

.....  
NCSU Libraries and North Carolina Center  
for Geographic Information & Analysis

---

# North Carolina Geospatial Data Archiving Project

In cooperation with the Library of Congress National  
Digital Information Infrastructure and Preservation  
Program

## Final Report

---

July 1, 2010

Principal Investigator: Steven P. Morris, NCSU Libraries

Co-Principal Investigator: Zsolt Nagy, NCCGIA

Project Coordinator: Jim Tuttle, NCSU Libraries





## TABLE OF CONTENTS

|  |    |
|--|----|
| Executive Summary.....   | i  |
| Project Overview .....   | 1  |
| Accomplishments and Lessons Learned.....                           | 3  |
| Unexpected Results of the Project .....                            | 11 |
| Plans for Ongoing Work.....  | 14 |
| Plans for Stewardship of Content Acquired Through the Project..... | 20 |
| Conclusion .....   | 21 |
| Appendix A: NGDAP Outreach Events .....                            | 22 |
| Appendix B: NCGDAP Publication List.....                           | 29 |
| Appendix C: NCCGIA and State Archives Collaboration.....           | 31 |
| Appendix D: Local and State Government Surveys.....                | 38 |
| Appendix E: GICC Archiving Report.....                             | 39 |
| Appendix F: GICC Data Sharing Report.....                          | 41 |
| Appendix G: NCGDAP Interim Report .....                            | 43 |

## Executive Summary

### Background

In October 2004 the NCSU Libraries and the NC Center for Geographic Information & Analysis entered into an agreement with the Library of Congress to pursue preservation of state and local digital geospatial data as part of the National Digital Information Infrastructure and Preservation Program (NDIIPP). The goal of the North Carolina Geospatial Data Archiving Project (NCGDAP) has been to inform development of a national digital preservation infrastructure through a “learning by doing” approach focused on identifying, acquiring, and preserving content within the context of the NC OneMap initiative and its framework of partnerships with state, local, and federal agencies. Although this three-year project was focused solely on the state of North Carolina, it is expected to serve as a demonstration project for data archiving and time series development elsewhere.

Digital geospatial data includes such data resources as geographic information systems (GIS) data sets, digitized maps, remote sensing data resources, and tabular data that are tied to specific locations. These complex data objects do not suffer well from neglect, and long-term preservation will involve some combination of format migration and retention of critical documentation. At the state and local government level geospatial data resources are created by a wide range of agencies for use in applications such as tax assessment, transportation planning, hazard analysis, health planning, political redistricting, and utilities management. These data resources are, in general, of greater detail and more current than data available from federal agencies, yet production points for these resources are diffuse—99 of 100 North Carolina counties have GIS, as do many cities—posing many challenges to the archive development process. Many of the targeted data resources are updated on a frequent basis—daily or weekly in some cases—yet data dissemination practices, for the most part, focus on providing access to current data.

Although often created with specific applications and functions in mind, these data resources are used in applications ranging far beyond those initially intended. End-user historical applications that might make use of historical and time series data include analyses of urbanization, environmental change, demographic change, land use change, and past uses of individual sites.

### Project Overview

The original project work plan was based on a three-year performance period from October 2004 through September 2007, and the project was subsequently extended through February 2010. NCGDAP was conceived as demonstration preservation experience in which the archive being developed is seen not so much as an end in itself as it is a catalyst for discussion among the various elements of spatial data infrastructure. That discussion, which includes libraries and archives, is centered not just on preservation processes and best practices but also on roles and responsibilities of the various players within the geospatial community.

NCGDAP focused less on technical architecture than it is on partnership building and on engagement with spatial data infrastructure. The purpose of the demonstration repository

developed for NCGDAP has been to: 1) to catalyze discussion within the geospatial data community about archive development, and 2) to generate learning experiences about domain-specific technical challenges associated with preserving geospatial data. To this end, a demonstration repository using Dspace was deployed, and over four terabytes of data have been acquired. A robust repository ingest workflow was developed to handle the transformation of complex multi-file, multi-formats formats into discrete digital repository items.

## Outreach and Engagement

While data preservation has been a low priority in the geospatial industry, growing interest in temporal data use created numerous, mostly unexpected opportunities to engage the data community. Key outreach and engagement outcomes included:

- Elements of spatial data infrastructure within the state, including the NC Geographic Coordinating Council (GICC) and its various subcommittees, were directly engaged in project work.
- Three surveys of current data archives practices (two for local government agencies, one for state government agencies) documented the current situation and helped to socialize the problem of data preservation within the data community.
- Through partnerships with EDINA (UK) and the National Archives and Records Administration, NCGDAP played a direct role in the formation of a Data Preservation Working Group within the geospatial standards organization: the Open Geospatial Consortium (OGC).
- The project led to the initiation of the NDIIPP-funded Geospatial Multistate Archiving and Preservation Partnership (GeoMAPP), initiated in November 2007 and extended through June 2011.
- The State Archives and Records Section was informally engaged in the project work and subsequently became co-lead, with NCCGIA, on the GeoMAPP multistate project.
- An Archival and Long-Term Access Committee was formed under the NC GICC, with representation from federal, state, regional, and local government agencies, resulting in formulation of a set of archiving recommendations to the data custodian community.

## An Evolving Content Domain

The geospatial data domain involves a complex mix of both data and services. In terms of understanding the evolving geospatial content domain, a number of learning experiences have emerged in the course of the project, including:

- PDF has emerged as a significant geospatial format. The ability of PDF to capture and preserve elements of cartographic representation makes it a powerful tool for capturing finished output in a way that the underlying datasets cannot, though underlying data intelligence is lost. Complex PDF documents, including geospatial PDF documents, present new preservation challenges of their own.
- There is significant local agency interest in resurrecting old analog maps for use in the digital environment. This interest creates a point of engagement and dialog around the issue of preserving current geospatial data for use in future historical analyses.
- The true counterpart to the old, preserved map is not the current GIS dataset but rather the cartographic representation that builds on that data. The representation is the result of a collection of intellectual choices and application of current methods with regard to

symbolization, classification, data modeling, and annotation. These representations typically occur in a complex proprietary project file format (difficult to preserve) or in an ephemeral web services interaction. Increasingly PDF is providing an option for static representations.

- Important data community documents such as inventories, standards, and policy or best practices documents must themselves be archived; in more than one case inventory information that had been retired from agency websites was retrieved from the Internet Archive.
- Non-spatial, place-based content such as building images and structural information, oblique imagery, and street-view imagery have become increasingly available on a comprehensive basis and are now an important part of the historical record.

### Engaging Spatial Data Infrastructure

At the outset it was understood that an infrastructure-based approach was needed to address preservation of geospatial data given the size and complexity of data resources and given the diffusion of production points at the state and local level. In terms of understanding the role that spatial data infrastructure might play in preservation, a number of learning experiences have emerged in the course of the project, including:

- Formal, structured data exchange networks, even if developed for other business reasons, support data archiving efforts by providing a low cost and routinized means to acquire data which is authenticated, documented, and for which rights have been clarified.
- The path to digital preservation may lead through other more compelling business problems. There is a significant overlap between the conjoined problems of business continuity and disaster preparedness and the lower priority problem of digital preservation.
- Regional efforts serve as building blocks for statewide infrastructure and provide diverse testbed environments for network development.
- State Archives and State Libraries have the potential to serve as significant components of state data infrastructures. Local records outreach and retention schedule processes serve as existing infrastructure which might be leveraged into geospatial data management.
- Emerging content exchange networks, such as NC StreetMap and the parcel data network being developed by the GICC Working Group for Seamless Parcels, can help to promote the use of content standards, the adoption of which will help to address semantic challenges that will arise with use of archival data in the future.

### Engaging Industry

There is a degree to which one might consider the geospatial industry to have been, in the recent past at least, to some extent “temporally-impaired.” In terms of understanding how to engage the geospatial community--both data producers and users--in the preservation challenge, a number of learning experiences have emerged in the course of the project, including:

- Promotion of temporal analysis opportunities and requirements indirectly promotes data preservation by cultivating demand for older data.

- Data is more likely to survive if users are made aware of the data's existence and the data is being actively sought and used.
- Software and data vendors are increasingly coming to see maintenance and use of temporal data as an important customer problem.

#### Technical Challenges: A Deeper Understanding

It was understood at the outset that geospatial data involves complex multi-file, multi-format content and is increasingly prone to be delivered by web services or to exist in spatial databases. In terms of addressing the technical challenges associated with geospatial data, a number of learning experiences have emerged in the course of the project, including:

- To the extent that geospatial metadata is available, it commonly needs to be synchronized to match the dataset at hand, normalized to a standard structure, and remediated to support discoverability via key fields.
- Mapping descriptive, technical, and administrative metadata elements to a single repository ingest spoke helped to refine the thinking about what discrete elements should be maintained within the project. The collective experiences of specific repository software communities, as represented in default metadata schemas, can help to shape metadata approach. It is possible that mapping to multiple repository ingest environments would help to evolve a more robust metadata approach within the project.
- Spatial databases may require a multi-pronged preservation approach, with data custodians managing temporal forward in time within the active database, with archival organizations managing database snapshots and possibly extracting selected subsets of data to be exported into more stable, archival formats.

#### Moving Forward

The GICC Archival and Long-Term Access Committee has established a set of recommended practices for data archiving that is already influencing data custodian practice within North Carolina. As a follow-up to NCGDAP, the GeoMAPP multi-state project has been initiated under the leadership of major stakeholders: state geospatial and state archives organizations. Each of these efforts have been closely aligned with the most immediate business needs of those stakeholders, with the “who, what, where, why, and how” of data archiving being addressed in very practical terms. The GeoMAPP partnership, initially comprising three states, has been expanded to include 10 additional states as informational partners, and is helping to deepen the national dialog about geospatial data preservation.

## Project Overview

Geospatial data resources created by state and local government agencies in the United States are, in general, of greater detail, more accurate, and more current than data available from federal agencies. These resources include key information such as cadastral (land parcel), building footprint, utility, and jurisdictional boundary data that are not available from other sources. Yet, since the production points for these resources are so diffuse—99 of 100 North Carolina counties have geographic information systems (GIS), as do a large number of municipalities—numerous challenges are posed to the process of archiving this data. Many of the data resources in question are updated on a frequent basis—daily or weekly in some cases—yet data dissemination practices, for the most part, focus on providing access to current information, especially with regard to vector data. Although often created with specific uses in mind, these data resources have come to be employed in applications ranging far beyond those initially intended and comprise a set of resources that are of cultural as well as economic value. While most GIS applications only require the most current data, end-user applications that can make use of historical and time series data include analyses of urbanization, environmental change, demographic change, land use change, and past uses of individual sites. Unfortunately much of this data is lost, either through data overwrite or technological obsolescence, or is simply made invisible by lack of access, and integration of state and local agency geospatial data into the appraisal and records retention processes of archives and records agencies has only recently begun.

### Project Background

In the year 2000 NCSU Libraries began to acquire and preserve North Carolina state and local geospatial data in response to rising user demand and in answer to a growing sense of long-term risk. Those early attempts to engage and archive content from well over 100 local agencies helped to cultivate an understanding of the need for an infrastructure-based approach to archive development. It became clear that a preservation effort could only scale by building from existing data infrastructure that has evolved as part of the National Spatial Data Infrastructure (NSDI). Spatial data infrastructure—which incorporates local, state, and federal government agencies as well as the private sector—had long focused on such issues as data standards, best practices, data sharing agreements, metadata production and harvesting, catalog development, and services integration, yet archiving and preservation had not yet become an area of focus in these efforts. The primary manifestation of spatial data infrastructure in North Carolina is NC OneMap, a combined state, federal, and local initiative that is focused on allowing users to view geographic data seamlessly across North Carolina, search for and download data for use on their own GIS, view and query metadata, and determine agency data holdings through an on-line data inventory.<sup>1</sup>

### Spatial Data Infrastructure in Support of Preservation Infrastructure

In 2004 NCSU Libraries and the NC Center for Geographic Information & Analysis entered into an agreement with the Library of Congress to pursue preservation of state and local digital geospatial data as part of the National Digital Information Infrastructure and Preservation Program (NDIIPP). The goal of the North Carolina Geospatial Data Archiving Project (NCGDAP) has been to inform development of a national digital preservation infrastructure

---

<sup>1</sup> North Carolina Geographic Information Coordinating Council. NC OneMap. Available: <http://www.nconemap.net> (accessed July 1, 2010)

through a “learning by doing” approach focused on identifying, acquiring, and preserving content within the context of the NC OneMap initiative and its framework of partnerships with state, local, and federal agencies.<sup>2</sup> Although the collection-building aspect of this project has been focused solely on the state of North Carolina, the project has served as a demonstration that informs data archiving and time series development more generally.

Working within the context of the NC OneMap network has provided an opportunity to engage content through traditional distribution channels as well as through emerging web services based modes of access. Prior to NCGDAP NC OneMap had primarily focused on data access and content standardization, with data archiving not being directly addressed, but the original NC OneMap vision statement did include the assertion that “Historic and temporal data will be maintained and available.”<sup>3</sup> This network has offered a means by which to engage a large number of local agencies in the digital preservation effort. NC OneMap is also closely tied to a national digital network through a series of connected initiatives devoted to access, integration, and utilization of geospatial data. Through their connections with spatial data infrastructure initiatives, NCGDAP and the more recent NDIIPP-funded Geospatial Multistate Archiving and Preservation Partnership (GeoMAPP), have worked to raise the profile of digital preservation and long-term access as issues to be addressed in these existing and emerging national geospatial networks.

## Project Objectives

NCGDAP was conceived as a demonstration preservation experience in which the archive being developed is not so much an end in itself as it is a catalyst for discussion among the various elements of spatial data infrastructure. That discussion, which includes libraries and archives as well as GIS agencies, is centered not just on preservation processes and best practices but also on roles and responsibilities of the various players within the geospatial community. The project was also seen as a way to generate learning experiences about domain-specific technical challenges associated with preserving geospatial data. While an extensive amount of data has been acquired as part of the project and a demonstration repository with associated workflows was developed, the overall project objective has been to shift archiving functions to relevant state agencies and the State Archives and Records Section as well as to cultivate an increased commitment to temporal data management on the part of the data producers and data custodians.

## Partnership Framework

NCSU Libraries, as lead institution in NCGDAP, has had a history of collaboration with key geospatial organizations within the state going back to the mid-1990’s. NCGDAP builds upon an existing statewide organizational framework, key components of which are the North Carolina Geographic Information Coordinating Council (GICC), the North Carolina Center for Geographic Information & Analysis (NCCGIA), and the NC OneMap Initiative. The GICC is established by legislation and is charged with improving the quality, access, cost-effectiveness and utility of North Carolina’s geographic information and promoting geographic information as a strategic resource for the state.<sup>4</sup> The Council creates policy and resolves technical issues related to North Carolina geographic

---

<sup>2</sup> North Carolina State University Libraries. North Carolina Geospatial Data Archiving Project. Available: <http://www.lib.ncsu.edu/ncgdap> (accessed July 1, 2010)

<sup>3</sup> North Carolina Geospatial Information Coordinating Council. May 14, 2003. NC OneMap Vision and Characteristics. Available: <http://www.nconemap.com/Portals/7/documents/visiondoc.pdf> (accessed July 1, 2010)

<sup>4</sup> North Carolina Geographic Information Coordinating Council. Available: <http://www.ncgicc.com> (accessed July 1, 2010)

information and GIS systems and fosters cooperation among government agencies, universities, and the private sector. NCCGIA is the primary state GIS agency, serves as staff to the GICC, and is responsible for implementing the goals and strategies of the GICC.<sup>5</sup> In the course of the project NCCGIA has played a coordinating role in a range of partnership and infrastructure development initiatives both within the state and in the national context.

## Accomplishments and Lessons Learned

Like the other initial NDIIPP collection building projects, NCGDAP was organized into four overlapping project phases: 1) Content Identification and Selection, 2) Content Acquisition, 3) Partnership Building, and 4) Content Retention and Transfer. Key project challenges, and responses to those challenges, are highlighted here.

### Challenge: Identifying Data Resources

One of the biggest challenges in archive development has been determining what data is available in the many counties and municipalities which produce data. It is often necessary to minimize the intrusion on the time of local agency data producers which, especially in rural counties, operate with small staffs which are often as small as one person. "Contact fatigue" arising from redundant requests for data or for information about data holdings can stretch local staffing resources and blunt enthusiasm for local participation in infrastructure-based approaches such as formal inventories. Furthermore, information acquired through informal methods is spotty, subject to transcription errors, and quickly becomes outdated.

Since 2000 NCSU Libraries has maintained directories of county and city geospatial resources in North Carolina, documenting agency contact points, data downloads, web mapping applications, and web services.<sup>6</sup> While these directories support the discovery needs of data seekers, they have also supported NCGDAP project work by making data access and contact information available in a single place. However, the project called for data holdings information that was more detailed, current, and complete.

### Response

To inform the archive development process NCGDAP utilized inventory information from current inventories as well as historical inventories. Early North Carolina surveys of local agency data in 1997 and 2003 informed initial data acquisition efforts, yet these surveys quickly became out of date and interest grew in the idea of developing a survey process that could operate on a continuous basis.<sup>7</sup> During the period of project work, the National States Geographic Information Council (NSGIC) decided to take an "all states" approach to the development of a survey instrument so that each state need not develop its own process and tools. The RAMONA (Random Access Metadata tool for Online National Assessment) inventory tool was developed, with the primary purpose of tracking the status of GIS in U.S. state and local government and aiding the planning and building of spatial data infrastructures.<sup>8</sup> RAMONA is designed to work

<sup>5</sup> North Carolina Center for Geographic Information & Analysis. Available: <http://www.cgia.state.nc.us/> (accessed July 1, 2010)

<sup>6</sup> North Carolina State University Libraries. County GIS Directory. Available: <http://www.lib.ncsu.edu/gis/counties.html> (accessed July 1, 2010)

<sup>7</sup> North Carolina State University Libraries. June 2008. North Carolina Geospatial Data Archiving Project, Interim Report. Available: [http://www.lib.ncsu.edu/hcgdap/documents/NCGDAP\\_InterimReport\\_June2008.pdf](http://www.lib.ncsu.edu/hcgdap/documents/NCGDAP_InterimReport_June2008.pdf) (accessed July 1, 2010)

<sup>8</sup> National States Geographic Information Council. GIS Inventory. Available: <http://www.gisinventory.net/> (accessed July 1, 2010)

in concert with the federal Geospatial One-Stop (GOS) portal, a metadata clearinghouse operated by the USGS National Geospatial Programs Office. Formalized, comprehensive, routine inventory processes which have been vetted by the data producers and stakeholders more efficiently serve general industry data discovery and access needs while also supporting archival efforts. An added benefit is that such inventory systems can produce basic metadata for documented data resources.

In 2006 RAMONA was implemented within the state as the inventory component of the NC OneMap program with support from NCCGIA staff participating in NCGDAP. After an initial push to market the inventory and recruit members, a snapshot of usage was taken in 2007. The initial report identified 96 local governments that had registered for the inventory, including 60 counties, 33 municipalities and 3 lead regional organizations. By 2007, more than 1,800 GIS datasets had been registered in the NC GIS Inventory, the most datasets registered by any state participating in the inventory. The most recent snapshot of the Inventory showed a continued growth in participation, with 224 organizations registered in the inventory, including 87 counties, 46 municipalities and 4 lead regional organizations. The number datasets has also grown significantly, with over 2,220 datasets in total being registered in the NC GIS Inventory. This number includes 887 total framework layers, including 115 orthoimagery collections and 78 cadastral datasets.

### Challenge: Data Acquisition

Another major challenge for the project has been to find sustainable and affordable ways by which data from 100 counties and as many as 140 municipalities could be acquired for archival purposes. Data acquisition experience prior to the project and especially in the early stages of the project made it clear that an infrastructure-based approach to data acquisition was needed in order to reduce acquisition costs and remove technical, legal, and financial barriers to data acquisition for the archive. Archiving aside, pressure within the data community for an infrastructure-based approach to data transfer was already arising from increasing local agency frustration with the volume of data requests received from state and federal agencies and other organizations such as universities.<sup>9</sup>

### Project Response

Initial project data acquisition plans were significantly modified in response to concerns within the data community about the volume of requests that state and federal agencies as well as others were imposing on local agencies. In addition, a dramatic increase in volume of data becoming available for acquisition forced a decision to only pursue a subset of available data while trying to maximize the learning experience, catalyze a community discussion about data preservation, and minimize negative impacts on statewide efforts to coordinate data acquisition efforts (the “first, do no harm” principle).

### Revised acquisition approach

The project proceeded on a two-track data acquisition approach that divided data sources into two groups: “low friction” and “high friction.” In low friction situations, where there are few or no technical or legal barriers to acquisition, data can be acquired at minimal cost. In the meantime,

---

<sup>9</sup> North Carolina Geospatial Information Coordinating Council. August 16, 2006. Local Government Committee, Requests by State Agencies for Data Produced by Local Governments, Report by the Local Government Committee of the Geographic Information Coordinating Council. Available: [http://www.ncgicc.com/Portals/3/documents/GICC\\_presentations\\_081606.zip](http://www.ncgicc.com/Portals/3/documents/GICC_presentations_081606.zip) (accessed July 1, 2010)

NCGDAP worked to support a variety of partnership efforts focused on increasing the number of “low friction” situations through the development of content exchange networks and the cultivation of open data sharing arrangements that have been explored by the state geographic information coordinating organizations.

#### Content exchange network development

Formal, structured data exchange networks, even if developed for other business reasons, support data archiving efforts by providing a low cost and routine means to acquire data which is authentic and well-documented, with clarified rights information. Against the background of increased local, state, and federal collaboration on development of data sharing infrastructure, data sharing mechanisms have begun to emerge in North Carolina, including the NC OneMap clearinghouse, the NCStreetMap.com street data sharing network, a digital orthoimagery “sneakernet” data sharing network, and a planned network for sharing of land parcel data.<sup>10</sup> NCGDAP played a supporting role in these efforts, providing feedback on the effectiveness of access and delivery mechanisms, and ensuring that archival use cases helped to inform requirements gathering. The archiving effort now benefits from the development of these infrastructures.

Content exchange networks such as NCStreetMap.com and the parcels data network being developed by the GICC Working Group for Seamless Parcels in North Carolina are beginning to help with another preservation challenge: inconsistently structured and coded data. These networks, in addition to facilitating streamlined access to data, are also beginning to cultivate and promote content standards according to which attribute information for commonly shared data layers such as roads and parcels begin to be structured using the same data attribute names and using shared classification schemes.<sup>11</sup> Consistency in attribute naming and coding across jurisdictional boundaries will help future users of the data to understand and interpret the data.

#### Data sharing agreements

Data that is openly shared without any use restrictions or copyright is much easier for archives to acquire, preserve, and make accessible. To promote open sharing of data the Local/State/Regional/Federal Data Sharing ad hoc Committee was created by the state geospatial coordinating council in February 2007 to study the problem of data sharing between government agencies and to develop specific recommendations that address the concerns of local, regional, state, and federal government agencies. NCGDAP project staff from NCCGIA and NCSU participated, and a final report was completed in November 2007,<sup>12</sup> with recommendations focusing on:

- Fostering partnership development across all organizations and levels of government.
- Avoiding wasteful duplication of effort.
- Optimizing the use of technical infrastructure to address business needs for information exchange.
- Ensuring effective and economical leveraging of geospatial resources for public benefit.

---

<sup>10</sup> North Carolina State University Libraries. June 2008. North Carolina Geospatial Data Archiving Project, Interim Report. Available: [http://www.lib.ncsu.edu/ngcdap/documents/NCGDAP\\_InterimReport\\_June2008.pdf](http://www.lib.ncsu.edu/ngcdap/documents/NCGDAP_InterimReport_June2008.pdf) (accessed July 1 2010)

<sup>11</sup> NC Statewide Mapping Advisory Council. May 2005. Content Elements for Statewide Publication of Core Geospatial Parcel Data, Version 1. Available: <http://www.ncgicc.com/Portals/3/documents/cadastral-standard.pdf> (accessed July 1, 2010)

<sup>12</sup> North Carolina Geospatial Information Coordinating Council, Ad Hoc Local/State/Federal Data Sharing Committee. Available: <http://www.ncgicc.com/Default.aspx?tabid=156> (accessed July 1, 2010)

A set of core practices were suggested to help data producers and content providers meet the intended goals for solving the issues with statewide data sharing. Included in that set of recommended practices was: “Establish a policy and procedure for the provision of access to historic data, especially for framework data layers.”

### Challenge: Determining Frequency of Data Capture

Many vector data layers are subject to ongoing update, the frequency of which may be a reflection of the pace at which the described features themselves change or a function of the operational processes of a particular agency. Cadastral data, for example, will tend to change on a fairly continuous basis in some agencies, while other agencies may handle updates in batch processes. Road and municipal boundary data also change but at a lower rate. One challenge faced by NCGDAP was to determine, with stakeholders, the frequency with which specific vector data layers should be acquired for archival purposes. Approaches to periodic data capture need to be cost effective while minimizing the amount of data loss between captures.

### Project Response

Early in the project, as outreach to local agencies was underway, it became clear that there was much to learn from individual agencies which were already creating data snapshots for their own business needs. While some anecdotal information about current practice had been acquired in the course of site visits and discussions with data custodians, it quickly became clear that there was a need to more formally and systematically gather data producer input on best (or, not bad) practices.

In 2006 the project elected to conduct a formal survey of local agency practice, with explicit focus on frequency of capture of key framework data layers.<sup>13</sup> An initial set of draft questions was developed by NCSU Libraries, NCCGIA, and the State Archives and Records Section, and then refined through discussions with the State Mapping Advisory Committee and the Local Government Committee (LGC). The survey targeted four framework data layers for detailed information gathering: parcels, street centerlines, jurisdictional boundaries, and zoning. A second local government survey as well as an initial survey of state agencies was executed in June 2008.<sup>14</sup> Additionally, a survey of data archiving practices of State Agencies was completed in 2008.<sup>15</sup>

In brief, about two-thirds of local government GIS coordinators are taking time to capture geospatial datasets, at least on an annual basis. It should be noted that, in formulating the surveys, it was very difficult to draw a distinction between regular data back-up for disaster recovery purposes and retention of geospatial records for archiving purposes, and it is expected that there were a number of false positives among those responses indicating archiving activity.

A by-product of the survey efforts has been that the process itself helped to socialize the problem of preservation within the state’s geospatial data community. The process of reviewing

---

<sup>13</sup> North Carolina Center for Geographic Information & Analysis. November 16, 2007. Frequency of Data Capture. Available: [http://www.nconemap.com/Portals/7/documents/NCOneMap\\_NDIIPLocalGovSurvey\\_1106.pdf](http://www.nconemap.com/Portals/7/documents/NCOneMap_NDIIPLocalGovSurvey_1106.pdf) (accessed July 1, 2010)

<sup>14</sup> North Carolina Center for Geographic Information & Analysis. February 1, 2009. 2008 Local Government Geoarchives Survey. Available: [http://www.nconemap.com/portals/7/documents/LocalGovt\\_GeoArchives\\_Survey\\_Results.pdf](http://www.nconemap.com/portals/7/documents/LocalGovt_GeoArchives_Survey_Results.pdf) (accessed July 1 2010)

<sup>15</sup> North Carolina Center for Geographic Information & Analysis and North Carolina State Archives. February 1, 2009. NC State Agency GeoArchives Survey. [http://www.geomapp.net/docs/StateAgency\\_GeoArchives\\_SurveyResults\\_NC.pdf](http://www.geomapp.net/docs/StateAgency_GeoArchives_SurveyResults_NC.pdf) (accessed July 1, 2010)

and refining the surveys together with various organizations as well as actual survey execution served to generate more awareness of the problem of digital preservation than any other outreach mechanism employed in the course of the project.

### Challenge: Engaging Data Producer Community

In order to engage the attention and resources of data producers, data custodians, and the state's geospatial data community it has been necessary to work towards building the business case for retention of historical data. Apart from acknowledging statutory drivers for data preservation, as outlined in Public Records Law and other requirements, administrative rules, laws, and policies, it has proven useful to promote and highlight business uses of historical data in a way that is tangibly related to the work and business problems faced by the data community since the path to digital preservation may lead through other more compelling business problems. For example, there is a significant overlap between the conjoined problems of business continuity and disaster preparedness and the lower priority problem of digital preservation.

### Project Response

Early project outreach focused on acquiring anecdotal use cases for older geospatial data, some of which include: land use change analysis, resolution of legal challenges, changes in the amount of impervious surfaces (which increase the propensity for flooding in a given area), shoreline change measurement, and site location analysis. In order to more systematically acquire information about business reasons for retaining older data a component of the Frequency of Capture surveys was designed to solicit county and municipal agency feedback on this topic. As a follow-up to these initial efforts, the Library of Congress-funded NDIIPP Geospatial Multistate Archive and Preservation Partnership (GeoMAPP) project has been working on developing a business case template that builds on a business case document developed by GeoMAPP project partners in Utah.<sup>16</sup>

Over the course of the project NCGDAP engaged the data producer community through outreach to individual producer agencies as well as geographic information coordinating organizations. Key points of engagement have included:

- Periodic updates to the NC Geographic Information Coordinating Council (GICC), which acted in an advisory capacity to the project.
- Regular project updates to the quarterly meetings of the State Mapping Advisory Committee.
- Presentations to a variety of stakeholder groups such as the GICC Local Government Committee and GIS Technical Advisory Committee.
- Participation in a statewide continuing education workshop program for land records managers.
- Presentations at statewide events such as the biennial North Carolina GIS conference and the annual meetings of the state chapter of the Urban and Regional Information Systems Association (URISA), the North Carolina Arc Users Group, and the North Carolina Property Managers Association.

---

<sup>16</sup> Geospatial Multistate Archive and Preservation Partnership. Interim Report. March 2010. Available: [http://www.geomapp.net/docs/GeoMAPP\\_InterimReport\\_Final.pdf](http://www.geomapp.net/docs/GeoMAPP_InterimReport_Final.pdf) (accessed July 1, 2010)

## Challenge: Industry Engagement

There is a degree to which, until relatively recently, one might have considered the geospatial industry to be to some extent “temporally-impaired” in terms of lack of access to older data, absence of tools for temporal analysis, and a predominant focus on applications and data services involving only the most current data.

## Project Response

NCGDAP took an aggressive approach to engaging the government and commercial data and vendor communities, going outside of the library and archives domains to present at high profile national and international industry meetings such as the Cambridge Conference (quadrennial meeting of heads of national mapping agencies), O’Reilly Where 2.0, the Urban and Regional Information Systems Association (URISA), the Geospatial Information & Technology Association (GITA), and GeoWeb. Project presentations were delivered in four consecutive years at the ESRI International Users Conference while NCGDAP also engaged in direct discussions with product development teams at ESRI to increase awareness of the data archiving challenge as a customer problem.

In terms of understanding how to engage the geospatial industry in the preservation challenge, a number of learning experiences have emerged in the course of the project, including:

- Promotion of temporal analysis opportunities and requirements indirectly promotes data preservation by cultivating demand for older data.
- Data is more likely to survive if users are made aware of the data’s existence and the data is being actively sought and used.
- Software and data vendors are increasingly coming to see maintenance and use of temporal data as an important customer problem.

One NCGDAP project objective had been to insert preservation use cases into the Open Geospatial Consortium (OGC) standards development and interoperability initiative processes.<sup>17</sup> The OGC defines standards covering a wide range of geospatial data interoperability and service scenarios, but preservation had not been in the scope of activity. After partnering with EDiNA on a separate geospatial data repository project in the UK,<sup>18</sup> NCSU Libraries teamed with EDiNA to present on the intersection of preservation issues with the OGC specification development space at the November 2005 OGC Technical Committee Meeting. A second thread of standards discussions arose within the NARA-led FGDC Historical Data Working Group, in which NCGDAP participated.<sup>19</sup> Representatives of NARA, NCSU Libraries, EDiNA, and others subsequently worked together towards the establishment of the OGC Data Preservation Working Group.<sup>20</sup>

The role of the working group, as outlined in the charter, is to address technical and institutional challenges posed by data preservation, to interface with other OGC working groups which

---

<sup>17</sup> Open Geospatial Consortium. Available: <http://www.opengeospatial.org> (accessed July 1, 2010)

<sup>18</sup> University of Edinburgh, EDiNA, Geospatial Repository for Academic Deposit and Extraction (GRADE). Available: <http://edina.ac.uk/projects/grade/> (accessed July 1, 2010)

<sup>19</sup> Federal Geographic Data Committee. Historical Data Working Group. Available: <http://www.fgdc.gov/participation/working-groups-subcommittees/hdwg> (accessed July 1, 2010)

<sup>20</sup> Open Geospatial Consortium, Data Preservation Working Group. Available: <http://www.opengeospatial.org/projects/groups/preservwg> (accessed July 1, 2010)

address technical areas that are affected by the data preservation problem, and to engage in outreach and communication with the preservation and archival information community. While working within the OGC has helped in the sense of inserting archival use cases into standards efforts, overall progress to date has been somewhat impeded by the fact that the OGC process only provides limited access to non-OGC members in the archival community. A further limitation is that the predominant focus of OGC efforts is on services and service-oriented architectures, with file-based scenarios receiving less attention. The challenge of supporting data persistence in a services-oriented environment presents a different set of problems to be addressed as part of standards efforts.

### Challenge: Metadata

The Federal Geographic Data Committee (FGDC) Content Standard for Geospatial Metadata has been widely adopted by state government agencies, yet the level of adoption is in general much lower among local government agencies. The state of North Carolina was an early adopter of the standard and NCCGIA has actively promoted the standard at the state and local level through grant-funded workshops and outreach. Metadata is frequently absent when data is acquired from local agencies, and to the extent that any metadata is received, it often needs to be enhanced in the following ways:

- Synchronization in order to improve concurrence of the data with the metadata.
- Normalization to adhere to a standard structure in order to support further metadata processing, including metadata element extraction as part of repository processing.
- Remediation to fix major errors and to enhance the suitability of key access fields for use in catalog and discovery environments.

In practice, the inconsistent nature of structure and content in received metadata makes this added value work very expensive. Project challenges and solutions with regard to metadata are outlined in detail elsewhere.<sup>21</sup>

### Project Response

A key component of spatial data infrastructure is the development and support of standards and best practices for data creation, data discovery and access, and metadata development. Wider adoption of standards and practices by data producers leads to a greater consistency of data and metadata received by the archive, making it possible to automate ingest workflows and lower archive development costs. Given the cost of processing heterogeneous metadata resources, the key to achieving efficient and cost effective handling of metadata may lie in the further development of formalized content exchange networks and data infrastructures in which the metadata is tightly bound to the data and flows within a standard framework that ensures metadata currency and authenticity while also promoting consistency in structure and content. Through metadata outreach, NC OneMap assists data providers in the construction of useful metadata documentation for common geospatial datasets and supports implementation of the FGDC metadata standard. Recently developed and emerging data inventories and networks are beginning to help meet these needs. Specific, recent operational examples of such infrastructure include:

---

<sup>21</sup> Steven P. Morris, James Tuttle, Robert Farrell. 2006. Preservation of State and Local Government Digital Geospatial Data: The North Carolina Geospatial Data Archiving Project, Proceedings of the IS&T Archiving Conference.

- The NC GIS Inventory, using RAMONA, which facilitates easy creation of at least minimal metadata by creating a metadata starter block that results from inventory submissions.
- NC OneMap metadata templates for key framework data layers, which promote consistency both in content and structure of metadata.<sup>22</sup>
- NCStreetmap, the new centerline data distribution system, which allows for at least minimum metadata to pass through the network in such a manner that the metadata is authenticated and consistently structured in such a way as to be suitable for automated ingest processes.<sup>23</sup>

## Challenge: Repository Development

Geospatial data involves complex multi-file, multi-format content and is increasingly likely to be delivered by web services or to be found within spatial databases. Development of a repository architecture that will accommodate such complex and heterogeneous content is a significant challenge. Ensuring that any particular archival approach is repository-agnostic and that archived content can be migrated to some unknown future system also raises formidable challenges.

## Project Response

NCGDAP initially elected to use the DSpace digital repository software not because it was expected to be an ideal platform for geospatial data but because the software met other requirements. The focus of NCGDAP has been on the social and organizational challenges associated with engaging the data producer and custodian community in data preservation. From the start, the project has been less focused on technical challenges associated with repository development, so an inexpensive repository solution was necessary. Since NCSU Libraries had already deployed DSpace for other repository projects it was possible to leverage that existing infrastructure for NCGDAP project needs at very low cost. A second reason for initially using DSpace arose from recognition of the fact that a large number of institutional repositories were already being deployed using DSpace, raising the question of whether geospatial data might fit in those repositories or whether domain-specific solutions were required.

To support the archive development process a demonstration repository using DSpace was deployed, and a robust repository ingest workflow was developed to handle the transformation of complex multi-file, multi-formats formats into discrete digital repository items. These processes are described in more detail elsewhere.<sup>24</sup>

In terms of addressing the technical challenges associated with geospatial data, two key data management issues have emerged in the course of the project:

### Items versus collections

While digital repository software such as that employed in the digital library community tends to distill digital content into discrete “items”, much geospatial data is easier to keep intact when

<sup>22</sup> North Carolina Geographic Information Coordinating Council. NC OneMap Metadata Templates. Available: <http://www.nconemap.com/Default.aspx?tabid=280> (accessed July 1, 2010)

<sup>23</sup> North Carolina Geographic Information Coordinating Council. NC StreetMap/CDDS. Available: <http://www.ncstreetmap.com/> (accessed July 1, 2010)

<sup>24</sup> Steven P. Morris, James Tuttle, Robert Farrell. 2006. Preservation of State and Local Government Digital Geospatial Data: The North Carolina Geospatial Data Archiving Project. Proceedings of the IS&T Archiving Conference.

managed as collections in either a file system or database due to the existence of interrelated and shared components within the collections. In the latter stages of the project, in response to the problem of complex data management, NCGDAP began to use a file system-based approach to managing data, additionally employing curation tools such as the open source Auditing Control Environment (ACE) software from University of Maryland, which is used to audit collections and identify missing or damaged files.<sup>25</sup>

### Spatial database curation

Spatial databases may require a three-pronged approach that involves: managing the content forward within the database over time, occasionally replicating the database in its entirety as an archival snapshot, and extracting individual database components (e.g., data layers) into easily preserved formats. The Shapefile format<sup>26</sup>, as a commercial yet openly-documented format has provided an intermediate safe zone in terms of managing vector data, but at the cost of lost content or functionality after extraction from spatial databases. The newer File Geodatabase format offers some potential for use in an archival context, and is being employed in an archival capacity by the states of Kentucky and Utah as part of the GeoMAPP project,<sup>27</sup> but the format is proprietary, although ESRI is currently developing an API that will open up access to File Geodatabases.<sup>28</sup>

The complete data is perhaps best retained by managing the temporal versions forward in time as part of the active production database maintained by the data producer, but that approach exposes the data to organizational risk in terms of long-term dependencies on that data producer organization. Transfer of database snapshots to a separate archive mitigates organizational risk to the record, but introduces technical risks since these snapshots may not be viable over the longer term. This approach also disconnects the temporal data in the snapshots from active discovery and use environments. Extract of individual datasets in the database into preservable forms such as Shapefiles mitigates both organizational and technical risk, but at the cost of immediate loss of data during the extract process.

## Unexpected Results of the Project

### Industry Engagement

Outreach and engagement with private industry occurred to a greater extent than anticipated in response to new industry interest in temporal and historical data.

User demand seems to indicate an increasing expectation that older data exists in digital form, and the industry as a whole has become more interested in enhancing the temporal aspect of geospatial work, with temporal and historical analysis functionality becoming built into desktop

---

<sup>25</sup> University of Maryland, ADAPT. ACE (Auditing Control Environment). Available: <https://wiki.umiacs.umd.edu/adapt/index.php/Ace:Main> (accessed July 1, 2010)

<sup>26</sup> Environmental Systems Research Institute. July 1998. ESRI Shapefile Technical Description, an ESRI White Paper. Available: <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf> (accessed July 1, 2010)

<sup>27</sup> Kentucky State Archives. Archival Procedures for Kentucky's Geospatial Data Assets. Available: [http://www.geomapp.net/docs/ky\\_geoarchives\\_procedures.pdf](http://www.geomapp.net/docs/ky_geoarchives_procedures.pdf) (accessed July 1, 2010)

<sup>28</sup> Environmental Systems Research Institute. March 24, 2010. Accessing Your Geodatabase Outside of ArcObjects. Available: <http://gisupdates.esri.com/video/devsummit2010/DEV37.wmv> (accessed July 1, 2010)

GIS software such as ArcGIS and digital globe environments such as Google Earth.<sup>29</sup>

NCGDAP was given the opportunity to present at a wide range of industry meetings, and it became clear that software firms and solutions providers have recognized the issue of data archiving and access as a customer problem. Environmental Systems Research Institute (ESRI), a major GIS software vendor and the provider of the software used by most North Carolina state and local GIS agencies, provided NCGDAP with the opportunity to meet with selected development teams to discuss preservation challenges and concerns, and NCGDAP gave presentations at four consecutive annual meetings of the ESRI International Users Conference. The formation of the OGC Data Preservation Working Group provided the opportunity to work directly with software providers, data vendors, major government agencies, and consulting firms in the context of standards development. Collaborative investigations into archiving issues together with private industry began as part of NCGDAP and have deepened through the course of the GeoMAPP project.<sup>30</sup>

### North Carolina GIS Inventory

In support of NCGDAP data inventory requirements, in 2006 NCCGIA led implementation and will be continuing with administration, reporting, and analysis operations related the NC GIS Inventory, a collection of information about GIS data creators and users from across the state and an inventory of the datasets that they create and maintain. The Inventory is powered by RAMONA, a geo-centric inventory database tool developed by the National States Geographic Information Council (NSGIC) to provide a central repository for states to track and manage GIS resources within their states. Despite solid participation from states such as Utah, Indiana and Wisconsin, North Carolina continues to lead the nation in datasets registered. Based on project experience using the RAMONA tool, both NCGDAP and GeoMAPP were able to provide NSGIC with feedback regarding desired functionality that would improve the utility of the inventory tool for archive development purposes. Adoption of the new RAMONA tool provided the opportunity to leverage infrastructure that was available at the national level and helped to inform continued development of that infrastructure according to archival requirements.

### GICC Archival and Long-Term Access Committee

A key, unexpected outcome of NCGDAP outreach efforts within the state was the formation of the Archival and Long Term Access ad hoc Committee, which was established by the North Carolina Geographic Information Coordinating Council (GICC) in November 2007 subsequent to the adoption of GICC Data Sharing Guidelines earlier in the same year. One recommendation in the Data Sharing Guidelines report included an action on long-term access to geospatial data. Specifically, the report recommended that “Data producers evaluate and publish their long term access, retention, and archival strategies for historic data.”<sup>31</sup> At the GICC meeting in which the Data Sharing Report was presented it was noted that the strong interest in archival and long term access guidelines indicated that more clarification and guidance was needed, resulting in

---

<sup>29</sup> Guy McGarva, Steve Morris, Greg Janee. Digital Preservation Coalition Technology Watch Report: Preserving Geospatial Data. May 2009. Available: <http://www.dpconline.org/advice/technology-watch-reports.html> (accessed July 1, 2010)

<sup>30</sup> Geospatial Multistate Archive and Preservation Partnership. Interim Report. March 2010. Available: [http://www.geomapp.net/docs/GeoMAPP\\_InterimReport\\_Final.pdf](http://www.geomapp.net/docs/GeoMAPP_InterimReport_Final.pdf) (accessed July 1, 2010)

<sup>31</sup> North Carolina Geospatial Information Coordinating Council, Recommendations for Geospatial Data Sharing, a report by the Local/State/Regional/Federal Data Sharing ad hoc Committee. Available: [http://www.ncgicc.org/Portals/3/documents/GICC\\_DataSharing\\_Final\\_11\\_07.pdf](http://www.ncgicc.org/Portals/3/documents/GICC_DataSharing_Final_11_07.pdf) (accessed July 1, 2010)

formation of an ad hoc Committee that included representatives from state, local, and federal agencies, as well as representatives from NCSU Libraries and the State Archives and Records Section. A draft report was reviewed and discussed by state and local agencies, with a final revised report submitted to the GICC in November 2008.<sup>32</sup> Although the work of the ad hoc Committee ended with the completion of the report and associated recommendations, an outcome has been that the issue of data archiving has been brought into the mainstream of ongoing discussions about data management and custodianship within the state's geospatial community, with the State Archives and Records Section also becoming deeply engaged in the issue of geodata preservation.

### Geospatial Multistate Archiving and Preservation Partnership (GeoMAPP)

From the time of the initial project proposal the intent of the North Carolina project had been to, through example, catalyze efforts in other states, working from the understanding that states frequently adopt innovations from each other and that innovations at the state level inform efforts at the national level. A major outcome of NCGDAP has been the initiation, starting in November 2007, of the NDIIPP Geospatial Multistate Archiving and Preservation Partnership (GeoMAPP) project.<sup>33</sup> Co-led by the North Carolina Center for Geographic Information & Analysis and the North Carolina State Archives and Records Section, this project is demonstrating, learning, and reporting on strategies to enable long term access and preservation of geospatial content. Kentucky and Utah are acting as partners, with involvement of state geospatial agencies as well as State Archives from each state. NCSU Libraries is participating in the project in an advisory and catalytic capacity.

In October 2009 GeoMAPP kicked off a new method of project engagement with the creation of an Informational Partners Program intended to share findings from the project and provide mentorship to participating states. The Informational Partners include: the District of Columbia, Georgia, Maine, Maryland, Minnesota, Montana, New York, Texas, Wisconsin, and Wyoming.

---

<sup>32</sup> North Carolina Geospatial Information Coordinating Council, GICC Archiving and Long Term Access Committee, Final Report. November 19, 2008. Available: [http://www.ncgicc.com/Portals/3/documents/Archival\\_LongTermAccess\\_FINAL11\\_08\\_GICC.pdf](http://www.ncgicc.com/Portals/3/documents/Archival_LongTermAccess_FINAL11_08_GICC.pdf) (accessed July 1, 2010)

<sup>33</sup> Geospatial Multistate Archiving and Preservation Partnership. Available: <http://www.geomapp.net> (accessed July 1, 2010)

## Plans for Ongoing Work

NCGDAP participation in a broader set of initiatives has ensured that preservation issues are being addressed in a broader context that involves significant government agency and industry support. Through partnership with established data infrastructures, benefits accrue to the preservation effort in terms of increased availability of data, greater efficiency in data acquisition, and improved consistency of data and metadata. In addition, utilization of the existing organizational infrastructure has allowed the project access to broad, ready-made audiences for preservation outreach.

### NC OneMap: Spatial Data Infrastructure

Working within the context of the NC OneMap network provides the opportunity to engage content through existing and developing organizational as well as technical infrastructures that are already self-sustaining within the state. Prior to NCGDAP, NC OneMap had not directly addressed data archiving. This network now offers a means by which to engage a large number of local agencies in the digital preservation effort on an ongoing, sustainable basis.

At the inception of the North Carolina Geospatial Data Archiving Project (NCGDAP) in the fall of 2004 NC OneMap was still in its infancy. During the subsequent five years NC OneMap grew from a limited demonstration data viewer web application to a mature GIS data clearinghouse and information portal. NC OneMap also offers a wealth of GIS and metadata documentation and links to major efforts taking place in North Carolina's GIS community. Another critical addition to the GIS community during this time period has been the implementation, marketing and management of the publicly accessible NC GIS Inventory. At the time of its official public launch in 2004, OneMap had enlisted 18 official partners participating in data sharing, and by the end of 2008 the number of partners had grown to 121. Current NC OneMap partners now include a diverse mix of federal, state and local government agencies and academic institutions. More than 80% of partners represent city or county government.

In addition to providing access to over 350 viewable datasets, made available as Web Map Service (WMS) end points, NC OneMap now has made a large collection of GIS datasets available for download. In 2005, due to demand from the GIS community, NC OneMap began hosting both vector and raster GIS datasets for download via FTP. By 2006, One Map had 101 vector and 6 raster datasets available for public download. By early 2010 those numbers had grown to 114 vector layers and 139 raster datasets. The large growth in raster holdings is largely due to the addition of county orthoimagery collections. Of the 139 raster datasets hosted by NC OneMap, 135 are local level orthoimagery series, representing 87 unique counties and municipalities. For 40 counties, NC OneMap has multiple series of orthoimagery including a snapshot of the most recently collected data and one or more superseded sets of imagery as well. The NC OneMap FTP site has become a highly utilized source for statewide and local GIS datasets, with the FTP site averaging over 5,500 visits and 325,000 hits per month. In addition to making superseded orthoimagery available on the FTP site, the NC OneMap database administrator keeps an internal archive of new and updated datasets received by NC OneMap. Recent holdings included 144 datasets managed in Personal GeoDatabases with almost 400 snapshots of NC OneMap data layers.

During the period of 2004 – 2009 North Carolina experienced a significant positive change in GIS resources available for public use. The creation and expansion of the NC One Map geospatial data clearinghouse and the NC GIS Inventory have helped increase the awareness of geospatial technologies and data in North Carolina. Both tools have become critical resources for the community at large to discover and access geospatial data. In addition to developing an archival capacity within the NC OneMap program itself, NCCGIA has begun to interface directly with the State Archives and Records Section to appraise NC OneMap data holdings, draft records retention and disposition schedules, and physically transfer data to State Archives for long-term custodianship.

## Drivers for Geospatial Data Archiving in North Carolina

The extent of ongoing investments in data archiving will to a great extent be a function of the existence of drivers that justify those investments. The NCGDAP team analyzed results from two 2008 surveys targeting local government and state agency GIS creators to supplement direct experience in technical service projects with local governments in order to better understand what the drivers for geospatial data archiving might be. Analysis of requirements, rules, laws, policies, and practices indicate that state and local geospatial data managers have four primary reasons or “drivers” for archiving datasets. These drivers are: 1) records retention policy related to the NC Public Records law, 2) popular demand for historic data, 3) information technology policy, and 4) administrative policy for access to superseded records. Each driver is further described.

### 1. Public Records

North Carolina General Statute 132 states that:

**“Public record”** or “public records” shall mean all documents, papers, letters, maps, books, photographs, films, sound recordings, magnetic or other tapes, electronic data-processing records, artifacts, or other documentary material, **regardless of physical form or characteristics**, made or received pursuant to law or ordinance **in connection with the transaction of public business** by any agency of North Carolina government or its subdivisions.

Clearly, geospatial datasets are public records and are subject to records retention requirements. In North Carolina, the Department of Cultural Resources has responsibility for records management and the selection and preservation of records considered essential. This includes efficient and economical management methods as well as standards, procedures and techniques for effective management of records. For geospatial data managers in state and local government agencies, public expectations for ready access to datasets, supported by the mandate, should be a strong incentive for effective data retention and retrieval processes.

### 2. User Requests for Historic Datasets and Analysis of Growth and Change

Local government geospatial datasets include aerial imagery, property boundaries, property tax valuation, and real estate ownership. Current data convey only part of a picture for legal, real estate, appraisal, and other private business purposes; historic datasets are invaluable for a range of questions relating to previous boundaries, ownership, zoning ordinances, service districts, and land uses. While historians are among the users of historic maps and datasets, private research is a more common driver for local records retention and retrieval. Data retention supports temporal analysis by local and regional

planners where changes in land use, population density, infrastructure, and other factors help identify trends and patterns that affect public policy and public decisions.

3. Information Technology Policy

Modern information technology practice includes regular replication of geospatial datasets at a minimum. Retention of a superseded instance of a dataset is not universal but does occur in some local GIS operations. In the local government survey, nearly a quarter of respondents indicated that information technology policy was a driver for geospatial record retention. However, it is likely that this practice is driven from business continuity requirements, which often focus on routine system back-ups and do not constitute true data archiving.

4. Administrative Policy

For local geospatial data management, tax administration policy drives archiving practice for more than one-fifth of respondents in the local survey. This reflects the intensive use of land records and associated tax information by public and private entities.

The results of surveys of local and state GIS data managers are displayed in Table 1. The survey results highlight some similarities in the primary drivers for archiving practices in state and local government, including historic mapping, archival policy, and change analysis. Several unique differences were identified as well. The differences appear to be related to the significance of tax administration for local governments, the influence of the Department of Cultural Resources on other state agencies' archival policies, and more common geospatial analysis in support of state programs. Records analysts from the State Archives and Records Section are becoming more aware of the importance of geospatial data and are asking about GIS datasets while creating or updating an agency's records schedule.

**Table 1. Business Rules and Policies That Drive Retention of Geospatial Datasets, North Carolina, 2008, Percent of Local Government and State Agency Respondents**

| <b>Business Rule or Policy</b>   | <b>Local</b> | <b>State</b> |
|--|--------------|--------------|
| Historic Mapping (including aerial imagery)                                  | 35.9         | 44.4         |
| Records Retention or Archival Policy   | 28.3         | 41.7         |
| Change Analysis (including land use, land cover, population, infrastructure) | 25.0         | 30.6         |
| Information Technology Policy  | 23.9         | 11.1         |
| Legal or Statutory Purposes  | 10.9         | 27.8         |
| Cultural Preservation  | 10.9         | 8.3          |
| Administrative Policy (including tax administration)                         | 21.7         | 0.0          |

## GICC Final Report and Recommendations

Recommendations provided in the final report of the GICC Archival and Long Term Access ad hoc Committee were intended as a beginning guide for initiating data retention and archiving practices and were not intended to comprise a comprehensive set of standards. Recommendations did not include the handling of older archived data, such as data already saved to tape or CD, but rather provided a “day forward” approach. A key organizing principle of the committee work was that recommended practices should not place an undue additional workload on state and local GIS professionals. It was understood that retention strategies should be easy to accomplish as part of the agencies' normal workflow, and that existing infrastructure should be employed as much as possible.

Key recommendations of the report included:

- The archiving schedule for geospatial data layers should be based on frequency of data update as well as business drivers identified by individual agencies. The report provided specific starting point recommendations for frequency of capture of key framework data layers.
- Participation in the NC OneMap Inventory was strongly recommended, and it was recommended that historic and superseded data be included in the inventory process.
- Online, networked storage of archival data was strongly recommended, and maintenance of at minimum three copies was recommended as an archival best practice.
- Format recommendations were based on a preference for publicly documented formats which are not proprietary, have no intellectual property restrictions, and benefit from broad support by software tools. The Shapefile format was recommended as an archival format for vector data, while GeoTIFF or TIFF (with georeferencing file) were recommended as the retention formats for raster data. Compression of archival data was not recommended.
- Consistent file naming schemes that make the characteristics of the data easily discernable from the name were recommended.
- Fully compliant FGDC metadata should be developed and should accompany the data.
- Data custodians should make archived data readily available to other agencies and the public.
- Periodic review of retention policies and as well as integrity of archived data was recommended.
- Organizations should publicize geospatial records retention schedules and archival practices online.

Recognizing the importance of integrating existing data custodianship infrastructure with archives and records processes under the purview of the State Archives and Records Section, the working group also recommended that North Carolina:

1. Develop records retention schedules with the Department of Cultural Resources to include GIS datasets and publicize the finalized schedule
2. Update the records retention schedule for NC OneMap
3. Develop the capacity and expertise at the Archives and Records Section to guide development of agency records retention schedules for geospatial content.
4. Develop a plan and implement steps at the Archives and Records Section to receive geospatial content as prescribed in emerging schedules, evaluating various options for getting and accepting geospatial content. A resource to consider is the NC OneMap clearinghouse and current work flows that move content from data stewards to the NC OneMap database.

As part of the GeoMAPP project, the State Archives and Records Section subsequently appraised the NC OneMap data collection and classified the majority of NC OneMap's holdings as "permanent" or "archival" records. Draft NC OneMap and local government records retention schedules were also subsequently produced as part of GeoMAPP.<sup>34</sup>

---

<sup>34</sup> Geospatial Multistate Archive and Preservation Partnership. Interim Report. March 2010. Available: [http://www.geomapp.net/docs/GeoMAPP\\_InterimReport\\_Final.pdf](http://www.geomapp.net/docs/GeoMAPP_InterimReport_Final.pdf) (accessed July 1, 2010)

The report acknowledged that the geospatial community is at the early stages of tackling the preservation problem and that the provided recommendations are intended to be initial, practical steps. The committee acknowledged that there are outstanding issues related to geospatial data preservation that are not addressed, or are only mentioned briefly, in the report and that an ongoing effort and additional guidelines will be needed, especially for spatial databases. It is worth noting that there is significant local agency interest in resurrecting old analog maps for use in the digital environment. This interest creates a point of engagement and dialog around the issue of preserving current geospatial data for use in future historical analyses.

## **Responding to an Evolving Content Domain**

The geospatial data domain involves a complex mix of data types and services, and much has changed in the domain since the inception of the project in 2004. Looking forward, it will be necessary for archiving efforts to take into account newer forms of content, a few of which are addressed here.

### **Non-spatial Place-based Content**

Place-based information, which is geo-located but not necessarily map-like in nature, is becoming an increasingly important component of the geospatial industry, with special importance in social networking, location-based services, and mobile applications. While geospatial datasets can provide a map- or image-based view of past landscapes, the non-spatial, location-based resources that are increasingly becoming available—especially in mass market geospatial data tools such as Bing Maps, Google Maps, and Google Earth—can provide an enhanced sense of place by making available images that provide street-level or oblique views of locations. Such information resources include Google Street View imagery, oblique (or “bird’s eye”) imagery products, three-dimensional building renderings such as those created with Google Sketchup, and tax assessment photography of buildings. The digital orthoimagery products used in traditional geographic information systems provide the straight-down view that is necessary to accurately locate features such as property lines or building footprints on the earth’s surface, yet all that one can see of a building is a rooftop. Non-spatial location-based imagery, in which buildings or places are seen from the side or ground level, provides a perspective of the landscape that is more fitting with how places are seen and remembered by people. Furthermore, three-dimensional information is becoming an increasingly important component of geospatial data, and some degree of convergence between the geospatial technology and Building Information Modeling (BIM) is occurring.

### **Digital Cartographic Products and PDF**

The true counterpart to the old, preserved map is not the current GIS dataset but rather the cartographic representation that builds on that data. The representation is the result of a collection of intellectual choices and application of current methods with regard to symbolization, classification, data modeling, and annotation. These representations typically occur in a complex proprietary project file format that is difficult to preserve, or in the context of an ephemeral web services interaction. Increasingly PDF is providing an option for static representations, since PDF can capture and preserve elements of cartographic representation, making the format a powerful tool for capturing finished output in a way that the underlying datasets cannot. Much underlying data intelligence is lost, but not to the extent that is the case with image snapshots. Complex PDF documents, including Geospatial PDF documents such as those that make use of the TerraGo GeoPDF method for geo-positioning, present new preservation challenges of their own. The GeoPDF approach for geo-positioning of documents

entered into the OGC open standards process in September 2008 and was voted into the status of an OGC Best Practice in February 2009.<sup>35</sup> Meanwhile Adobe's method for geo-positioning entered into the ISO standards process. Geospatial PDF documents are being used as an alternative preservation and access format by some partners within the GeoMAPP project.

## The GeoMAPP Multistate Project

NCGDAP began as a single state initiative with the understanding that each state presents its own myriad challenges and has its own unique organizational and technical infrastructures for geospatial data, though states commonly adopt innovations from each other while also adopting standards or best practices developed at the national level. Continuing and extending the work begun by NCGDAP to a wider set of states, GeoMAPP is exploring ways to expand the capabilities of state governments to provide long-term access to geospatial data. Geospatial and archival staff in the participating states are working together to identify, preserve, and make available data with ongoing research or other value. A key project component involves the testing of geographically dispersed content-exchange networks for the replication of state and local geospatial data among agencies within states as well as transfers between states to promote preservation and access. Project activity has included structured facilitation, collaboration among the geospatial and archival community, network building and outreach to other state partners and stakeholder associations.

Key GeoMAPP activities to date have included:

- Acquainting archives staff and geospatial agency staff with the nature of each others work, organizations, and technology.
- Evaluating and putting into place data inventory processes.
- Appraising geospatial records and establishing records retention schedules for geospatial data.
- Identifying existing and potential flows of data within states.
- Establishing procedures for routine transfer of data from geospatial agencies to archives.
- Investigating archiving challenges and requirements for complex GIS projects.
- Developing the business case for preserving geospatial data.
- Engaging key software vendors in discussions about customer requirements related to archival data management.

The GeoMAPP project accommodates a diversity of state organizational structures and technical environments, allowing each state to implement archival practices in a way that is fitting with the local infrastructure, while at the same time identifying commonalities in experience and identifying possible shared best practices and tools. The original GeoMAPP work plan was based on a two-year performance period from November 2007 through December 2009 and the project will be continued under a new work plan through mid-2011.

In North Carolina, NCCGIA and the State Archives and Records Section formed a joint team to investigate geospatial data archiving issues. Initial efforts focused on providing cross training on both GIS and archival technologies and terminology. Once joint familiarity with terminology and practices had been established, the team shifted focus to the creation of new geo-centric draft records retention schedules for local government and NC OneMap, while also assessing

---

<sup>35</sup> Open Geospatial Consortium. April 16, 2009. GeoPDF Encoding Best Practice. Available: [http://portal.opengeospatial.org/files/?artifact\\_id=33332](http://portal.opengeospatial.org/files/?artifact_id=33332) (accessed July 1, 2010)

the requirements for building a demonstration repository at State Archives. [see Appendix C for more detail on NCCGIA and State Archives collaborations]

Current economic conditions resulting from deep recession and a subsequent budget crisis within state government has limited the funds and resources available to fully implement a sustainable production geospatial data archiving system. Competing priorities and consolidation of department IT resources into the State CIO's office has also complicated the allocation of Department of Cultural Resources IT staff to assist with design and implementation of the system and has also introduced questions about potential management fees associated with the demonstration SAN space, not previously identified earlier in the project. Another challenge is the potential exponential growth of an archival system potentially ingesting NC One Map, local government and state agency geospatial data. Due to these factors, a full implementation of a data archiving system managed by State Archives with assistance from NCCGIA will require a staged approach.

## Plans for Stewardship of Content Acquired Through the Project

Data acquired by NCSU Libraries as part of the NC Geospatial Data Archiving Project will continue to be maintained by NCSU Libraries, while primary responsibility for stewardship of the state's geospatial data will fall to the NC Center for Geographic Information & Analysis, in a coordinating and data clearinghouse role, and to the State Archives and Records Section, in a long-term archival capacity.

### NCSU Libraries

Data under the stewardship of NCSU Libraries will continued to be maintained as part of the library's digital repository collections, which include the NCSU collection of electronic theses and dissertations, technical report series, scholarly publications, and digitized special collections. While the library's DSpace infrastructure has been used to support the NCGDAP project work, alternative approaches involving file system-based approaches such as those employed in California Digital Library (CDL) Microservices are being investigated for use with data archiving. At the same time, the Libraries will be exploring other opportunities to participate in pilot initiatives centered on cloud storage.

### Spatial Data Infrastructure and State Archives

As a result of the catalytic activities of NCGDAP, the recommendations of the GICC Archival and Long-Term Access Committee, and the initial work of the GeoMAPP multistate project the NC Center for Geographic Information & Analysis and the Archives and Records Section have begun to work closely together to put into place infrastructure and services. Draft records retention schedules have been updated to accommodate geospatial materials, capacity and expertise have begun to be put into place to implement processes to handle ingest of geospatial content as prescribed in emerging schedules.

## Conclusion

As a consequence of NCGDAP outreach and engagement with the data community, new geospatial data preservation initiatives have begun under the leadership of key stakeholders within the data community, with NCSU Libraries continuing to support these efforts in an advisory and catalytic capacity. It is expected that these efforts will be closely aligned with the most immediate business needs of those stakeholders, with the “who, what, where, why, and how” of data archiving being addressed in very practical terms and in a manner that is designed to maximize the likelihood of active participation of the data custodian community in the preservation effort. From the start the intent of the project work has been to catalyze discussion about data preservation and archiving within the data producer and custodian community and to engage spatial data infrastructure in this effort. The path to sustainability for this effort is seen to involve making data archiving part of the lifecycle process of the data. The key is to leverage existing infrastructure which could come to encompass data archiving and temporal data management functions without the need for prohibitively expensive additional investments.

From the beginning of the NCGDAP work it had been expected that the demonstration project would lead to a more formal involvement of the State Archives and Records Section in the state’s geospatial data archiving effort. State Archives functions such as local records outreach and retention schedule development serve as existing infrastructure which might be leveraged into geospatial data management. A key outcome of the NCGDAP effort has been to initiate the integration of the State Archives into the spatial data infrastructure of the state. After collaborating informally in NCGDAP project work, the State Archives and Records Section became co-lead in the GeoMAPP project and is now assuming a direct leadership role in the data archiving effort.

## Appendix A: NGDAP Outreach Events

| Event Name                              | Location          | Dates                   | Scope         | Presentations  |
|---|-------------------|-------------------------|---------------|--|
| All Hands Kickoff Meeting               | Washington D.C.   | Jan. 12-13, 2005        | National      | <a href="#">NDIIPP Project: Collection and Preservation of At-Risk Digital Geospatial Data</a>   |
| South Carolina CIO Office Visit         | Raleigh, NC       | Feb. 10, 2005           | State         | Overview of project  |
| GICC Technical Advisory Committee Visit | Raleigh, NC       | Feb. 9, 2005            | State         | Overview of project  |
| NC GIS 2005 Conference                  | Winston-Salem, NC | March 3-4, 2005         | State         | <a href="#">Long Term Preservation of Geospatial Data</a> (poster) & <a href="#">Long-Term Preservation of At-Risk Digital Geospatial Data: A Cooperative Agreement with Library of Congress</a> |
| UCSB Workshop                           | Santa Barbara, CA | March 7-8, 2005         | National      | <a href="#">Content and Practice: Background to the NC Geospatial Data Archiving Project</a>   |
| UCSD SuperComputer Visit                | San Diego, CA     | April 14, 2005          | National      | Discussion of possible points of collaboration   |
| DLF NDIIPP Panel                        | San Diego, CA     | April 15, 2005          | National      | <a href="#">NDIIPP Project: North Carolina Geospatial Data Archiving Project</a>   |
| ESRI Visit                              | Redlands, CA      | April 17-18, 2005       | National      | Meeting with Geodatabase, Cartography, Metadata, and Portal development teams  |
| NARA/FGDC Phone Conference              |                   | May 3, 2005             | National      | Participation in FGDC Historical Data Committee discussion   |
| IASSIST                                 | Edinburgh, UK     | May 25, 2005            | International | <a href="#">North Carolina Geospatial Data Archiving Project/NDIIPP: Collection and preservation of at-risk digital geospatial data</a>  |
| EDINA Meetings                          | Edinburgh, UK     | May 23 and May 30, 2005 | International | Meetings to discuss collaboration in the JISC-funded GRADE project   |

|  |                        |                  |               |   |
|--|------------------------|------------------|---------------|---|
| Digital Curation Centre Database Group | Edinburgh, UK          | May 30           | International | <a href="#">Collection and Preservation of At-Risk Digital Geospatial Data: the North Carolina NDIIPP Project</a>                                     |
| Partners Meeting                       | Warrenton, VA          | July 11-13, 2005 | National      | <a href="#">Key Issues and Findings in Work to Date</a>   |
| NDSAB Meeting                          | Washington D.C.        | July 14, 2005    | National      | <a href="#">North Carolina GDAP Project Overview</a>  |
| ESRI Users Conference                  | San Diego, CA          | July 23-28, 2005 | International | Project discussions   |
| GRADE Kickoff Meeting                  | Edinburgh, UK          | Sept. 28, 2005   | International | Project presentation and discussion   |
| CWRU GIS Symposium                     | Cleveland, OH          | Oct. 13-14, 2005 | National      | <a href="#">Map Portals and Geoarchiving: New Opportunities in Geospatial Information Service</a>   |
| NC ArcGIS Users Group                  | Wrightsville Beach, NC | Oct. 26-28, 2005 | State         | <a href="#">Cooperative Project with Library of Congress on Preservation of Digital Geospatial Data</a>   |
| Albemarle Regional GIS Meeting         | Camden County, NC      | Nov. 3, 2005     | State         | Project presentation and discussion   |
| NDIIPP Panel at DLF                    | Charlottesville, VA    | Nov. 7, 2005     | National      | Participation in panel discussion   |
| OGC Technical Committee Meeting        | Bonn, Germany          | Nov. 8-10, 2005  | International | <a href="#">Long-term preservation of digital geospatial data: challenges for ensuring access and encouraging reuse</a>                               |
| Charlotte/Mecklenburg GIS Day Exhibit  | Charlotte, NC          | Nov. 16, 2005    | State         | <a href="#">GIS Day 2005</a> (poster)   |
| CNI Task Force Panel                   | Phoenix, AZ            | Dec. 5-6, 2005   | National      | <a href="#">NDIIPP Project Briefing</a>   |
| NARA Technology Team Meeting           | Washington, DC         | Dec. 14, 2005    | National      | <a href="#">Preservation of Digital Geospatial Data: Challenges and Opportunities</a>   |
| LC Brown Bag and G&M Visit             | Washington, DC         | Dec. 15, 2005    | National      | <a href="#">Collection and Preservation of At-Risk Digital Geospatial Data: North Carolina Geospatial Data Archiving Project (NDIIPP Partnership)</a> |
| Partners Meeting                       | Berkeley, CA           | Jan. 9-11, 2006  | National      | <a href="#">NCGDAP Update</a> (poster)  |
| South Carolina GIS Conference          | Charleston, SC         | Jan. 23-25, 2006 | State         | Project presentation and discussion   |
| Transportation Research Board          | Washington D.C.        | Jan. 24, 2006    | International | Project presentation and discussion   |
| Meeting with State Archives            | Raleigh, NC            | Feb. 24, 2006    | State         | <a href="#">Collection and preservation of at-risk digital geospatial data</a>  |

|  |                   |                   |               |   |
|--|-------------------|-------------------|---------------|---|
| NSDI Partnership Office Webex Meeting    |                   | March 1, 2006     | National      | <a href="#">Collection and preservation of at-risk digital geospatial data</a>  |
| Eastern Carolina GIS Users Group         | New Bern, NC      | March 8, 2006     | State         | Project presentation and discussion   |
| Applied Research Associates Visit        | Raleigh, NC       | March 9, 2006     | State         | Project discussion  |
| Meet with State Archives                 | Raleigh, NC       | March 17, 2006    | State         | Project discussion; Introduction to NC OneMap   |
| GSA Southeastern Section                 | Knoxville, TN     | March 23-24, 2006 | National      | Presentation on NCGS maps project   |
| State Government Digital Archiving       | Wilmington, NC    | March 27-28, 2006 | National      | <a href="#">Workflow: Tools and Resources;</a><br><a href="#">Collection and preservation of at-risk digital geospatial data;</a><br><a href="#">Identification, Selection, and Appraisal within the North Carolina Geospatial Data Archiving Project (NCGDAP);</a><br><a href="#">Metadata Handling in the North Carolina Geospatial Data Project (NCGDAP);</a><br><a href="#">Collection Building Processes within the North Carolina Geospatial Data Archiving Project (NCGDAP);</a><br><a href="#">Preservation Strategies in the North Carolina Geospatial Data Archiving Project (NCGDAP)</a> |
| GITA                                     | Tampa, FL         | April 23-26, 2006 | National      | <a href="#">GITA 2006 poster presentation</a>   |
| NCPMA NC OneMap Outreach - Winston-Salem | Winston-Salem, NC | April 24, 2006    | State         | Project presentation and discussion   |
| Local Government Committee Telecon       |                   | April 25, 2006    | State         | Project presentation and discussion   |
| NCPMA NC OneMap Outreach – Tarboro       | Tarboro, NC       | May 4, 2006       | State         | Project presentation and discussion   |
| Joint NDIIPP-JISC Workshop               | Washington, DC    | May 7-9, 2006     | International | <a href="#">NCGDAP Project Overview</a>   |
| NCPMA NC OneMap Outreach – Jacksonville  | Jacksonville, NC  | May 12, 2006      | Local         | Project presentation and discussion   |
| Archiving 2006                           | Ottawa, CA        | May 23-26, 2006   | International | <a href="#">Preservation of State and Local Government Digital Geospatial Data</a>  |
| JCDL                                     | Chapel Hill, NC   | June 11-15, 2006  | International | <a href="#">NDIIPP Preservation Network: Progress, Problems, and Promise</a>  |

|   |                   |                        |               |  |
|---|-------------------|------------------------|---------------|--|
| Where 2.0   | San Jose, CA      | June 13-14, 2006       | International | <a href="#">The Disappearing Data Problem</a>  |
| July Partners Meeting   | Washington D.C.   | July 19-20, 2006       | National      | <a href="#">Preserving North Carolina Legacy Geologic and Topographic Maps</a>   |
| Joint Meeting of NAGARA/CoSA/SAA                                | Washington D.C.   | July 30 – Aug. 5, 2006 | National      | Participation in panel discussion  |
| ESRI Annual Users Conference                                    |                   | Aug. 7-11, 2006        | International | <a href="#">Spatial Data Infrastructure and Data Preservation in North Carolina</a>  |
| NC Property Mappers Assn. Fall Meeting                          | Raleigh, NC       | Oct. 11-13, 2006       | State         | <a href="#">GIS Data Preservation: Partnership with Library of Congress</a>  |
| DCC Workshop on Maintaining Long-term Access to Geospatial Data | Edinburgh, UK     | Oct. 27, 2006          | International | <a href="#">Maintaining Long-Term Access to Geospatial Data</a>  |
| GRADE 2nd Project Meeting                                       | Edinburgh, UK     | Oct. 30, 2006          | International |  |
| DLF 2006 Fall Forum   | Boston, MA        | Nov. 8, 2006           | National      | <a href="#">Collection and Preservation of At-Risk Digital Geospatial Data: NDIIPP Project Update on the NC Geospatial Data Archiving Project</a>                      |
| CRADLE  | Chapel Hill, NC   | Nov. 17, 2006          | State         | <a href="#">Preserving Digital Geospatial Data: The NC Geospatial Data Archiving Project</a>   |
| OGC Dec. 2006 Technical Meeting                                 | Tysons Corner, VA | Dec. 12, 2006          | International | Co-chaired Working Group session   |
| NDIIPP Winter 2006/2007 Partners Meeting                        | San Diego, CA     | Jan. 17-19, 2007       | National      | <a href="#">NDIIPP Project Update: NC Geospatial Data Archiving Project (NGDAP)</a>  |
| NCGIS 2007  | Winston-Salem, NC | March 1-2, 2007        | State         | <a href="#">Twenty Years of Spatial Vision, But What Does 1987 Look Like in Your GIS? Emerging Issues, Hindsight and Insights from the NC Preservation Partnership</a> |
| March 2007 PI Meeting   | Washington, DC    | March 23, 2007         | National      |  |
| OGC Technical Committee Meeting                                 | Ottawa, ON        | April 16-19, 2007      | International | Co-chaired Working Group session   |
| DigCCurr 2007   | Chapel Hill, NC   | April 18-20, 2007      | National      | <a href="#">Curation and Preservation of Complex Data: North Carolina Geospatial Data Archiving Project</a>  |
| ESRI Annual Users Conference                                    | San Diego, CA     | June 18-22, 2007       | International | <a href="#">Data Snapshot Archiving: A Frequency of Capture Survey</a>   |

|   |                    |                    |               |  |
|---|--------------------|--------------------|---------------|--|
| American Library Annual Meeting NDIIPP Symposium        | Washington, DC     | June 25, 2007      | National      | <a href="#">Preserving State and Local Government Digital Geospatial Data</a>  |
| NDIIPP Partners Meeting                                 | College Park, MD   | June 26-27, 2007   | National      | <a href="#">Preserved Digital Content: Collections, Value, and Stewardship</a>   |
| OGC Technical Committee Meeting                         | Paris, France      | July 9-12, 2007    | International | Co-chaired Working Group session   |
| Cambridge Conference                                    | Cambridge, UK      | July 18, 2007      | International | <a href="#">Geospatial Data Preservation Challenges at the Sub-National Level: The North Carolina Experience</a>   |
| National Digital Strategic Advisory Board Meeting       | Washington, DC     | July 24, 2007      | National      | <a href="#">Preserved Digital Content: Value to Public Policy Decision Making Now and in the Future</a>  |
| URISA Annual Meeting                                    | Washington, DC     | August 21-22, 2007 | International | <a href="#">Preserving State and Local Agency Digital Geospatial Data</a>  |
| OGC Technical Committee Meeting                         | Boulder, CO        | Sept. 17-20, 2007  | International | Co-chaired Working Group session and presented on quality, functionality, and sustainability factors   |
| LC Presentation to GICC Meeting (hosted)                | Raleigh, NC        | Nov. 7, 2007       | State         | <a href="#">Library of Congress Partnerships for Managing Geospatial Data</a>  |
| NDIIPP PI Meeting                                       | Washington, DC     | Nov. 5, 2007       | National      |  |
| OGC Technical Committee Meeting                         | Stresa, IT         | Dec. 10-13, 2007   | International | Co-chaired Working Group session   |
| NDIIPP Multi-state Geospatial Project Kickoff           | Salt Lake City, UT | Jan. 22-23, 2008   | National      | <a href="#">Archiving State and Local Agency Digital Geospatial Data: Overview of the Problem Area ; Archiving State and Local Agency Digital Geospatial Data: Looking for Solutions</a> |
| NC GICC Archival and Long-Term Access Committee Kickoff | Raleigh, NC        | Feb. 29, 2008      | State         | <a href="#">Archiving State and Local Agency Digital Geospatial Data: An Overview of the Problem Area</a>  |
| NCCGIA/State Archives Metadata Meeting                  | Raleigh, NC        | Mar. 4, 2008       | State         | <a href="#">Metadata Handling in the North Carolina Geospatial Data Project (NCGDAP)</a>   |
| OGC Technical Committee Meeting                         | St. Louis, MO      | Mar. 25, 2008      | International | Co-chaired Working Group session and presented on the Multi-state project and disaster response scenarios  |

|  |                  |                |               |  |
|--|------------------|----------------|---------------|--|
| Library of Congress Workshop: Preservation Issues Related to the Geospatial Data | Washington, DC   | Apr. 21, 2008  | National      | <a href="#">Part 1: Introduction to the Geospatial Data Content Area</a><br><a href="#">Part 2: Preservation Issues Related to the Geospatial Data</a> |
| DLF Fall Forum   | Minneapolis, MN  | Apr. 28, 2008  | National      | <a href="#">The Disappearing Data Problem: Preserving Today's Geospatial Data to Meet Tomorrow's Temporal Analysis Needs</a>                           |
| Indiana State University Digital Preservation Summit                             | Terre Haute, IN  | May 21, 2008   | State         | <a href="#">Challenges and Solutions for Digital Geospatial Data Preservation</a>  |
| NDIIPP 2008 Partners Meeting   | Ballston, VA     | July 10, 2008  | National      | <a href="#">Collecting Digital Content Going Forward: Lessons Learned and New Initiatives, NCGDAP</a>  |
| GeoWeb   | Vancouver, BC    | July 23, 2008  | International | <a href="#">Putting time into the GeoWeb: Data persistence in a web services environment</a>   |
| 2008 CURISA & CGITA GIS in the Carolinas   | Concord, NC      | Sept. 9, 2008  | State         | <a href="#">NCSU Libraries County and City Data Directory: Everything You've Always Wanted to Know</a>   |
| National States Geographic Information Council (NSGIC) Annual Meeting            | Keystone, CO     | Sept. 10, 2008 | National      | <a href="#">Preservation of Digital Geospatial Data Resources: A Team Climb</a>  |
| North American Cartographic Information Society (NACIS)                          | Missoula, MT     | Oct. 10, 2008  | National      | <a href="#">Next Generation Archives: The NC Geospatial Data Archiving Project</a>   |
| State Government Users Committee (NC)  | Raleigh, NC      | Oct. 12, 2008  | State         | <a href="#">Archiving Geospatial Data: Background to the Problem Area</a>  |
| DLF Fall Forum   | Providence, RI   | Nov. 12, 2008  | National      | <a href="#">Content Transfer: Getting Data Moved Around the Network, the NCGDAP experience</a>   |
| Workshop on Archiving Digital Cartography and Geoinformation                     | Berlin, Germany  | Dec. 4, 2008   | International | <a href="#">Preserving State and Local Government Digital Geospatial Data: North Carolina Partnerships</a>   |
| Coastal GeoTools   | Myrtle Beach, SC | Mar. 3, 2009   | National      | <a href="#">Preservation of Coastal Community Geospatial Content: What's Your Long Term Care Plan For Aging Data?</a>                                  |
| Indo-US Workshop on Trends in Digital Preservation                               | Pune, India      | Mar. 24, 2009  | International | <a href="#">Preserving Geospatial Data: Challenges and Opportunities</a>   |

|  |                    |                  |               |   |
|--|--------------------|------------------|---------------|---|
| Library of Congress Geospatial Content Strategies Meeting        | Washington, DC     | June 12, 2009    | National      |   |
| NDIIPP 2009 Partners Meeting                                     | Washington, DC     | June 24-26, 2009 | National      |   |
| Federation of Earth Science Information Partners (ESIP) Workshop | Santa Barbara, CA  | July 8, 2009     | National      | <a href="#">State and Local Agency Digital Geospatial Data: The North Carolina Experience</a>                           |
| ESRI International Users Conference                              | San Diego, CA      | July 14, 2009    | International | <a href="#">Are Geodatabases a Suitable Long-Term Archival Format?</a>  |
| NC Arc Users Group Fall Conference                               | Carolina Beach, NC | Sept. 17, 2009   | State         | <a href="#">The NC Geospatial Data Archiving Project: An Overview of Accomplishments</a>                                |
| Sun Persistent Archives Special Interest Group (PASIG)           | San Francisco, CA  | Oct. 7-9, 2009   | International |   |
| Library of Congress Digital Preservation GeoSUMMIT               | Washington, DC     | Nov. 13, 2009    | National      | <a href="#">Models for Shared Responsibility: Collaboration and Engagement with the NCGDAP and GeoMAPP Partnerships</a> |
| Open Geospatial Consortium Technical Committee Meeting           | Mountain View, CA  | Dec. 6, 2009     | International | <a href="#">Update on Geospatial Data Preservation Efforts</a>  |

## Appendix B: NCGDAP Publication List

Steve Morris, James Tuttle, and Jefferson Essic. A Partnership Framework for Geospatial Data Preservation in North Carolina. *Library Trends*, Vol. 57, No. 3, Winter 2009. Available:  
[http://www.lib.ncsu.edu/ncgdap/documents/Morris\\_LibraryTrends\\_09.pdf](http://www.lib.ncsu.edu/ncgdap/documents/Morris_LibraryTrends_09.pdf)

Guy McGarva, Steve Morris, and Greg Janee. Preserving Geospatial Data, Digital Preservation Coalition Technology Watch Report 09-01. May 2009. Available:  
<http://www.dpconline.org/vendor-reports/download-document/363-preserving-geospatial-data-by-guy-mcgarva-steve-morris-and-gred-greg-janee.html>

Steven P. Morris, Preserving Geospatial Data: Challenges and Opportunities Proceedings Indo-US Workshop on International Trends in Digital Preservation, March 24-25, 2009. Available: <http://ndpp.in/download/Indo-US-DP-Proceedings-C-DAC-2009.pdf>

North Carolina Center for Geographic Information & Analysis. 2008 Local Government Geoarchives Survey. February 2009. Available:  
[http://www.nconemap.com/portals/7/documents/LocalGovt\\_GeoArchives\\_Survey\\_Results.pdf](http://www.nconemap.com/portals/7/documents/LocalGovt_GeoArchives_Survey_Results.pdf)

North Carolina Center for Geographic Information & Analysis. North Carolina State Agency GeoArchives/Retention Study. February 2009. Available:  
[http://www.geomapp.net/docs/geomapp\\_survey\\_state.pdf](http://www.geomapp.net/docs/geomapp_survey_state.pdf)

North Carolina Center for Geographic Information & Analysis. North Carolina State Agency Results. February 2009. Available:  
[http://www.geomapp.net/docs/StateAgency\\_GeoArchives\\_SurveyResults\\_NC.pdf](http://www.geomapp.net/docs/StateAgency_GeoArchives_SurveyResults_NC.pdf)

Steven P. Morris, Zsolt Nagy, Jim Tuttle. North Carolina Geospatial Data Archiving Project Interim Report. June 2008. Available:  
[http://www.lib.ncsu.edu/ncgdap/documents/NCGDAP\\_InterimReport\\_June2008.pdf](http://www.lib.ncsu.edu/ncgdap/documents/NCGDAP_InterimReport_June2008.pdf)

North Carolina Center for Geographic Information & Analysis. Frequency of Capture Survey Report. September 2006. Available:  
[http://www.nconemap.com/Portals/7/documents/NCOneMap\\_NDIIPLocalGovSurvey\\_1106.pdf](http://www.nconemap.com/Portals/7/documents/NCOneMap_NDIIPLocalGovSurvey_1106.pdf)

North Carolina Center for Geographic Information & Analysis. Frequency of Capture Survey Questions. September 2006. Available:

[http://www.nconemap.com/Portals/7/documents/Appendix%20NCOneMap\\_NDIIPLocalGovQuestions\\_\\_1106.pdf](http://www.nconemap.com/Portals/7/documents/Appendix%20NCOneMap_NDIIPLocalGovQuestions__1106.pdf)

Steven P. Morris, Geospatial Web Services and Geoarchiving: New Opportunities and Challenges in Geographic Information Services, *Library Trends*, Vol. 55, No. 2, Fall 2006. Available:

<http://www.lib.ncsu.edu/ncgdap/documents/MorrisLibraryTrendsFall2006.pdf>

Steven P. Morris, James Tuttle, and Robert Farrell. Preservation of State and Local Government Digital Geospatial Data: The North Carolina Geospatial Data Archiving Project. *Proceedings of Archiving 2006, The Society for Imaging Science and Technology*. Available: <http://www.lib.ncsu.edu/ncgdap/documents/archiving2006.pdf>

NCGDAP, CGIA and NC State Archives:  
*A Short History of Geoarchiving in North Carolina*

NC Center for Geographic Information & Analysis  
April 2009

The North Carolina Geospatial Data Archiving Project (NCGDAP) has been a significant catalyst for discussion about the need for archiving geospatial content in North Carolina. NCGDAP and subsequent parallel efforts with the NC Geographic Information Coordination Council (GICC) and the Geospatial Multistate Archives and Preservation Partnership (GeoMAPP) have increased awareness about the archival of geospatial records, helped form a partnership between State Archives and the NC Center for Geographic Information and Analysis (CGIA), and have launched efforts to develop GIS-inclusive records retention schedules and the build of a demonstration Geoarchives storage system.

**Pre NCGDAP**

Prior to the Fall 2004 launch of the NCGDAP project, archiving of geospatial datasets in North Carolina was largely a localized effort driven by GIS data producers, often tied to regular back-up management practices or informal archives including personal collections of data stored cds and tapes. There was little statewide awareness of the issue. Records retention schedules, which were adept at handling paper records, made little mention of digital geospatial records. CGIA's records retention schedule, last updated in 1997, makes mention of GIS data for the "Corporate Geographic Information System" and provided guidance for inclusion of this data in a regular back up cycle, but the actual disposition for this data was to "erase or update in office recorded information when administrative value ends." CGIA project files and applications were to be "erase(ed) in office after 3 years."

**GICC Involvement**

The GICC's early interest in preserving and providing access to superseded geospatial content was made evident in the May 2003 vision statement for the development of NC OneMap. The council stated that "Historic and temporal data will be maintained and available" as one of the characteristics of NC OneMap. Data archiving was highlighted again in the findings of the 2007 Local/State/Regional/Federal Data Sharing ad hoc Committee. The Data Sharing Committee was comprised of members from the local, state and federal government and university communities and was focused on addressing demands on local government to provide their locally created data to multiple requesting agencies. Of the group's ten recommendations, recommendation #8

stated “Data producers should evaluate and publish their long term access, retention, and archival strategies for historic data.”<sup>36</sup>

Based on the findings of the Data Sharing Committee, the GICC created the Archival and Long Term Access Ad hoc Committee in November of 2007 to further investigate the issue of archiving of geospatial data. This committee also had a diverse membership, drawing from all levels of government and included participants from State Archives, CGIA, NC State Libraries and the Library of Congress. In November of 2008 the group presented its findings to the GICC.<sup>37</sup> In addition to specific recommendations about data format, storage media, metadata and frequency of capture of data, the committee made the following key recommendations:

1. Update the records retention schedule for NC OneMap
2. Develop the capacity and expertise at the Archives and Records Section to guide development of agency records retention schedules for geospatial content.
3. Develop a plan and implement steps at the Archives and Records Section to handle ingest of geospatial content as prescribed in emerging schedules. This plan should evaluate various geospatial content harvesting and ingest scenarios, including the leveraging of the NC OneMap clearinghouse and data work flow as a content transfer point between data stewards and the Archives and Records Section.

### **CGIA/ State Archives Relationship**

Following the recommendations of the GICC’s Archives committee and to address project deliverables from both the NCGDAP and GeoMAPP Library of Congress projects, CGIA and State Archives formed a team to investigate geospatial archiving issues and began meeting weekly in June of 2008. One of the initial goals of these meetings was to provide cross training on both GIS and archives technologies and terminology.

Key archives terminology and processes imparted to CGIA staff included:

- what is a Chief Records Officer and their role in state government archives
- details on records retention scheduling
- records appraisal and determination dispositions and frequency of capture
- how data is managed with the MARS cataloging system

To increase the State Archives staff’s familiarity with GIS, CGIA set up a half day hands on training session using ArcGIS software in August 2008. Archives staff also had the software installed on their machines after the training. Other presentations to further the archives staff knowledge about GIS included a description of the Sustainable Sandhills land use modeling

---

<sup>36</sup> The full data sharing report can be found here: <http://www.ncgicc.com/LinkClick.aspx?link=156&tabid=306&mid=547>

<sup>37</sup> The full data sharing report can be found here:

[http://www.ncgicc.com/Portals/3/documents/Archival\\_LongTermAccess\\_FINAL11\\_08\\_GICC.pdf](http://www.ncgicc.com/Portals/3/documents/Archival_LongTermAccess_FINAL11_08_GICC.pdf)

project and its underlying datasets and analysis as well as a hands on demonstration of the NC GIS Inventory.

In June of 2008 the NC OneMap database administrator presented an in depth review of NC OneMap. Topics covered included: the types of data hosted on NC OneMap, differences in hosted GIS data formats, naming conventions, and the ingest process of data into NC OneMap. Also presented was the DBA's current informal process of how superseded datasets are archived internally.

The team also traveled to NCSU in June to meet with the NCGDAP team and review the NCSU team's Geoarchives system. Items such as data ingest, metadata management, file system layout, and storage considerations were noted for potential inclusion into the future demonstration Geoarchives environment to be implemented at State Archives

As both groups became more comfortable with each others terminology and practices, in depth work on a review of existing records retention schedules began. The team began looking at the effective dates of the existing schedules of state agency GIS producers, as well as reviewing specific state agency and local government schedules. To help bridge information from archives and GIS inventories, the team extracted and cross referenced information from the NC GIS Inventory and spatial items from existing local and state government schedules to find potential matches to help refine the schedules.

Much of the team's efforts during the summer of 2008 was focused on the design and release of several surveys that focused on archiving practices for geospatial data. Two surveys targeted NC local government and state agency GIS data creators while another two targeted State Archivists and state GIS coordinators nationally.

## **Schedule Writing**

By the fall of 2008, the CGIA/ State Archives project team had moved beyond the initial cross training and familiarization phase and had shifted focus to create new geo-centric draft records retention schedules for local government and NC OneMap, while also beginning the process of assessing the requirements for building a demonstration Geoarchives repository at State Archives.

The drafting of new local government schedules that included geospatial datasets began with a formal review of existing local government datasets. Some geospatial datasets or their paper map counterparts existed in some disparate local government agency schedules, but there was no general schedule that specifically handled digital geospatial records in a consolidated fashion. The initial draft schedule included key local government data sets such as parcel, zoning, boundary, road and orthoimagery data in one consolidated general schedule and referenced local government agencies that typically produce or manage this data. The frequency of capture of these datasets was typically an annual snapshot.

The disposition instructions included two options for permanent storage of the superseded data:

Either:

Transfer snapshot to NCOneMap according to established procedures, complying with standards and procedures adopted by the North Carolina Geographic Information Coordinating Council.

Or,

If retained in office permanently, your agency must comply with standards (for metadata, file naming, data sharing, and long term preservation) and procedures adopted by the North Carolina Geographic Information Coordinating Council.

This proposed disposition approach defines NC OneMap as the initial recipient for archived geospatial records. The NC OneMap team would be responsible for aggregating these local government datasets and then transferring them to State Archives, thus acting as a conduit between local governments and State Archives. This approach was driven by the key recommendations from the GICC's Archives and Long Term Access committee. The team's strategy for addressing geospatial records schedule drafting was presented to the GICC on February 11<sup>th</sup> 2009. The GICC gave the project team permission to engage GICC committees to further the drafting and review process.

In preparation for engaging the GICC's Local Government Committee, the team met with Wake County's database administrator to review the local government schedules and receive feedback. The team also presented the schedules to the GeoMAPP partners during an extended web seminar and received additional valuable feedback. On April 21 2009, the modified draft schedules were presented to the Local Government Committee. The committee took the action to form a smaller working group to further investigate the schedules and to gather additional feedback from local government GIS creators that are not directly involved with the GICC.

While efforts to build and refine the local government schedules were underway, the project team was also working with NC OneMap staff to develop a records retention schedule for NC OneMap datasets. The first step in this process was to develop an inventory of all of the datasets that were hosted on NC OneMap and were available for download by FTP. The inventory included several fields to help identify and organize the available data to help aid in the appraisal process of each of the datasets. Inventory items included:

- Layer Name
- ISO 19115:2003 Category
- ISO 19115:2003 Keyword
- Source Agency
- Creation Date
- Update Frequency
- Notes (a details section to help further describe the dataset)
- Data Format (Shapefile, MrSID, etc)
- File size

The inventory template was then populated with NC OneMap's downloadable holdings. Datasets were further delineated between framework and non framework layers. In total, 225 datasets were added to the inventory including 23 vector and 117 raster framework layers and 85 non framework layers.

The project team used the information in the inventory to formally appraise each dataset for its archival worthiness and to determine frequency of capture. All but one dataset were determined to be either permanent or archival in their disposition. The next step was to take the appraisal findings and create a records retention schedule that captured these diverse data.

In order to prevent the NC OneMap records retention schedule from having 225 entries to address each dataset hosted, the team created a classification strategy to reduce the number of specific datasets referenced and to provide flexibility for datasets to be added in the future. NC OneMap currently has 16 datasets that it hosts that are regularly updated by the data creator. These datasets were individually referenced in a Dynamic Datasets portion of the schedule and have an annual capture from an archives perspective.

NC OneMap hosts a large number of orthoimagery datasets. Due to the large number of ortho series in NC OneMap's holdings acquired from multiple counties and the unique nature of the data (static dataset, but superseded when new imagery is created several years later), orthoimagery is captured as a unique schedule general category. Orthoimagery datasets are archived when they are superseded by newer imagery.

The final category in the NC OneMap schedule is the Thematic/Categorical general category. This section was set up to capture static GIS data derived from projects or other activities that are a single snapshot of a phenomenon in time. Since these datasets are not regularly updated, they only need to be archived once, upon receipt by NC OneMap. Instead of referencing each unique dataset, the schedule captures these holdings more generally by ISO Category to provide flexibility. In the schedule, all 19 ISO categories are listed and described.

In addition to the CGIA and State Archives project team's efforts to create geo-centric local government and NC OneMap records retention schedules, the team also reached out to other state agencies to raise awareness and to identify future agency schedule changes to include information about digital geospatial records. Specific outreach included presenting at the GICC's State GIS Users group meeting in October of 2008 and agency specific meetings held by State Archives staff with the Department of Transportation, North Carolina Emergency Management and the Department of Environment and Natural Resources' Division of Land Quality. The GICC also tasked the project team with following up with the State GIS Users Group's executive committee to review the NC OneMap and local government schedules as well as plans moving forward with state government. State government GIS schedules would likely contain similar disposition language to the local government schedules, leaving departments the option to permanently retain critical datasets internally or to hand them off to NC OneMap to consolidate and transfer to State Archives.

## **Geoarchives System Design and Procurement**

A secondary effort addressed by the project team has been to design, procure and implement a demonstration system to ingest and manage archived geospatial datasets at State Archives. Initial discussions about system architecture requirements began in the Fall of 2008 and continued through the Spring of 2009 with a systems group including staff from CGIA/ NC OneMap, State Archives, NC State Libraries and Department of Cultural Resources IT department.

The initial challenge faced by the team was sizing of the database environment. To follow the recommendations from the GICC's Archives committee, the group decided on using network attached SAN storage for fast online access to the data. Data sizing estimates were based largely on the NC OneMap data inventory exercise where dataset sizing was captured. The NC OneMap downloadable data holdings totaled 13.9 terabytes of uncompressed data. Uncompressed orthoimagery comprised almost this entire storage block, with uncompressed vector data only totaling 7.3 gigabytes. In addition to SAN space, several portable hard drives were needed as well to help out with data transfer activities.

The system specification and procurement analyses were actually fairly simple. SAN space in excess of 14 terabytes was needed to ingest the uncompressed NC OneMap holdings, several portable hard drives were needed for data transfer and funds were limited to what was available with the GeoMAPP project grant. In the end, \$27,000 of available grant funds were used to purchase 15 terabytes of SAN disk and 4 portable drives totaling 7 terabytes in size. In addition to the SAN and portable drives, an application server has been allocated to the project that will be directly attached to the SAN and will be used for managing the archived data. Folder lay-out and ingest and data management practices were still ongoing at the time of writing of this paper, but initial plans aim to set up an ISO 19115:2003 category based file structure hierarchy in which individual shapefiles and TIFF images will be stored and managed. The team also hopes to utilize check sum and hashing tools to help perform data validation as well as data manifest creating software to track file contents and folder structures during the transfer processes.

## **Next Steps**

While CGIA's participation in the NCGDAP project is winding down in April 2009, the overall effort to address the archiving of geospatial datasets continues. With heightened awareness from the GICC and continued funding support from the GeoMAPP project, the project team will continue its work with schedule writing and the build of the Geoarchives environment at State Archives.

Current economic conditions resulting from deep recession facing the country and a subsequent budget crisis within state government has limited the funds and resources available to fully implement a sustainable production Geoarchives system. Competing priorities and consolidation of department IT resources into the State CIO's office has also complicated the allocation of Department of Cultural Resources IT staff to assist with design and implementation of the

system and has also introduced questions about potential management fees associated with the demonstration SAN space, not previously identified earlier in the project. Another challenge is the potential exponential growth of a Geoarchives system potentially ingesting NCOne Map, local government and state agency geospatial data. Due to these factors, a full implementation of a Geoarchives system managed by State Archives with assistance from CGIA will require a staged approach.

The next stage will be to further review the current draft records retention schedules with the GIS community and continue to enhance the schedules while getting buy in and sharing the message of the value of archiving critical geospatial data. This stage will also involve finalizing development and implementation of the demonstration Geoarchives environment at State Archives. The initial load of this environment would include all NC OneMap vector holdings, superseded orthoimagery from NC OneMap's holdings, a sampling of targeted local government datasets, and GIS project files and digital output products such as digitized maps. As part of GeoMAPP project aims, data transfers within the state and between partner state members will be used to document and refine these processes.

Future stages could include development of business case documentation to help sell the idea of a formal Geoarchives program to decision makers. Plans could include acquisition of "study" funds or additional grant funds to further investigate the issue, with the end goal of establishing recurring funds to support an ongoing program.

### **2008 Survey of North Carolina State Agency Archiving Practice**

North Carolina Center for Geographic Information & Analysis. North Carolina State Agency GeoArchives/Retention Study. February 2009. Available:  
[http://www.geomapp.net/docs/geomapp\\_survey\\_state.pdf](http://www.geomapp.net/docs/geomapp_survey_state.pdf)

North Carolina Center for Geographic Information & Analysis. North Carolina State Agency Results. February 2009. Available:  
[http://www.geomapp.net/docs/StateAgency\\_GeoArchives\\_SurveyResults\\_NC.pdf](http://www.geomapp.net/docs/StateAgency_GeoArchives_SurveyResults_NC.pdf)

### **2008 Survey of North Carolina Local Agency Archiving Practice**

North Carolina Center for Geographic Information & Analysis. 2008 Local Government Geoarchives Survey. February 2009. Available:  
[http://www.nconemap.com/portals/7/documents/LocalGovt\\_GeoArchives\\_Survey\\_Results.pdf](http://www.nconemap.com/portals/7/documents/LocalGovt_GeoArchives_Survey_Results.pdf)

### **2006 Survey of North Carolina Local Agency Archiving Practice**

North Carolina Center for Geographic Information & Analysis. Frequency of Capture Survey Report. September 2006. Available:  
[http://www.nconemap.com/Portals/7/documents/NCOneMap\\_NDIIPLocalGovSurvey\\_1106.pdf](http://www.nconemap.com/Portals/7/documents/NCOneMap_NDIIPLocalGovSurvey_1106.pdf)

North Carolina Center for Geographic Information & Analysis. Frequency of Capture Survey Questions. September 2006. Available:  
[http://www.nconemap.com/Portals/7/documents/Appendix%20NCOneMap\\_NDIIPLocalGovQuestions\\_\\_1106.pdf](http://www.nconemap.com/Portals/7/documents/Appendix%20NCOneMap_NDIIPLocalGovQuestions__1106.pdf)

## Appendix E: GICC Archiving Report

The Archival and Long Term Access ad hoc Committee was established by the Council in 2007. It was a direct outcome to one of the “Ten Recommendations in Support of Geospatial Data Sharing” adopted by the Council at its November 7, 2007 meeting. That recommendation stipulated that “data producers should evaluate and publish their long term access, retention, and archival strategies for historic data.”

Chaired by GICC member Anne Payne, the ad hoc committee, representing all government sectors, framed the issue. While key feature data layers such as land records, street centerlines, jurisdictional boundaries, and zoning are constantly changing, current data management practices commonly involve overwriting of older versions of data which are then no longer available. If retained, the data could serve several business purposes, such as, historical/cultural interests, support of legal proceedings, enforcement of environmental regulations, and aid in analysis of trends. Retention and preservation requirements and schedules, if they exist, are not considered nor included in up-front data life-cycle planning, budgeting, nor in work flow development, by local and state agencies. To the extent that data snapshots are retained, the archived data does not tolerate neglect: Long-term preservation will involve migration of data to supported data formats, media refresh, and retention of critical documentation.

The ad hoc committee considered numerous aspects of this community-wide problem and suggested guidelines, and made three recommendations involving records retention. Archiving geospatial public information should become part of an agency’s records retention plan.

The Final Report was accepted by the Council at its November 19, 2008 meeting. Available: [http://www.ncgicc.com/Portals/3/documents/Archival\\_LongTermAccess\\_FINAL11\\_08\\_GICC.pdf](http://www.ncgicc.com/Portals/3/documents/Archival_LongTermAccess_FINAL11_08_GICC.pdf)

Committee website: <http://www.ncgicc.com/Default.aspx?tabid=306>

| <u>Name</u>       | <u>Organization</u>                    |
|-------------------|--|
| Anne Payne, Chair | Wake County                            |
| Kathryn Clifton   | City of Salisbury                      |
| Mark Crane        | USDA                                   |
| Kelly Eubank      | NC Dept of Cultural Resources          |
| John Gallimore    | Davie County                           |
| Tracey Glover     | City of Fayetteville                   |
| Amy Keyworth      | NC DENR - Water Quality                |
| Bill Lefurgy      | Library of Congress                    |
| Butch Lazorchak   | Library of Congress                    |
| Scott Miller      | Western Piedmont Council of Government |
| Tom Morgan        | Secretary of State - Land Records      |
| Steve Morris      | NCSU Libraries                         |
| Zsolt Nagy        | NCCGIA                                 |
| Doug Newcomb      | US Fish and Wildlife Service           |
| Thomas Parrish    | NC Dept of Cultural Resources          |

Joe Sewash  
Ed Southern  
Rebecca Troutman

NCCGIA  
NC Dept of Cultural Resources  
NC Association of County Commissioners

## Appendix F: GICC Data Sharing Report

The Geographic Information Coordinating Council adopted ten recommendations for geospatial data sharing at the November 7, 2007 meeting. The ad hoc Committee worked through several drafts, soliciting wide public comment, before presenting the final recommendations.

Full report available :

[http://www.ncgicc.org/Portals/3/documents/GICC\\_DataSharing\\_Final\\_11\\_07.pdf](http://www.ncgicc.org/Portals/3/documents/GICC_DataSharing_Final_11_07.pdf)

Summary of recommendations and core best practices available:

[http://www.ncgicc.org/Portals/3/documents/2007GICC\\_DataShare\\_2pgSum.pdf](http://www.ncgicc.org/Portals/3/documents/2007GICC_DataShare_2pgSum.pdf)

The ten recommendations include:

- Avoid formal agreements between agencies
- Web access for data handling and distribution
- Establish secure access sites when necessary
- Free data exchange between government agencies
- Publish single point of contact for data within each organization
- Employ regional solutions for data collection and data sharing when appropriate
- Acquire data only from original sources and official outlets
- Establish archive and long term data access strategies
- Display NC OneMap brand on collaborative websites
- Outreach to local governments and state agencies

The attachments referenced in the recommendations, include:

- Attachment A: Requests by State Agencies for Geospatial Data Produced by Local Government
- Attachment B: Guidelines for Providing Appropriate Access to Geospatial Data in Response to Security Concerns
- Attachment C: NC OneMap Implementation: Initial Data Layers to Serve
- Attachment D: Data Sharing Committee Business Case Summaries
- Attachment D1: American Forests and CITYGreen, "Calculating the Value of Nature"
- Attachment D2: Parcel Data and Hurricane Isabel: A Case Study

Background on the ad hoc Committee

At the August 16, 2006 meeting of the Geographic Information Coordinating Council (GICC), the chair of the Local Government Committee (LGC) presented a report describing issues related to state government agencies' requests to local governments for local government data, such as land ownership (parcel) data and orthophotography. The LGC report documented the problems; described the burden that these requests place on local governments with limited staff and resources; and summarized the major issues involved, including:

- Redistribution of Data by a State Agency—Local and state policies vary
- Data Currency
- Space Limitations
- Documentation
- Formal Agreements

The LGC initially recommended “...that the state designate a single state agency to serve as a clearinghouse for all data requests by state government agencies to local governments,” but acknowledged that “a mix of policy, process, and technology solutions will be required to solve the problem.”

The GICC established the ad hoc committee on February 7, 2007 to study the problem and develop specific recommendations. The committee, which includes local, state, and federal government representatives and others from the NC GIS community, first met on March 17. Chaired by GICC member Bill Holman, the group presented a draft set of recommendations at the GICC’s August 8 meeting. These were extensively reviewed by the Local Government Committee and other user committees. A final version incorporating that feedback was presented for adoption at the November 7, 2007 GICC meeting.

All parties involved acknowledge that there is a better way to handle data sharing. We are confident that together we can resolve this issue as the GIS community.

**Members of the 2007 ad hoc committee on Data Sharing:**

Bill Holman, Chair, Duke University Visiting Scholar  
 Anne Payne, Wake County GIS  
 Colleen Sharpe, City of Raleigh  
 Chris Koltyk, Moore County  
 John Spurrell, NC League of Municipalities  
 Rebecca Troutman, NC Association of County Commissioners  
 Jake Petrosky, Capital Area Metropolitan Planning Organization  
 Steve Strader, USGS National Geospatial Programs Office  
 Doug Newcomb, US Fish and Wildlife Service  
 John C. Farley, NC Department of Transportation  
 Allan Sandoval, NC Department of Commerce  
 Tim Johnson, Center for Geographic Information and Analysis  
 Zsolt Nagy, Center for Geographic Information and Analysis  
 Jim Dolan, NC Office of State Budget & Management  
 Richard Taylor, NC Wireless 911 Board  
 Steve Morris, NC State University Libraries

## Appendix G: NCGDAP Interim Report

The NCGDAP Interim Report, released June 1, 2008, documents the initial 3.5 years of the project and addresses technical and organizational components of the project in detail.

Available: [http://www.lib.ncsu.edu/ncgdap/documents/NCGDAP\\_InterimReport\\_June2008.pdf](http://www.lib.ncsu.edu/ncgdap/documents/NCGDAP_InterimReport_June2008.pdf)

### Interim Report Contents

#### Executive Summary

#### Project Overview

- Introduction
- Engaging Spatial Data Infrastructure (SDI)
- Original Work Plan
- Project Response to a Changing Environment.
- Project Phases

#### Content Identification and Selection

- Targeted Content
- Digital Geospatial Data Types
- Risks to Geospatial Data
- Source Organizations
- Data Inventories in Content Identification and Selection
- Present-Day NC GIS Inventory
- Frequency of Capture

#### Content Acquisition

- Data Acquisition Summary
- Data Formats
- Developing Data Archiving Formats
- Geospatial Metadata
- Data Acquisition and Transfer Technical Methods
- Legal and Intellectual Property Issues
- Rights Management
- Emergent Content Acquisition Methods

#### Partnership Building

- Project Approach to Partnership Building
- Advisory Committee
- Interagency Collaboration on Data Acquisition
- Emerging Content Exchange Networks
- Outreach and Engagement
- Outreach and Engagement to Private Industry
- Collaboration with State Archives
- Engaging Standards Organizations

## Participation in NDIIPP Network Development

### Content Retention and Transfer

- Digital Repository Software Environment
- Repository Ingest Workflow Challenges
- Current NCGDAP Ingest Workflow
- Repository Ingest Process
- Data Access
- Transfer of Data to Library of Congress

### Moving Forward

### **Appendices**

- Appendix A: Frequency of Geospatial Data Capture Survey
- Appendix B: Data Acquisition Summary
- Appendix C: NC Geologic Survey Maps
- Appendix D: Geospatial Metadata Investigation Summary
- Appendix E: Rights Coding Scheme
- Appendix F: Local/State/Federal Data Sharing Committee
- Appendix G: NCGDAP Outreach and Engagement Events
- Appendix H: Points of Engagement with Standards Efforts
- Appendix I: Storage and Backup System
- Appendix J: Repository Ingest Workflow
- Appendix K: Transfer Set Seed Files
- Appendix L: Project Extension Work Plan
- Appendix M: DSpace Qualified Dublin Core Mapping