of Minnesota can lead this effort because it has expertise with curating spatial data, a strong desire to have access to historic data (being one of the largest consumers of the spatial data), and a mission to serve the interests of the citizens of the state. Discussions between University of Minnesota and MnGeo are in progress to explore funding needs and sources and the formation of a working group to investigate a pilot project including statewide aerial imagery from MnGeo and data sets from MetroGIS. The next section describes how this framework encourages collaboration and highlights key players for success.

Collaboration and Communication

By partnering with both the data producers and the data consumers from various government agencies, important decisions relating to data curation and retention will be better informed. As Bethune, Lazorchak, and Nagy (2009) put it, "the ongoing engagement between different divisions in each state helps to break down barriers in comprehending the issues that each

face" (54). The University of Minnesota has been meeting regularly with representatives from different parts of the state to begin publishing its data assets on the Geospatial Commons, another way in which the Commons has lowered the barrier to systematic archiving.

The geospatial community in Minnesota is strong and expands throughout the state at all levels of government. Next, we identify the initial stakeholders who will collaborate in making geospatial archiving a reality in Minnesota and how they are committed to participate.

MnGeo

The chief geospatial information officer acts as the voice of enterprise GIS in the state and coordinates MnGeo's activities with advice from the SGAC and the now defunct State Government Geospatial Advisory Council, both of which were formed at the same time as the agency. SGAC is comprised of members from all levels of government including state, city, county, and regional entities, as well as representatives from the private sector and the two major public university systems. Funded by legislative appropriation and internal cost recovery, MnGeo hosts and collaborates with all government agencies in the efficient collection of geospatial data. For example, MnGeo provides leadership for acquiring aerial imagery for the State. Even with limited resources, MnGeo officials are keenly interested in potentially collaborating with the University of Minnesota on a data archiving endeavor. Launched in early 2015, but many years in the making, the Minnesota Geospatial Commons[1] is a centralized point of distribution for geospatial data produced across the state. The Commons' consolidation of distribution for Minnesota's geospatial data immediately reduces the cost of data curation by decreasing the amount of time required for data acquisition, harmonization, and most importantly, documentation. Another key feature of the Commons is an application programming interface (API), which would be integral to automating the data capture aspects of the archiving process, a source of further cost savings. Above all else, the Commons instills an ethos of sharing amongst Minnesota's geospatial data resource producers, an ethos that lends itself to supporting data archiving activities. MnGeo's success is evident by the development and launch of the Geospatial Commons only five years into its existence, all the more remarkable as MnGeo carries on LMIC's legacy of being under-resourced. MnGeo would play a key role in managing the relationships necessary for data archiving.

---

1 See https://gisdata.mn.gov

MetroGIS

In a state where a major metropolitan region contains nearly two thirds of the population, regional governments can lead the way in spatial data production, use, and management. Formulated by the Metropolitan Council, which is the regional governmental agency of the Twin Cities, MetroGIS is a largely volunteer-based collaborative (it has a paid staff of one) made up of local and regional governments, with partners at the state and federal levels, academia, the nonprofit world, and the private sector ("History and Development of MetroGIS"). Its purpose is to provide a regional forum to promote and facilitate widespread sharing of geospatial data ("MetroGIS Overview"). In many ways MetroGIS is granted the same authority in the Twin Cities metropolitan region as is given to MnGeo statewide. In 2002, MetroGIS launched DataFinder.org, an award winning, comprehensive catalog of data resources for the metropolitan region, with free downloads and thorough metadata. For the past several years it has been a state leader in urging counties to open their data, a goal that would make data archiving more attainable ("Free and Open Data"). While not currently engaged in any specific data archiving activities, MetroGIS is an enthusiastic potential partner for a data archiving project and would lend its highly respected voice in support.

State Archives

The Minnesota State Archives (part of the Minnesota Historical Society) has a statutory charge to collect and preserve the state's historical government records. In the late 1990s, the Archives started a digital records program to cope with the anticipated influx of digital records from agencies as they moved from paper-based to electronic systems. Almost immediately, staff identified a kinship with the geospatial community, particularly in advocating for the use of standardized metadata for discovery and description, and formed an educational partnership with LMIC. In 2001, this partnership was highlighted when the Archives led an interagency effort to create a state recordkeeping metadata standard, one that crosswalked to the Minnesota Geographic Metadata Guidelines, the standard for government geospatial data ("Electronic Records Management Guidelines, Metadata").

In 2002, the Archives received its first transfer of geospatial data, which came from LMIC on four CDs along with paper files documenting the work of the 2000–2001 Governor's Citizen Advisory Commission on Redistricting. Given the significance of the Commission, the Archives accepted the geospatial data into its collections with the plan that interested researchers could obtain copies of the CDs, which were stored with the paper files. While a handful of state archives around the country have developed expertise in

working with geospatial data, the Minnesota State Archives has no capacity for providing anything but the most basic levels of access to these sets, namely file download for users. This is analogous to the Archives offering images of diary pages in a foreign language—the Archives can provide the content, but the user must translate it into something understandable and is responsible for the interpretation.

While the Archives may itself have limited capacity for providing access to geospatial data, it is interested in offering its digital preservation expertise to others with the goal of maintaining and sharing data sets over the long term. The Archives will be testing possible dark archive service models with agencies that have historical records with ongoing value but that are actively used. In this scenario, the Archives (perhaps along with other repositories) would preserve master sets of records, while the originating agency or a partner organization (such as the University of Minnesota) provides access to copies.

University of Minnesota

The Morrill Land Act of 1862 set aside 30,000 acres of public land per Congressional representative for the establishment of institutions of higher learning across the United States (McDowell 2001). Having land grant status, as the University of Minnesota does, has grown beyond its agriculture-focused roots to a more general emphasis on public service. CURA, the home of MLMIS, was founded in 1968 "in keeping with the University of Minnesota's land-grant mission" ("History of CURA | Center for Urban and Regional Affairs"). In CURA's newsletter, the creation of MLMIS was cited as an example of computers serving the public interest (*CURA Reporter* 1975). Archiving the state's land record information within the University Libraries would continue the tradition of the University of Minnesota fulfilling its land grant mission of public service through the support of public geospatial data management. Several groups within the university are particularly well situated to assist in geospatial data archiving.

As the Minnesota Geospatial Commons took shape, the University of Minnesota Libraries began to address the intertwined problem of data preservation, curation, and archiving. These developments include provisions for the handling of geospatial data, which, combined with the University's land grant status, suggest that the Libraries would be an ideal home for archiving data created by government agencies throughout the state. In October 2014, the Libraries launched the Data Repository for University of Minnesota (DRUM), a home for research data produced by the University and its affiliates. Experiences curating GIS data for DRUM help identify potential areas of trouble for any potential archiving plans. For example, experience from DRUM indicates that acquiring adequate metadata is oftentimes the greatest hurdle to comprehen-

sive data curation, which in turn suggests that Minnesota, with its strong record of metadata creation and the metadata requirement for datasets contributed to the Commons, is well positioned to undertake systematic archiving.

The John R. Borchert Map Library at the University of Minnesota, named for the esteemed geographer who was director of CURA when MLMIS was first conceived, serves the University community and the State of Minnesota, including other academic institutions, corporate patrons, and the general public. The Map Library has a long history of collaboration with partners involved in the production and use of geospatial data resources, including departments and other academic units at the University, as well as local, regional, and statewide government entities.

In 2011, the University launched U-Spatial as a network to support geospatial research activities across the University. Similar to many other R1 universities today, before 2011, many of the services offered by U-Spatial were fragmented and usually offered informally by a variety of groups. Having a central point of contact for assistance makes it easy for people to get the help they need, as well as freeing up the time of other labs that previously provided *ad hoc* support.

The Map Library is well positioned as an intermediary between the agents of the state and the University due to the long running relationships with many of the state's geospatial data producers, as well as its close relationship with many University of Minnesota departments and U-Spatial. In collaboration with U-Spatial, the library has implemented a platform for geospatial data discovery that incorporates metadata from the Commons and a number of other sources, a potentially valuable experience in data acquisition and metadata workflow development. With the University Libraries providing the expertise in archive and preservation practices, the Borchert Map Library's history of working with state agencies, and U-Spatial's capability with hardware and software, the University of Minnesota is well suited to the task of archiving the state's spatial data.

FRAMEWORK INFRASTRUCTURE

While creating a collaborative infrastructure is key to our proposed archiving framework, we recognize that there are certainly other, more technical facets that need to be addressed to make the initiative successful.

Preservation and Curation

Archiving data is not a one-time action. As Sweetkind-Singer (2009) puts it, although "it is easy to grab digital content and bring it in house, it is

entirely a different matter to make sure that access is provided now and into the future as securely as any book we pull off our shelves" (40). The need for ongoing curation in order to preserve the same level of access to a record is the greatest difference between archiving analog materials and digital data. With regards to spatial data, terminological slipperiness complicates the matter. For example, those working as spatial data creators and users often refer to routine backups as archiving, even when those backups may overwrite data to conserve space make sure the most current data is preserved for disaster recovery purposes (Bethune, Lazorchak, and Nagy 2009). Other challenges highlighted include workflows and metadata. Specifically, developing a "structured workflow" for the data life cycle and the frequent lack of accompanying metadata with deposited datasets (40).

A goal to strive for is a comprehensive workflow as seamless as the process for paper-based materials (Sweetkind-Singer 2009). While this may seem common-sense, it is a solid reminder that we are not acquiring, curating, archiving, and preserving geospatial data in a vacuum—there are valuable precedents for workflows and metadata that exist for print resources and many of these workflows can be utilized or adapted for handling spatial data. As highlighted above, the State Archives possesses a wealth of knowledge in this area, and drawing upon their experience will enormously improve our workflow development. Specifically for spatial metadata, the Geospatial Commons has a base requirement for all contributed datasets, resulting in a centralized, uniform set of metadata records for data from all over the state. Having resources centralized and documented following accepted metadata standards drastically reduces the potential cost of archiving.

Determining "Value" and Justifying Costs

In the context of a discussion of data archiving at a meeting of the SGAC, one representative from greater Minnesota (that is, outside of the Twin Cities metropolitan region) related a problem that will in all likelihood constitute the "first wave" of data archiving problems to be addressed. With remote-sensing data, such as LiDAR (Light Detection and Ranging) being collected every few years, and the considerable size of LiDAR data (many terabytes for a statewide collection), pressure on storage resources is already building. Regarding the value of imagery (aerial, satellite, and many ways in which LiDAR is used), the member described an initial period of very high value, followed by a steep decline until the data reaches several decades in age, when it then begins to appreciate once again. Since most users need the most recent imagery, the general perceived "value" of imagery tends to decrease over time. However, that same "value" tends to rise steeply after a number of decades due to research needs (see Figure 1).

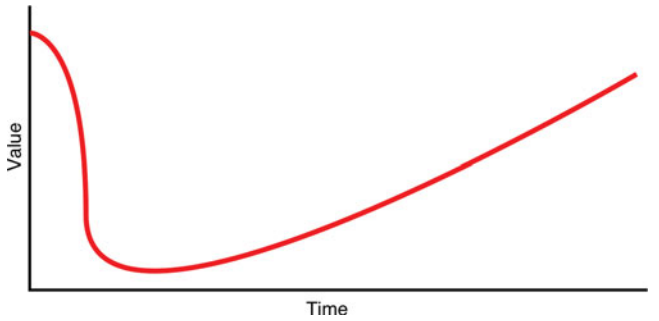

FIGURE 1 Perceived value of remotely sensed imagery over time.

The most challenging aspect of establishing an archiving program will be getting past this "valley" of lower perceived value, yet at the same time these specific perspectives strengthen our belief that many data producers across Minnesota do not need to be convinced of the value of archiving data. To secure any state funds for an archiving project, however, demonstrating the value to policy decision makers will be crucial.

A cost benefit analysis can be useful for demonstrating value, although costs for archiving are "fairly straightforward" to identify and calculate, whereas the benefits require more explanation and justification (*Geoarchiving Comprehensive Cost-Benefit Analysis Guidance* 2011, 4). The U.K. Data Archive[2] provides a thorough questionnaire[3] for estimating the costs of data archiving. The tool is intended for use with research data, but the principles it suggests are still relevant and useful for governmentally produced data. The costliest tasks identified by the questionnaire relate to data cleanup and documentation, tasks that are already part of the workflow for agencies adding data to the Commons. Using these tools shows we can leverage actions already taken for the Commons, thereby greatly reducing the costs of archiving.

The potential future need for recreating data can be seen by examining any current research project that involves creating geospatial data from historical maps or documents. While we cannot necessarily predict which data will be most useful for research purposes in the future, this data has already been created, and preserving data is less costly than attempting to recreate it. In the past we have had the luxury of relying on historical maps or documents as the source for creating geospatial data; however, much of the data created today is born digital, and recreating the data again in the future would mean starting from scratch rather than from a print resource, a costly and difficult endeavor (North Carolina Center for Geographic Infor-

[2] See http://www.data-archive.ac.uk/
[3] See http://www.data-archive.ac.uk/media/247429/costingtool.pdf

mation and Analysis and North Carolina Department of Cultural Resources 2011).

Creating a robust data archive addresses both concerns of value and cost. This framework provides data producers a central location to deposit their data at little or no cost. There is cost for the University to archive and preserve the data, but incrementally each data set takes few resources. The University Libraries will provide curation services based on professional practices, utilizing staff already in place, and processes already initiated with the Geospatial Commons alleviate the traditionally high cost of documentation and standardization.

Bureaucracy

One of the main challenges will be convincing city and county government agencies to make their geospatial data freely available (and more readily available for archiving). Charging for access to geospatial data (via a cost-recovery model) is still standard practice for smaller government agencies. In recent years, however, there have been successful efforts to change these practices, and Minnesota is increasingly well positioned in this area. Thanks in large part to the efforts of MetroGIS, nearly all Twin Cities metropolitan region counties have passed resolutions opening their data and eliminating cost-recovery fees associated with access. Counties outside the metro are also adopting open data policies. Once again, the advent of the Geospatial Commons is the engine driving the development and sharing of Minnesota's spatial data.

The move towards open data improves the preservability of those data, but there remain legal and ethical concerns about archiving private, restricted, or sensitive governmental data. An example is much utility data, such as pipeline locations, are generally not publicly shared due to security concerns. While a "dark" archive may be realized in the future in coordination with the plans of the State Archives, our proposed framework addresses only publicly shareable data. The complexity introduced by private and/or sensitive data could inadvertently derail the early, tentative stages of archiving.

CONCLUSION

Thanks to the nascence of the Geospatial Commons and the promise of active participation from the State Geospatial Advisory Council and the University of Minnesota, the outlook for preserving Minnesota's geospatial data is encouraging. The history of GIS in Minnesota shows that the state and University of Minnesota worked in tandem. Moving forward, collaboration of this kind will form the backbone of our efforts. Already underway is the

entrance of the University of Minnesota as a data publisher in the Geospatial Commons. As a data publisher, the University will share the fruits of research completed using the state's freely distributed data. In addition to the spirit of reciprocity, becoming a node in the Commons network provides localized access to all of its datasets. The University's local mirror of Commons resources updates daily, making it trivial to capture data snapshots as we determine dataset retention schedules. Alongside this ongoing technical work, representation on the State Geospatial Advisory Council enables us to pursue consensus on the course of action on the policy question. Potential sources of funding are also being explored. Much work remains to be done to reach the goal of statewide systematic archiving, but the increasingly prevalent climate of cooperation across Minnesota makes reaching that goal far more feasible.

ACKNOWLEDGMENTS

The authors would like to acknowledge the assistance provided from MetoGIS and MnGeo.

REFERENCES

About the Minnesota Geospatial Commons. *Minnesota Geospatial Commons*. https://gisdata.mn.gov/content/?q=about. Accessed March 10, 2015.

Arbeit, David. 2007. *Implementing an enterprise strategy for Minnesota's spatial data infrastructure*. 06HQAG0104. 50 States Initiative Project. https://www.fgdc.gov/grants/2006CAP/Reports/104-06-3-MN-FinalReport

Arbeit, David. 2008. Functional transformation kick-off meeting. May 13. http://www.gis.state.mn.us/committee/MSDI/dte/D2E_kickoff_arbeit.pdf

Arbeit, David, Fred Logman, Randall Johnson, Glenn Radde, and Sonia Dickerson. 2004. *A foundation for coordinated GIS Minnesota's spatial data infrastructure: A strategic plan*. Minnesota Governor's Council on Geographic Information. http://www.gis.state.mn.us/pdf/MN%20IPlan%20Consolidation%20%20Final%20_04OCT13.pdf

Bethune, Alec, Butch Lazorchak, and Zsolt Nagy. 2009. GeoMAPP: A geospatial multistate archive and preservation partnership. *Journal of Map & Geography Libraries* 6(1): 45–56.

Borchert, John R., George W. Orning, Joseph Stinchfield, and Les Maki. 1970. *Minnesota's lakeshore: Summary report of the Minnesota lakeshore development study*. Minneapolis: University of Minnesota.

Center for Urban and Regional Affairs (CURA) and Minnesota State Planning Agency. 1976. *Overview of the Minnesota land management information system*. Minneapolis: University of Minnesota Center for Urban and Regional Affairs.

Craig, William J. 1989. A plethora of GIS use in Minnesota? *The Operational Geographer* 7(4): 3–5.

CURA Reporter. 1975. Computers and public service. *5*(1). http://www.cura.umn.edu/publications/catalog/reporter-05-1.

Electronic Records Management Guidelines, Metadata. *Minnesota Historical Society*. http://www.mnhs.org/preserve/records/electronicrecords/ermetadata.php

Erwin, Tracey, and Julie Sweetkind-Singer. 2009. The National Geospatial Digital Archive: A collaborative project to archive geospatial data. *Journal of Map & Geography Libraries 6*(1): 6–25.

Erwin, Tracey, Julie Sweetkind-Singer, and Mary Lynette Larsgaard. 2009. The National Geospatial Digital Archives—Collection development: Lessons learned. *Library Trends 57*(3): 490–515.

Executive Order 12906: Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure. 1994. http://www.archives.gov/federal-register/executive-orders/pdf/12906.pdf

Free and Open Data. *MetroGIS*. http://www.metrogis.org/projects/free-open-data.aspx

Geoarchiving Comprehensive Cost-Benefit Analysis Guidance. 2011. GeoMAPP. http://www.geomapp.net/docs/03_Geoarchiving_Cost-Benefit_Analysis_Guidance_20111231.pdf

History and Development of MetroGIS. *MetroGIS*. http://www.metrogis.org/about-metrogis/history-development.aspx

History of CURA | Center for Urban and Regional Affairs. http://www.cura.umn.edu/about/history

History of MnGeo and LMIC. *Minnesota Geospatial Information Office*. http://www.mngeo.state.mn.us/About/history.htm

Kne, Len, Randy Knippel, Mark Kotz, Ryan Mattke, Cory Richter, and Dan Ross. 2014. The 4 W's and 1 H of data archiving. Presented at the Minnesota GIS/LIS Annual Conference, October 3. https://uspatial.umn.edu/node/670

Kollen, Christine, Cynthia Dietz, Joy Suh, and Angela Lee. 2013. Geospatial data catalogs: Approaches by academic libraries. *Journal of Map & Geography Libraries 9*(3): 276–295.

Kozar, Kenneth, and John Schmitt. 1974. *A guide to the CRT data entry system*. Center for Urban and Regional Affairs. Box 1. Technical Reports. 4001. Land Management Information Center (Minn.). Records. Minnesota Historical Society. State Archives.

Larsgaard, M. 1995. Spatial data in digital form: An overview. *Association of Canadian Map Libraries and Archives Bulletin 95*: 12–19.

Lauriault, Tracey P., and D. R. Fraser Taylor. 2014. The preservation and archiving of geospatial data and cybercartography as a proactive preservation process. In *Developments in the theory and practice of cybercartography applications and indigenous mapping*, Tracey P. Lauriault and D.R. Fraser Taylor, eds., 2nd ed., 325–342. Modern Cartography Series 5. New York: Elsevier. doi: 10.1016/B978-0-444-62713-1.00021-0.

McDowell, George R. 2001. *Land-grant universities and xxtension into the 21st century: Renegotiating or abandoning a social contract*. Iowa City, IA: Iowa State University Press.

MetroGIS Overview. *MetroGIS*. http://www.metrogis.org/about-metrogis/overview.aspx. Accessed March 12, 2015.

Minnesota Governor's Council on Geographic Information. 1997. *Laying the foundation for a geographic data clearinghouse*. State Planning Agency.

Minnesota Governor's Council on Geographic Information. 2005. *Minnesota state GIS enterprise conceptual architecture design*. http://www.mngeo.state.mn.us/pdf/MNGISConceptualArchitectureDesign.pdf. Accessed March 12, 2015.

Morris, Steven P. 2009. The North Carolina Geospatial Data Archiving Project: Challenges and initial outcomes. *Journal of Map & Geography Libraries* 6(1): 26–44.

MRC Report. 1969. 11–1015. Minnesota Resources Commission. http://archive.leg.state.mn.us/docs/2011/other/111015.pdf

NGDA Members. *NGDA*. http://www.ngda.org/participants.html

North Carolina Center for Geographic Information and Analysis, and North Carolina Department of Cultural Resources. 2011. *GeoMAPP Final Report: 2007–2011*. http://www.geomapp.net/docs/GeoMAPP_FinalReport_final_20111231.pdf

Orning, George W., and Les Maki. 1972. *Land management information in Northwest Minnesota: The beginning of a statewide system*. 1. Minnesota Land Management Information System Study. Minneapolis: University of Minnesota, Center for Urban and Regional Affairs.

Orning, George W., Linda Sjoberg, and Les Maki. 1975. Historical perspective on the Minnesota land management information system study. Box 2. LMIC Progress Reports, Development and Overviews, 1970–1986. Land Management Information Center (Minn.). Records. Minnesota Historical Society. State Archives.

Ramapriyan, Hampapuram K., Gilbert L. Rochon, Ruth Duerr, Robert Rank, Stefano Nativi, and Erich Franz Stocker. 2010. Advances in spatial data infrastructure, acquisition, analysis, archiving and dissemination. In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pp. 1980–1983. IEEE. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5653074. doi: 10.1109/IGARSS.2010.5653074

Robinette, Alan. 1984. Institutional innovations: Moving from academia to serving government needs. In: *Seminar on the multipurpose cadastre: Modernizing land information systems in North America*, 155–166. Wisconsin Land Information Report #1. Madison, WI: University of Wisconsin-Madison.

Robinette, Alan. 1988. Development and implementation of a state land information system: Impediments after the technology is mastered. In: *Seminar on multipurpose land information systems–Applications of information technology for natural resource planning, management, and monitoring*. Niemann, Bernard J., Jerome G. Sullivan (eds.), University of Wisconsin–Madison, and Institute for Environmental Studies, pp. 125–146. Wisconsin Land Information Report 3. Madison, WI: Institute for Environmental Studies, University of Wisconsin–Madison.

Sweetkind-Singer, Julie. 2009. Unchartered territory: Building a network for the archiving of geospatial images and data. *Against the Grain 21*(2): 6.

Tessar, Paul, William J. Craig, John Scoggins, Bernard J. Niemann, and John Antenucci. 1988. Plenary 5: Geographic information systems management and implementation. In: *International Geographic Information Systems (IGIS) Symposium: The Research Agenda* 1:141–162. Arlington, VA: [Washington, DC]. http://hdl.handle.net/2027/umn.31951d00052783t