

**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**

Interim Report

June 1, 2008

Principal Investigator: Steven P. Morris, NCSU Libraries

Co-Principal Investigator: Zsolt Nagy, NCCGIA

Project Coordinator: Jim Tuttle, NCSU Libraries





TABLE OF CONTENTS

| | |
|---|----|
| Executive Summary..... | 1 |
| Project Overview | 5 |
| Introduction..... | 5 |
| Engaging Spatial Data Infrastructure (SDI)..... | 6 |
| Original Work Plan..... | 6 |
| Project Response to a Changing Environment..... | 7 |
| Project Phases..... | 8 |
| Content Identification and Selection..... | 9 |
| Targeted Content | 9 |
| Digital Geospatial Data Types | 10 |
| Risks to Geospatial Data..... | 12 |
| Source Organizations..... | 15 |
| Data Inventories in Content Identification and Selection..... | 17 |
| Present-Day NC GIS Inventory..... | 19 |
| Frequency of Capture | 20 |
| Content Acquisition..... | 24 |
| Data Acquisition Summary..... | 24 |
| Data Formats | 26 |
| Developing Data Archiving Formats..... | 29 |
| Geospatial Metadata | 31 |
| Data Acquisition and Transfer Technical Methods..... | 36 |
| Legal and Intellectual Property Issues..... | 38 |
| Rights Management..... | 41 |

| | |
|--|-----|
| Emergent Content Acquisition Methods..... | 42 |
| Partnership Building..... | 45 |
| Project Approach to Partnership Building..... | 45 |
| Advisory Committee..... | 48 |
| Interagency Collaboration on Data Acquisition..... | 48 |
| Emerging Content Exchange Networks..... | 50 |
| Outreach and Engagement..... | 52 |
| Outreach and Engagement to Private Industry..... | 55 |
| Collaboration with State Archives..... | 57 |
| Engaging Standards Organizations..... | 59 |
| Participation in NDIIPP Network Development..... | 61 |
| Content Retention and Transfer..... | 62 |
| Digital Repository Software Environment..... | 62 |
| Repository Ingest Workflow Challenges..... | 63 |
| Current NCGDAP Ingest Workflow..... | 65 |
| Repository Ingest Process..... | 66 |
| Data Access..... | 68 |
| Transfer of Data to Library of Congress..... | 71 |
| Moving Forward..... | 72 |
| Appendix A: Frequency of Geospatial Data Capture Survey..... | 74 |
| Appendix B: Data Acquisition Summary..... | 96 |
| Appendix C: NC Geologic Survey Maps..... | 100 |
| Appendix D: Geospatial Metadata Investigation Summary..... | 101 |
| Appendix E: Rights Coding Scheme..... | 102 |
| Appendix F: Local/State/Federal Data Sharing Committee..... | 106 |

⋮

| | |
|---|-----|
| Appendix G: NCGDAP Outreach and Engagement Events | 116 |
| Appendix H: Points of Engagement with Standards Efforts | 121 |
| Appendix I: Storage and Backup System | 122 |
| Appendix J: Repository Ingest Workflow | 123 |
| Appendix K: Transfer Set Seed Files | 130 |
| Appendix L: Project Extension Work Plan | 132 |
| Appendix M: DSpace Qualified Dublin Core Mapping..... | 141 |

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 is 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 has since been extended through March 2009. 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 is 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 4 terabytes of data have been acquired to date. 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, emerging industry 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.
- A survey of current local government data archives practices 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 Multi-State Geospatial Content Transfer and Archival Demonstration and Project (started in November 2007).
- State Archives was informally engaged in the project work and will be formally engaged in the project extension and Multi-State work.
- A new Archival and Long-Term Access Committee was formed under the NC GICC, with representation from federal, state, regional, and local government agencies.

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 those in the proprietary GeoPDF format, 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.

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.

Engaging Industry

There is a degree to which one might consider the geospatial industry to be 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.
- The best outreach and engagement efforts may be those that are indirect in nature. For example, the process of working with the data community on reviewing and refining a survey on current data retention practices served to socialize the problem of digital preservation more than any other outreach mechanism employed in the course of the project.

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.

Moving Forward

The new GICC committee and the Multi-State geospatial project have been initiated under the leadership of the stakeholder community, and 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. With the more immediate and more conventional problems being addressed in the new GICC committee and the Multi-State geospatial project, there will be an opportunity for NCGDAP to focus on and take leadership in investigations and efforts that are not so directly tied to immediate business needs but which do contribute to the solution of the longer-term challenges of data preservation.

The project work has been organized by the four project phases: 1) Content Identification and Selection, 2) Content Acquisition, 3) Partnership Building, and 4) Content Retention and Transfer. This report details project findings and accomplishments during the initial three year work period according to those phases.

Project Overview

Introduction

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).¹ The goal of the North Carolina Geospatial Data Archiving Project (NCGDAP) is 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.² As a component of the National Map³, NC OneMap provides an opportunity to engage content through traditional distribution channels such as data download as well as through emerging web services based modes of access. Although this project is 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” consists of digital information that identifies the geographic location and characteristics of natural or constructed features and boundaries on the earth. Such data resources include geographic information systems (GIS) data sets, digitized maps, remote sensing data resources such as digital aerial photography, 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.⁴ 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 processes, environmental change, demographic change, land use change, and past uses of individual sites.

¹ National Digital Information Infrastructure and Preservation Program (NDIIPP): <http://www.digitalpreservation.gov/>

² The NC Geospatial Data Archiving Project (NCGDAP): <http://www.lib.ncsu.edu/ncgdap/>; and NC OneMap: <http://www.nconemap.net>

³ National Map: <http://nationalmap.gov/>

⁴ An ongoing inventory of state and local agency geospatial data is available through the NC GIS Inventory: <http://www.nconemap.net/GISInventory/tabid/288/Default.aspx>; A directory of county and city GIS data resources and services is available at <http://www.lib.ncsu.edu/gis/counties.html>

Engaging Spatial Data Infrastructure (SDI)

NCGDAP builds on earlier efforts by NCSU Libraries, beginning in the year 2000, to acquire and preserve state and local geospatial data. That effort began in response to two factors: 1) rising user demand for newly emerging local data, and 2) a growing sense of long-term risk to this newly emerging content. The challenge of scalability in terms of engaging and archiving content from well over 100 local agencies cultivated 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 geospatial data infrastructures that are evolving under the auspices of the National Spatial Data Infrastructure (NSDI), National Geospatial Programs Office (NGPO), Federal Geographic Data Committee, and Geospatial One-Stop (GOS).⁵ Spatial data infrastructure—which incorporates local, state, and federal government agencies as well as the private sector—had already been focused on such issues as data standards, best practices, data sharing agreements, metadata production and harvesting, catalog development, and services integration. However, 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.⁶ Included in the NC OneMap vision statement is the assertion that “Historic and temporal data will be maintained and available⁷.” While primarily focused on access and content standardization, NC OneMap has offered a means by which to engage a large number of local agencies in the process of creating a digital preservation infrastructure.

Original Work Plan

The original project work plan was based on a three-year performance period from October 2004 through September 2007. The project has since been extended through March 2009 [See Appendix L: Project Extension Work Plan]. Key original objectives of the project included:

- Identification of available resources through existing statewide data inventory processes
- Acquisition of at-risk geospatial data, including static data such as digital orthophotos as well time series data such as local land records and assessment data
- Development of a digital repository architecture for geospatial data, using open source software tools such as DSpace
- Enhancement of existing geospatial metadata with additional preservation metadata, using Metadata Encoding and Transmission Standard (METS) records as wrappers
- Investigation of automated identification and capture of data resources using emerging Open Geospatial Consortium specifications for client interaction with data on remote servers
- Development of a model for data archiving and time series development

⁵ National Spatial Data Infrastructure (NSDI): <http://www.fgdc.gov/nsdi/nsdi.html>; National Geospatial Programs office (NGPO): <http://www.usgs.gov/ngpo/>; Federal geographic Data Committee (FGDC): <http://www.fgdc.gov/>, Geospatial One-Stop (GOS): <http://www.geodata.gov/>

⁶ About NC OneMap: <http://www.nconemap.com/Default.aspx?tabid=289#initiative>

⁷ NC OneMap Vision and Characteristics: <http://www.nconemap.com/Portals/7/documents/visiondoc.pdf>

NCGDAP was conceived as 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, is centered not just on preservation processes and best practices but also on roles and responsibilities of the various players in what the geospatial community.

Project Response to a Changing Environment

The original project proposal was completed in November 2003, and the project work plan was finalized in December 2004. In the ensuing years it was necessary to make a number of modifications to the project work plan and to some extent modify project focus.

Circumstances which triggered changes in approach included:

- Organizational and political considerations within the data producer community (e.g., a need to re-think method and scope of data acquisition in light of other state and federal efforts)
- Changes in content domain (e.g., new content forms and new distribution methods)
- Emergent partnership opportunities (e.g., exploring niche or peripheral problem areas with key partners as opportunities presented themselves)
- A deepened understanding of the technical challenges associated with preserving geospatial data (e.g., a fuller understanding of the complexities of repository ingest workflow with geospatial data)

Key project shifts in focus from the original work plan include:

- More national, international, and private industry outreach and engagement than expected
- Less focus on the use of METS for content packaging than planned
- More focus on complex repository ingest workflow issues than expected
- Unexpected involvement with formal standards processes
- More selectiveness in data acquisition than initially planned
- Unexpected collaboration on development of data exchange infrastructure
- More focus on county data and less focus on state data than initially planned

In terms of populating the archive there was a reduction in efforts to acquire data in “high friction” situations, and an increase in efforts to increase the instances of “low friction” data acquisition through cultivation of common infrastructure in cooperation with statewide partners.

Project Phases

The project work has been organized by the four project phases which were conceived by Library of Congress as an organizational framework for developing digital preservation infrastructure. These phases include:

1. Content Identification and Selection
2. Content Acquisition
3. Partnership Building
4. Content Retention and Transfer

The following report details project findings and accomplishments during the initial three year work period according to those phases.

Content Identification and Selection

Targeted Content

This project focuses on state and local geospatial data for a number of reasons. This data was seen to be particularly at risk, given the diffusion of production and custodianship points, and given the ephemeral nature of much of the data. State and local data was seen as particularly valuable, since it is typically more detailed, current, and accurate than federal sources, and since it includes thematic content not available through other sources. Furthermore, this data is not directly addressed by archiving efforts at the federal level. A focus on the geospatial domain provides an opportunity to engage and learn from existing, relatively well-developed spatial data infrastructures, which center on states and aggregate upward to the federal level.

The following generalized checklist characterizes the process for determining which data resources fall within the domain of this project:

- Is the data geospatial in nature? (GIS data, remote sensing imagery, digital maps, georeferenced, tabular data, georeferenced ancillary data)
- Is it a product produced by or for North Carolina? (state/local content, university/NGO content, commercial content, regional federal products)
- If the data is in a digital map form, does it lend itself well to georegistration (not exceptionally distorted)? (Digital map resources with extensive coverage and for which no alternate georegistered information sources already exist would be prioritized)
- If the data is tabular in form, is it of sub-county resolution? (e.g., land records)

General principles that were expected to apply as a default framework for acquisition efforts included:

- “Time-sensitive” data—those used to create time series or at short-term risk of disappearing—would be acquired first. Targeted resources would include vector data that are routinely updated.
- Digital orthophotos, while “at risk,” are not “time-sensitive;” acquisition would be largely deferred to years two and three of the project. Since orthophotos comprise the bulk of total collection file size, this would allow time for the redundant storage model to be fully implemented and tested.
- County and state data would be targeted from the first year of the project, since there is greater knowledge about data availability for these agencies and more comprehensive benefit from initiation of time series development. Municipal, lead regional organization, not-for-profit, and university data would be targeted in later phases.
- Geographic gaps in existing agency acquisition efforts would be targeted for early acquisition, with a particular focus on Western North Carolina, which had not been thoroughly addressed in earlier acquisition efforts.

Digital Geospatial Data Types

Digital geospatial data takes a wide variety of forms, including vector data, imagery, and tabular data as well as other information types. Following is a brief description of the most common types.

Vector Data

Vector data resources model features on the earth's surface as points, lines, or polygons. For example, a well location or a school may be modeled as a point; a stream or street centerline may be modeled as a line; and a land parcel or school district may be modeled as a polygon. A vector data set may form a "data layer," such as a streets dataset covering a county. State agency vector data is typically of larger scale (more detailed) than federal or national-level data. Even more detailed vector data is available at the county and municipal government level. If available as snapshots in time, vector data will be able to provide future researchers with historical information about human and environmental processes. Unfortunately state and local agencies typically make only the current data versions available.

County and municipal data resources are in many ways analogous to the Sanborn Fire Insurance Maps published at the turn of the last century. Those maps, while created with a very narrow purpose in mind, survived by virtue of their relatively stable analog form and the intervention of interested organizations, including the Library of Congress. The new local geospatial data, while initially created for very specific administrative and operational purposes, already find uses in a wide range of applications beyond the intended uses.

Digital Orthophotography

A conventional aerial photograph contains image displacements caused by camera lens distortion, camera tip and tilt, terrain relief, and scale. The effects of camera tilt and terrain relief may be removed through a rectification process to create a digital orthophoto, which is a uniform scale photographic image—essentially a photographic map. Digital orthophotos are usually produced in TIFF, GeoTIFF, or BIL image format, with MrSID, JPEG, or JPEG 2000 compressed version created for image delivery. These images are georeferenced and can be used within GIS systems.

The mix of orthophoto resources available from federal, state, and local government agencies is beginning to make possible the analysis of change over time. The increasing frequency of local government orthophoto flights are further enabling such analyses, though the agencies in question typically focus on providing access to the most current images. County government orthophoto data is generally of much higher resolution than the state/federal data, with data resolution ranging from six inches ground surface per pixel in urban areas to two feet per pixel in some rural areas. Statewide orthophotos at a lower resolution (one meter) were created through a combined state and federal effort for the years 1993 (black and white) and 1998 (color infrared). More recently statewide efforts have been supplanted by a combination of individual county flights and regional collaborative flights. Orthophoto flights are increasingly carried out under federal cost sharing programs. In North Carolina 99 counties currently have digital orthophotos and counties typically conduct orthophoto flights roughly every two to five years. File sizes for

an individual county flight can total in the hundreds of gigabytes, and 79 counties have had multiple flights.⁸ Overall frequency of orthophoto flights is increasing as is the quantity of data generated in each subsequent flight.

Digital Maps

A wide variety of digital maps are also being acquired. These maps represent intellectual content and meaning beyond that found in the underlying datasets, as the maps result from the combination of a number of components including: data layer selection and ordering, symbolization, classification, output of data models, and annotation. Some of these digital maps are georeferenced or produced in a multi-page atlas format. The rapid increase in production of PDF maps at the county level since project start was a major surprise and introduced unexpected project challenges and opportunities in the area of PDF selection and curation.

Key Finding: 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. At the same time, complex PDF documents, including those in the proprietary GeoPDF format, present new preservation challenges of their own.

There has been a surge of interest on the part of local agencies in the issue of making their older, analog content available in digital form either for their own project development or for public consumption.⁹ Local agencies are increasingly building time into their applications, and the general public is keenly interested in seeing this historic information. To the extent that this content already exists in digital form the content will be targeted for acquisition by NCGDAP. More to the point, the local interest in historic content provides an opportunity to connect with those agencies on the issue of preserving current digitally-born content.

Key Finding: 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. Historical map imagery also helps to create an emotional connection to the issue of data preservation.

Tabular Data

Tabular data are numeric or textual data stored in database, spreadsheet, comma separated value, or other like formats. A given geographic feature may be assigned many different attribute values based on tabular data that might be associated with that feature. Some tabular data may be associated with geographic features such as land parcels or census tracts. This project focuses solely on non-federal tabular data occurring at a granularity greater than county level. This largely concerns tax assessment data, which associates land parcel records with such attributes as property value, purchase price, purchase date, building type, construction date, square footage, zoning, land use, and

⁸ Inventory of County Orthophoto Flights in NC: http://www.nconemap.com/Portals/7/documents/ortho_dates.pdf

⁹ Local agency interest in digitizing older analog content was documented in the NCGDAP Frequency of Capture Survey: http://www.nconemap.com/Portals/7/documents/NCOneMap_NDIIPLocalGovSurvey_1106.pdf

owner name. These data are produced by county tax assessment agencies and are made publicly available according to public records law, often bundled together with land parcel vector data.

Risks to Geospatial Data

While key feature data layers such as land records, street centerlines, jurisdictional boundaries, and zoning are constantly changing, current data management practice commonly involves overwriting of older versions of data which are then no longer available for historical or trends analysis. Emerging web services- or API-based technologies pose further challenges to the archive development process as it becomes easier to get and use data without creating a local copy—secondary archives often being in part a by-product of providing data access.

Even if the data has been saved, there is a chain of possible failure events that can impede permanent access to data:

- To the extent that such data is saved, it may be stored in such a way that it is not discoverable.
- If the data is discoverable, policies may not have addressed the issue of what sort of access should be provided to older versions of data.
- If the data is accessible, there is a possibility that the storage media will no longer be readable.
- If the media is readable, the data files themselves may be corrupt.
- If the files are not corrupt, it is possible that the files will be in a format that is no longer supported by current software.
- If the format is useable, it is possible that the documentation needed to use and understand the contents of the data will not exist.

Unlike vector data, digital orthophotography is not typically at risk of overwrite, yet data from older flights are known to have become less discoverable and less accessible.

While digital geospatial data inherits preservation challenges that apply to digital resources in general, this content area also presents a number of domain-specific challenges to the preservation process.

Unique Data Formats

Geospatial vector data file formats are highly complex and extremely sensitive to both format migration and software display environment, and backwards compatibility between versions of the same software can be problematic. The absence of widely adopted, open vector formats ensures that a preponderance of vector data exists in proprietary formats. Due to the complexity of the content, migration between formats can lead to unacceptable data distortion and data loss.

Data resources are often available from the data producers in multiple formats. Content may be available in “early stage” (e.g., pre-rectification orthoimagery), “middle stage” (e.g., rectified and quality-controlled orthoimagery), and “late stage” (e.g., compressed orthoimagery for delivery) formats. Data formats at one or more of the stages may be more

suitable for long-term preservation. Information may be both added and lost during the early-to-late progression. For example, sometimes only the delivery version of a parcel data set includes attached assessment data. Conversely, delivery versions of images may be in lossy compressed formats in which information is lost. Data in late stages is usually easier to acquire than data in early stages.

Spatial Database Complexity

The emergence of spatial databases has further complicated the preservation of digital geospatial data. Spatial databases may consist of multiple individual datasets or “data layers,” while also storing components such as behaviors, relationships, classification schemes, data models, or annotations that are external to or in addition to the datasets themselves. The whole of the spatial database is greater than the sum of the parts, as database components that build on the individual data layers add value. These complex databases can be difficult to manage over time due to the complexity of data models, uncertainty over long-term support of proprietary database models, and reliance on specific database back ends for data storage. Local agencies are increasingly turning to spatial databases to manage geospatial data.

Fragility of Cartographic Representation

The true counterpart to the old, preserved map is not the current GIS dataset but rather the cartographic representation that builds on that data. The representation is the result of a collection of intellectual choices and application of current methods with regard to symbolization, classification, data modeling, and annotation. Unfortunately this representation is typically stored either: a) in proprietary project file for which there is no preservation-safe alternative; b) in a complex PDF document in which the underlying data linkages have been severed; or c) as part of a web services-driven interface for which the displayed results are ephemeral. Exporting or converting these complex documents to preservable image formats captures the data view but loses the underlying data intelligence. There are semantic concerns as well, as the symbologies employed have particular meanings within particular contexts at particular points in time. While the original proposal focused more exclusively on the data itself, growing awareness of the long-term value of these representations led to a focus on this issue that was not envisioned in the original work plan. Clearly, however, any preservation of cartographic representation should occur in addition to—not instead of—preserving the underlying data.

Key Finding: 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.

Semantic Issues

Heterogeneous approaches to dataset naming, attribute naming, and attribute classification schemes create both short- and long-term barriers to understanding and use of content.

Data producers are discovering that naming and coding inconsistencies complicate the process of data sharing even in the context of present day use. While good metadata can make it possible to interpret these components, such metadata is unfortunately often absent or may not include the data dictionaries associated with names and codes found in the data. “Framework data” content standards provide some hope for improved consistency in the content and structure of geospatial data.¹⁰

Time-Versioned Content

At the local level many vector data resources are continuously or at least periodically updated. County cadastral (land parcel) datasets, for example, are typically updated on a daily or weekly basis. Such time-versioned content, if preserved, can form the basis of time series analyses such as land use change analysis. Time-versioned content presents three distinct challenges to the archiving process:

- The updated data in many cases is simply over-written or otherwise modified with no digital knowledge of the historic version maintained.
- Even if a data provider captures historic information, the absence of a standard identifier scheme, such as associated with serial publications, makes it difficult to relate data versions outside of a local data collection context.
- An optimal capture frequency is difficult to determine for any particular type of data given the significant variation in update frequencies among data producers.¹¹

Metadata Unavailability or Inconsistency

In the United States, the geospatial metadata standard since 1994 has been the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata, commonly referred to as FGDC metadata.¹² Geospatial metadata presents several distinct challenges:

- In terms of government data, while FGDC metadata is mandated at the federal level, it is less common at the state level, and only rarely available at the local level.¹³
- In cases where metadata is absent, an archive may be able to populate some FGDC record sections, yet only the data producers have the information needed to populate sections such as data quality and lineage.
- Even if metadata exists, the metadata information is often asynchronous with the data (e.g., the metadata may not have been updated to reflect format or datum change) or the metadata may simply be incorrect.

¹⁰ Framework data comprises seven themes of geospatial data (geodetic control, orthoimagery, elevation, transportation, hydrography, governmental units, and cadastral information) used by most GIS applications. These data include an encoding of the geographic extent of the features and a minimal number of attributes needed to identify and describe the features. From: FGDC Framework: <http://www.fgdc.gov/framework>. Framework data standards and practices are also addresses at the state level, such as in the case of the NC GICC Standards: <http://www.ncgicc.com/Default.aspx?tabid=141>

¹¹ NCGDAP Frequency of Capture Survey Results: http://www.nconemap.com/Portals/7/documents/NCOneMap_NDIIPLocalGovSurvey_1106.pdf (See Appendix A)

¹² FGDC Content Standard for Geospatial Metadata: <http://www.fgdc.gov/metadata/geospatial-metadata-standards>

¹³ According to the 2003 NC Local Government GIS Data Inventory only 26% of local agencies were creating FGDC-compliant metadata: <http://www.nconemap.com/Portals/7/documents/local-inventory-pack.zip>. In the NCGDAP Frequency of Capture Survey 25% of responding local agencies indicate that they were archiving FGDC-compliant metadata with their data.

- Since the original FGDC standard was a content standard for which no standard encoding was defined; existing metadata commonly requires some degree of structural normalization in order for the metadata to be interoperable with a repository.
- The FGDC standard, while extensive, does not provide container spaces for the added technical and administrative metadata elements needed for archival processes. Examples of metadata not supported by the standard include: data of data acquisition by the archive, rights of the archive vis-à-vis the data, plans for future transfers, technical information about method of transfer, and method of assuring integrity of the data.

Content Packaging

Geospatial data is characterized by complex, multi-file formats. In addition, datasets are often accompanied by metadata and ancillary documentation or data files which need to be bundled with the core dataset files. Furthermore, archival technical or administrative metadata elements not accommodated by FGDC records, such as non-producer rights information, must be bundled up with the data in some other way. Unfortunately, the geospatial industry has not adopted a standard content packaging scheme.

Ancillary files include metadata records, data dictionaries, additional data documentation, legend files, data licenses, disclaimers, and associated images. In many cases an individual ancillary file will be shared by many or all datasets in a given collection, creating the requirement that the ancillary files either be replicated for bundling with individual datasets or referenced separately through a persistent access mechanism.

Other Archiving Challenges

Other preservation challenges include:

- Securing and adequately defining archival and use rights for content
- Providing long-term support of coordinate systems and datums
- Maintaining the independence of the preserved content from any particular repository software environment

Source Organizations

The scope of NCGDAP includes state and local geospatial data in North Carolina. Data produced at the state level is typically more detailed and current than data produced at the federal level but not so detailed and current as data produced at the county and municipal level. Data producing roles at the state versus the local level do vary somewhat state to state within the U.S., with some states following a decentralized data production pattern, as found in North Carolina, and other states having data production more strongly centralized at the state level. The scope of data production at the different levels of government in North Carolina are outlined here.

State Agencies

In addition to NCCGIA, over 20 state agencies in North Carolina have active GIS programs that are involved with geospatial data production and use.¹⁴ Key agencies include the Department of Transportation (NCDOT), the Department of Environment and Natural Resources (NCDENR), the North Carolina Flood Mapping Program, the Department of Agriculture, and the Department of Health and Human Services. Both NCDOT and NCDENR include several sub-departments with their own GIS programs. Many of the geospatial data resources created by these state agencies are included in the NC OneMap viewer and data download systems managed by NCCGIA¹⁵, yet many more data resources are not available in a central state repository. Data from these agencies are produced directly by state agencies, or are produced by private contractors, with the data coming into the ownership of the state agency under public records law. Some resources are developed in collaboration with or under a cost sharing agreement with federal agencies (e.g., orthophotography and surface hydrography).¹⁶

County Agencies

Geospatial data development has typically been initiated by the county tax assessment (or land records) office and focuses on development of resources such as high-resolution digital orthophotos, street centerlines with addresses, and municipal boundaries, all of which are needed for tax assessment operations. Digital orthophotos provide the base material for creation of vector layers such as land parcels and so are typically the first data resource created. This data is usually much more detailed and more current than data available from the state or federal level. Much of the data is initially produced by private contractors for the county, which then takes ownership of the data under public records law and continues data maintenance. Base data, including orthophotography, is often taken by other county departments in order to create additional data layers for the county. As county GIS programs mature, the data comes to be employed by a wide range of agencies outside of tax operations, and more data layers representing features such as school locations, utility lines, elevation, and land use zones are created. To meet broader demand for GIS services, operations are often later moved under the IT department or into a separate GIS department in order to provide services to the entire county government.

Municipal Agencies

Many municipalities in North Carolina have developed GIS systems, which are used as part of ongoing work in areas such as land use planning, zoning, utilities management, park and open space planning, and emergency response. City governments typically acquire the relevant county data as a base resource and then develop additional, needed data layers as a result of ongoing operations.¹⁷

¹⁴ State agency data resources are documented in the NC GIS Inventory:

<http://www.nconemap.com/GISInventory/tabid/288/Default.aspx>

¹⁵ NC OneMap Map Services Catalog: <http://www.nconemap.com/Default.aspx?tabid=298>

¹⁶ Cost Share Opportunities and NC OneMap: <http://www.nconemap.com/Home/CostShare/tabid/279/Default.aspx>

¹⁷ NCSU Libraries NC City GIS agency list: <http://www.lib.ncsu.edu/gis/cities.html>; municipal data resources are listed in the NC GIS Inventory: <http://www.nconemap.com/GISInventory/tabid/288/Default.aspx>

Lead Regional Organizations

Lead Regional Organizations include Council of Governments (COG) and Municipal Planning Organizations (MPO). There are 17 COGs in North Carolina, with each county in the state having membership in one of these regional groups, which are voluntary associations of county and municipal governments. These regional associations address issues which are better addressed at the regional rather than local level, such as economic development and environmental protection. Many of the COGs provide GIS services to their constituent counties and municipalities.¹⁸ The 17 MPOs in North Carolina¹⁹ have responsibility for planning, programming and coordination of federal highway and transit investments. Regional agencies sometimes acquire data from constituent counties and municipalities in order to assemble seamless regional datasets or create new datasets that use the local data as a base.

Other Data Producing Organizations

Data is also available from university research groups. This data is often derived from other state, federal, or local data resources. Facilities data from university campuses is available in some cases. Non-governmental organization data is available from groups such as Triangle Land Conservancy and private land-holding organizations also maintain extensive GIS datasets pertaining to assets.

Data Inventories in Content Identification and Selection

One of the biggest challenges in archive development is determining what data is available in the 100 counties and many municipalities in North Carolina. Earlier experience had shown that a content identification approach based on individual phone calls to agencies does not scale and also threatens to damage the overall network by contributing to data producer contact fatigue. The information acquired in this manner is spotty, subject to transcription errors, and quickly becomes outdated. Formalized, comprehensive, routinized 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 inventory systems can produce basic metadata for documented data resources.

Key Finding: A major challenge of the archive development process lies in minimizing 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.

Historical Inventories

Since the mid-1990s, a number of efforts have been undertaken to track the spread of GIS activity among local government agencies. These inventories typically have been

¹⁸ GIS Capabilities of North Carolina Councils of Government: <http://www.tjcog.dst.nc.us/gis/giscaps.htm>

¹⁹ Image Map of Councils of Government and Lead Regional Organizations: <http://www.tjcog.dst.nc.us/nccogs.shtml>

incomplete given the complexity of the task of surveying 100 counties and 140 cities. All of these inventory results became obsolete quickly in the absence of an ongoing process for update.

In the early stages of the project NCGDAP undertook an analysis of these older surveys in order to find out what information is available. The project also sought to learn from past inventory experiences in order to provide input into future data survey efforts. Following is a list of key state and federal data inventories or surveys:

| Survey | Year | Scope | Response Rate |
|--|---------------|-------|---|
| FGDC National Geospatial Data Framework Survey ²⁰ | 1997 | US | Over 200 state, local, and federal agencies responded to the NC component of the survey. |
| NC County GIS Survey ²¹ | 1997 | NC | 94 out of 100 NC counties responded to this survey. |
| NCSU Libraries Local Government Data Acquisition Project ²² | 2000 | NC | Information was informally gathered for roughly half of NC counties. |
| NC Floodplain Mapping Program ²³ | 2000-2001 | NC | Information was gathered for all NC counties in order to support the flood mapping effort. |
| NC Dept of Transportation Surveys | 2000?-present | NC | Selected information is maintained for all counties to support NCDOT needs. |
| NC Dept of Agriculture Surveys | 2001?-present | NC | Selected information is maintained for all counties to support Dept. of Agriculture needs. |
| NC Local Government GIS Data Survey ²⁴ | 2003-2004 | NC | The initial survey gathered data for 91 counties and 92 (out of 141) municipalities. NC Flood Mapping Project survey information was used to fill gaps. |
| NC GIS Inventory ²⁵ | 2006-present | US | The NC GIS Inventory (ongoing) comprises the NC portion of RAMONA. 64 counties and 36 municipalities had responded as of Dec. 2007. |

2003 NC Local Government GIS Data Survey

Initially, the 2003 NC Local Government GIS Data Survey served as the primary content identification resource for the project, and results from the survey were incorporated into various project reports characterizing content within the project domain. The survey was extremely detailed, with 218 questions covering such topics as contact information, rights and sharing statements, and data layer availability. Specific survey information of use in the project content identification and selection process included:

- Contact information

²⁰ 1997 FGDC Framework Data Survey results available from Internet Archive:

http://web.archive.org/web/20050415063150/http://www.fgdc.gov/framework/survey_results/download.html

²¹ 1997 NC County survey results available from Internet Archive:

<http://web.archive.org/web/20061102190756/http://www.cgia.state.nc.us/cosurvey97/index.html>

²² Archived data and inventory information available from the NCSU Libraries County and City GIS Directories:

<http://www.lib.ncsu.edu/gis/counties.html> and <http://www.lib.ncsu.edu/gis/cities.html>

²³ NC Floodplain Mapping Program: <http://www.ncfloodmaps.com/>

²⁴ 2003 NC Local Government GIS Data Survey <http://www.nconemap.net/Portals/7/documents/local-inventory-pack.zip>

²⁵ NC GIS Inventory: <http://www.nconemap.net/GISInventory/tabid/288/Default.aspx>

- Data holdings information
- Information about rights issues
- Format, coordinate system, datum, etc.
- Distribution method

In order to inform future inventory efforts the project also considered possible survey improvements that would support preservation efforts. One key suggestion included providing more granular information about data formats for different data layers in order to generate data for use in measuring format utilization as a risk assessment metric.

At the project outset it was anticipated that the various earlier surveys would point out resources that did not surface in the survey. For example, county agencies will sometimes omit older orthophotos that were originally produced as hard copies (“mylar”) and later scanned and georeferenced since these resources are older and of lower quality than more current products. Orthophoto flight years recorded in the NC OneMap inventory information was compared with NCSU, NCDOT, and NCDA records as well as the '97 NC County survey. Discrepancies, particularly in the way of omission of older flights, were reported for inclusion in the NC OneMap orthophoto flight history.

Possible Future Work with Inventories

Historical surveys and inventories might be used for future research in terms of time series analysis focused on:

- Shifts in use of formats over time
- Changes in agency responsibility for geospatial data management and services
- Data availability
- Trends in access and distribution policies
- Shifts in use of commercial software packages

Key Finding: Key data community documents such as inventories, standards, and policy or best practices documents must themselves need to be archived. In more than one case inventory information that had been retired from agency websites was retrieved from the Internet Archive.

Present-Day NC GIS Inventory

The 2003 NC Local Government GIS Data Survey was completed making use of the SurveyMonkey commercial Internet service. The original NCGDAP work plan proposed development of a next-generation inventory instrument that would support an ongoing inventory process for NC One Map. However, in the period between the original proposal and the completion of the final project work plan in December 2004 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 NSGIC-led effort made a North Carolina-specific solution unnecessary.

NSGIC, of which NCGDAP co-PI Zsolt Nagy was outgoing President at the time, developed the RAMONA (Random Access Metadata tool for Online National Assessment) inventory tool with funding from the National Oceanic and Atmospheric Administration, the Department of Homeland Security, and the Federal Emergency Management Agency

(FEMA).²⁶ RAMONA's primary purpose is to track the status of GIS in US state and local government, aiding the planning and building of Spatial Data Infrastructures. RAMONA provides one consistent platform for the nation that is designed to work in concert with the federal Geospatial One-Stop (GOS) portal, a metadata clearinghouse operated by the USGS National Geospatial Programs Office. FEMA has direct access to the national database as a resource supporting emergency management operations.

In 2006 RAMONA was implemented within the state as the inventory component of the NC OneMap Program. In support of NCGDAP data inventory requirements, NCCGIA led implementation and will be continuing with administration, reporting, and analysis operations related to the RAMONA tool's implementation within North Carolina as the NC OneMap GIS Inventory.²⁷ Individual data producing agencies within the state are responsible for inputting and updating information about data holdings and data projects. Availability of the inventory information supports acquisition efforts by lowering contact and selection costs and minimizing impact on data custodians. The federal GOS portal automatically harvests the metadata building blocks generated in the NC OneMap GIS Inventory. As of November 14, 2007 a total of 206 NC agencies had registered, including 164 data producing agencies. An interim report summarizing data availability and related results was released in February 2008.²⁸

Frequency of Capture

Many vector data layers are subject to ongoing update, the frequency of which may be a reflection of the frequency with which the described features themselves change or a function of the operational processes of a particular agency. Cadastral or property data, for example, will tend to change on a fairly continuous basis in some agencies, while other agencies may handle updates in batch processes. On the other hand, 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. Such a plan would have to be both cost effective and minimize the amount of data loss between captures.

Motivation for a Survey

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 engage data producer input. In 2006 NCGDAP elected to conduct a formal survey of local agency practice, with explicit focus on frequency of capture of key framework data layers, an issue not addressed by other data survey efforts. Input from this survey will provide some idea of what capture frequency makes sense for each framework data layer from the point of view of data update cycles and local agency uses of temporal data.

²⁶ RAMONA (Random Access Metadata Tool for Online National Assessments): <http://www.gjsinventory.net/>

²⁷ NC OneMap GIS Inventory: <http://www.nc.gjsinventory.net/>

²⁸ NC OneMap GIS Inventory Report: <http://www.nconemap.com/Portals/7/documents/RAMONAdb012908.pdf>

Survey Design Process

An initial set of draft questions was developed by NCSU Libraries, NCCGIA, and State Archives, and then refined through discussions with the State Mapping Advisory Committee and the Local Government Committee (LGC). The survey process itself was vetted through the NC Geographic Information Coordinating Council, with a draft survey instrument submitted to the Local Government Advisory Team (“A Team”) for further review and testing. Following final revisions, NCCGIA configured the questions within the SurveyMonkey online tool for final review. The survey targeted the following framework data layers for detailed information gathering: parcels, street centerlines, jurisdictional boundaries, and zoning. Notably, in a subsequently released report by the LGC it was indicated that these are the four vector data layers that are most frequently requested by external data users.²⁹ A challenge for the survey was to distinguish between regular data back-up for disaster recovery purposes and retention of geospatial records for archiving purposes.

Survey elements included:

- Questions relating to key framework data layers (parcels, centerlines, jurisdictional, zoning), including frequency of capture, format, conversion process, and inclusion of attribute data
- General questions about archiving practice (when started, inclusion of metadata, storage environment, public access)
- A question about applications that require older data
- Questions about handling of superseded orthophotos and digitization of analog resources

Survey Response

The survey was sent to a list of local government GIS contacts covering all 100 counties and 25 of the largest municipalities. 58% of targeted agencies, 61 of 100 counties and 11 municipalities, responded to the survey following an initial call for response and one call to non-respondents. This was a strong response to an online survey. It had been decided in advance that only one reminder would be sent due to: 1) the issue of contact fatigue among local agencies related to surveys and data requests, and 2) potential competition with other key surveys such as RAMONA.

Survey Results

In brief, about two-thirds of local government GIS coordinators are taking time to capture geospatial datasets, at least on an annual basis. For those who capture data more often than annually, the frequency varies from weekly to semi-annually. Cadastral data are most commonly archived among the respondents (41 of the 47 who retain geospatial data). Archiving occurs to a somewhat lesser extent for street centerlines (28 of the archiving respondents), jurisdictional boundaries (28), and zoning (26). As a share of all survey respondents, geospatial records are archived for cadastral (57 percent), street centerlines (39 percent), jurisdictional boundaries (39 percent), and zoning (36 percent).

²⁹ “Requests by State Agencies for Data Produced by Local Governments, Report by the Local Government Committee of the Geographic Information Coordinating Council,” August 16, 2006:
http://www.ncgicc.com/Portals/3/documents/GICC_presentation_081606.zip

According to the survey results, there are several business rules and needs that drive retention, including: historic mapping, tax administration rules, information technology policies, records for resolution of legal issues, records retention policies, and land use change analysis. Storage formats tend to be consistent with the dominant GIS vendor among local North Carolina governments: ESRI. Storage environments vary, with servers and CDs the most common. Offsite storage (or both onsite and offsite) is used by nearly half of the respondents. In addition to vector data, 65 of the 72 respondents store digital orthophotos. The popularity of historic images has resulted in scanning and geo-referencing of hardcopy aerial photos among about one-third of the respondents. Survey results and a final survey report may be found in Appendix A.³⁰

Survey Outcomes

Survey results will be used to:

- Guide selection and acquisition processes for the NCGDAP project
- Inform future State Archives thinking about optimal practice for retention schedule processes in the future
- Inform development of a set of best practices for data management by local data producers (possibly working in concert with the local records outreach program of the State Archives)

The NCGDAP extension work plan includes a reassessment of county and municipal practice with regard to creating snapshots of geospatial data layers. Results from a follow-up survey will be compared with 2006 returns and report materials will be posted to the public website.

One notable aspect of the survey has been the manner in which the effort has socialized the problem of preservation within the state's geospatial data community. The process of reviewing and refining the survey together with various organizations as well as the actual survey process itself served to generate more awareness of the problem of digital preservation than any other outreach mechanism employed in the course of the project. The survey effort also helped to solidify the role of State Archives as an additional key partner in the NCGDAP effort.

Key Finding: The best outreach and engagement efforts may be those that are indirect in nature. The process of reviewing and refining the survey together with various organizations as well as the actual survey process itself served to socialize the problem of digital preservation more than any other outreach mechanism employed in the course of the project.

A paper on the topic of the survey was presented at the 2007 ESRI International Users Conference "Metadata and Data Management" Track. This outreach effort spawned further discussion with individuals involved with RAMONA development, exploring the idea of incorporating similar questions into future iterations of that inventory tool. As a follow up to the survey, NCCGIA will be leading efforts to gather additional information on business driver references identified in the survey results in order to get a better understanding of what the business cases are that lead individual agencies to choose to start managing

³⁰ "Frequency of Data Capture," November 16, 2007:
http://www.nconemap.com/Portals/7/documents/NCOneMap_NDIIPLocalGovSurvey_1106.pdf

temporal versions of their content. Preservation requirements for data inventories were discussed in a teleconference with RAMONA planners in December 2007.

Content Acquisition

In the course of the project the data acquisition plan was significantly retooled in order to maximize the learning experience, minimize short term loss of key data resources, and more fully engage the data producer community.

Data Acquisition Summary

Preliminary data acquisition plans were significantly modified in response to three organizational and technical factors:

- Concerns within the data community about the volume of requests that state and federal agencies as well as others were imposing on local agencies.³¹
- Development of a repository ingest workflow for geospatial data proved to be more complicated than initially anticipated, with the complex multi-file nature of the data and the inconsistency of provided metadata creating formidable technical challenges.
- A dramatic increase in volume of data becoming available for acquisition forced a decision to only go after 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”).

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 there were few or no technical or legal barriers to acquisition data could be acquired at minimal cost. In the meantime, NCGDAP would work to support a variety of partnership efforts focused on decreasing the number of “low friction” situations through the development of content exchange networks and the cultivation of open data sharing arrangements being explored in the GICC Ad Hoc Local/State/Federal Data Sharing Committee.³²

Legal or financial factors that would contribute to a situation being considered high friction include:

- Requirement that a formal agreement be signed
- Requirement of payment to cover the costs of data transfer
- Data being held by a secondary owner and in which case rights for redistribution are not clearly established

³¹ “Requests by State Agencies for Data Produced by Local Governments, Report by the Local Government Committee of the Geographic Information Coordinating Council,” August 16, 2006:

http://www.ncgicc.com/Portals/3/documents/GICC_presentations_081606.zip

³² GICC Ad Hoc Local/State/Federal Data Sharing Committee: <http://www.ncgicc.com/Default.aspx?tabid=156>

Technical factors that would contribute to a situation being considered high friction include:

- Lack of network access to the data
- Complicated data extraction interfaces that make it costly or impossible to acquire complete data coverage via network access

NCGDAP has acquired vector data for 60 counties and orthophotos for all 100 counties, with 322 total county/flight year combinations. (See Appendix B for county data acquisition summary.) Municipal data has been acquired for eight cities to date (additional municipal data is packaged with county data). State agency data acquisition has mostly focused on the 125 statewide data layers assembled and documented on the NC OneMap data download site. This data originates from several state agencies and is as assembled for single point access by NCCGIA.³³ State agency data acquisition will expand during the project period.

The NCGDAP effort has spawned a variety of individual, often unplanned partnership efforts focused on enhancing permanent access to data. These unplanned acquisition efforts were pursued in cases where one or more of the following types of opportunities came into play:

- Engagement with content that is of exceptional value or provides exceptional learning experiences
- Engagement with key partners which bring added energy and insight to the project and partnership
- Engagement that might provide a valuable learning experience

Two examples of special, targeted collections are highlighted below:

Special Collection: NC Geologic Survey Maps

While the original proposal and finalized work plan did not address the issue of georegistering acquired digital map content, an exception was made for a collection of content provided by the NC Geologic Survey (part of the NC Dept. of Environmental and Natural Resources or DENR), which has been scanning and creating metadata for geologic and historic topographic maps that are not widely available. The following factors made this an attractive addition to the NCGDAP work:

- Procedures for georegistering the existing digital map content were provided by NCGS, providing an opportunity for technology transfer to the project
- The content was of high value: the legacy geologic and topographic maps have no digital counterparts and paper copies are scarcely accessible
- The high demand content, given its broad appeal, can provide a test case for exposing archived content through the NC OneMap viewer and discovery system
- The pilot project provided insight into georegistration production costs, informing future funding efforts to support georegistration of other digital map content acquired by NCGDAP

³³ NC OneMap Data Download includes statewide data layers from several state agencies, documented with full FGDC metadata: <http://www.nconemap.net/Default.aspx?tabid=286>

While the TIFF images and world files (text files used to support georeferencing of images) became part of the NCGDAP archive, the data are being disseminated through the North Carolina Geological Survey's Geologic Map Catalog³⁴, the NCSU Libraries' campus-wide server³⁵ pointing at the NCGS' Geologic map catalog URL, and by contribution to NOneMap, the National Geologic Map Database and the USGS National Geologic Map Image Library.³⁶

Special Collection: Historical Municipal Boundary Data

In order to assess the problems associated with extracting legacy data from existing data storage systems, NCCGIA completed an effort to retrieve a set of historic municipal boundary datasets from agency archives. In addition to making available a prominent historical resource for the NC OneMap system, this effort will provide some illumination about challenges involved in retrieving data that has been archived according to standard IT archiving procedures over the course of several years.

Data Formats

In North Carolina, vector data are typically delivered in commercial data formats such as the ESRI Shapefile and Coverage formats.³⁷ These data are increasingly managed by the data producers in the ESRI Geodatabase format, a form of spatial database.³⁸ Non-proprietary exchange file formats such as the Spatial Data Transfer Standard (SDTS) have not taken root in the industry, though open standards for web services client access to data servers are gaining ground.³⁹

Image products are typically produced as uncompressed TIFF or GeoTIFF files, with earlier imagery sometimes available in formats such as BIL. Derivative MrSID, JPEG, or JPEG 2000 files are typically made for ease of distribution due to the large file size of the uncompressed imagery.

Format Preservation Factors

While the data originally acquired will be retained in the archive, as a safeguard it is beneficial to, additionally, migrate the content to one or more preservation-friendly formats. This is particularly true of data in closed, proprietary formats. For example, if MrSID data is acquired as only resort, it may be best to create a TIFF copy for long term preservation and use.

In the case of compound formats such as the ESRI Geodatabase or desktop GIS project file formats, the sum of the parts (mostly individual datasets) is not equal to the whole. Yet these complex data structures are less likely to weather long-term preservation and so, as a hedge against loss, these larger entities may also be disaggregated into individual

34: <http://wfs.enr.state.nc.us/NCGeologicMaps/>

35 <http://www.lib.ncsu.edu/gis/geolmaps.html>

36 <http://ngmdb.usgs.gov/>

37 ESRI Shapefile Technical Description: <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>

38 ESRI Geodatabase: <http://www.esri.com/software/arcgis/geodatabase/index.html>

39 OpenGIS Specifications (Standards): <http://www.opengeospatial.org/standards>

component parts as individual repository items to be redundantly archived. Unfortunately, compound format migration is a one-to-many operation and the relationships between migrated files as maintained by the compound file structure are largely lost in these extractions. Individual shapefiles and raster datasets may be extracted, but the totality of the original database can only be retained in its native binary format or in its XML export format.

The general project approach to format handling could be described as follows:

| Decision | Rationale |
|---|---|
| <i>Retain the original data in original format regardless of conversion plans</i> | Data translations will often lead to unacceptable data loss. |
| <i>Costs permitting, also translate the data into an acceptable preservation format</i> | Although the data translation may result in some loss of quality or functionality, an alternate format version provides a hedge against loss of support of the original format. |
| <i>Use the ESRI Shapefile as a preservation format for vector data</i> | The Shapefile format is widely supported, relatively simple in structure, and openly documented. |
| <i>Export Geodatabase feature classes to Shapefile in addition to doing an XML export of the entire database.</i> | Long-term support of the Geodatabase XML export is unclear (in the future a GML-based export option will be available). <i>See Appendix J for a detailed description of Geodatabase handling in the ingest workflow.</i> |
| <i>Acquire image data in it's original, uncompressed format (usually TIFF or GeoTIFF)</i> | Data compression, even at zero compression, introduces data loss and undermines data integrity. |
| <i>Use the TIFF format as a preservation format for image data</i> | The TIFF format is open and widely supported. Until 2006 there was no standard way to georeference JPEG 2000 images. As the new GMLJP2 standard for georeferencing ⁴⁰ gains market share and support, JPEG 2000 will be reconsidered as a target format. |

While developing a migration strategy for handling various data formats is seen as an important part of longer-term efforts to support data preservation, in practice little conversion is currently occurring within the current project workflow due to:

- The costs, in workload terms, of data conversions, and the implications those costs have for rate of growth of the data archive
- The complexities that conversions introduce into the repository ingest, quality control, and metadata development processes
- The questionable short- and medium-term value of the converted data files in the absence of any urgent short-term threat associated with the original data format

Quality and Functionality Factors

In considering quality and functionality factors for preservation of formats, such as has been outlined by Library of Congress for some audio, video, and text formats,⁴¹ project experience

⁴⁰ OpenGIS GML in JPEG 2000 for Geographic Image Encoding Specification:
<http://www.opengeospatial.org/standards/gmljp2>

⁴¹ Sustainability of Digital Formats, Planning for Library of Congress Collections, Format Descriptions:
<http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml>

suggests that in the case of vector geospatial data these factors may not be the same for specific data formats but in fact may be different for various framework data layers.⁴² There also are many other content types to consider, but framework data merits special attention because of the volume of data available, the demand for that data, and the national- and state-level efforts to devise and implement consistent content standards for those content types. At minimum line, point, and polygon vector data may need to be analyzed as different content types because functionality factors for things like geometric connectivity and topology affect them differently.

Format Registry Development Issues

NCGDAP has carried out discussions about format strategy with a variety of organizations, including the Library of Congress, NARA, ESRI, the Open Geospatial Consortium (OGC) Data Preservation Working Group, and the UC Santa Barbara/Stanford NDIIPP project. Some of these discussions have touched on geospatial requirements that might feed into a Global Digital Format Registry⁴³ development process. Given the large number of existing geospatial data formats⁴⁴ the position of NCGDAP has been that a methodology is needed for deciding which of the many formats should be prioritized for initial inclusion in a format registry effort. From the project's perspective, factors that would favor prioritizing a geospatial format for inclusion in a format registry effort include:

- Occurrence of that format in the target domain, as indicated in RAMONA or Geospatial One-Stop inventory data; the 1997 FGDC inventory provides some opportunity for longitudinal analysis of format adoption
- Consideration of a format as a migration target for preservation purposes (e.g., Shapefile, GeoTIFF, or JPEG2000)
- An acute sense of risk for the format, typically deriving from a small rate of adoption (e.g., the OASIS or BSB formats)
- For a national project such as NDIIPP, occurrence of the format within the United States

In connection with format registries, there is a longer-term interest in the development of ongoing market assessment tools to bring dynamic, quantitative analysis to measure--as a sustainability factor--growth or decline in use in a format. A primary source of data would be inventory data from RAMONA or Geospatial One-Stop. Commercial desktop GIS support of formats for import, export, and creation is also an important metric. If available, clearinghouse access log data might also be used to track user preferences when offered competing format options for download.

Commercial Role in Format Management

A number of commercial firms, notably including Safe Software, specialize in geospatial data translation tools and services, some of which are licensed to major GIS desktop and server software products. Extensive, detailed format descriptions are already made

⁴² FGDC Framework: <http://www.fgdc.gov/framework>; and NC GICC Standards: <http://www.ncgicc.com/Default.aspx?tabid=141>

⁴³ Global Digital Format Registry (GDFR): <http://hul.harvard.edu/gdfr/>

⁴⁴ Safe Software documents 235 geospatial data formats: <http://www.safe.com/products/fme/formats-supported/index.php>

available, for example, on the Safe Software website.⁴⁵ In considering long-term approaches to format registry sustainability it may be worth considering how to leverage the business interests of such firms in the ongoing effort to maintain intelligence about data formats. With format information already publicly provided as part of a commercial service there may be ways to leverage that information in a way that is mutually beneficial. For example, it is possible to imagine a scenario in which a format registry uses the commercially-provided information to provide an assessment of risk, leading users of the registry service to, optionally, use commercial translation tools or services when available free tools are not sufficient.⁴⁶ Migration of geospatial content from format to format comes with choices and trade-offs between information components and functionality that can be lost. While free and open alternatives for data translation exist and are growing in number, commercial expertise may also prove valuable for complex or large scale transformations.

Developing Data Archiving Formats

The absence of reliable, open vector formats is a stumbling block to preservation of vector data in particular. SDTS (Spatial Data Transfer Standard), while an open standard, has proven problematic and is not in wide use. The initial plan of the NCGDAP project involves retention of the data objects in the format received, while also exporting the content into a safer commercial vector format and buying time until a reliable, open alternative emerges. It is considered preferable to retain the content in a widely understood and supported commercial format rather than to rely solely on a migration of the content to an open format that may not be widely supported and which may involve subjecting the content to some unfortunate transformations and data loss. The ESRI Shapefile, for now, serves as something of a safe zone due to the simplicity of the format, the widespread nature of support and adoption, and the fact that the format is openly documented.⁴⁷

Prospects for an Archival Profile of the Geography Markup Language (GML)

The absence of an open and widely-supported archival format for geospatial vector data has prompted some interest in the archives community in a solution based on the Geography Markup Language (GML), which is an open specification developed by the Open Geospatial Organization (OGC). NARA, which is leading the Federal Geographic Data Committee Historic Data Working Group⁴⁸, is committed to use of open, non-proprietary formats and has been interested in using GML.⁴⁹ The challenge lies in the fact that, while GML is open, in text form, and documented by a schema, it is not so much a format as it is a means to define something like a format in the way of a specific GML application schemas that adhere to specific GML profiles.⁵⁰ As is the case with many XML-based standards, GML 3.x provides a great deal of flexibility in order to meet different application needs. GML Profiles make it possible for specific implementer communities to work with a constrained set of

⁴⁵ Safe Software documents 235 geospatial data formats: <http://www.safe.com/products/fme/formats-supported/index.php>

⁴⁶ GDAL, Geospatial Data Abstraction Library, is an open source translator library for raster geospatial data formats: <http://www.gdal.org/>

⁴⁷ ESRI Shapefile Technical Description: <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>

⁴⁸ Federal Geographic Data Committee Historic Data Working Group: <http://www.fgdc.gov/participation/working-groups-subcommittees/hdwg>

⁴⁹ NARA guidelines for transfer of geospatial data may be found in "Expanding Acceptable Transfer Requirements, Transfer Instructions for Permanent Electronic Records, Digital Geospatial Data Records: <http://www.archives.gov/records-mgmt/initiatives/digital-geospatial-data-records.html?>

⁵⁰ GML profiles are listed at: <http://www.opengeospatial.org/standards/gml>

GML, lowering the barrier to implementation.⁵¹ Permanent access to GML-based data will be complicated by the diversity of profiles and application schemas, bringing into question the likelihood of long-term tool support for any particular GML variant. The GML Simple Features Profile, finalized in 2006 as part of an effort to lower the barrier to vendor adoption of GML, might provide the basis to develop a widely-supported GML-based solution for longer-term maintenance of vector data.⁵² An archival profile of GML would be roughly analogous to what PDF/A is to PDF.⁵³ The core challenge in developing an archival profile of GML would lie in minimizing the quality and functionality tradeoffs which might have to be made in favor of sustainability.

Emerging Formats

Since 2005 the geospatial industry experienced something of a revolution in terms of emergence of new mainstream, mass market applications which reach a vast new audience. Implementers and participants in the “neogeography” (or “neogeo”) community have to a great extent come from outside the geospatial industry, and a cross-over of mainstream IT professionals into the geo space and seeking light-weight, open tools has also spurred an explosion in open source activity, culminating in the formation of the Open Source Geospatial Foundation (OSGeo) in February 2006⁵⁴. In terms of data, a focus on needs for lightweight, integrated solutions has helped to foster “good enough” approaches to data functionality and quality. Interest in rapid development of “mashups” and AJAX-based applications that draw from many sources has spurred an interest in nimble data formats which function well as highly interoperable network payloads for which there are low technical barriers to use. Notable formats include:

KML - Originally known as Keyhole Markup Language, but now just referred to as KML, this format was developed by a company called Keyhole, which was acquired by Google in 2004. KML is the chief format used in Google Earth, and support for reading and generating KML is now common in various geospatial software packages. In December 2006 KML was introduced into the OGC standards development process, with a formal KML Standards Working Group being formed in 2007 to develop KML 2.2.⁵⁵

GeoRSS – GeoRSS comprises RSS and Atom feeds that have been provided with coordinate information. There are currently two encodings of GeoRSS. GeoRSS-Simple is meant as a very lightweight format that can be easily added to existing feeds with little effort. GeoRSS GML provides a more feature-rich option and is represented by a formal GML Application Profile.⁵⁶

GeoJSON – JSON (JavaScript Object Notation) is a lightweight, text-based data-interchange format that is designed to be easy for humans to read and write and easy for machines to parse and generate, providing an alternative to XML-based representation. GeoJSON refers to efforts to encode geographic content in JSON.⁵⁷

⁵¹ Ron Lake, “GML Profiles and Applications Build the GeoWeb,” GeoWorld, September 2005

⁵² GML 3.1.1 Simple Features Profile: http://portal.opengeospatial.org/files/?artifact_id=15201

⁵³ PDF/A-1, PDF for Long-term preservation, Use of PDF 1.4: <http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml>

⁵⁴ Open Source Geospatial Foundation (OSGeo): <http://www.osgeo.org/>

⁵⁵ KML 2.2 Standards Working Group: <http://www.opengeospatial.org/projects/groups/kml2.2swg>

⁵⁶ GeoRSS Geographically Encoded Objects for RSS Feeds: <http://georss.org/>; and OGC GeoRSS White Paper (OGC Document 06-050r3): <http://www.opengeospatial.org/pt/06-050r3>

⁵⁷ GeoJSON: http://wiki.geojson.org/Main_Page

TMS – While not so much a format as reconfigured existing OGC Web Mapping Services as static image tile sets optimized for high performance use, Tile Map Services represent an emergent content form of possible interest in an archiving context.⁵⁸ (See pp. 43, “Tile Map Services and Archiving” for more information of the implications of TMS for archiving web services-based content)

While these emergent forms are typically intended to function as transient network payloads, often derived from more traditional geospatial content types, the increasingly ubiquitous nature of this content suggests a need for attention from a preservation perspective.

Geospatial Metadata

Metadata plays a central role in facilitating discovery as part of searchable or browseable indexes. Metadata also supports use of geospatial data by informing the user about data structure, content, georeferencing system used, data lineage (or processing history), rights, and recommended use. Additional ancillary documentation such as data dictionaries for attributes (e.g., land use codes for land use polygons) may also be required in order to properly use the data. The Federal Geographic Data Committee (FGDC) published the Content Standards for Digital Geospatial Metadata (CSDGM) in 1994, and federal agencies were mandated to begin using the standard in 1995.⁵⁹ The standard, which reached version 2 in 1998, has since been widely adopted at the state government level, with a lower level of adoption at the local level. 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. In the future, the current standard will be supplanted by the emerging North American Profile of the ISO 19115 metadata implementation specification for geographic information, using the ISO 19139 XML schema implementation.⁶⁰

Initial project work included an investigation into a range of issues related to geospatial metadata domain in order to determine metadata strategy (See Appendix D). Topics of investigation included determining:

- Availability of metadata for local agency data
- Whether metadata in various clearinghouses is synchronous
- Consistency in structural of metadata
- Whether substantive changes in data trigger revision of local agency metadata records
- Whether the new ISO 19115 geospatial metadata standard should be adopted in place of the current FGDC standard for project work
- Whether the ESRI Profile of the FGDC standard provides additional archival functionality
- Whether ESRI Profile elements could be successfully stripped in order to create vanilla FGDC records
- How it may be possible to link dataset versions through persistent identifiers for serial entities
- What tools are available for metadata processing
- What alternate sources for metadata are available if no formal metadata is provided

⁵⁸ Tile Map Service Specification: http://wiki.osgeo.org/index.php/Tile_Map_Service_Specification

⁵⁹ FGDC Geospatial Metadata Standards: <http://www.fgdc.gov/metadata/geospatial-metadata-standards>

⁶⁰ ISO Metadata Standard: <http://www.fgdc.gov/metadata/geospatial-metadata-standards#isometadata>

Key Geospatial Metadata Handling Decisions

Following initial investigations into metadata availability, standards, and tools the following decisions were arrived at regarding metadata handling in the short term:

- Select authoritative sources, ideally the producing agencies, for metadata in order to maximize the likelihood of metadata synchronization with the data;
- Defer adoption of ISO 19115 and ISO 19139 until such a time as widespread adoption, support, and tool development occurred within the community
- Use the ESRI Profile where appropriate in order to take advantage of added technical metadata elements and synchronization functionality (*actual implementation of the ESRI Profile has been deferred to a later time due to automation requirements*)
- Use the “mp” and “cns” (chew-and-spit) tools for raw metadata cleanup and preparation⁶¹
- Use the ArcCatalog toolset (cataloging functionality available within ESRI’s ArcGIS desktop software) for metadata creation and synchronization
- Employ “Transfer Set Seed Files” to capture administrative, technical and descriptive metadata relating to an entire data collection, allowing the metadata elements to propagate to the item level (See Appendix K: Transfer Set Seed Files)
- Use metadata templates as cores to batch create metadata from scratch

Metadata is often absent when data is acquired from local agencies. To the extent that existing metadata is received with the data, the metadata 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
- Remediation to fix major errors and to enhance the suitability of key access fields for use in catalog and discovery environments

Key Finding: 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.

In practice, the heterogeneous nature of received metadata in terms of structure and content makes this added value work very expensive. In the longer term the following approaches might be pursued.

- Remediate key access fields that enable discovery (disambiguate title, standardize agency names, and utilize ISO topical keywords)
- Use of ArcCatalog for use in batch preparation of metadata (requires additional Visual Basic for Applications (VBA) or Python development)
- Synchronize metadata elements using ArcCatalog

⁶¹ “mp” checks the structure of a metadata record against the FGDC standard, indicating discrepancies. It also re-expresses the metadata in several useful alternative formats. “cns” is used to restructure text metadata records into an indented format that can be properly parsed by “mp: <http://geology.usgs.gov/tools/metadata/>

Procedures were developed for each of these tasks though implementation has to date occurred only a demonstration basis. (See Appendix J for a detailed outline of metadata workflow)

Automated Processing of Orthophoto Metadata

A separate metadata process was created for digital orthophotography collections. Whereas vector data layers require specialized handling for each data layer, making automation difficult, typical orthophoto collections involve several hundred images to which a single metadata template can easily be applied and supplemented with auto-extracted, image-specific metadata elements. The orthophoto metadata process was designed to accommodate several scenarios:

- Data accompanied by acceptable metadata
- Data with unacceptable metadata
- Data that lacks metadata entirely

A flexible template-based process driven by a Python script allows for application of collection level and generation of image specific metadata elements.

Orthophoto indexes are generated with ArcGIS and visually inspected to identify missing or duplicate images in a collection. In cases of missing images, a set of policies describe the process for obtaining replacements from the contact from whom the collection was obtained, other data collection points, or the data producer. The indexing process generates an attribute table containing the bounding coordinates of each image in geographic coordinates. The attributes table is exported as a comma separated value file and used to apply bounding coordinates to each image when metadata records are created. The indexing process also builds individual image dataset titles by appending the file name to the collection title. Transfer size or size of the image, and bounding coordinates are also captured in order to populate the metadata record.

Orthophoto templates are created using information gathered from several sources including the website and documentation of the agency responsible for the data, metadata accompanying the data, and information gathered in previous surveys. A basic template adhering to the 1998 FGDC DTD (<http://www.fgdc.gov/metadata/fgdc-std-001-1998.dtd>) is used as the basis for building a template for all collections without metadata. For each collection, a collections-specific templates is then created for use in spawning individual FGDC metadata records for each image.

Lineage Statement for Altered Metadata

NCGDAP elected to provide a lineage statement describing the supplemental metadata generated by the project. There was some discussion over authorship of metadata and creating agency especially in the case of collections without metadata and in cases where collections were migrated to ESRI formats by the project. In lieu of claiming authorship of data or metadata a lineage statement describing processing by NCGDAP is included in the Supplemental Information FGDC field.

The following is an example of a lineage statement:

“The North Carolina Geospatial Data Archiving Project, a joint effort between the North Carolina State University Libraries and North Carolina Center for Geospatial Information and Analysis, has added supplemental information to this metadata record available from the [fill in] County website. Changes include the following: adding the image name to the title; addition of bounding coordinates; and alteration of image format information.”

The Role of Spatial Data Infrastructure in Metadata

Given the cost of processing heterogeneous metadata resources, the key to achieving efficient and cost effective handling of metadata may lie in the 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. Recently developed and emerging data inventories and networks are beginning to help meet these needs. Specific, recent operational examples of such infrastructure include:

- 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 submissions⁶²
- NC OneMap metadata templates for key framework data layers, which promote consistency both in content and structure of metadata⁶³
- 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 processes⁶⁴

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. Training as well as on-demand support are provided, and metadata templates are provided.

Content Packaging and Metadata Wrappers

While the FGDC metadata standard is quite extensive with regard to describing data content, structure, lineage, and georeferencing, it does not address some critical preservation metadata requirements, including:

- Bundling the various components that make up a complex data object, including: the multiple data files making up a data object, georeferencing files, metadata files, and ancillary documentation

⁶² NC GIS Inventory: <http://www.nconemap.com/GISInventory/tabid/288/Default.aspx>

⁶³ Metadata Templates for Local Government available from NC OneMap Metadata:
<http://www.nconemap.com/Home/Metadata/tabid/280/Default.aspx>

⁶⁴ Working Group for Roads and Transportation: <http://www.ncgicc.com/Default.aspx?tabid=162>

- Providing a container for administrative metadata that is external to the FGDC metadata, including data acquisition history, acquisition process, rights pertaining to the archive (as distinct from rights as expressed by the data producer)
- Linking to services that operate on the data, potentially providing an end user or harvester with a pointer to a web map or feature service operating on a different copy of the same data
- Providing Submission Information Package (SIP), Archival Information Package (AIP), and Dissemination Information Package (DIP) functions in a digital repository context

These requirements may be addressed through the use of content packaging approaches such as XML-based wrapper formats that bundle all of the content and metadata for a digital object into a single package. In practice Zip files commonly function as rudimentary content packages for multi-file datasets or groups of related datasets. Such Zip files typically lack data intelligence about file relationships and functions within the data bundle.

Metadata Wrapper and Content Packaging Approach in Project

The initial project proposal included plans to fold the FGDC records into Metadata Encoding and Transfer Standards (METS) records as descriptive metadata sections and possibly extract selected elements to populate administrative or technical metadata sections of the METS record as well.⁶⁵ Additional technical and administrative metadata elements would be developed in connection with the acquisition and ingest workflow. The implementation of METS is seen not as an end in itself so much as a way in which to catalyze discussion of content packaging in the geospatial community, which has not adopted a content packaging approach. Later, with the emergence of the PREMIS (Preservation Metadata Implementation Strategies) standard, it became apparent that there is an opportunity to utilize PREMIS within METS records to formalize the structure of added technical and administrative metadata elements.⁶⁶

A draft METS structure for received geospatial was formulated, incorporating FGDC elements as well as technical and administrative metadata deriving from the acquisition and ingest work flow processes. In practice, however, the heterogeneous nature of geospatial metadata currently being received makes complex metadata element processing very expensive in light of the human intervention needed to properly handle unstructured and inconsistent metadata. In a project environment involving over 100 agencies using different metadata production approaches, manual intervention substantially raises processing costs per item while also introducing new possibilities for human error.

While PREMIS offers a more consistent way in which to structure added technical and administrative metadata elements within METS, best practices for implementation of PREMIS within METS were still in the early stage of development at the time of developing workflow in the project. The use of METS and possibly PREMIS will be revisited during the later stages of the project.

In light of metadata inconsistency, the early state of METS/PREMIS practices, and the need to initiate and accelerate content acquisition, the following project approaches to content packaging were taken:

⁶⁵ Metadata Encoding and Transmission Standard: <http://www.loc.gov/standards/mets/>

⁶⁶ PREMIS , PREservation Metadata Implementation Strategies: <http://www.oclc.org/research/pmwg/>

- For the short-term, focus on using the item bundling functionality of the selected digital repository software environment (Dspace)
- Map selected technical, administrative, and descriptive metadata elements to Dspace Qualified Dublin Core metadata records as a short term expediency for repository population⁶⁷
- Monitor evolution of METS and PREMIS best practices in the preservation community
- Work with data producer partners to cultivate more consistency in metadata content and structure in order to make METS record creation more affordable
- Introduce content packaging and metadata wrapper discussions within the geospatial community through the OGC Data Preservation Working Group⁶⁸

Future Directions in Content Packaging for Geospatial Data

Content packaging and metadata wrappers have been near the forefront of issues being presented for discussion in efforts to engage the OGC standards organization on preservation issues. The issue of content packaging has emerged in Geo Rights Management discussions in the context of bundling rights expression with data objects.⁶⁹ More recently, a manner of lightweight metadata wrapper called the Metadata Exchange Format (MEF) has been implemented within the open source GeoNetwork data discovery environment developed by the Food and Agriculture Organization of the United Nations.⁷⁰ ArcCatalog support has already been developed for MEF.

Data Acquisition and Transfer Technical Methods

In response to the variety of distribution methods provided by local and state agencies, a number of different approaches were expected to be used to acquire data. The following table indicates the degree to which these methods were used:

| Transfer Method | Original 2003 Plan Compared with Actual Implementation |
|--|---|
| <i>Direct Download</i> | Plan: At time of proposal in 2003, 14 counties, 7 cities, and several state agencies in North Carolina allowed direct FTP or web download from their servers. This access method was expected to be used on a moderate basis. |
| | Implementation: As of Apr. 2008, 42 counties and 7 cities made data available online for download (some cities also make data available through the relevant county site). Direct download has provided a key method of access for “low friction” acquisition. Emergent data exchange networks such as NCStreetmap will make direct download a more widely available option. |
| <i>Data Upload to NCSU Libraries Servers</i> | Plan: It was suggested that NCSU Libraries servers will be equipped to allow incoming file transfer of data by state and local agencies. |
| | Implementation: This method has not been pursued since it puts the burden of transfer on the local agency and introduces potential security problems. |

⁶⁷ DSpace Dublin Core with Qualifiers, from DSpace Metadata: http://www.dspace.org/index.php?option=com_content&task=view&id=141
⁶⁸ OGC Data Preservation Working Group: <http://www.opengeospatial.org/projects/groups/preservwg>
⁶⁹ OGC Web Services, Phase 3: <http://www.opengeospatial.org/projects/initiatives/ows-3>
⁷⁰ Metadata Exchange Format: <http://trac.osgeo.org/geonetwork/wiki/MEF>

| | |
|---|---|
| <i>External hard drive transfer of data</i> | Plan: This method was expected to become more common as agencies include external drive support in their technical infrastructure. Onsite visits for external drive transfer from CD-ROM would be conducted on an as-needed basis. |
| | Implementation: This method is commonly used, particularly in cases where large volumes of orthophotography are transferred. A routing slip-based “sneakernet” system for distribution of county imagery among state agencies has been implemented and is administered by NCCGIA. Some local agencies require new, unused drives for data transfer, citing security concerns. |
| <i>Web extract from agency interfaces</i> | Plan: Several local agencies have developed elaborate Web interfaces to allow for download of individual data files for specific locations, while not enabling bulk download. It was suggested that “Agent applications may be constructed for automated parsing of individual web sites, although this approach is not sustainable on a wide scale due to the site-specific nature of the tools developed.” |
| | Implementation: Although this approach has been used in a few select cases, it is not a sustainable method for capturing data given the variety of web structures encountered. Complex web interfaces for data extraction are considered a technical impediment that constitute a “high friction” data acquisition situation. |
| <i>Feature server extraction</i> | Plan: As of late 2003 many local agencies maintained web map servers, some of which supported extraction of the underlying data. |
| | Implementation: In practice this backdoor method to data extraction has often not been intentionally provided by the county and could contribute to unanticipated server load on the agency site. Also, WFS, which intentionally streams vector data, has not yet been widely deployed. Some counties explicitly support feature streaming using ArcXML. ⁷¹ |
| <i>CD or DVD transfer</i> | Plan: Data would also be transferred by CD and DVD where necessary. Orthophoto holdings for individual counties may require between 10 and several hundred CDs for full data transfer. |
| | Implementation: CD or DVD is still commonly used for transfer of vector data and compressed imagery. |

Content Authentication Issues and Data Source Selection

In the original proposal vector and tabular data were expected to be acquired directly from producing agencies in order to clarify rights and ensure integrity of the data. In an effort to extend time series into the past, it was suggested that vector and associated tabular data would also be acquired--where possible--from intermediary state agencies who have in the past acquired the data, with appropriate provenance metadata to be created. Due to the size of uncompressed orthophoto collections and the cost of their transfer, these imagery collections would be acquired, where possible, from intermediaries such as NCCGIA, NCDOT, and NC Department of Agriculture rather than directly from the local agencies. Otherwise, uncompressed orthophotos would be acquired directly from producing county agencies.

In practice acquiring vector or tabular data from intermediate agencies has so far proven to be possible but impractical from the point of securing clarified rights. Aside from content authentication issues, local data producers have been uncomfortable with having data

⁷¹ An example of a county explicitly supporting ArcXML-based feature streaming is Henderson County, as documented at: <http://www.hendersoncountync.org/gis/GISdownload.htm>

redistributed by the state agencies that they have provided data to. Getting individual clearances from local agencies for such secondary distribution, where possible, would not be cost effective. Emerging content exchange systems such as the NCStreetmap are expected to provide a means by which individual data uploads will be made available to many agencies.⁷² Orthophoto exchange through secondary sources has occurred extensively, however, since the cost savings are high and since it is easier to establish the age of the data.

Legal and Intellectual Property Issues

State and local government data resources in North Carolina are subject to public records law and as such must be made available to the public “free or at minimal cost unless otherwise specifically provided by law,” with a further stipulation that “minimal cost is defined as ‘the actual cost of reproducing the public record or public information.’”⁷³ In practice, many local agencies freely redistribute data, while others charge significant sums of money for data transfer. While this data is in the public domain, there are a number of rights-related issues that can complicate preservation, as described below.

Variations in Interpretation of Public Records Law

Public Records Law varies from state to state. Even within a single state, such as North Carolina, interpretation varies widely. For example, the 2003 NC OneMap Data Survey indicated that 53.9% of local agencies charged for geospatial data, 28.9% sometimes charged, and 15.8% did not charge for data (1.3% were not sure). Furthermore, 48.7% of local NC agencies restricted redistribution, and 29.6% restricted access.⁷⁴ At statewide GIS meetings open debate about data access policy is common. Some agencies choose to go to the extent of copyrighting and charging large sums of money for their data, while others provide free download. In general there has been a trend towards more open access to data in recognition of: a) the societal benefit that derives from free data access and, b) the costs of local agency workload related to mediated or fee-based data request handling.

Concerns About Commercial Use and Resale

According to NC General Statutes Section 132-10, “Qualified exception for geographical information systems,” agencies reserve the right to restrict resale of the data and to restrict commercial use, with some exceptions for the real estate industry. In practice, freedom of access to data varies greatly. Some counties provide free download and unrestricted use and redistribution, including resale. Other counties charge considerable sums of money for commercial access to data and restrict redistribution and resale. On occasion the records law is interpreted to mean that commercial resale is not allowed by law.

⁷² GICC Working Group for Roads and Transportation: <http://www.ncgicc.com/Default.aspx?tabid=162>

⁷³ NC General Statutes Section 132-10, “Qualified exception for geographical information systems: <http://www.ah.dcr.state.nc.us/e-records/ncgs/ncgs132.htm#gs132-10>

⁷⁴ From the 2003 NC Local Government GIS Data Survey: <http://www.nconemap.net/Portals/7/documents/local-inventory-pack.zip>

Local data producer positions on the issue of data vendor requests are widely divergent, as exemplified by postings on this topic to the NC Local Government GIS listserv:

“I have found it very beneficial to not concern myself with what they may do with my data, as long as the understanding that no warranties are implied.”

“If they resell it, even as-is, that results in fewer direct requests for the data (which is free for public download from our website.)”

Or;

“I always send these [requests to carry out commercial resale] to our County attorney. He replies to them that this is a violation to the North Carolina General Statute 132-10. We do not send the data.”

Restrictions on commercial use or resale of data have implications for archive development since such restrictions implicitly restrict open secondary redistribution and introduce significant costs and barriers in terms of implicit requirements to screen users based on intended usage. In November 2006 the GICC formed and charged a new Public/Private Partnership Working Group to address the variety of issues including commercial access to and use of data.⁷⁵

Sensitive Data and Privacy

Some state agency data is restricted for privacy reasons (e.g., public health and livestock disease data). Other data resources describing endangered species or cultural heritage resources must be restricted in order to discourage poaching or theft of cultural objects. Such data is typically made available only very selectively, through a formal screening process or by virtue of established agency relationships. Including such data in an archive would introduce the requirement for stringent screening and data protection measures. In the case of land records data, some local agencies choose to mask or filter information related to particular individuals such as key personnel, victims of stalkers, witness relocation program participants, etc. There is no way for the archive to know whether records which have been filtered from the public display application have also been filtered from the underlying raw data acquired by the archive.

Post 9/11 Security Considerations

Since 9/11 some geospatial data resources have been subject to restricted access. The FGDC has published guidelines for assessment of data resources as security risks and these guidelines have been widely distributed and promoted among state and local data producers.⁷⁶ In the case of state and local map services, selected content has in some cases been altered for public display (for example, “fuzzing out” the image of a gas tank farm or nuclear power plant in a digital orthophoto on a map server). Similar processing does not typically occur in the raw data acquired for the archive.

⁷⁵ The Nov. 2, 2007 final report of the Public Private Partnership Working Group is available at: http://www.ncgicc.com/Portals/3/documents/GICC_presentations_110707_FINAL.zip

⁷⁶ FGDC Guidelines for Providing Appropriate Access to Geospatial Data in Response to Security Concerns: <http://www.fas.org/sgp/othergov/fgdc0605.pdf>

Provision of Disclaimer

Liability is a significant concern of county agencies in particular. While some counties require individuals accepting data to sign a liability waiver as part of formal data agreements, many are satisfied with providing a liability statement within the metadata or a creating a click-through disclaimer on the agencies data access website. It should be noted that older versions of content are sometimes seen as a particular source of liability when there is a perceived risk that archived content will be misconstrued as current. The desire to ensure that the user encounters the liability statement and receives the most current version of the data tends to discourage secondary redistribution and places a burden on the archive to provide such disclaimer functionality within archive metadata and access mechanisms. The disclaimer click-through scenario was a simple yet very important component of early GeoRM-related interoperability tests carried out by the OGC as part of OGC Web Services, Phase 3 effort.⁷⁷

Restrictions on Redistribution

County restrictions on redistribution of data complicate the archive development process. However, the general experience of data acquisition and partnering over the past three years has been that, as counties become more comfortable with data distribution, relevant policies become more relaxed. The growing openness towards making data freely available is reflected in the increasing number of counties providing free public download.⁷⁸ On the issue of the requirements of the public records law when a local government provides data to the state, the NC Geographic Information Coordinating Council sought and received an opinion from the Attorney General's Office. According to that opinion, the state must redistribute this data, if asked. The state can not refuse a request, but the state agency can emphasize that the requestor should contact the original source of that data.⁷⁹

Data Agreements

Individual data agreements, where required, place a significant burden on the archive in a situation where 240 separate local agencies are involved. Individual agreements are costly to execute and frequently sometimes employ unique legal language that complicates data acquisition. The project is working within the framework of the NC OneMap partnership agreements, where available, in order to avoid legal costs and acquisition barriers associated with individual agreements. In recognition of the costs and barriers implicit in formal agreements, one of the ten recommendations of the GICC Local/State/Federal Data Sharing Committee was that "written agreements that unnecessarily restrict the free exchange of geospatial data will be avoided unless absolutely necessary."⁸⁰

⁷⁷ The OWS-3 effort, including GeoDRM components, is described at: <http://www.opengeospatial.org/projects/initiatives/ows-3>

⁷⁸ Whereas in Nov. 2004 14 counties provided public download of data, by Nov. 2007 37 counties provided this service, as documented through the NCSU County GIS Directory: <http://www.lib.ncsu.edu/gis/counties.html>

⁷⁹ Minutes, Geographic Information Coordinating Council, November 7, 2007:
http://www.ncgicc.com/Portals/3/documents/GICC_Minutes_110707.pdf

⁸⁰ GICC Ad Hoc Local/State/Federal Data Sharing Committee: <http://www.ncgicc.com/Default.aspx?tabid=156>

Rights Management

Since 2004 the OGC has led efforts to define a rights management approach for the geospatial industry. The GeoRM Working Group was formed to assess industry rights management requirements, with initial efforts informed by a survey of rights management approaches in other communities (ODRL, METS DRM, JISC DRM work, etc.). While the GeoRM effort has been very much driven by web services use cases, static files scenarios have also been addressed. The GeoRM WG⁸¹ has developed the Geospatial Digital Rights Management Reference Model (GeoDRM RM), an abstract specification for the management of digital rights in the area of geospatial data and services.⁸² The GeoRM Standards Working Group is to create an OGC Standard for the rights enablement of OGC Web Services. It will deliver an OGC "GeoRM Common" Standard that extends the OGC Web Services Common Specification.⁸³

The term "DRM" has in practice carried negative connotations in terms of association with technical protection mechanisms that restrict content use and potentially undermine longer term preservation. These negative connotations have confused the rights management discussion. Perhaps with this confusion in mind, the GeoDRM Working Group was renamed as the GeoRM effort in 2007.

Rights Expression Implementation in the Project

For the short- and medium-term, NCGDAP has adopted a very light-weight locally-developed rights expression approach in order to support rights management for data acquired:

- Rights are recorded in the Workflow Management Database⁸⁴ and in the DSpace Qualified Dublin Core⁸⁵ as a metadata element containing: 1) a simple rights class value which can be script interpreted, and 2) more verbose rights text which more fully describes the rights conditions.
- Where appropriate, license files, disclaimers, or other ancillary documents are included as files associated with individual repository items.

The simple rights expression coding system supports the following functionality:

- Assigns rights for various uses by various user classes
- May be script interpreted and is designed to be extended to include additional user classes and uses
- Indicates availability of a formal agreement or disclaimer
- May be script-translated into a simple English-language rights statement

The scheme was developed to meet contractual obligations to Library of Congress in terms of maintaining a rights management database and in order to allow automated subsetting of archived data based on rights class. The simple rights expression scheme is detailed in Appendix E.

⁸¹ OGC GeoRM Working Group website: <http://www.opengeospatial.org/projects/groups/geormwg>

⁸² OGC Geospatial Digital Rights Management Reference Model: <http://www.opengeospatial.org/standards/as/geodrmrm>

⁸³ OGC GeoRM Standards Working Group: <http://www.opengeospatial.org/projects/groups/georm1.0swg>

⁸⁴ The Workflow Management Database is discussed in the Content Retention and Transfer Phase report section

⁸⁵ The DSpace Qualified Dublin Core Mapping is available in Appendix 0

Emergent Content Acquisition Methods

Over the past decade geospatial web services and API technologies have played an increasing role in data delivery, allowing users to stream data into their applications directly from the Internet without downloading data sets. Open methods include the OGC Web Map Service (WMS), Web Feature Service (WFS), and Web Coverage Service (WCS).⁸⁶ Google's KML is also being cultivated as an open standard within the OGC standards process. Proprietary methods such as ESRI's ArcIMS image and feature services also exist, and access to commercial data services is available by APIs or SOAP (Simple Object Access Protocol) connections.⁸⁷ Web services- or API-based access to geospatial data and functionality is attractive since geospatial data resources tend to be large in size and, in the case of some data resources, subject to frequent update. Furthermore, complex functions such as geocoding and routing may be provided through services, obviating the need to develop such functionality in a client. While increasing reliance on web services access may inhibit the development of secondary data archives (i.e., there may not be a need to acquire the data locally), these technologies may also provide an opportunity to automate the processes of remote data inventory and acquisition and make the process of data archive development more sustainable.

Geospatial Web Services in North Carolina

Local and state agencies provide access to web mapping services that allow users to interact with geospatial data and create maps using an ordinary web browser—without downloading the actual data. These applications satisfy the public access requirements of North Carolina public records law and allow for more efficient allocation of county staff time as office visits are reduced. In late 2003, at time of project proposal, 58 North Carolina counties, 14 cities, and several state agencies provided access to web mapping services. As of April 2008, 87 counties and 24 cities provided access to such services.⁸⁸ Interoperable OGC WMS services had only been implemented by eleven local governments at time of proposal, but active development of the NC OneMap network led to WMS deployment in 44 counties, and 25 municipalities by October 2007.

Exploring Web Services Harvesting for Archive Development

The original 2003 project work plan proposed an investigation into web services-based capture of the data. WMS, which is widely deployed, does not allow capture of the underlying data that feeds the web service. WFS does allow access to the underlying vector data, which is streamed as GML.⁸⁹ While it was anticipated that WFS would be more widely deployed during the project period, in fact the number of deployments has been limited. Reasons for low adoption of WFS by local agencies might include:

- The WFS specification is more complicated than WMS, presenting a higher technical barrier to implementation.
- The WFS specification has not been around as long as WMS (WFS 1.0 was released two years after the release of WMS 1.0 in 2000).

⁸⁶ OpenGIS Specifications (Standards): <http://www.opengeospatial.org/standards>

⁸⁷ ArcWeb Services: <http://www.esri.com/software/arcwebservices/index.html>

⁸⁸ NCSU County GIS Directory: <http://www.lib.ncsu.edu/gis/counties.html>

⁸⁹ OpenGIS Web Feature Service Implementation Specification: <http://www.opengeospatial.org/standards/wfs>

- WFS transactions can introduce more load to the local agency server (or the server of an application service provider employed by the agency).
- Not all agencies want the underlying data delivered directly to the user.

WMS has been deployed on a much wider scale. WMS is essentially a light-weight web service by which simple “Get Map” requests against a WMS server results in transmission of a simple map image based on a request for a specific data layer (or set of layers), at a specific spatial extent, represented in a specific way. Individual map clients may blend data layers from different remote services, and “cascading map servers” may integrate, through a single map interface, multiple remote services.⁹⁰ The National Map, in its initial web services-based conception, is composed of hundreds of individual WMS services originating from national, state, regional, and local agencies.⁹¹ The NC OneMap system, which is the North Carolina component of the National Map, draws from state and local WMS services as well as federal services.⁹² One catalog has documented over 1,217 WMS server links⁹³, and another catalog counts over 600,000 total WMS data layers.⁹⁴

While the primary digital preservation target is obviously the underlying data that supports the WMS, preservation of WMS representations is of interest for the following reasons:

- Many of these resources may never be captured any other way, with data sources being too numerous to handle or with transfer of very large underlying datasets not being feasible within technical constraints. WMS captures, on the other hand, could be automated, with periodic capture of content that is subject to change.
- Users are increasingly making decisions based on online map portal systems such as NC OneMap and there is a need to document the basis of decisions or to represent the dynamic information as viewed at particular points of time. The OGC Web Map Context Specification⁹⁵ has defined a means to capture service state as a sort of spatial bookmark, but this mechanism does not capture data state.
- Whereas some of the underlying vector content, and certainly the cartographic representation of such content, may be technically difficult to preserve over time, WMS captures could be preserved as simple graphic files.

Tile Map Services and Archiving

A core technical challenge faced by WMS harvesting efforts lies in defining a tiling scheme by which to capture content. Image atlas captures from WMS services would not have a user market without standard methods in which to tile images, describe the tiling system used, and request tiled images. Skepticism about the value of arbitrary image captures has deterred any project efforts to capture images in this manner.

Industry efforts in the direction of standardized tiling systems for static representations of WMS services may in the future provide some opportunity to exploit emerging tiling schemes to capture temporal snapshots of these web services. Beginning in 2005 efforts to integrate emerging AJAX applications and web mashups with existing WMS services made

⁹⁰ OpenGIS Web Map Service (WMS) Specification: <http://www.opengeospatial.org/standards/wms>

⁹¹ USGS National Map: <http://nationalmap.gov/>

⁹² NC OneMap Map Service Catalog: <http://www.nconemap.net/Default.aspx?tabid=298>

⁹³ WMS Links in the GIDB Portal: <http://columbo.nrlssc.navy.mil/ogcwms/servlet/WMSServlet?REQUEST=ServiceLinks>

⁹⁴ MapDex: <http://www.mapdex.org/search/>

⁹⁵ OpenGIS Web Map Context (WMC) Specification: <http://www.opengeospatial.org/standards/wmc>

clear the need for static, tiled derivatives of WMS services which could easily be consumed by new Web 2.0 applications which asynchronously request a large number of data objects from multiple networked sources, expecting rapid response. While applications such as Google Maps are built on internally defined tiled content, other applications such as NASA World Wind have been able to attain high rates of performance on top of existing WMS services by creating tiled representations.⁹⁶ Many different approaches to tiling have been taken. In the interest of cultivating interoperability among tiled content, the Open Source Geospatial Foundation (OSGEO) incubated development of an informal Tile Map Service specification in late 2006.⁹⁷ Reference implementations such as the TileCache software are already in place.⁹⁸ More recently, WMS Tiling proposals have been submitted as discussion papers within the OGC.⁹⁹

Should Tile Map Services become standardized, temporal snapshots of web services could potentially be created and help to solve the problem of how to capture and document web services interactions at point of time and at point of decision making. It is possible that such an approach could feed into industry efforts to develop standards around decision support systems.

It should be clear that archives of tiled images would supplement rather than substitute for archives of the underlying data sources.

⁹⁶ NASA World Wind: <http://worldwind.arc.nasa.gov/>

⁹⁷ Tile Map Service Specification: http://wiki.osgeo.org/index.php/Tile_Map_Service_Specification

⁹⁸ TileCache: <http://www.tilecache.org/>

⁹⁹ The OGC Web Map Service Revision Working Group has listed WMS tiling as an area of future standards development: <http://www.opengeospatial.org/projects/groups/wmswmg>; see also Keith Pomakis, "OpenGIS Tiled WMS Discussion Paper," OGC 07-057r2: http://portal.opengeospatial.org/files/?artifact_id=23206

Partnership Building

Project Approach to Partnership Building

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 (CGIA), and the NC OneMap Initiative. NCSU Libraries, as lead institution in NCGDAP, has also collaborated in the development of data infrastructure within the state.

North Carolina has proven to be an ideal test ground for a geospatial data preservation infrastructure initiative the following reasons:

- An active and robust coordination structure—the GICC and its committees—is in place
- Successful cooperation between state, federal, and local governments and the academic community is already a reality as formal agreements and partnerships for sharing geospatial are already in place
- A resource for accessing and sharing geospatial data – NC OneMap – is already operational; the infrastructure and institutional framework of NC OneMap provides a focal point and core mechanism for preserving digital geospatial data

The project proposal is tied closely to NC OneMap, a statewide framework of geographic information based on partnerships between municipal, county, regional, state, federal agencies, utilities and others. The project indirectly serves to complement and advance North Carolina's component of several national geospatial initiatives including the National Map, GeoSpatial One-Stop, the National Spatial Data Infrastructure (NSDI), and also the efforts of the Federal Geographic Data Committee (FGDC). Of particular importance is NC OneMap's role as North Carolina's chief expression of the National Map, a consistent framework for geographic information, providing public access to high-quality, geospatial data and information from multiple partners to help support decision-making by resource managers and the public. The National Map is the product of a consortium of federal, state, and local partners who provide geospatial data to enhance the nation's ability to access, integrate, and apply geospatial data at global, national, and local scales.

NC OneMap is closely tied to a national digital network through a series of connected initiatives devoted to access, integration, and utilization of geospatial data. State and local data provides much of the highest resolution and most current data available within such networks. This project, through its connections with those initiatives, is--by example, through the development of practices, and through existing partnerships--exploiting an opportunity to raise the profile of digital preservation and long-term access as issues to be addressed in these existing and emerging national geospatial networks.

NC Geographic Information Coordinating Council

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. The Council creates policy and resolves technical issues related to North Carolina geographic information and GIS systems and fosters cooperation among government agencies, universities, and the private sector. The GICC includes 33 members representing the GIS community statewide. Ten members are from local government, and the academic community is represented by the President of the University of North Carolina system and the President of the NC Community College System. The legislation also established six committees that support the GICC, including an active Local Government Committee.

NC Center for Geographic Information and Analysis

The NCCGIA is the primary state GIS agency and serves as staff to the GICC. In this role, NCCGIA is responsible for implementing the goals and strategies of the GICC. Established in 1977, NCCGIA also operates a GIS service program and provides GIS services—application development, data development, spatial analysis, system planning, image analysis, and general GIS technical assistance—to users in North Carolina. In the course of the project NCCGIA has been active in playing a coordinating role in a broad range of partnership and infrastructure development initiatives both within the state and in the national/federal context. NCGDAP participation in these initiatives ensures that preservation issues are addressed in much larger project contexts that involve significant government agency and industry support. 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. Making use of the existing organizational infrastructure has allowed the project access to broad, ready-made audiences for preservation outreach.

NCSU Libraries

NCSU Libraries, which is lead organization for NCGDAP, has partnered with state and local agencies on data access issues for the past 15 years. While the main focus of the library GIS data services program is on providing data access and support to NCSU faculty, staff, and students, the library has also partnered at the statewide level on such issues as map server development, data directory management, and data archiving. NCSU Libraries represents the University of North Carolina system on the State Mapping Advisory Committee and has participated in a wide range of statewide committees and working groups. The library also provides NCSU representation to the Open Geospatial Consortium (OGC) standards organization.

NC OneMap Implementation

In 2003 the GICC adopted as its priority the design and implementation of NC OneMap as a comprehensive statewide geographic data resource. NC OneMap serves the basic information requirements for decision-making in the community, statewide, and in support of national priorities. The NC OneMap website (<http://www.nconemap.net>) serves as a portal to geospatial data resources and activities in the state and serves as a point of coordination for spatial data infrastructure development and network building within the state.

NC OneMap: Key Components and Value to NCGDAP

| Component | Function | Value to NCGDAP |
|---|---|--|
| Partners | Partnership is not limited to data interchange but also addresses commitment to adherence to geospatial data standards, keeping data current and maintained, and providing data access through on-line mapping services (OGC web services). | The NC OneMap network provided the organizational context for NCGDAP. NCSU Libraries and Library of Congress are listed on the NC OneMap website as "Supporters and Collaborators" of the NC OneMap network. |
| NC GIS Inventory | The statewide NC OneMap GIS Inventory offers a catalog of GIS data produced by government agencies (and others) as well as a Who's Who in GIS. Participating agencies are provided with a metadata building block for each data set cataloged. The inventory is an ongoing process, with data producers able to update information at any time. | The inventory provides a key source of information for NCGDAP, dramatically lowering costs associated with data identification. |
| Data Download | Free public download of a wide range of state agency data is provided. Some local agency data is also now available. FGDC metadata is made available with data downloads. | The NC OneMap data download provides a convenient point of aggregation for archival data acquisition of state agency data. Data are presented in a consistent manner with full FGDC metadata. |
| Web Services Catalog | Interoperable web services published by state and local agencies form the "geospatial backbone" of the state. NC OneMap partners establish OGC-standard WMS services for use in the NC OneMap Viewer or in other environments. | The NC OneMap Web Services Catalog has provided the basis for NCGDAP testing of WMS services reliability and might provide the basis for explorations of WMS harvesting. |
| NC OneMap Viewer | Partner WMS services are accessible through the NC OneMap Viewer. Multiple years of orthoimagery are available by WMS. | NCGDAP work in the future will focus on the issue of getting more temporal data into the NC OneMap Viewer and web services access environment in order to help socialize the problem of data preservation. |
| Metadata Outreach and Support | NC OneMap provides support for metadata creation. Template records are made available for the following key framework data layers: Building, Cadastral, Municipal Boundaries, School Attendance Districts, and Street Centerlines. | NC OneMap metadata outreach efforts promote metadata availability for the archive. Template records and the NC OneMap starter block help to promote structural consistency and metadata quality. |
| Cost Sharing Arrangements | The cost of flying and producing orthoimagery (aerial photography) has been traditionally borne by local governments. As part of the NC OneMap initiative, state and federal government agencies work with local governments to leverage cost-share opportunities and | Imagery produced through the cost share program is made available to NCGDAP as part of the "orthophoto sneakernet." Since the imagery has been produced under cost share open rights are guaranteed to the |

| | | |
|--|---|--|
| | <p>offset the financial burden to each organization. The program helps to ensure that current and past imagery exists statewide to meet the business processes of local, state, and federal agencies.</p> | <p>archive. Participation in the cost sharing program requires adherence to a set of minimum best practices which includes registering data with the statewide inventory, producing metadata, and making data available to the public.</p> |
|--|---|--|

Advisory Committee

As this project is closely associated with the state geospatial data infrastructure, the issue of forming an advisory committee was brought to the NC Geographic Information Coordinating Council (GICC) for consideration at project inception in October 2004. The GICC suggested that elements of the State Mapping Advisory Committee (SMAC) and the GIS Technical Advisory Committee (TAC) serve this function, with state, local, federal, university, and commercial representation on hand. These groups meet on a quarterly basis and are well positioned to solicit community input.

Project updates are provided to the SMAC at the regular quarterly meetings, with opportunities for discussion, and presentations to the Local Government Committee (LGC) as well as the Chair of the TAC have been completed in the course of the project. The GICC and its subcommittees have been approached and activated as needed for major components of the project requiring stakeholder guidance. For example, the Frequency of Capture Survey was developed in discussions with the SMAC and the LGC and then vetted with the GICC, with the LGC Advisory Team being activated to test the draft survey instrument.

Interagency Collaboration on Data Acquisition

A major challenge for the project has been to find a sustainable and affordable way by which data for 100 counties and as many as 140 municipalities could be made available to a central archive. 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 arose from increasing local agency frustration with the volume of data requests received from state and federal agencies and other organizations such as universities.

State and federal agencies acquire local data to support their own operations and in order to support data improvement and enrichment efforts at the state and federal levels. For example:

- The U.S. Census Bureau acquires local data in order to improve the Census TIGER mapping data.
- The NC Dept. of Transportation uses local agency street data to build and enhance a statewide street network dataset.

- The North Carolina Department of Agriculture & Consumer Services, Emergency Programs Division, in support of their Multi-Hazard Threat Database, maintains a database of local government data in the in order to provide rapid emergency response and planning capability to the emergency response community.

By 2006 the issue of efficient content exchange among government agencies as well as academic and commercial consumers had become increasingly prominent in the geospatial data community. A wide range of federal and state agencies were requesting geospatial data from local (county and municipal) agencies, resulting in contact fatigue on the part of local agencies, which typically have small staffs and are not prepared to handle the volume of requests. Furthermore, while the data volumes available from individual local agencies can be quite substantial (hundreds of gigabytes per orthophoto flight), and many of the local agencies lack significant technical or network infrastructure. The problem of content exchange was highlighted at the August 2006 NC Geographic Information Coordinating Council Meeting, at which the Chair of the Local Government Committee released a brief report describing issues related to state government agencies' requests to local government for local government data.¹⁰⁰ The LGC report identified several issue areas, and 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..."

In that report, LGC directed the problem toward:

"...the lack of communication among state agencies..." yet acknowledged that "...issues are complicated and the inconsistent policies at the local government level contribute to the problem."

The LGC also acknowledged that:

"A mix of policy, process, and technology solutions will be required to solve the problem..." and suggested that "recommendations to address the issues should be consistent with the vision and characteristics of NC OneMap."

The preservation challenge came to be increasingly intertwined with the more general problem of coordinating data sharing at the local, regional, state and federal levels. The LGC report on the content exchange problem recognized the role of NCGDAP in addressing the issue, stating as one of its recommendations that the State Mapping Advisory Committee, in searching for a solution, should consider "... the activities of the Library of Congress partnership between NC State University Library and CGIA. This project, which focuses on archival and preservation of geospatial data, may provide insights that are applicable to the data sharing problem."

Local/State/Federal Data Sharing Committee

The Local/State/Regional/Federal Data Sharing ad hoc Committee was created in February 2007 to address issues brought forward by the Local Government Committee (LGC). The North Carolina Geographic Information Coordinating Council (GICC) referred the issue to

¹⁰⁰ "Requests by State Agencies for Data Produced by Local Governments, Report by the Local Government Committee of the Geographic Information Coordinating Council," August 16, 2006:
http://www.ncgicc.com/Portals/3/documents/GICC_presentation_081606.zip

the Chair of the Statewide Mapping Advisory Committee (SMAC) and an ad hoc committee was appointed to study the problem and develop specific recommendations that address the concerns of local, regional, state, and federal government agencies. NCGDAP was involved in the work of the committee through representation from NCCGIA and NCSU Libraries.

Recommendations of the Local/State/Regional/Federal Data Sharing ad hoc Committee were provided in a draft report to the GICC in August 2007, with a subsequent revision in November 2007. (See Appendix F) The recommendations focus 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

The committee identified ten data sharing recommendations for consideration by the GICC.¹⁰¹ The recommendations, when adopted in final form, will be publicized and used to encourage cooperation among all government agencies. Among the recommendations was the item:

“Data producers should evaluate and publish their long term access, retention, and archival strategies for historic data.”

In addition, seven core practices were suggested to help data producers and content providers meet the intended goals for solving the issues with statewide data sharing, including the suggestion that custodians should:

“Establish a policy and procedure for the provision of access to historic data, especially for framework data layers.”

Emerging Content Exchange Networks

Against the background of increased local, state, and federal collaboration on development of data sharing infrastructure, individual data sharing mechanisms beginning to emerge.

NCStreetmap (Street Centerline Data Distribution System)

In March 2006 the Working Group for Roads and Transportation (WGRT), operating under the State Mapping Advisory Committee (SMAC), was formed to address the specific challenge of transportation data transfer. Thirteen state agencies are represented along with federal and local agencies as well as NCSU Libraries, representing archiving concerns. In late 2007 the work of the WGRT culminated in the development of NCStreetmap, which will make street centerline data for participating counties available to federal, state, regional, and university data users. NCCGIA is completing technical development of the exchange service and is hosting the application. The benefit to NCGDAP of this effort lies in having a more efficient method to acquire centerline data. Data acquired through networks such as

¹⁰¹ GICC Ad Hoc Local/State/Federal Data Sharing Committee: <http://www.ncgicc.com/Default.aspx?tabid=156>

this is easier to identify, can be acquired in a more routine and automated fashion, is well documented, and is associated with clearly established rights information.

Key Finding: 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.

Although this initiative is focused initially only on transportation data, it is expected that other resources such as parcel data will be addressed in similar efforts.¹⁰²

Digital Orthophotography Sneakernet

The August 2006 GICC Local Government Committee (LGC) report on data exchange cited particular challenges with regard to digital orthophotography transfers. Orthophoto collections are quite large in size, making network transfer unfeasible. Data from a single county's orthophotography flight can exceed 200 GB in size and some counties have data for several years. From an archive perspective, convincing an individual agency to both agree to share the data and to commit staff time to transfer hundreds of gigabytes of data to an external storage device presents a barrier to acquisition. Key points in the report include:

- "Uncompressed aerial imagery cannot be easily distributed over a network. The time required to transfer aerial imagery for a county may be hours and transmission failure is common."
- "Currently no single state government agency has the capacity to store aerial imagery for all or even most of the counties in North Carolina. State agencies that acquire aerial imagery from counties cannot easily store and redistribute the data to other state or federal agencies."
- "Some counties, even those that provide their vector data through a download capability, simply do not fulfill requests for aerial imagery due to the burden on staff and computer resources."¹⁰³

In 2006 NCCGIA initiated administration of a routing slip approach to transfer of county aerial imagery among state organizations and NCSU Libraries. Through this exchange system, local government orthoimagery that is subject to federal cost sharing is automatically made available to a range of agencies and organizations, including NCGDAP, dramatically lowering acquisition costs and effectively removing both technical and rights barriers to data sharing.

Other Collaboration with State Agencies on Local Data Acquisition

NCGDAP has collaborated with various state agencies, including the Department of Agriculture & Consumer Services, Emergency Programs Division and the Department of Transportation, on data acquisition issues. Securing specific rights for redistribution is often an acquisition barrier in these cases.

¹⁰² Working Group for Roads and Transportation: <http://www.ncgicc.com/Default.aspx?tabid=162>

¹⁰³ "Requests by State Agencies for Data Produced by Local Governments, Report by the Local Government Committee of the Geographic Information Coordinating Council," August 16, 2006: http://www.ncgicc.com/Portals/3/documents/GICC_presentation_081606.zip

Business Continuity and Disaster Preparedness as Content Transfer Catalysts

The conjoined problems of business continuity and disaster preparedness are more compelling than data archiving, and in the aftermath of Hurricane Katrina there has been a resurgent interest on the part of local agencies regarding the issue of off-site replication of geospatial data. NCSU's own initial local government data collection efforts were a response to the impacts of Hurricane Floyd on local data access in 1999, when a perception of risk to the local data spurred initial acquisition efforts, laying the groundwork for the later initiation of NCGDAP as a larger scale project. Disaster preparedness and data improvement activities have data requirements that directly overlap with archiving and preservation, including: 1) the need for routine, cost effective mechanisms for transfer of content, 2) the need for mechanisms to manage rights related to use of the replicated content, and 3) a disposition plan for the replicated content.

Key Finding: The path to digital preservation may lead through other more compelling business problems. There is a significant overlap between the conjoined problem of business continuity and disaster preparedness and the lower priority problem of digital preservation.

Outreach and Engagement

From the outset a key component of NCGDAP was expected to be outreach and engagement to state, local, and national elements of spatial data infrastructure, with a focus on state and local sphere in North Carolina. In practice the interest in this project has been such that there has been an abundance of outreach opportunities at the national and international level. (See Appendix G for a listing of engagement efforts.)

State and Local Outreach and Engagement

State and local outreach has involved a combination of site visits, conference presentations, participation in key committees, and surveys. Highlights have included:

- Quarterly updates to the State Mapping Advisory Committee
- Presentations to the Geographic Information Coordinating Council, the Technical Advisory Committee, and the Local Government Committee
- Presentations, posters, and panel sessions at the biennial NC GIS Conferences in 2005 and 2007¹⁰⁴
- Preservation outreach through a regional series of workshops offered by the NC Property Mappers Association, part of continuing education for property mapping professionals
- A poster session in the 2006 Charlotte/Mecklenburg GIS Day Event
- Presentations to several regional GIS meetings across the state
- Hosting a presentation by Bill Lefurgy of Library of Congress at the November 2007 meeting of the GICC

¹⁰⁴ "Long-Term Preservation of Geospatial Data" (poster), NCGIS 2005 Conference. March 03, 2005 http://www.lib.ncsu.edu/ncgdap/presentations/NCGIS2005_Poster_v1.6.pdf; "Long-Term Preservation of At-Risk Digital Geospatial Data: A Cooperative Agreement with Library of Congress" (presentation), NCGIS 2005. Mar. 4, 2005 http://www.lib.ncsu.edu/ncgdap/presentations/NCGIS05_NCSULibraries.ppt

- Completion of the Frequency of Capture Survey

In local and regional site visits and meetings it became clear very early on that there was much to be learned from local agencies, many of which had already started to put into practice some manner of data archiving. These agencies already have important insights into what the business drivers and requirements might be for archive development and management. Questions asked of data producers and custodians have included:

- What are current uses cases for historic and temporal data?
- What is current practice for archiving or maintaining temporal data?
- What rights or other issues are envisioned for content transfer to the archive?
- What are preferred technical approaches to content transfer?
- What would the local agencies like to see come out of NCGDAP?

The importance of regional groups, typically centered on Council of Governments, also became clear. Some of these groups have already begun to tackle the issue of building collective infrastructure and are actively implementing content exchange frameworks that address the issues of data transfer, rights management, and promotion of content standards.

Key Finding: Regional efforts, such as those coordinated by Lead Regional Organizations, serve as building blocks for statewide infrastructure and provide diverse testbed environments for network development.

Regional GIS meetings such as those hosted by Lead Regional Organizations provide an environment of familiarity that is conducive to open discussion. Such meetings have provided opportunities to constructively engage multiple agencies as part of outreach efforts.

National Outreach and Engagement

An objective of NCGDAP is to, by example, influence development of geospatial data infrastructures in other states and at the national level by raising awareness of preservation issues. NCCGIA continues to play a key role in engaging geospatial industry leadership at the federal and all-states level, promoting the preservation effort while working in the context of broader discussions. Key activities have included:

- A March 2006 teleconference presentation to the NSDI Partnership Office representatives throughout the country
- Highlighting NDIIPP activities in an issue of the National States Geographic Information Council (NSGIC) Newsletter¹⁰⁵
- A presentation in the Spatial Data Infrastructure track of the 2006 ESRI International Users Conference (while the venue is vendor-specific, the ESRI market is such that the annual users meeting has become an important venue for SDI discussions)¹⁰⁶
- Hosting NDIIPP presentations by Library of Congress representatives at NSGIC meetings in 2005 and 2007

¹⁰⁵ Zsolt Nagy, "Preservation of At-Risk Geospatial Data, Will Your Records be Mined in 2109?", NSGIC News (Winter 2006)

¹⁰⁶ "Spatial Data Infrastructure and Data Preservation in North Carolina":
http://www.lib.ncsu.edu/ncgdap/presentations/esri06_essic.ppt

- A presentation in the partnership track of the Urban and Regional Information Systems Association (URISA) Annual Meeting in 2007
- Cooperation with NACo (National Association of Counties), which is collaborating with the Library of Congress' Map Division on a project to "enhance the library's collection of county cartography products"¹⁰⁷
- Cooperation with the Homeland Security Infrastructure Program (HSIP), which has already made use of the NC OneMap Data Download service to acquire North Carolina data for use in their data integration efforts
- Participation in the Federal Geographic Data Committee (FGDC) Historical Data Working Group¹⁰⁸

In January 2008, Zsolt Nagy (co-principal investigator on NCGDAP and principal investigator on the Multi-State initiative) and Dennis Goreham (co-principal investigator on the Multi-State Initiative) were named to the new 28 member National Geospatial Advisory Committee. The Committee will provide advice and recommendations on federal geospatial policy and management issues and provide a forum to convey views representative of partners in the geospatial community.¹⁰⁹

International Outreach and Engagement

In the course of the project NCSU Libraries become an Associate Partner in the UK-wide JISC-funded project "Geospatial Repository for Academic Deposit and Extraction (GRADE)" based out of EDINA at the University of Edinburgh.¹¹⁰ Participation in this project allowed NCGDAP to engage more fully on the issue of preserving academic content, which plays a more critical role in the UK where the national data is subject to tight usage restrictions. Participation in GRADE made it possible for NCGDAP to: a) collaborate with EDINA on engaging with the Open Geospatial Consortium on a range of issues; b) engage with the Digital Curation Centre (for which EDINA plays an administrative role); and c) engage with EDINA on investigations into suitability of mainstream repository software—such as DSpace—for geospatial content; and d) cross-fertilize with similar efforts in the UK and Europe.

Specific activities have included:

- Participation in GRADE project meetings in Edinburgh in September 2005 and October 2006¹¹¹
- Presentation to the Database Research Group of the Digital Curation Centre in May 2005¹¹²
- Presentation as part of the Digital Curation Centre Workshop on Maintaining Long-Term Access to Geospatial Data in October 2006¹¹³

¹⁰⁷ Pedro E. Flores, "NACo, Library of Congress Preserve County Cartography", County News Online 9Jan. 15, 2007): <http://www.naco.org/CountyNewsTemplate.cfm?template=/ContentManagement/ContentDisplay.cfm&ContentID=22262>

¹⁰⁸ FGDC Historical Data Working Group: <http://www.fgdc.gov/participation/working-groups-subcommittees/hwdg>

¹⁰⁹ U.S. Dept. of Interior News Release, Secretary Kempthorne Names Members of the National Geospatial Advisory Committee, Jan. 28, 2008: http://www.fgdc.gov/ngac/NGAC_Press_Release12908.pdf

¹¹⁰ Geospatial Repository for Academic Deposit and Extraction (GRADE): <http://edina.ac.uk/projects/grade/>

¹¹¹ GRADE Project Meetings: <http://edina.ac.uk/projects/grade/meetings.html>

¹¹² Long-Term Preservation of Digital Geospatial Data: Challenges for Ensuring Access and Encouraging Reuse : <http://www.lib.ncsu.edu/ncgdap/presentations/DCCGeospatialMorris.ppt>

¹¹³ Maintaining Long-Term Access to Geospatial Data:

<http://www.lib.ncsu.edu/ncgdap/presentations/DCCGeospatialMorris.ppt>

- Participation in an evaluation of the GRADE repository implementation, using Dspace
- Participation in other GRADE survey and stakeholder activities
- Joint presentation with EDINA at the November 2005 OGC Technical Committee Meeting, introducing the geospatial data preservation issue to that venue and starting the process leading to the formation of an OGC Data Preservation Working Group¹¹⁴
- Inclusion of an EDINA-led discussion of a Geospatial Application Profile for repositories at the December 2007 OGC Technical Committee Meeting

As a spin-off of participation in the GRADE project NCGDAP entered into discussions with the UK National Archives, which was also investigating geospatial data preservation issues. As an outcome of these discussions NCGDAP was invited to present at the 2007 Cambridge Conference, a highly-exclusive, invitation-only quadrennial meeting of heads of national mapping agencies. NCSU presented on preservation issues along with Carleton University (Canada), University of Portsmouth (UK), and the UK National Archives.¹¹⁵ In an indication of how important the data archiving issue has become at the national level, the data archiving workshop received the second-highest level of attendance out of five workshops offered to national mapping agency heads on that day.

Outreach and Engagement to Private Industry

Outreach and engagement with private industry has occurred to a greater extent than anticipated in response to new industry interest in temporal and historical data. It might be argued that the geospatial industry has been “temporally-impaired” from the point of view of poor availability (and management) of older data and poor software support for temporal analysis. It also might be argued that the atemporal orientation of the field has, in the past at least, been reinforced by GIS education, which often encourages students to focus on projects for which data exists, often precluding projects of a historical nature. Yet anecdotal evidence based on user demand seems to indicate an increasing expectation that older data exists in digital form, and there are broader indications that the industry as a whole has become more interested in enhancing the temporal aspect of geospatial work.

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

Software vendors

GIS software vendors have increasingly come to see maintenance and use of temporal data as an important customer problem. Software vendors have become more interested in offering analytical tools that support historical and time series analysis and are showing more interest in the data archiving problem.

¹¹⁴ Long-term preservation of digital geospatial data: challenges for ensuring access and encouraging reuse:
<http://www.lib.ncsu.edu/nsgdap/presentations/Architecture%20WG%20OGC%20Bonn%209th%20Nov%20Robertson%20Morris%20public.ppt>

¹¹⁵ Cambridge Conference presentation:
http://www.cambridgeconference.com/2007_conference_information/Conference%20proceedings/p3_3_morris.pdf

Key Finding: Software and data vendors are increasingly coming to see maintenance and use of temporal data as a customer problem that creates both commercial opportunities and vendor obligations.

In April 2005 the 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. A full day was spent with the Geodatabase team, and additional time was spent with teams or individuals in the areas of Cartography, Metadata, and Portals. Discussion topics included: archiving and preserving Geodatabase contents, managing dataset versions, preserving cartographic representation, rights management, linking data objects with services, and data bundling mechanisms.

Data vendors

NCGDAP was invited to present at the O'Reilly Where 2.0 conference in April 2006, a high profile opportunity to stress the value of temporal data as a valuable commercial resource. Vast quantities of data, especially high resolution satellite imagery, digital aerial photography, and oblique imagery are now produced by commercial vendors and remain under commercial license. In nearly all cases only the most current ("latest and greatest") data is marketed for sale. Older data is typically moved into the storage where it becomes subject to long-term risk due to the lack of provisions for discovery, access, and use. There are perhaps two mutually exclusive approaches that archives might take in engaging this commercial data: 1) convince the vendor that the older data has little market value or replacement value for current data, and acquire the data at low cost, or 2) help the vendor discover market value in the older data and exposes it for sale and use. In the latter case, the hope is that the vendor will have incentive to keep the data alive and that ongoing consumer interest in, awareness of, and demand for that data will make it more likely that it survives.

Some data vendors are beginning to see value in older data. In February 2006 the major data vendor GlobeExplorer began to actively market its older imagery, offering a web-based tool allowing for temporal comparisons for individual locations. A possible advantage of the vendor-centered approach is that it comes with a built-in economic sustainability model: market value funds the data storage and encourages permanent access. In a scenario where a library or archive acquires the data it will be necessary to obtain funding for large scale storage on an ongoing basis. It will also be difficult to replicate the marketing power and exposure that the commercial firm will bring to bear in the course of doing normal business.

Key Finding: 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.

Consulting and Contracting Firms

Consulting firms increasingly see data archiving both as a business opportunity and as a requirement to be considered when completing application development and data management projects for client agencies. Consulting firms have landed a number of data archiving projects with major federal agencies, and at industry outreach events

representatives of consulting firms have often been among the most interested in the work of the project. NCGDAP was invited to meet with research and development staff from two major firms, Lockheed Martin and Applied Research Associates, to discuss data preservation challenges and approaches. Major consulting firms have also been represented in the activities of the OGC Data Preservation Working Group.

Collaboration with State Archives

NCSU and NCCGIA have begun active collaboration with the State Archives and State Library of North Carolina, with staff from the State Archives regularly attending project meetings and participating in project discussions. At the outset it had been hoped that the project might lead to a more formal involvement of the State Archives in the state's geospatial data archiving effort, leveraging existing expertise and organizational capacity within the realm of state government. In addition to having a formal role in development of a central archive, the State Archives already conducts local records outreach¹¹⁶ and leads the definition of records retention practices across the state. NCSU Libraries has been developing a data archive as a catalyst for discussion and as a way to generate a learning experience in support of the NDIIPP effort, yet preservation of statewide data is not formally part of the NCSU Libraries' mission. A key outcome of the NCGDAP effort would be for the State Archives to become an integral part of the spatial data infrastructure of the state.

The collaboration with State Archives was formally initiated with a series of meetings involving staff from NCSU Libraries, NCCGIA, State Archives and the State Library:

- In February 2006: NCCGIA hosted a joint meeting at which NCSU presented on NCGDAP to Archives and Library representatives for discussion
- In March 2006: NCCGIA followed with a presentation on the NC OneMap Initiative
- In June 2006: the State Archives presented on their activities in archiving digital content

For this collaboration State Archives has formed a core geospatial team that includes representatives from the Electronic Records Unit, Information Technology, and Local Records. The original NCGDAP work plan included a work item focused on completing planning documents that might inform future archiving outreach efforts. In fact a better approach may be to inform existing State Archives outreach efforts, leveraging existing infrastructure.

Key Finding: 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.

Collaboration with State Archives also provides the opportunity to cross-fertilize with archiving efforts in other states. When the State Library of North Carolina hosted the first "Digital Preservation in State Government: Best Practices Exchange" in March 2006, NCGDAP was invited to offer eight separate presentations on geospatial data archiving, covering such topics as content selection, content acquisition, metadata workflow, ingest workflow, and repository architecture.

¹¹⁶ Government Records Branch of North Caroline, Workshops : <http://www.ah.dcr.state.nc.us/records/workshops.htm#local>

State Archives and the NDIIPP Multi-State Geospatial Initiative

Collaboration with State Archives, which occurred informally during the initial project period, will be formalized in the project extension period. A major outcome of NCGDAP has been the initiation of the Multi-State Geospatial Content Transfer and Archival Demonstration and Learning Project. This multi-state partnership, co-led by the North Carolina Center for Geographic Information & Analysis and the North Carolina State Archives, will demonstrate, learn, and report on a) state-to-state transfer of geospatial content using spatial data infrastructure, b) replication of content, and c) strategies to enable long term access and preservation of geospatial content. Kentucky and Utah will act as partners, with involvement of state geospatial agencies as well as State Archives from each state.

The project will seek to identify core requirements and existing capacity of partners using a 'business continuity' theme as an enabling purpose for the demonstration. Replication and movement of content between states are of interest to the Library of Congress because a robust state-to-state content exchange network can serve as a point of access to non-federal content, and therefore help meet the nation's requirements for preservation of at-risk geospatial data. The project will further serve digital preservation interests through the assessment of digital archives and collaborators. The project is of interest to State geospatial coordination offices because the practice of moving content in an organized way across jurisdictional boundaries furthers state interests in national spatial data infrastructure, which supports many business processes requiring access to geospatial content. Project activity will include structured facilitation, collaboration among the geospatial and archival community, network building and outreach to other state partners and stakeholder associations. A core committee of partners and stakeholders will guide the work and provide valuable feedback on action plans. Project reporting will support the Library of Congress' mandate to report to Congress in 2010 on the status and needs for digital information preservation.

NC GICC Archival and Long-Term Access Archival Committee

In response to increasing interest and awareness of the data archiving problem within the geospatial data producer and custodian community, the NC GICC formed a new Archival and Long-Term Access Committee in February 2008. The committee, which includes members from state agencies (5), local agencies (4), federal agencies (3), as well as one regional and one academic organization, is chaired by Anne Payne, GIS Database Administrator of Wake County Geographic Information Systems. The working group will develop a set of recommended practices for data archiving, addressing some of the following issues:

- What?: What content should be preserved?
- When?: How often should data snapshots be captured?
- Where?: Where should the archived data be stored and made accessible?
- How?: What data formats, compression formats, and media should be used? Should joined attribute data be included?
- Who?: Who should be responsible for creation and long-term storage of archived data?
- Why?: What are the business reasons for retaining and providing access to older data?

To establish a framework for the efforts of the working group, a proposed draft set of guiding principles for the formulation of recommended practices has been assembled by the working group Chair. These proposed principles include:

- Recommended practices should not place an undue additional workload on state and local GIS professionals. Retention strategies should be easy to accomplish as part of the agencies' normal workflow.
- An organized and structured approach for life cycle creation, management and sharing of geospatial content brings order and efficiencies to the retention and archival process.
- Technical approaches recommended should be designed to minimize the risk of loss of data over time.
- Archiving practices should be consistent with all other GICC-approved standards and recommendations (data sharing recommendations, security guidelines, etc.).
- Recommendations should be consistent with electronic records guidelines, policies and requirements published by the NC Archives – Archives and Records Section.
- Existing retention policies and schedules of local and state agencies should be considered in the development of recommendations.
- Existing infrastructure should be employed as much as possible (example - NC One Map Inventory).
- Recommendations should address geospatial data that are currently not digital (example– scanned imagery of old aerials and orthophotography).

Engaging Standards Organizations

One NCGDAP project objective had been to insert preservation use cases into the Open Geospatial Consortium (OGC) specification development and initiative process. 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. While most OGC activities focus on web services scenarios, inserting a temporal component into those services (interoperability across time, not just across systems) was seen as important. Furthermore, emerging geospatial web services may provide a means to develop archives in a more efficient, automated fashion.

Approaching the Open Geospatial Consortium (OGC)

Engagement with the OGC, leading to the eventual formation of a Data Preservation Working Group, occurred along two simultaneous tracks, with NCGDAP involved in both. In the first case, as an offshoot of participation in the EDINA GRADE preservation project in

the UK¹¹⁷, 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 in Bonn. At this event, a set of seven points of intersection between the digital preservation problem and existing OGC specification development activities were outlined.¹¹⁸ A second thread of discussion focused on the NARA-led FGDC Historical Data Working Group, in which NCGDAP has participated, conducted a series of discussions related to the use of GML for archiving data.¹¹⁹ Following a NARA-led preservation session at the February 2006 ESRI Federal Users, representatives of NARA, NCSU Libraries, ESRI, USGS, and others engaged in ongoing archiving discussions leading to the idea of approaching OGC with the issue of creating an archival profile of GML.

Specific points of engagement with the OGC included:

- Use of GML in archiving (possibly with an archival profile)
- Preservation use cases for Geo Rights Management (an OGC working group area)
- Content packaging and data bundling
- Use of web services in automated archive development
- Persistent identifiers and time version management
- Content transfer mechanisms and distributed storage
- Preservation of cartographic representation and snapshot mechanisms for geospatial web services (See Appendix H for a more complete listing of potential points of contact with the OGC)

Subsequent events leading to the formation of the Data Preservation Working Group included:

- October 2006 OGC Technical Committee Meeting (Tysons Corner, VA): NCSU Libraries, NARA, and the OGC Interoperability Institute (OGCII) led a discussion of archiving issues in an ad hoc Historical Data Working Group session. At this meeting it was decided that a formal working group should be formed.¹²⁰
- December 2006 OGC Technical Meeting (San Diego, CA): A draft working group charter, prepared by NCSU Libraries and NARA with the support of the OGCII was discussed in an ad hoc session and then approved in a vote of the full plenary.
- April 2007 Technical Meeting (Ottawa, ON): Steve Morris (NCSU) and Brett Abrams (NARA) were elected co-chairs of the Data Preservation Working Group. David Arctur of the OGC Interoperability Institute was subsequently added as an additional co-chair.

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 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. In particular, the working group will create and invite dialog with the

¹¹⁷ Geospatial Repository for Academic Deposit and Extraction (GRADE): <http://edina.ac.uk/projects/grade/>

¹¹⁸ Long-term preservation of digital geospatial data: challenges for ensuring access and encouraging reuse: <http://www.lib.ncsu.edu/ncgdap/presentations/Architecture%20WG%20OGC%20Bonn%2009th%20Nov%20Robertson%20Morris%20public.ppt>

¹¹⁹ FGDC Historical Data Working Group: <http://www.fgdc.gov/participation/working-groups-subcommittees/hdwg>

¹²⁰ Introduction to Archives and Geospatial Issues: <http://www.lib.ncsu.edu/ncgdap/presentations/HDWG%20agenda-200610SM.ppt>

broad spectrum of geospatial community and archival community constituents which have a stake in addressing data preservation issues.

Participation in NDIIPP Network Development

While NCGDAP is focused on the issue of bringing data preservation into the realm of spatial data infrastructure development, it has also become clear in the course of the project that network development experience of the geospatial community would provide learning experiences of value to the Library of Congress-led effort to develop a sustainable, national digital preservation infrastructure. NCGDAP has actively participated in a range of formal and informal collaborations within the NDIIPP network, and these events and venues have provided an opportunity to cross-fertilize between projects and to cultivate new collaborations. NCGDAP participation in NDIIPP-sponsored activities has included:

- Participation in biannual NDIIPP Partners Meetings (providing project updates, delivering presentations on selected topics, and taking part in breakout group discussions)
- Participation in NDIIPP principal investigators meetings to discuss issues related to development of the NDIIPP Network
- Presentations on project progress and emerging issues to the NDIIPP advisory board meetings
- Participation in NDIIPP panels in a variety of venues including three Digital Library Federation Forums, a Coalition for Networked Information (CNI) Task Force Meeting, a Joint Conference on Digital Libraries Meeting (JCDL), and an SAA Annual Meeting (Society of American Archivists)
- Completion of an NCGDAP entry point on the NDIIPP Content Portal; and
- Participation in planning for transfer of content to Library of Congress

Other NDIIPP-related activities have included:

- Participation in project meetings of the UC Santa Barbara/Stanford NDIIPP project: National Geospatial Digital Archive (NGDA)
- Sharing of ingest workflow procedures with NGDA
- Providing consultation to the MARC Standards Office on geospatial data standards issues
- Providing consultation to Library of Congress format specialists on geospatial data format issues
- Participating in a Library of Congress Geography & Maps Division survey of geospatial data services practices

Content Retention and Transfer

NCGDAP is focused less on technical architecture and archive development than it is on partnership building and on engagement with spatial data infrastructure. Nonetheless, it is necessary to actually acquire and then attempt to preserve data in order to learn what the real organizational and technical challenges are. The purpose of the demonstration repository developed for NCGDAP is: 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.

Digital Repository Software Environment

NCGDAP is initially making use of DSpace repository software.¹²¹ A major consideration in this choice is that of cost: the project is able to leverage an existing organizational investment in DSpace for other projects. Another consideration in the selection of a widely deployed, domain-neutral repository software package such as DSpace is a desire to assess the interplay between geospatial data and a mainstream digital repository software environment. Institutions which are already pursuing more general repository programs are increasingly interested in folding geospatial data into those efforts and it remains an open question whether or not domain-specific repositories are the best or only reasonable approach to handling this type of content. Also, digital repository software packages are still relatively young and individual software product communities benefit from efforts to put different types of content into each platform. As such, initial ingest into Dspace isn't intended to test use of Dspace in particular for geospatial content, but rather to test the more generic problem of putting geospatial content into a repository.

One project goal is for the preservation package (the combination of data files and various metadata components) to be completely independent of the repository environment. An initial mapping of content and metadata to DSpace ingest objects and associated Qualified Dublin Core¹²² metadata is seen as just the first spoke in what is expected to be a multi-repository process in the longer term. The core set of metadata will be maintained in a structure that is independent of the separate repository environments.

Factors guiding selection of Dspace included:

- Lower implementation costs deriving from the ability to piggyback on technical infrastructure being developed for other NCSU digital repository projects
- Compatibility with the NCSU Libraries technical environment: Java/JSP, Oracle/PostgreSQL, Solaris, Handle
- Membership in a vibrant open source development community, with accompanying network effect benefits
- Alignment with other NDIIPP projects using Dspace
- Object storage in file system, enabling retrieval in case of system corruption

¹²¹ DSpace: <http://www.dspace.org/>

¹²² Qualified Dublin Core: http://en.wikipedia.org/wiki/Dublin_Core#Qualified_Dublin_Core

Although Dspace had not been widely used for geospatial data management, there are an increasing number of projects using the repository software for that purpose. Recently some groups have begun to build access layers on top of Dspace, for example integrating Dspace with the open source OpenLayers software package.¹²³

Software Configuration and Deployment

Archived data is managed by the DSpace 1.4.2 repository software. The NCGDAP DSpace instance and Postgres database are run from a Solaris 10 virtual zone. The Solaris server is maintained by the Libraries Information Technology department and are subject to the stringent backup policies of the Libraries' Information Technology department.

Storage System Deployment

Primary data storage is provided by two Nexsan ATABeast storage arrays. Each ATABeast is partitioned into 1 TB slices and are made available with several networking protocols including NFS, Samba, and SSH. The current backup schedule calls for monthly full backups and nightly incremental backups. Backup images are retained for 3 months. Off-site backups are maintained at Poe Hall on the Campus of NC State. (See details of the storage and backup system in Appendix I).

Repository Ingest Workflow Challenges

Geospatial data is characterized by a number of complicating factors that require careful redress within the ingest workflow. (See Appendix J for a detailed outline of the NCGDAP workflow)

Format Recognition and Complex Object Bundling

Complex, multi-file objects are the norm for many types of geospatial data. The individual dataset may be composed of files of different mime-types, and formats can often have optional file components. As a result the ingest process must include an item grouping process that combines automated and manual approaches. Further complicating the ingest process is the existence of ancillary information such as license files, documentation, data models, scripts, and data dictionaries. Associating these entities with data items for repository ingest can be a challenging process that requires human intervention. Individual ancillary data files may apply to multiple separate datasets, requiring that the ancillary files be replicated across items prior to ingest.

Object Conversions

Much data is received in formats that are not archive-friendly by virtue of their complexity or impending obsolescence. Therefore, formats requiring immediate conversion in response to an immediate sense of risk must be identified early in the ingest process. These formats include the spatial database, topological vector data, and certain raster formats. Datasets of this sort will be archived in their native format as well as in a more archive friendly format.

¹²³ John Preston post to the Dspace-tech listserv on February 22, 2008: <http://www.mail-archive.com/dspace-tech@lists.sourceforge.net/msg03544.html>

Spatial databases provide a good example of the challenges faced in ingest workflow. A spatial database may be archived as is, but long term access to such content is not reliable given the complexity of the systems and the closed, proprietary nature of commercial implementations. In some cases XML export from these databases presents a more reliable preservation approach, but long-term software support of the XML files is open to question. Individual datasets may be extracted for retention in more stable forms, but database elements which span across datasets are lost in the process. In the defined NCGDAP workflow a mix of approaches is taken, with binary databases, XML exports, and dataset extracts all being retained or cultivated. Extracted datasets become items unto themselves and receive ingest and metadata preparation attention at a level of detail that is higher than that given to the parent database.

Key Ingest Workflow Choices

In developing an ingest workflow some key decisions needed to be made with regard to degree of ingest automation, degree of dependence on the initial repository software environment, and degree of trust in received metadata.

Balancing automation with human intervention: NCGDAP has no control over the manner in which received data is organized, and designing automated ingest processes for heterogeneous and irregular content is very difficult. Hands-on data processing remains important for some parts of the ingest process, raising ingest costs and increasing the possibility of introducing human error. However, as experience is accumulated it may become possible to automate additional portions of the ingest process. In the longer term, partnership efforts focused on developing routine data transfer processes are hoped to increase the ability of the archive to automate ingest and lower cost of ingestion. In the current situation, the costs of hands-on intervention have deterred implementation of some item management procedures.

Repository-agnostic architecture: In order to reduce long-term reliance on a particular digital repository software platform the project has attempted to abstract repository ingest processes, developing more generalized ingest workflows in hopes of making the data more adaptable to other repository environments. This adaptability will be necessary to support repository software upgrades, migration to other repository software, and inter-archive exchange. Ultimately, however, it can be difficult to completely prevent the repository organizational model from imprinting on the data.

Metadata remediation: Supporting long-term access to the data requires both retention of original metadata--as an archival object--and remediation of that metadata to increase the likelihood of successful use of that data. Remediation steps include error correction, functional improvements to key access fields, structural normalization, and synchronization of the metadata record with the dataset at hand. Some of the defined metadata processes that apply to pre-existing metadata have not yet been fully implemented due to the costs of hands-on intervention.

Current NCGDAP Ingest Workflow

Data Receipt Processes

The repository ingest workflow involves a multi-stage process that starts with receipt of heterogeneous data collections from a variety of data producers and ends with content that has been distilled into discrete data items that can be handled by digital repository software. It should be acknowledged that this is only one approach to data preservation, with *in situ* management of complex data within native software environments being another approach. For example, the Maine GeoArchives project has conducted investigations into issues related to maintaining spatial database content over time.¹²⁴

Data Acquisition

Geospatial data is received by the project either as a data download or as a set of files delivered on optical or magnetic media. To ensure file integrity MD5 checksums are generated and recorded as early in the process as possible. Data transferred to local systems from media are checksum validated as part of the copy procedure.

An initial metadata “Transfer Set Seed File”, including collection information, is generated on transfer of all data. Collection-level information such as acquisition date, acquisition circumstances, and transaction-specific rights information are captured in the seed file, which then informs later metadata processes as individual technical and administrative metadata elements propagate to the item level. The seed file metadata is entered into a form that encapsulates it in well-formed XML (See Appendix K for more information about the Seed File). Using a simple, internally-developed rights expression scheme, rights are registered in both human-readable and machine-actionable forms to support the possibility of fine-grained access control in the future. (See Appendix E for more information about the rights encoding process.)

Dataset Reorganization

Given the lack of control over of the manner in which data is delivered to the archive, some degree of remediation of data organization is required prior to ingest. Data files must also be extracted from archive file formats such as tar, zip, and the ESRI e00 format. Extraction of archive formats requires some manual intervention due to the unpredictable nature and hierarchy of the contents. In addition, particularly complicated formats such as Geodatabases are disaggregated into constituent elements in order to form discrete repository items.

Dataset Validation and Threat Analysis

All data are scanned for viruses and executability as identified by magic numbers using the UNIX ‘file’ utility, and infected and executable files will be deleted and reported to the contributing agency. Harvard’s JHOVE utility is used to validate formats that it recognizes.¹²⁵ Support for geospatial formats is currently limited, but validation of the ESRI

¹²⁴ Maine GeoArchives : <http://www.maine.gov/sos/arc/GeoArchives/geoarch.html>

¹²⁵ JHOVE, JSTOR/Harvard Object Validation Environment : <http://hul.harvard.edu/jhove/>

Shapefile format is a planned feature of JHOVE2.¹²⁶ In the case of digital orthophoto collections, image extent grids are developed in order to identify any gaps in transfer sets. Gaps in county orthophoto collections have occurred in the following situations:

- The county has chosen not to collect imagery for a portion of the county because there is no tax assessment activity in that area (e.g., military bases and wild life reserves)
- A portion of the imagery collection has been omitted from the transfer set
- A portion of the imagery collection has been incorrectly georeferenced

Following initial remediation, data sets are cataloged in a machine-readable manifest used to track changes and validate integrity. The path, size and checksum from the set manifest are compared to the data set after each processing stage to flag unintended changes to the data set. Intended changes, such as format migrations, are recorded in the provenance metadata for each file. The manifest accompanies the dataset through the ingest process and is used to ensure validity throughout the pre-ingest workflow as well as post-ingest.

Repository Ingest Process

The processed geospatial metadata together with the seed file metadata and the technical metadata generated during pre-ingest processing all form a superset of metadata which can be used to inform creation of Submission Information Packages for various systems as well as to populate a metadata database used to separately track project content.

The Metadata “Hub”

In order to facilitate transformations to alternate archive Submission Information Packages and in order to flexibly acquire content through different pre-ingest processes; the project has adapted the “hub-and-spoke” model employed by the University of Illinois ECHO DEpository NDIIPP project.¹²⁷ The premise behind this metadata transformation model is that each transformation is a spoke connected to a central hub in which metadata exists in a special profile from which it may be transformed into other schemas. An independent metadata database stores metadata independent of the DSpace archive and may be used to interoperate with other discovery systems, including the local data discovery environment. The external metadata database also provides a simple means to generate reports concerning the contents of the archive. The current NCGDAP process is composed of a hub and single spoke, yet additional ingest spokes are envisioned for the future.

Key Finding: 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.

¹²⁶ JHOVE2: A Next-Generation Architecture for Format-Aware Digital Object Preservation Processing: <http://hul.harvard.edu/jhove/JHOVE2-proposal.doc>

¹²⁷ ECHO DEpository Project: <http://www.ndiipp.uiuc.edu/>

Workflow Management Database

The Workflow Management Database (WMD) was designed both as an expression of the repository agnostic approach of the project and as an intermediary through which data access might be provided. The elements in the WMD were selected from metadata elements available from among the following sources:

- The FGDC metadata record, which (ideally) arrives with the data
- The transfer set seed file, which is generated at the outset of the ingest process
- Ingest workflow file manifest, which is generated during the ingest process
- DSpace Qualified Dublin Core, which is formulated for DSpace ingest

Elements were selected based on their capacity to support one or more of the following functions:

- Integration of NCGDAP-acquired data into the existing GIS data retrieval system at NCSU Libraries. The combination of keywords, titles, formats, and unique identifiers highly simplify integration into the library data discovery system.
- Generation of simple resource for generating reports on data processed by the project. Statistics concerning agencies, themes, formats, and rights may be easily collected from the WMD.
- Reconstruction of datasets independently from DSpace and its metadata database. The elements in the WMD may be used to rebuild the structure and filenames of datasets in the event of a DSpace failure.

The WMD is a simple PostgreSQL database with one table containing the following fields:

| Element Name | Description | Sample Value |
|-----------------|---|------------------------|
| Id | Database assigned identifier. Auto-incrementing integer. Provides a unique identifier for accessing records. | 1 |
| community_name | The name of the DSpace community to which a record belongs. Defines creating agency of dataset. | Wake County |
| collection_name | The name of the DSpace collection to which a record belongs. Defines collection dataset belongs to. Collections belong to communities. | Orthophotos |
| collection_id | Integer identifier of collection to which record belongs. DSpace internal identifier necessary to script actions involving collections. | 2 |
| Formats | Comma delimited list of file extensions in dataset. Provides detailed information about file formats in dataset. | tif, tif.xml, tfw, aux |
| Rights | Unix-style, machine actionable rights code. Provides a forward compatible, fine-grained access control structure. ¹²⁸ | M00N111P111L111 |

¹²⁸ The NCGDAP rights coding scheme is described in Appendix G

| | | |
|---------------|--|---|
| Handle | CNRI Handle persistent identifier. Persistent identifier used to resolve item location for http access. ¹²⁹ | 1840.2/105 |
| Noid | Nice Opaque Identifier Persistent identifier used as foreign key between WMD and DSpace databases. ¹³⁰ | Ncgdap8415 |
| Title | Title of dataset Provides brief description of dataset to support access. | Wake County, NC 2005 Orthophotos Image File: 0607_005.tif |
| Keywords | ISO 19115 Topic Category keywords. Provide controlled vocabulary to facilitate access, thesaurus mapping. | imageryBaseMapsEarthCover |
| transfer_info | Unstructured text field describing conditions under which the data is received. | please return USB drive to CGIA by Friday. |
| media_label | Unstructured text field to record information on media label. | Wake County, CD 1 of 6, Nov 2005 |
| Notes | Unstructured text field to record miscellaneous information. | Tiffs generated from Sid images |

Data Access

While it was recognized that it is difficult to prove the success of any digital preservation effort without providing access, access was not made an explicit part of the initial work plan for the following reasons:

1. The rights situation associated with many datasets was unclear at the outset. Many local agencies are uncomfortable with secondary redistribution of data and may have concerns about liability with regard to older datasets. Excluding access from the initial work plan provided more flexibility with regard to data acquisition and resolution of other organizational and technical issues.
2. There was a desire to avoid building a separate data access infrastructure from the existing NC OneMap framework, and it was recognized that a demonstration repository at NCSU was not necessarily the ideal long-term location of the collection given the interest in more formally involving State Archives in spatial data infrastructure.

At the outset it was expected that the project would trigger a community-wide discussion about access to older data would help to clarify project directions vis-à-vis data access. In fact, during the initial three years of work the project has substantially catalyzed discussion about data sharing and about roles with regard to access to older data.¹³¹ Formal engagement of State Archives in project work and movement towards inclusion of historical data in the NC OneMap framework will provide avenues towards more open access to

¹²⁹ Handle System: <http://www.handle.net/>

¹³⁰ NOID (nice opaque identifier): <http://www.cdlib.org/inside/diglib/noid/>

¹³¹ Data archiving issues and access to temporal data became significant issues in the in the NC GICC Ad Hoc Local/State/Federal Data Sharing Committee discussions and final reporting in 2007. The Nov. 2007 "Working Discussion Document" is available in Appendix H and online at: http://www.ncgicc.com/Portals/3/documents/GICC_presentations_110707_FINAL.zip

archived data. In the meantime most acquired data is made available to NCSU faculty, staff, and students since acquisition arrangements generally support that university use.¹³²

Enhancement of access to historical data through the NC OneMap framework is a key component of ongoing project work through March 2009. NC OneMap is cultivating data access in a variety of ways: 1) as interoperable web mapping services, 2) through data download mechanisms, and 3) through semi-formal sneakernet arrangements for transfer of large data collections from agency to agency via external drive exchanges.

Web Map Services (WMS)

The “NC Geospatial Backbone” provided by NC OneMap relies on the OGC Web Map Services (WMS) specification to bring data together from multiple hosts.¹³³ NC OneMap partners establish a WMS from their servers. The WMS standard allows data to be created and stored in numerous proprietary software configurations yet be viewable by all. Web Map Services are a two-way exchange: both to and from NC OneMap, and partners can choose to stream data from NC OneMap into a desktop application in their workplace. A benefit for GIS professionals is that raster and vector data from another web service can be loaded into GIS desktop software that supports WMS and used as a static backdrop graphic. The advantage to using a WMS map service from NC OneMap is that the data is served “live” and should be the most current data available. One component of the NCGDAP effort is to investigate how historic data might be integrated into the viewer. To this end NCSU Libraries recently began serving out historical census information into the NC OneMap system via WMS. CGIA will also be investigating including historical municipal boundary data in this system as a first step towards inclusion of temporal vector data in the NC OneMap Viewer. Multiple years of orthoimagery covering the entire state are already available via WMS.

NC OneMap infrastructure development involves several key initiatives and components:

Linking Services: NCCGIA continues to lead efforts to recruit additional counties, cities, state agencies, and regional organizations into the NC OneMap Framework. Bringing additional entities into the framework makes more data available for sharing, promotes further production of metadata, promotes adherence to content standards, and increases opportunities to engage producers directly on data preservation issues. Technical support is provided to bring these additional agencies online.

Database Administration: NCCGIA maintains a database of web services connections to inform the NC OneMap Viewer application. This information is also made available to the public via the NC OneMap web services catalog.¹³⁴ Connections in this catalog can be used to access and use shared data sets without needing a copy that physically resides on the users computer.

WMS Service Checker: The NC OneMap viewer draws on data from individual state, local, regional, and federal agencies via WMS services, so that a single map interface draws data simultaneously from across agencies. One component of NCGDAP has been an

¹³² Local agency data holdings are available to NCSU users through the County GIS Data Directory: <http://www.lib.ncsu.edu/gis/counties.html>

¹³³ NC OneMap web services: <http://www.nconemap.net/Default.aspx?tabid=287>

¹³⁴ NC OneMap Map Service Catalog: <http://www.nconemap.net/Default.aspx?tabid=298>

investigation into capture of map imagery via WMS. To this end, initial tests to capture and analyze service metadata, which takes the form of WMS Capabilities Files, was conducted. In these tests the services metadata was found to be sometimes inconsistent in structure, and the services themselves were found to be subject to spotty availability. Particularly in response to concerns about service availability, NCCGIA has developed an application to monitor WMS service uptime and make it possible to establish overall metrics about service availability and improve response time in case of service failure. This effort would support possible future NCGDAP efforts in the area of experimental WMS harvesting to develop historical Tile Map Services.

NC OneMap Data Download Service

As a counterpart to the NC OneMap Viewer, CGIA recently made available a service which makes a wide range of state agency data and some local agency data available for public download. While serving public and industry needs, this service also streamlines NCGDAP efforts to acquire state agency data. This service also supports the needs of federal agencies. For example, the National Geospatial-Intelligence Agency (NGA) has been able to use this service as part of its effort to secure state and local data for the Homeland Security Infrastructure Program (HSIP). NGA has then validated the acquired data, provided some extended value to the data, and made the resulting product available back to state and local governments.

NCSU Libraries County and City Data Directories

Since 2000 NCSU Libraries has maintained directories of county and city geospatial resources in North Carolina.¹³⁵ These directories document agency contact points, data downloads, web mapping applications, and WMS services. In the course of the project these directories have been enhanced in a number of ways including: a) development of a database to support directory update, b) addition of WMS and database connection links, c) addition of new counties and municipalities, and d) quality control of existing links. Additionally, the directories have been enhanced to provide public access to inventory information about data archived as part of the project. While these directories support the discovery needs of data seekers, they also support NCGDAP project work by making data access and contact information available in a single place.

Connectivity between the city and county pages was created so that users searching for links to a particular county's website will be able to see that a municipality within that county also has an online GIS presence. The link database created for the county and municipal pages is also used for individual county web pages, which are linked from the counties page and provide reference information, local links, and NCGDAP data holdings for each county in the state. These pages are generated dynamically using a PHP script template in tandem with databases for links and data holdings.

The county directory in particular is widely used, and commercial and public interest is such that it is the third most highly used entry point into the entire NCSU Libraries website. The directories are now also linked from the NC OneMap website as one of the data access options. NCGDAP has also experimented with making the data access links indexable and

¹³⁵ NCSU Libraries County GIS Directory: <http://www.lib.ncsu.edu/gis/counties.html>; and NCSU Libraries City GIS Directory: <http://www.lib.ncsu.edu/gis/counties.html>

discoverable as KML placemarks that have been exposed for indexing and discovery in the Google Earth environment via KML Search.

Instability of the hundreds of URL's providing links to county and municipal websites and services makes directory maintenance a challenge. To enable easier link maintenance, the directories use a system that stores the links in an online database and references individual sites through persistent identifiers. The use of identifiers allows the site and service references to be used in many locations while maintained in a single place. These identifiers can be used by other agencies to reference sites from their own environments. For example, the NC State Property Office makes use of NCSU-maintained identifiers and the associated resolution service to provide links to county map services from their own directory of county links.¹³⁶ The use of identifier resolution services also makes possible the capture of data about spatial differentiation of user demand for data across the state.

Transfer of Data to Library of Congress

In spring 2008 NCGDAP will transfer data to the Library of Congress, which will then maintain copies of the data. This effort will provide experience in solving the technical challenges associated with large scale transfer of content. The transfer will also help develop experience with managing rights arrangements with regards to transfer of content to secondary archives. Machine readable rights expression metadata will be transferred with the data along with any metadata, agreements, and data disclaimers. In order to simplify the transfer and focus on basic levels of interoperability with regard to data exchange, the data will be transferred with a "bag it and tag it" approach by which the receiving repository does not attempt to process received data at the item level. Data for which copyright, privacy or other necessary rights cannot be cleared will be held out of scope and will not be transferred to the Library of Congress.

¹³⁶ NC State Property Office "County Links": <http://www.doa.state.nc.us/spo/county.htm>

Moving Forward

From the start the intent of the project 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. Evidence that some progress is being made in making data preservation a part of spatial data infrastructure includes:

1. The introduction of the archiving problem as a topic of common discussion within the data producer and custodian community;
2. The formation of the new GICC Archival and Long-Term Preservation Working Group; and
3. The initiation of the new NDIIPP multi-state geospatial initiative.

The new GICC working group and the Multi-State geospatial project have been initiated under the leadership of the stakeholder community, and 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. With the more immediate and more conventional problems being addressed in the new GICC committee and the Multi-State geospatial project, there will be an opportunity for NCGDAP to focus on and take leadership in investigations and efforts that are not so directly tied to immediate business needs but which do contribute to the solution of the longer-term challenges of data preservation.

Specific areas of opportunity for ongoing NCGDAP work include:

- Engagement with standards organizations on issues such as content packaging and archival data profiles
- Experimentation with large scale data transfers and distributed archive management
- Investigation of methods for incorporating temporal and historical data in data discovery and access environments
- Exploration of possible approaches for managing temporal content within web services-based decision support environments
- Investigation of the implications of emergent geospatial PDF content for long-term preservation
- Exploration of preservation issues related to emergent place-based, non-spatial information resources

- Conduct of inter-organizational and inter-repository content exchange tests
- Exploration of implementation of established and emerging digital library technologies such as METS, PREMIS, and OAI-ORE in connection with geospatial data preservation
- Further developing the business case for preservation of geospatial data
- Providing transitional support to the new NDIIPP multistate geospatial initiative

Efforts in these areas will provide opportunities for cross-fertilization back into the NDIIPP multi-state geospatial initiative as well as the work of the GICC Archival and Long-Term Preservation Working Group.

Appendix A: Frequency of Geospatial Data Capture Survey

Following is the final report of the Frequency of Data Capture Survey carried out by NC Center for Geographic Information & Analysis as part of the NCGDAP partnership effort. The report is available at: http://www.nconemap.com/Portals/7/documents/NCOneMap_NDIIPLocalGovSurvey_1106.pdf

North Carolina Geospatial Data Archiving Project

Frequency of Geospatial Data Capture

Survey conducted by

NC Center for Geographic Information & Analysis

Under a Partnership with

The Library of Congress

National Digital Information Infrastructure and Preservation Program (NDIIPP)

and

North Carolina State University Libraries

November 15, 2006



North Carolina Geospatial Data Archiving Project

Frequency of Geospatial Data Capture

Survey of Local Government GIS Contacts

Project Introduction

As described on www.digitalpreservation.gov: The joint project of the North Carolina State University Libraries and the North Carolina Center for Geographic Information and Analysis in partnership with The Library of Congress, National Digital Information Infrastructure and Preservation Program will focus on collection and preservation of digital geospatial data resources from state and local government agencies in North Carolina.

The objectives of the project include:

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

Survey Overview

The North Carolina Geospatial Data Archiving Project (NCGDAP) is in the process of obtaining archival snapshots of county and city geospatial vector data layers and is seeking guidance about frequency of capture. At the same time, there is interest in defining a set of best practices with regard to maintenance of data archives at the local agency level, including periodic capture of vector data and associated attributes. In addition, State Archives has expressed preliminary interest in establishing a connection between records retention scheduling processes and any elaborated best practices for data archiving. NCSU, CGIA, and State Archives collaborated to develop a survey instrument to assess current data archiving practice at the county and municipal level. The survey was administered between September 13 and September 28, 2006 using the SurveyMonkey.com web service.

Survey Objective

The objective of the survey was to document current practices among county and municipal GIS practitioners relating to the frequency of capture of geospatial data for purposes of long-term retention. The survey distinguished between regular data back-up for disaster recovery purposes and retention of geospatial records for archiving purposes.

Response Status

The survey was sent to a list of local government GIS contacts covering all 100 counties and 25 of the largest municipalities. In North Carolina, local geospatial framework datasets are produced and managed by counties for the most part. Land records (cadastral data) are managed by counties as well as street centerlines and orthophotos. Municipalities typically have a role in managing geospatial representations of jurisdictional boundaries as well as planning and permit-related datasets.

A total of 72 respondents represented 61 of 100 counties and 11 municipalities. This was a strong response to an online survey. The county respondents included a range of sizes (population) and GIS capabilities. The counties that responded had somewhat larger resident populations than the counties that did not respond as shown in Table 1.

• Table 1: Population of Jurisdiction by Response Status

| Status of County | Population | mean | median |
|----------------------|------------|--------|--------|
| Responded (N = 61) | 5,669,527 | 92,943 | 48,498 |
| No response (N = 39) | 2,379,786 | 61,020 | 36,348 |

Results

Two-thirds of the respondents indicated that they create and retain periodic snapshots of any vector datasets for long-term retention and archiving. The responding counties that do not capture any vector data for long-term retention tend to be smaller in terms of population as shown in Table 2. Lists of jurisdictions that capture and do not capture data are shown in Tables 3 and 4.

• Table 2: Population of Jurisdiction by Status of Data Capture Practices

| Status of County | Population | mean | median |
|------------------------------|------------|---------|--------|
| Capture data (N = 43) | 4,810,592 | 111,874 | 59,648 |
| Do not capture data (N = 18) | 858,935 | 47,719 | 29,967 |

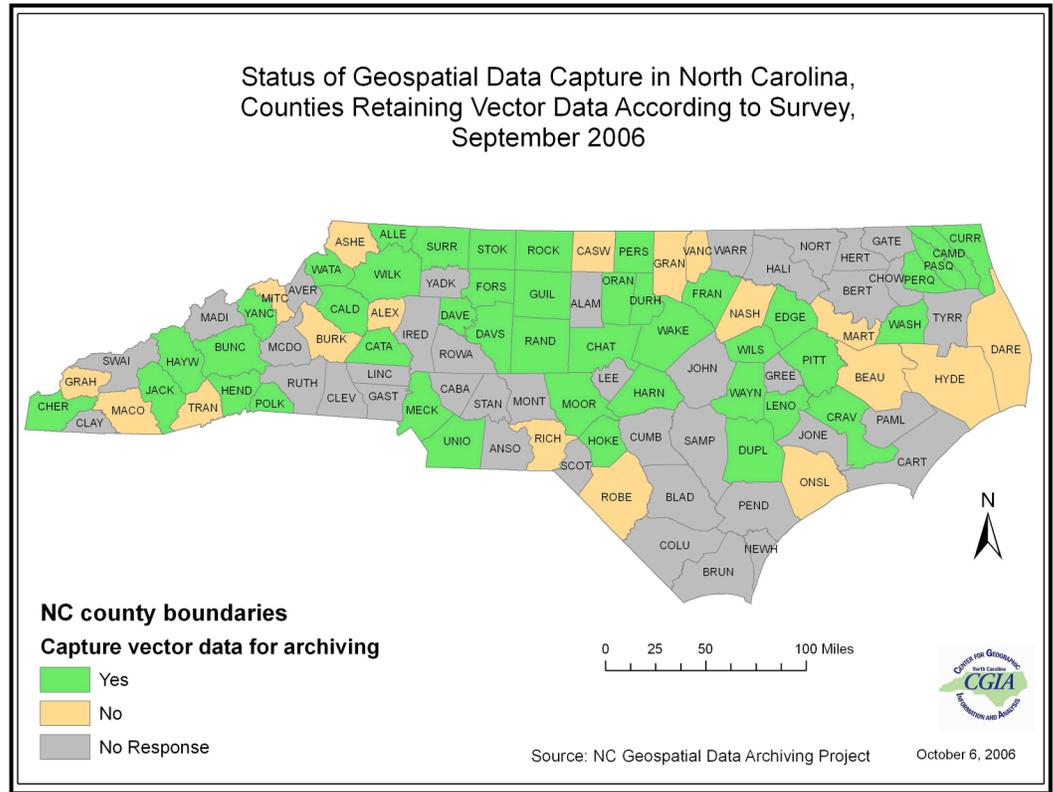
• Table 3: Jurisdictions That Capture Vector Data for Long-Term Retention

| NAME | TYPE | NAME | TYPE |
|------------|-----------|-------------|-----------|
| Alleghany | County | Jackson | County |
| Asheville | Municipal | Lenoir | County |
| Buncombe | County | Mecklenburg | County |
| Caldwell | County | Moore | County |
| Camden | County | Morrisville | Municipal |
| Cary | Municipal | Orange | County |
| Catawba | County | Pasquotank | County |
| Chatham | County | Perquimans | County |
| Cherokee | County | Person | County |
| Craven | County | Pitt | County |
| Currituck | County | Polk | County |
| Davidson | County | Randolph | County |
| Davie | County | Rockingham | County |
| Duplin | County | Stokes | County |
| Durham | County | Surry | County |
| Edgecombe | County | Union | County |
| Forsyth | County | Wake | County |
| Franklin | County | Washington | County |
| Greensboro | Municipal | Watauga | County |
| Guilford | County | Wayne | County |
| Harnett | County | Wilkes | County |
| Haywood | County | Wilson | County |
| Henderson | County | Yancey | County |
| Hoke | County | | |

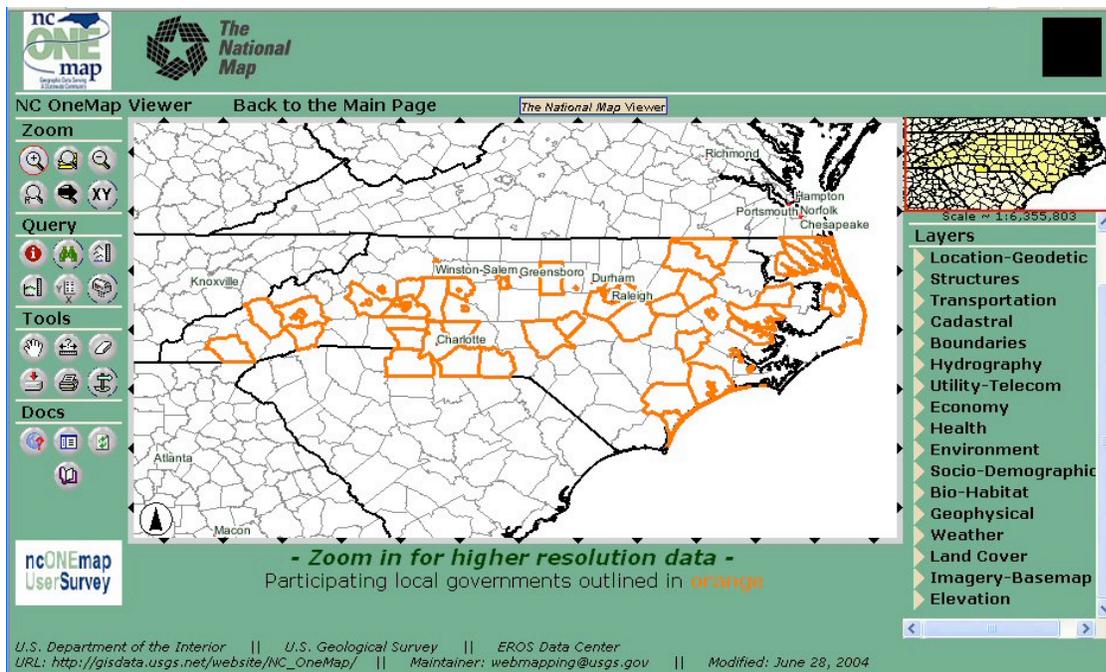
• Table 4: Jurisdictions That Do Not Capture Vector Data for Retention

| NAME | TYPE |
|------------|-----------|
| Alexander | County |
| Ashe | County |
| Beaufort | County |
| Boone | Municipal |
| Burke | County |
| Burlington | Municipal |
| Carrboro | Municipal |
| Caswell | County |
| Charlotte | Municipal |
| Dare | County |
| Graham | County |
| Granville | County |
| Hyde | County |

A map of the status of data capture by county is shown in Figure 1. The patterns of archiving and survey participation are a bit different than the current participation (data serving) in the NC OneMap map viewer (Figure 2).



• Figure 1: Data Capture Status by County



• Figure 2: NC OneMap Participating Local Governments, October 2006

In brief, about two-thirds of local government GIS coordinators are taking time to capture geospatial datasets, at least on an annual basis. For those who capture data more often than annually, the frequency varies from weekly to semi-annually. Cadastral data are most commonly archived among the respondents (41 of the 47 who retain geospatial data). Archiving occurs to a somewhat lesser extent for street centerlines (28 of the respondents), jurisdictional boundaries (28), and zoning (26). As a share of all survey respondents, geospatial records are archived for cadastral (57 percent), street centerlines (39 percent), jurisdictional boundaries (39 percent), and zoning (36 percent).

There are several business rules and needs that drive retention, including historic mapping, tax administration rules, information technology policies, records for resolution of legal issues, records retention policies, and land use change analysis.

Storage formats tend to be consistent with the dominant GIS vendor among local governments (ESRI). Storage environments vary, with servers and CDs the most common. Offsite storage (or both onsite and offsite) is used by nearly half of the respondents.

In addition to vector data, 65 of the 72 respondents store digital orthophotos. The popularity of historic images has resulted in scanning and geo-referencing of hardcopy aerial photos among about one-third of the respondents.

Responses to 28 questions are summarized in Table 5.

• Table 5. Summary of Survey Results

1 Do you create and retain periodic snapshots of any vector datasets for long-term retention and archiving?

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Yes | 47 | 65.3% | 65.3% |
| No | 25 | 34.7% | 34.7% |
| Total Respondents | 72 | 100.0% | 100.0% |
| (skipped this question) | 0 | 0.0% | 0.0% |

2 How often are snapshots of PARCEL geometry made for long-term retention? (choose frequency closest to your practice)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Annually | 20 | 42.6% | 27.8% |
| Every 6 Months | 4 | 8.5% | 5.6% |
| Quarterly | 4 | 8.5% | 5.6% |
| Monthly | 7 | 14.9% | 9.7% |
| Weekly or Daily | 6 | 12.8% | 8.3% |
| Not Saved | 6 | 12.8% | 8.3% |
| Total Respondents | 47 | 100.0% | 65.3% |
| (skipped this question) | 25 | | 34.7% |

3 In what formats are PARCEL geometry snapshots saved? (check all that apply)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Shapefile | 32 | 76.2% | 44.4% |
| Geodatabase | 15 | 35.7% | 20.8% |
| Arc Coverage | 12 | 28.6% | 16.7% |
| Arc Interchange (e00) | 3 | 7.1% | 4.2% |
| Other | 4 | 9.5% | 5.6% |
| Total Respondents | 42 | 100.0% | 58.3% |
| (skipped this question) | 30 | | 41.7% |

4 Does this involve a data conversion from the native format? (e.g. Geodatabase feature class export to shapefile)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Yes | 23 | 52.3% | 31.9% |
| No | 21 | 47.7% | 29.2% |
| Total Respondents | 44 | 100.0% | 61.1% |
| (skipped this question) | 28 | | 38.9% |

5 Are PARCEL attributes (such as tax record information) saved with the PARCEL geometry data?

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|----------------------------|----------------|------------------------|--------------------------|
| Yes as attached attributes | 27 | 61.4% | 37.5% |
| Yes in a separate table | 10 | 22.7% | 13.9% |
| No | 7 | 15.9% | 9.7% |
| Total Respondents | 44 | 100.0% | 61.1% |
| (skipped this question) | 28 | | 38.9% |

6 How often are snapshots of STREET CENTERLINE geometry made for long-term retention? (choose frequency that is closest to your practice)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Annually | 13 | 28.3% | 18.1% |
| Every 6 Months | 3 | 6.5% | 4.2% |
| Quarterly | 4 | 8.7% | 5.6% |
| Monthly | 5 | 10.9% | 6.9% |
| Weekly or Daily | 3 | 6.5% | 4.2% |
| Not Saved | 18 | 39.1% | 25.0% |
| Total Respondents | 46 | 100.0% | 63.9% |
| (skipped this question) | 26 | | 36.1% |

7 In what formats are STREET CENTERLINE geometry snapshots saved? (check all that apply)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|-------------------|---------------------------|--------------------------------|
| Shapefile | 24 | 77.4% | 33.3% |
| Geodatabase | 8 | 25.8% | 11.1% |
| Arc Coverage | 8 | 25.8% | 11.1% |
| Arc Interchange (e00) | 3 | 9.7% | 4.2% |
| Other | 3 | 9.7% | 4.2% |
| Total Respondents | 31 | 100.0% | 43.1% |
| (skipped this question) | 41 | | 56.9% |

8 Does this involve a data conversion from the native format? (e.g. MapInfo format to Shapefile)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|-------------------|---------------------------|--------------------------------|
| Yes | 12 | 34.3% | 16.7% |
| No | 23 | 65.7% | 31.9% |
| Total Respondents | 35 | 100.0% | 48.6% |
| (skipped this question) | 37 | | 51.4% |

9 Are STREET attributes saved with the STREET CENTERLINE geometry?

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|----------------------------|-------------------|---------------------------|--------------------------------|
| Yes as attached attributes | 29 | 82.9% | 40.3% |
| Yes in a separate table | 2 | 5.7% | 2.8% |
| No | 4 | 11.4% | 5.6% |
| Total Respondents | 35 | 100.0% | 48.6% |
| (skipped this question) | 37 | | 51.4% |

10 For which of the following JURISDICTIONAL BOUNDARY datasets do you create snapshots? (check all that apply)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|--------------------------------|----------------|------------------------|--------------------------|
| County Boundaries | 19 | 40.4% | 26.4% |
| Municipal Boundaries | 25 | 53.2% | 34.7% |
| Extraterritorial Jurisdictions | 16 | 34.0% | 22.2% |
| None | 19 | 40.4% | 26.4% |
| Total Respondents | 47 | 100.0% | 65.3% |
| (skipped this question) | 25 | | 34.7% |

11 How often are snapshots of JURISDICTIONAL BOUNDARY data made for long-term retention? (choose frequency closest to your practice)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|---|----------------|------------------------|--------------------------|
| Any time an official boundary change occurs | 13 | 39.4% | 18.1% |
| Annually | 10 | 30.3% | 13.9% |
| Every 6 Months | 3 | 9.1% | 4.2% |
| Quarterly | 1 | 3.0% | 1.4% |
| Monthly | 5 | 15.2% | 6.9% |
| Weekly or Daily | 1 | 3.0% | 1.4% |
| Total Respondents | 33 | 100.0% | 45.8% |
| (skipped this question) | 39 | | 54.2% |

12 In what formats are JURISDICTIONAL BOUNDARY data snapshots saved?

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Shapefile | 24 | 72.7% | 33.3% |
| Geodatabase | 10 | 30.3% | 13.9% |
| Arc Coverage | 7 | 21.2% | 9.7% |
| Arc Interchange (e00) | 2 | 6.1% | 2.8% |
| Other | 5 | 15.2% | 6.9% |
| Total Respondents | 33 | 100.0% | 45.8% |
| (skipped this question) | 39 | | 54.2% |

13 Does this involve a data conversion from the native format? (e.g. Geodatabase feature class to shapefile)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Yes | 11 | 34.4% | 15.3% |
| No | 21 | 65.6% | 29.2% |
| Total Respondents | 32 | 100.0% | 44.4% |
| (skipped this question) | 40 | | 55.6% |

14 How often are snapshots of ZONING geometry made for long-term retention? (choose frequency closest to your practice)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Annually | 16 | 34.8% | 22.2% |
| Every 6 Months | 3 | 6.5% | 4.2% |
| Quarterly | 2 | 4.3% | 2.8% |
| Monthly | 4 | 8.7% | 5.6% |
| Weekly or Daily | 1 | 2.2% | 1.4% |
| Not Saved | 20 | 43.5% | 27.8% |
| Total Respondents | 46 | 100.0% | 63.9% |
| (skipped this question) | 26 | | 36.1% |

15 In what formats are ZONING geometry snapshots saved? (check all that apply)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Shapefile | 20 | 66.7% | 27.8% |
| Geodatabase | 8 | 26.7% | 11.1% |
| Arc Coverage | 6 | 20.0% | 8.3% |
| Arc Interchange (e00) | 2 | 6.7% | 2.8% |
| Other | 5 | 16.7% | 6.9% |
| Total Respondents | 30 | 100.0% | 41.7% |
| (skipped this question) | 42 | | 58.3% |

16 Does this involve a conversion from the native format? (e.g. MapInfo to Shapefile)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Yes | 11 | 35.5% | 15.3% |
| No | 20 | 64.5% | 27.8% |
| Total Respondents | 31 | 100.0% | 43.1% |
| (skipped this question) | 41 | | 56.9% |

17 Are ZONING attributes saved with ZONING geometry?

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|----------------------------|----------------|------------------------|--------------------------|
| Yes as attached attributes | 25 | 78.1% | 34.7% |
| Yes in a separate table | 2 | 6.3% | 2.8% |
| No | 5 | 15.6% | 6.9% |
| Total Respondents | 32 | 100.0% | 44.4% |
| (skipped this question) | 40 | | 55.6% |

18 Please list or summarize any OTHER DATA layers you are archiving for long-term retention if any.

| | | | |
|-------------------------|----|--------|-------|
| Total Respondents | 24 | 100.0% | 33.3% |
| (skipped this question) | 48 | | 66.7% |

19 How far back does your archive of vector data snapshots go?

| | | | |
|-------------------------|----|--------|-------|
| Total Respondents | 39 | 100.0% | 54.2% |
| (skipped this question) | 33 | | 45.8% |

20 What METADATA types are saved with the snapshot data?

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|----------------------------------|----------------|------------------------|--------------------------|
| FGDC format | 12 | 25.5% | 16.7% |
| Locally defined metadata | 4 | 8.5% | 5.6% |
| NC OneMap metadata starter block | 3 | 6.4% | 4.2% |
| None | 28 | 59.6% | 38.9% |
| Total Respondents | 47 | 100.0% | 65.3% |
| (skipped this question) | 25 | | 34.7% |

21 In what STORAGE environment are the snapshot data saved?

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|--------------------------|----------------|------------------------|--------------------------|
| Tape | 10 | 21.3% | 13.9% |
| CD | 20 | 42.6% | 27.8% |
| DVD | 8 | 17.0% | 11.1% |
| External Hard Drive | 4 | 8.5% | 5.6% |
| Server or Online Storage | 27 | 57.4% | 37.5% |
| Other | 0 | 0.0% | 0.0% |
| Total Respondents | 47 | 100.0% | 65.3% |
| (skipped this question) | 25 | | 34.7% |

22 Where are the snapshot data stored?

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Onsite | 26 | 55.3% | 36.1% |
| Offsite | 3 | 6.4% | 4.2% |
| Both Onsite and Offsite | 18 | 38.3% | 25.0% |
| Total Respondents | 47 | 100.0% | 65.3% |
| (skipped this question) | 25 | | 34.7% |

23 What local business RULES and/or USES drive the long-term retention of vector data in your jurisdiction? (check all that apply)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------------|----------------|------------------------|--------------------------|
| Information technology policy | 9 | 19.6% | 12.5% |
| Records retention policy | 8 | 17.4% | 11.1% |
| Tax administration rules | 11 | 23.9% | 15.3% |
| Land use change analysis | 5 | 10.9% | 6.9% |
| Resolution of legal issues | 9 | 19.6% | 12.5% |
| Historic mapping | 26 | 56.5% | 36.1% |
| Other | 13 | 28.3% | 18.1% |
| Total Respondents | 46 | 100.0% | 63.9% |
| (skipped this question) | 26 | | 36.1% |

24 Do you provide PUBLIC ACCESS to snapshots of vector data?

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Yes all of the files | 14 | 30.4% | 19.4% |
| Yes some of the files | 6 | 13.0% | 8.3% |
| None | 26 | 56.5% | 36.1% |
| Total Respondents | 46 | 100.0% | 63.9% |
| (skipped this question) | 26 | | 36.1% |

25 How are historic (superseded) digital ORTHOPHOTOS being stored? (check all that apply)

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|--------------------------|----------------|------------------------|--------------------------|
| Tape | 4 | 5.7% | 5.6% |
| CD | 30 | 42.9% | 41.7% |
| DVD | 23 | 32.9% | 31.9% |
| External Hard Drive | 13 | 18.6% | 18.1% |
| Server or Online Storage | 46 | 65.7% | 63.9% |
| Other | 10 | 14.3% | 13.9% |
| Not Stored | 5 | 7.1% | 6.9% |
| Total Respondents | 70 | 100.0% | 97.2% |
| (skipped this question) | 2 | | 2.8% |

26 Have you created digital versions of any of the following:

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|---|----------------|------------------------|--------------------------|
| Historic hardcopy maps scanned only | 11 | 15.5% | 15.3% |
| Historic hardcopy maps scanned and geo-referenced | 7 | 9.9% | 9.7% |
| Aerial photos scanned only | 6 | 8.5% | 8.3% |
| Aerial photos scanned and geo-referenced | 19 | 26.8% | 26.4% |
| None | 39 | 54.9% | |
| Total Respondents | 71 | 100.0% | 98.6% |
| (skipped this question) | 1 | | 1.4% |

27 Please add any additional comments clarifications or questions:

| | | | |
|-------------------------|----|--------|-------|
| Total Respondents | 15 | 100.0% | 20.8% |
| (skipped this question) | 57 | | 79.2% |

28 Would you like to participate in FORUMS concerning preservation of local geospatial data?

| | Response Total | Percent of Respondents | Percent of Survey Takers |
|-------------------------|----------------|------------------------|--------------------------|
| Yes | 26 | 38.2% | 36.1% |
| Not sure | 26 | 38.2% | 36.1% |
| No | 16 | 23.5% | 22.2% |
| Total Respondents | 68 | 100.0% | 94.4% |
| (skipped this question) | 4 | | 5.6% |

Other data that are captured for long-term retention were described in responses to question 18.

18. Please list or summarize any OTHER DATA layers you are archiving for long-term retention, if any.

Comment:

Address points, driveway line segments, emergency service boundaries, school districts, contour line data, census boundaries, historic districts, railroad centerline, township boundaries, subdivision boundaries, voting precincts, watershed overlay districts

I made an archive of my data when we converted from coverages to geodatabase. That is typically the only time I do an archived backup. I do a weekly backup of existing data but it is overwriting the previously saved data

Fire districts, Fire stations, Building footprints, Boundary, CPL Areas, Easements, Lot lines, Utility easements, Child care, FEMA & regular flood, Government services, Pump stations, Schools, Senior citizen areas, Zoning, Airport boundary, County Offices, Tax Parcels, Roads

We had been periodically archiving impervious surface data for 'history' during the development process. At this time (now that the stormwater bills have gone out) we plan to track reductions/additions through HTE software and through our Appeals database (.mdb). For most other enterprise data we're relying upon ITS recovery mechanisms. I have also archived project/analysis level data using our CD Index.

Annexations, Zip codes, Pavement Management, Land use

Fire Tax Districts

Land Use

All of our data is archived daily, then weekly, then monthly, and yearly.

Flood Plain shapefiles ETJ shapefiles

Airport, bridges, churches, city limits, creeks, structures (houses), landmarks, land use, railway, soils, tar river, utility service areas, and cross road information.

Ortho Photos, Topographic maps, Planimetrics

All layers included on our Public Access CD (produced every 6 months). We keep one of each issue as an historic archive.

Ortho Photography, topographic maps, planimetrics, building footprints

Fire Districts Sanitary Districts Watersheds Federal Lands Land Cover Soils
Topography Water lines Sewer lines Survey Monuments Voting Districts
Flood Census

Imagery

Critical watershed areas

Address points

We Archive All Property Ownership Records Along With Chain Of Title. We Have Scanned All 1993 Photo, 1974 Photo. We Also Have Scanned All Recorded Survey Plats. All Sheriff, EMS, Fire, Zip, Township, Phone Messages, Zoning

Subdivisions Electoral Districts - precincts, congressional districts, NC House districts, Judicial districts, polling places

Address points

Address points

Those not currently being archived are not yet in existence. We are only an emerging GIS. But it is my intention that ALL data will be archived.

Respondents were asked for comments about archiving in question 27.

27. Please add any additional comments, clarifications or questions:

Response

1985 - blue hardcopy maps scanned and geo-referenced 1999 - Digital B&W saved on DVD and online 2002 - Digital color saved on DVD and online 2006 - new acquisition this year; The soil and water group here locally has some aerial imagery (B&W) hardcopy going back to the 1950s. I would like to have that scanned and geo-referenced. Are there any funds (cost share) that would help with that effort? We'd like to be able to put them online for our users to view. Thanks!

We are looking into having old orthophotos scanned and geo-referenced.

Getting ready to implement this type of archiving of data.

I have not done this, but it does seem like a good idea!

All of our data is kept monthly for 1 year; i.e., September 2006 tape will be overwritten September 2007.

Have old B/W photos (1990) that have been scanned to disk and original's are stored in Master File cabinet. Also new photo's stored on DVD.

We have taken our original 1990 aerial photography, scanned and geo-referenced the images. We will be adding other archive aerial imagery as available from NRCS.

I do not see why this can not be incorporated with disaster recovery. Don't you think you would foster greater support?

We have sold CD copies of our annual data to the public for that current year, from 99 to the present.

We do not have any historic DIGITAL orthophotos. Our historic records are just hardcopy. The digital orthophotos that we are using are our only ones we have.

We have hard copies of a historic Aerial Photo but no resources to scan

No emphasis on historical data here. We just try to keep from losing data completely. Very minimal hardware to work with and no money.

Our current digital orthophotography (flown spring of 2004) has not been superseded.....yet.

Have not yet begun scanning old Aerials but intend to do so soon.

The survey questions are attached as Appendix A.

Conclusion

The survey results are encouraging for the NC Geospatial Data Archiving Project and provide insights into frequency of capture and current practices. In addition, the survey brought attention to the archiving issue for those GIS coordinators who are not capturing geospatial data for long-term retention. This presents an opportunity for NCGDAP to influence the breadth and quality of archived geospatial information in North Carolina.

Appendix B: Data Acquisition Summary

Vector Data Acquisition by County as of February 2008

| | | | | |
|-------------|------|------|------|-----------|
| Alamance | 2003 | 2007 | | |
| Anson | 2006 | 2007 | | |
| Bertie | 2006 | 2007 | | |
| Bladen | 2006 | | | |
| Brunswick | 2001 | 2004 | | |
| Buncombe | 2004 | 2006 | 2007 | |
| Burke | 2007 | | | |
| Cabarrus | 2000 | 2002 | 2006 | 2007 |
| Carteret | 2001 | | | |
| Catawba | 2007 | | | |
| Chatham | 2000 | 2006 | 2007 | |
| Cherokee | 2007 | | | |
| Columbus | 2001 | 2006 | 2007 | |
| Craven | 2000 | 2002 | 2006 | 2007 |
| Cumberland | 2000 | 2007 | | |
| Dare | 2000 | | | |
| Davie | 2007 | | | |
| Duplin | 2007 | | | |
| Edgecombe | 2006 | 2007 | | |
| Forsyth | 2000 | 2001 | 2006 | 2007 |
| Franklin | 2001 | | | |
| Gaston | 2006 | 2007 | | |
| Granville | 2004 | | | |
| Guilford | 2002 | 2007 | | |
| Halifax | 2006 | 2007 | | |
| Harnett | 2000 | 2007 | | |
| Haywood | 2006 | 2007 | | |
| Henderson | 2002 | 2006 | 2007 | |
| Iredell | 2007 | | | |
| Johnston | 2002 | 2004 | 2006 | 2007 |
| Lee | 2000 | 2006 | 2007 | |
| Lenoir | 2000 | | | |
| Macon | 2006 | 2007 | | |
| Mecklenburg | 1997 | 1999 | 2000 | 2001 2007 |
| Montgomery | 2004 | | | |
| Moore | 2000 | 2006 | 2007 | |
| Nash | 2001 | | | |
| New | | | | |
| Hanover | 1999 | 2002 | 2006 | 2007 |
| Onslow | 2000 | 2006 | 2007 | |

| | | | | | | | | | | | | | | | | | | | | | |
|------------|------|------|------|------|------|------|------|------|------|------|--|--|--|--|--|--|--|--|--|--|--|
| Orange | 2001 | 2006 | 2007 | | | | | | | | | | | | | | | | | | |
| Pamlico | 2001 | | | | | | | | | | | | | | | | | | | | |
| Pasquotank | 2002 | 2007 | | | | | | | | | | | | | | | | | | | |
| Pender | 2006 | 2007 | | | | | | | | | | | | | | | | | | | |
| Pitt | 2000 | 2007 | | | | | | | | | | | | | | | | | | | |
| Randolph | 2000 | | | | | | | | | | | | | | | | | | | | |
| Richmond | 2005 | | | | | | | | | | | | | | | | | | | | |
| Rockingham | 2001 | | | | | | | | | | | | | | | | | | | | |
| Rowan | 2002 | 2006 | 2007 | | | | | | | | | | | | | | | | | | |
| Sampson | 2006 | | | | | | | | | | | | | | | | | | | | |
| Scotland | 2006 | 2007 | | | | | | | | | | | | | | | | | | | |
| Stanly | 2000 | 2006 | 2007 | | | | | | | | | | | | | | | | | | |
| Union | 2002 | 2006 | 2007 | | | | | | | | | | | | | | | | | | |
| Vance | 2007 | | | | | | | | | | | | | | | | | | | | |
| Wake | 1996 | 1997 | 1999 | 2000 | 2001 | 2002 | 2004 | 2005 | 2006 | 2007 | | | | | | | | | | | |
| Warren | 2007 | | | | | | | | | | | | | | | | | | | | |
| Washington | 2000 | | | | | | | | | | | | | | | | | | | | |
| Wayne | 2000 | 2006 | 2007 | | | | | | | | | | | | | | | | | | |
| Wilkes | 2007 | | | | | | | | | | | | | | | | | | | | |
| Wilson | 2000 | | | | | | | | | | | | | | | | | | | | |

Digital Orthophoto Data Acquisition by County as of February 2008

| | | | | | | | | | | | | | | | | | | | | | |
|-----------|------|------|------|------|------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| Alamance | 2002 | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | |
| Alexander | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Alleghany | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Anson | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Ashe | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Avery | 1998 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Beaufort | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Bertie | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Bladen | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Brunswick | 1988 | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | |
| Buncombe | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Burke | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Cabarrus | 2000 | 2002 | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | |
| Caldwell | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Camden | 2005 | 2006 | | | | | | | | | | | | | | | | | | | |
| Carteret | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Caswell | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Catawba | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Chatham | 1997 | 2002 | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | |
| Cherokee | 1998 | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | |
| Chowan | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Clay | 2005 | 2006 | | | | | | | | | | | | | | | | | | | |
| Cleveland | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |
| Columbus | 2004 | 2005 | 2006 | | | | | | | | | | | | | | | | | | |

| | | | | | | | | |
|-------------|------|------|------|------|------|------|------|--|
| Craven | 2000 | 2004 | 2005 | 2006 | | | | |
| Cumberland | 1995 | 2004 | 2005 | 2006 | | | | |
| Currituck | 2004 | 2005 | 2006 | | | | | |
| Dare | 1997 | 2004 | 2005 | 2006 | | | | |
| Davidson | 1996 | 2005 | 2006 | | | | | |
| Davie | 2004 | 2005 | 2006 | | | | | |
| Duplin | 2004 | 2005 | 2006 | | | | | |
| Edgecombe | 2004 | 2005 | 2006 | | | | | |
| Forsyth | 1997 | 2004 | 2005 | 2006 | | | | |
| Franklin | 2004 | 2005 | 2006 | | | | | |
| Gaston | 2004 | 2005 | 2006 | | | | | |
| Gates | 2004 | 2005 | 2006 | | | | | |
| Graham | 2005 | 2006 | | | | | | |
| Granville | 2004 | 2005 | 2006 | | | | | |
| Greene | 2004 | 2005 | 2006 | | | | | |
| Guilford | 1999 | 2004 | 2005 | 2006 | | | | |
| Halifax | 2004 | 2005 | 2006 | | | | | |
| Harnett | 2004 | 2005 | 2006 | | | | | |
| Haywood | 2005 | 2006 | | | | | | |
| Henderson | 2004 | 2005 | 2006 | | | | | |
| Hertford | 2004 | 2005 | 2006 | | | | | |
| Hoke | 2004 | 2005 | 2006 | | | | | |
| Hyde | 2005 | 2006 | | | | | | |
| Iredell | 2004 | 2005 | 2006 | | | | | |
| Jackson | 2004 | 2005 | 2006 | | | | | |
| Johnston | 2001 | 2004 | 2005 | 2006 | | | | |
| Jones | 2004 | 2005 | 2006 | | | | | |
| Lee | 1996 | 1997 | 1998 | 1999 | 2004 | 2005 | 2006 | |
| Lenoir | 2004 | 2005 | 2006 | | | | | |
| Lincoln | 1999 | 2004 | 2005 | 2006 | | | | |
| Macon | 1992 | 2004 | 2005 | 2006 | | | | |
| Madison | 2005 | 2006 | | | | | | |
| Mecklenburg | 1997 | 1999 | 2001 | 2004 | 2005 | 2006 | | |
| Mitchell | 2005 | 2006 | | | | | | |
| Montgomery | 2000 | 2004 | 2005 | 2006 | | | | |
| Moore | 2004 | 2005 | 2006 | | | | | |
| Nash | 1997 | 2004 | 2005 | 2006 | | | | |
| New Hanover | 1997 | 2002 | 2004 | 2005 | 2006 | | | |
| Northampton | 1998 | 2004 | 2005 | 2006 | | | | |
| Onslow | 1996 | 2004 | 2005 | 2006 | | | | |
| Orange | 1998 | 2004 | 2005 | 2006 | | | | |
| Pamlico | 2004 | 2005 | 2006 | | | | | |
| Pasquotank | 1997 | 2004 | 2005 | 2006 | | | | |
| Pender | 1998 | 2004 | 2005 | 2006 | | | | |
| Pitt | 2004 | 2005 | 2006 | | | | | |
| Polk | 2004 | 2005 | 2006 | | | | | |
| Randolph | 1998 | 2004 | 2005 | 2006 | | | | |
| Richmond | 1993 | 2004 | 2005 | 2006 | | | | |
| Robeson | 2000 | 2004 | 2005 | 2006 | | | | |
| Rockingham | 1998 | 2004 | 2005 | 2006 | | | | |

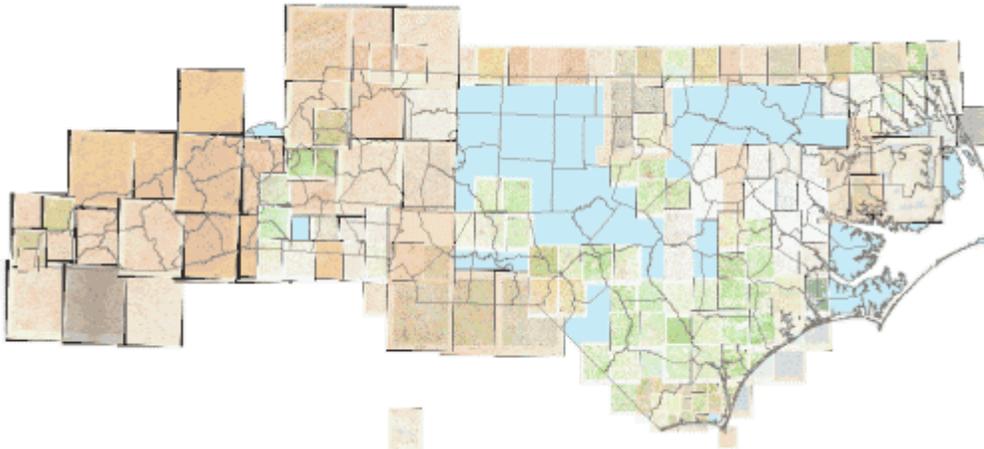
| | | | | |
|--------------|------|------|------|------|
| Rowan | 2004 | 2005 | 2006 | |
| Rutherford | 2004 | 2005 | 2006 | |
| Sampson | 2004 | 2005 | 2006 | |
| Scotland | 2004 | 2005 | 2006 | |
| Stanly | 2004 | 2005 | 2006 | |
| Stokes | 2001 | 2004 | 2005 | 2006 |
| Surry | 1997 | 2004 | 2005 | 2006 |
| Swain | 2005 | 2006 | | |
| Transylvania | 2004 | 2005 | 2006 | |
| Tyrell | 2005 | 2006 | | |
| Union | 2004 | 2005 | 2006 | |
| Vance | 2005 | 2006 | | |
| Wake | 1999 | 2004 | 2005 | 2006 |
| Warren | 2004 | 2005 | 2006 | |
| Washington | 2005 | 2006 | | |
| Watauga | 2004 | 2005 | 2006 | |
| Wayne | 2005 | 2006 | | |
| Wilkes | 2004 | 2005 | 2006 | |
| Wilson | 1995 | 2004 | 2005 | 2006 |
| Yadkin | 2004 | 2005 | 2006 | |
| Yancey | 2005 | 2006 | | |

Appendix C: NC Geologic Survey Maps

Geologic maps, including collars, had been scanned by NCGS scanned to 300 dpi TIF files with a large format (42-inch) HP Designjet 815 mfp scanner-plotter device. Using ArcMap 9.1's georeferencing extension, at least four geographic locations were interactively selected from the TIF based on coordinates and grids printed on the maps. ArcMap creates a table of these selected coordinate values, and with U.S. Army Corps of Engineers Corpscon software, each coordinate pair was converted to NC Stateplane NAD83 meters and then appended to the table. ArcMap then creates a TFW world file and transforms the TIF image so that it is represented in the data view in Stateplane Coordinates (NAD83). Each image file was rectified with ArcInfo workstation and compressed with MrSID and FGDC metadata was developed

The TIF images and world files became part of the NCGDAP archive. The inventory consists of 101 U.S. Geological Survey geologic maps, 130 North Carolina Geological Survey geologic maps, 47 maps from theses and dissertations, 8 N.C. Department of Transportation maps and 165 legacy 15-minute topographic maps.

The topographic maps are available online from the NCSU Libraries at: <http://www.lib.ncsu.edu/gis/historictopos.html>. Coverage of the historical topographic maps is illustrated in the map below:



Data are being disseminated through the North Carolina Geological Survey's Geologic Map Catalog, (<http://wfs.enr.state.nc.us/NCGeologicMaps/>), the NCSU Libraries' campus-wide server (<http://www.lib.ncsu.edu/gis/geolmaps.html>) pointing at the NCGS' Geologic map catalog URL, and by contribution to NCOneMap (<http://www.nconemap.com>), the National Geologic Map Database and the National Geologic Map Image Library (U.S. Geological Survey, <http://ngmdb.usgs.gov/>).

Appendix D: Geospatial Metadata Investigation Summary

| Metadata Issue | Findings |
|---|--|
| Determine availability of metadata for local agencies | FGDC metadata is available for data originating from several state agencies, and is comprehensively available for NC OneMap data layers. Availability of FGDC metadata for counties and cities is less comprehensive (roughly one quarter of counties having metadata in 2004, half of these with metadata available online), yet this number is increasing through outreach |
| Determine whether metadata in various clearinghouses such as GOS, NC OneMap, and the NC Data Clearinghouse is synchronous | Various discrepancies between GOS metadata and NC OneMap metadata were found, with examples including: a) same metadata reference dates, but different information; b) different dates and different information; and c) GOS metadata exists when neither NCOneMap or the local agency website contains current metadata. |
| Determine consistency of structure of metadata | Metadata formats encountered included HTML, txt, xml, and MS Word. Only rarely were the tested metadata fully FGDC CSDGM compliant. However, nearly all instances adhered in spirit to the FGDC metadata structure. |
| Determine whether substantive changes in data such as switching to a new format or using a new datum also trigger revision of local agency metadata records | Anecdotal cases were found in which format conversions (e.g., Arc Coverage to GeoDatabase) or datum conversions (e.g., NAD 27 to NAD 83) had taken place yet the metadata had not been updated. |
| Determine whether the new ISO 19115 geospatial metadata standard should be adopted in place of the current FGDC standard for project work | A North American Profile of the ISO 19115 geospatial metadata standard, using the ISO 19139 encoding standard, yet implementation has awaited finalization of the North American Profile of the ISO standards which will effectively replace FGDC version 2 |
| Determine whether the ESRI Profile of the FGDC provided additional archival functionality | The ESRI Profile, which had been designed to enhance software interaction with the metadata and to support automated catalog development, does provide a variety of additional technical metadata elements which are of archival value. Some of these elements facilitate synchronization of the metadata with the dataset. |
| Determine whether ESRI Profile elements could be successfully stripped in order to create vanilla FGDC records. | The FGDC 'mp' parser throws warnings rather than errors when encountering ESRI elements. NCSU participated in tests of a new ISO standard metadata parser, providing records that were stripped of ESRI elements either manually or using the NPS Metadata Toolkit. |
| Investigate method to link dataset versions through a persistent identifier | This issue has not been addressed in the industry any consistent or standard way |
| Identify tools available for metadata processing | 'cns' and 'mp' were found to be useful for raw metadata processing and lend themselves well to batch processing. ArcCatalog allows synchronization of data with metadata and has a variety of export options, but batch processing was very complicated. The NPS Metadata Toolkit had some export options not provided by ArcCatalog. |
| Determine alternate sources for metadata if no metadata is provided | A set of alternate sources was identified, including agency websites, the various data inventories, and direct agency contacts. Metadata acquired in this manner needs to be qualified in the repository. Batch extraction of various technical and descriptive metadata elements directly from the data is also possible using tools such as ArcCatalog. |

Appendix E: Rights Coding Scheme

North Carolina Geospatial Data Archiving Project Rights Codes Implementation

July 26, 2006

Purpose: Define a basic set of codes to hold dataset rights information in a script-actionable form. To assign related text for use in constructing brief rights statements.

Structure: Codes are assigned on a fixed string position basis. Rights assigned to particular user types are grouped after a flag character for that user group.

Initial User Groups:

- NCSU Faculty/Staff/Students (Code “N”)
- General Public (Code “P”)
- Library of Congress (Code “L”)

Additional user groups can be added later (e.g. “S” for state agencies)

Elements:

The element list is composed of two parts:

1. A group of flags indicating certain conditions (method of acquisition, presence/absence of agreement, and presence/absence of disclaimer. This group could be grown over time (e.g., presence/absence of copyright)
2. A group sets of rights for each of three user groups (NCSU, General Public, Library of Congress). Additional rights could be added over time. Additional groups could also be added.

Initial Rights Types:

- Use
- Redistribute
- Commercial Use

This group could be expanded at a later time. Candidate additional uses include: “web mapping” and “commercial sale”. These rights do not need to be defined on a dataset-by-dataset basis within our current acquisition domain (in our domain “web mapping” is an assumed “allowed” and “commercial sale” is an assumed “not allowed”). These rights would need to be defined for content outside of the current acquisition domain.

Sample Record:

M01N110P110L110

Interpretation: This dataset was acquired in a mediated transaction directly from the data producer (acquired on media or via arranged download). There is no data agreement but there is a data disclaimer. NCSU, General Public, and LC all can use and redistribute the data but commercial use is not allowed.

Element breakdown by position:

- 1) Method of access:
 - a. "M" for mediated acquisition from the data producer
 - b. "U" for unmediated access from the data producer
 - c. "T" for acquisition from a third party.
- 2) Presence/absence of signed agreement:
 - a. "1" for data agreement
 - b. "0" for no data agreement
- 3) Presence/absence of disclaimer
 - a. "1" for signed disclaimer
 - b. "0" for no signed disclaimer
- 4) NCSU rights positional indicator
 - a. "N"
- 5) NCSU use rights
 - a. "1" for use allowed
 - b. "0" for use not allowed
- 6) NCSU redistribution rights
 - a. "1" for redistribution allowed
 - b. "0" for redistribution not allowed
- 7) NCSU commercial use rights
 - a. "1" for commercial use allowed
 - b. "0" for commercial use not allowed
- 8) General Public rights positional indicator
 - a. "P"
- 9) General Public use rights
 - a. "1" for use allowed
 - b. "0" for use not allowed
- 10) General Public redistribution rights
 - a. "1" for redistribution allowed
 - b. "0" for redistribution not allowed
- 11) General Public commercial use rights
 - a. "1" for commercial use allowed
 - b. "0" for commercial use not allowed
- 12) Library of Congress rights positional indicator
 - a. "L"
- 13) Library of Congress use rights
 - a. "1" for use allowed
 - b. "0" for use not allowed
- 14) Library of Congress redistribution rights
 - a. "1" for redistribution allowed
 - b. "0" for redistribution not allowed
- 15) Library of Congress commercial use rights

- a. "1" for commercial use allowed
- b. "0" for commercial use not allowed

Coding considerations:

The user interface for the seed file form allows the ingest operator to:

- select method of acquisition
- indicate presence/absence of signed agreement
- indicate presence/absence of disclaimer
- select from among a handful of common rights configurations

Selecting for each of these four elements will result in formation of a complete 15-character rights code.

Common rights configurations:

1. N110P110L110 – NCSU, public, and LC can all use and redistribute but commercial use is not allowed.
 - Applicable in cases where data has been acquired without any signed or informal agreement restricting use, redistribution
2. N100P100L100 – NCSU, public, and LC can all use but redistribution and commercial use are not allowed.
 - Applicable in cases where data has been acquired with signed or informal agreement restricting redistribution
3. N100P000L000 – Only NCSU can use (*no other uses*)
 - Applicable in cases where data has been acquired with signed or informal agreement restricting use to NCSU (*redistribution to other agencies can be vetted on a case-by-case basis with the producer*)

Translating Codes Into Agreement Text:

Method of Access Flag:

- "M" – "Acquired directly from the producer in a mediated transaction."
- "U" – "Acquired directly from the producer in an unmediated transaction."
- "T" – "Acquired from a third party."

Text position: This is the last sentence in the rights statement.

Agreement Flag:

- "1" – "Please refer to the data sharing agreement." (*NOTE: informal data sharing agreement information is added as a text ancillary file*)
- "0" - <empty>

Text position: First sentence after rights code text.

Disclaimer Flag:

- "1" – "Please refer to the disclaimer."
- "0" - <empty>

Text position: Sentence after rights code text and agreement text.

Rights Codes - N:

- "N111" – "NCSU faculty, staff, and students may use, redistribute, and use for commercial purposes."

- “N110” - “NCSU faculty, staff, and students may use and redistribute, but commercial use is not allowed.”
- “N100” – “NCSU faculty, staff, and students may use, but redistribution and commercial use are not allowed.”
- “N000” – “No use of this data is allowed.”

Text position: First sentence.

Rights Codes - P:

- “P111” – “The general public may use, redistribute, and use for commercial purposes.”
- “P110” - “The general public may use and redistribute, but commercial use is not allowed.”
- “P100” – “The general public may use, but redistribution and commercial use are not allowed.”
- “P000” – <empty>

Text position: Second sentence.

Rights Codes - L:

- “L111” – “Library of Congress may use, redistribute, and use for commercial purposes.”
- “L110” - “Library of Congress may use and redistribute, but commercial use is not allowed.”
- “L100” – “Library of Congress may use, but redistribution and commercial use are not allowed.”
- “L000” – <empty>

Text position: Third sentence.

Coding-to-text example:

M01N110P110L110

“NCSU faculty, staff, and students may use and redistribute, but commercial use is not allowed. The general public may use and redistribute, but commercial use is not allowed. Library of Congress may use, but redistribution and commercial use are not allowed. Please refer to the disclaimer. Acquired directly from the producer in a mediated transaction.”

Appendix F: Local/State/Federal Data Sharing Committee

(Available:

http://www.ncgicc.com/Portals/3/documents/GICC_presentations_110707_FINAL.zip)

Recommendations for Geospatial Data Sharing

North Carolina Geographic Information Coordinating Council

A Report by the Local/State/Regional/Federal Data Sharing *ad hoc* Committee

Bill Holman, Chair

Revised, November 7, 2007

Background

The Local/State/Regional/Federal Data Sharing *ad hoc* Committee was created by the North Carolina Geographic Information Coordinating Council (GICC) to address issues brought forward by the Local Government Committee (LGC). At the August 16, 2006 meeting of the GICC the Chair of the Local Government Committee presented a report describing issues related to state government agencies' requests to local government for local government data. In the report (ATTACHMENT A), the LGC identified several issue areas, and 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...”

In that report, LGC directed the problem toward:

“...the lack of communication among state agencies...” but acknowledged that “...issues are complicated and the inconsistent policies at the local government level contribute to the problem.”

The LGC also acknowledged that:

“A mix of policy, process, and technology solutions will be required to solve the problem...” and suggested that “recommendations to address the issues should be consistent with the vision and characteristics of NC OneMap.”

The GICC referred the issue to the Chair of the Statewide Mapping Advisory Committee (SMAC) and on February 7, appointed the *ad hoc* committee to study the problem and develop specific recommendations that address the concerns of local, regional, state, and federal government agencies. Mr. Bill Holman, as Committee Chair, convened the *ad hoc* committee in four meetings from March through June. Recommendations of the Local/State/Regional/Federal Data Sharing *ad hoc* Committee are provided in this report.

Vision

The recommendations offered by the committee for the sharing of geospatial information are consistent with the GICC’s vision for *NC OneMap*. The committee acknowledges that each government sector invests significantly in geospatial data and each sector brings value to the statewide community. The recommendations: a) foster partnership development across all organizations and levels of government; b) avoid wasteful duplication of effort; c) optimize the use of technical infrastructure to address business needs for information exchange; and d) ensure effective and economical leveraging of geospatial resources for public benefit. The recommendations support the Council’s vision for *NC OneMap*, which include the following:

“The (NC OneMap) framework will promote the maintenance of economic vitality in our communities, public health and safety, and the quality of life for all North Carolinians. Our citizens will take the availability of comprehensive geographic information for granted.”

“...NC OneMap will serve the basic information requirements for decision-making in the community, statewide, and in support of national priorities. NC OneMap will provide information to support the daily business processes of numerous organizations and their functions. While any user may have a unique view of the resource and it ostensibly may be physically distributed and maintained by a variety of data producers, it will appear to users as consolidated and integrated.”

Ten Recommendations for Data Sharing

The committee identified ten data sharing recommendations for consideration by the Council. The recommendations, and associated best practices, should be publicized and used to encourage cooperation among all government agencies.¹³⁷

1. Avoid Formal Agreements

Written agreements that unnecessarily restrict the free exchange of geospatial data will be avoided. Exceptions could include circumstances that involve:

- a. Records that are protected by General Statute, such as those under the authority of the State Veterinarian; or are otherwise deemed confidential by appropriate authorities;
- b. Records that could pose a public safety or security risk, as written in law, or when appropriately restricted as part of a structured decision-making process guided by the “Guidelines for Providing Appropriate Access to Geospatial Data in Response to Security Concerns,” adopted by the Council (ATTACHMENT B).

In the rare exception, if a formal agreement is necessary the data *producer* is the appropriate authority to decide on the need for that agreement.

2. Web Access

Local, regional, state and federal geospatial content providers will make every effort to make data available to other local, regional, state, and federal entities through Internet technology, by uploading data or linking local services through *NC OneMap*. This will help meet a desired increase in efficient handling and distribution of geospatial data. Also, it will enable all users to access data and reduce the number of inquiries and data requests to staff in each individual organization.

3. Secure Access

In order to facilitate the distribution of certain data among government organizations additional services, including a secure site, may be necessary. Use of secure sites should not hamper or prevent the free sharing of data among public agencies. The road centerline data distribution tool, currently under development by the SMAC-Working Group for Roads and Transportation (WGRT), is an emerging example of secure ways to facilitate an organized approach to loading and distribution of data among public organizations.

4. Free Data

If local, regional, state, or federal data providers do not choose to make their data available on *NC OneMap*, it is recommended that providers supply data, including metadata, to other local, regional, state, and federal governmental organizations free of charge.

¹³⁷ Government agencies include federal, state, regional and local agencies and state-supported universities.

5. Single Point of Contact

Local, regional, state, or federal government organizations are to consolidate and identify point of contacts for acquiring and distributing data. This will streamline and alleviate the number of inquiries and requests to each organization. Contacts should be registered on the *NC OneMap* inventory and contact information should be publicized, regardless of whether or not the organization releases data as part of *NC OneMap*.

6. Regional Solutions

Regional approaches for data collection and data sharing through *NC OneMap* should be employed where beneficial and appropriate, typically when the local agency does not have the technical capability or available resources to distribute its own data.

7. Official Outlets

Because data are most current and accurate at the original data source, data will be acquired only from original sources, *NC OneMap*, or through an official outlet, named by the *producer* of that data. Consumers that request data from secondary sources shall be directed to primary outlets. Secondary sources of data shall not redistribute original data, except at the request or permission of the originator. This does not apply if secondary sources have added value in some way to the original content and provide disclaimers/metadata indicating that they are not the original source.

8. Archive and Long Term Access

Data producers should evaluate and publish their long term access, retention, and archival strategies for historic data.

9. NC OneMap Brand

The *NC OneMap* logo (Service Mark) should be displayed on web sites among collaborating partners. The *NC OneMap* ‘brand’ conveys to visitors that the agencies are working together in a collaborative network on mutual interests to meet the GICC goals and *NC OneMap* vision.

10. Outreach

Outreach and education on the recommendations above and the “Best Practices” below are vital components to the success and benefits of cooperative data sharing among government agencies. Appropriate material should be developed to communicate the recommendations and best practices. Suggested strategies for distribution of this information include:

- The NC Association of County Commissioners (NC ACC) and the NC League of Municipalities (NC LM) should formally endorse the final report, as approved by the GICC.

- NC ACC and NC LM should proactively promote the policy or recommendations that are adopted by the GICC, through newsletter articles, organization conferences and other means. It is important that county and municipal managers and elected officials be aware of the support by these organizations of the data sharing policies and recommendations.
- The GICC committees, specifically the Local Government Committee, the State Government GIS Users Committee, and the Federal Interagency Committee should proactively promote the policies and recommendations adopted by the GICC.
- State government departments should formally endorse or issue directives that staff will adhere to the policies and recommendations related to data sharing adopted by the GICC. It is true that many of the state government departments are represented on the GICC and the adoption of the final recommendations by the GICC will imply the tacit approval by these departments of the policies and recommendations in the report. However, without proactive support and promotion within all state departments, the policies and recommendations may not be adopted.

The Role of NC OneMap

Activities already underway to organize access to statewide geospatial data, such as *NC OneMap*, *NC OneMap* Inventory, and *NC OneMap* FTP Services, will be used as a data sharing framework. Additional services will be required to address “secure access” capabilities, such as the service currently under development for statewide road centerline sharing. *NC OneMap* is a collection of central and distributed services organized within a network of local, regional, and state agency stakeholders. *NC OneMap* is the ‘official’ statewide geospatial data clearinghouse. With full participation, users will be able to find information from across the state and be directed to appropriate on-line services and content providers.

Participating organizations are required to register in the *NC OneMap* Inventory and to create and maintain metadata that accompany transactions. The *NC OneMap* Inventory and metadata are important ‘best practices’ because those actions yield vital information about each organization and describe the availability of data to the statewide data sharing community.

Unless otherwise noted in statute or by agreement with the data producer, data that are part of *NC OneMap* are accessible in the public domain and can be freely redistributed. *NC OneMap* is an ‘official’ outlet for partner data.

No undue burden beyond the practice of sound data management principles will be placed on any one organization to participate in *NC OneMap* data sharing.

The Role of Center for Geographic Information and Analysis

The Center for Geographic Information and Analysis (CGIA) is the managing and coordinating agency in the state for data distribution as described by NC General Statute §143-725(b). Under the direction of the GICC, CGIA and its partners will seek to fulfill this role through development of the appropriate technical infrastructure and practices as part of the GICC's *NC OneMap* program. If additional resources are required for CGIA to serve in this role, the GICC, its members, and other stakeholders should seek and advocate for additional ways to fund and resource the program.

NC General Statute §143-725(b)

The Role of CGIA - The Center for Geographic Information and Analysis (CGIA) shall staff the Geographic Information Coordinating Council and its committees. CGIA shall manage and distribute digital geographic information about North Carolina maintained by numerous State and local government agencies. It shall operate a statewide data clearinghouse and provide Internet access to State geographic information. (2001–359, s. 1; 2004–129, s. 44.)

Scope of Initial Data Sharing Efforts

Initial efforts to implement the recommendations above will focus on “key” framework layers, such as geodetic control, orthoimagery, road centerlines, parcels, surface waters, county and municipal jurisdictional boundaries, parcels, and local zoning. In total, the initial efforts will include the 37 data layers identified by the GICC to be part of *NC OneMap* (ATTACHMENT C) and also include leaf-on imagery from the National Agricultural Imagery Program. Data that are aggregated from original sources for these data layers and contain added-value content are also acknowledged as part of initial efforts.

Core Best Practices

Members of the committee developed the following core practices to help data producers and content providers meet the intended goals for solving the issues with statewide data sharing. The list does not reflect all of what can be done, but represents a minimum set.

1. Data producers and content providers should register on the *NC OneMap* Inventory and complete agency profile and data content information. Maintain the information on a regular basis.
2. Write and maintain complete Federal Geographic Data Committee-compliant geospatial metadata for all datasets. Publish the metadata for discovery, and distribute it with the dataset.

3. Provide access to geospatial data for free via the Internet, through your agency's data download page, FTP site, and/or via *NC OneMap* FTP site (or combination of above).
 - ✓ Determine the data layers most needed/most requested; a minimum list should be the framework layers;
 - ✓ Periodically review the content available to see if layers should be added, removed, or replaced;
 - ✓ Periodically review the data formats provided and modify/update based on current industry standards; and
 - ✓ Include aggregated or value-added data in your data sharing process.
4. Make sure the location of the data for download and instructions for downloading are clear and posted prominently on your agency's web site. Provide "Help Desk" capability for the public and others accessing your data.
5. Publish your web services and catalog through *NC OneMap*. Connect or enhance your Web Map Services to the *NC OneMap* viewer by contacting the *NC OneMap* staff at CGIA. If your organization does not utilize web services, then contact the staff about using the services from a regional partner or from the *NC OneMap* servers.
6. Establish a policy and procedure for the provision of access to historic data, especially for framework data layers.
7. Provide alternate methods to share/use restricted data between key approved partners in preparation of emergency, security, and hazard events.

Value and Benefit of Data Sharing

The committee agrees that the benefits of a shared and organized approach to geospatial information technology investments are far greater in the aggregate than from the sum of each individual result. The benefits from sharing data among partners increase with participation. Five business cases were identified that exemplify how efficient and open sharing of geospatial data among organizations is of benefit and yields a collective return on investment. Each case is highlighted below. Additional information about each case is provided in ATTACHMENT D.

Case #1: At least \$130,000 will be saved annually upon implementation of an on-line statewide road centerline collection and distribution service. Producers and users will be able to access an on-line service to load or access state and local centerline datasets. Benefits are realized when all stakeholders participate.

Case #2: The value of cost avoidance and efficiencies in the sharing of surface waters data among stakeholders in a joint development project is over \$6,000,000. The benefits are accrued by numerous agencies, including Department of Transportation, Ecosystem

Enhancement Program, NC Wildlife Resources Commission, US Geological Survey, local governments, and the development community, among others.

Case #3: The Natural Heritage Program, Ecosystem Enhancement Program, Division of Water Quality Wetlands Unit, and Division of Forest Resources are just some of the organizations that benefit from the acquisition of summer “leaf-on” imagery via the National Agricultural Imagery Program (NAIP). The base product is made available for free by the United States Department of Agriculture with buy-up options offered to help meet specific requirements. Municipalities and counties could benefit from the NAIP imagery. As one example, the City of Salisbury could have applied the data in a program called CITYGreen (see ATTACHMENT D1) to determine the ‘value’ of externality costs derived from the reduction of pollutants by the amount of tree cover in a given area. Externality costs are calculated as indirect societal impacts, such as rising health care costs. Another CITYGreen example shows the ‘value’ of tree cover in the reduction of storm water volume which correlates to potential lower costs for storm water infrastructure. In these examples commercially available licensed imagery was used by American Forests for the work but this licensed imagery was not available to the City of Salisbury for further work following the completion of these examples. The use of NAIP imagery in these projects in place of licensed imagery would have provided the City of Salisbury the opportunity to continue the work beyond the completed American Forests program, utilizing the same base NAIP imagery for temporal and technical consistency.

Case #4: Benefits are realized when the US Department of Agriculture, Animal and Plant Health Inspection Service, Veterinary Services (USDA APHIS VS) is called in to aid in the response to animal disease outbreak. When USDA is activated, the Multi-Hazard Threat Database is critical to the response team because it is pre-loaded with shared data from state agencies, local governments, and from the various animal industries. Those data can be loaded to the USDA Emergency Management Response System, thus allowing quicker deployment of USDA surveillance teams in the field to support incident management. While no quantitative data has been found to specifically show dollars saved through quick and decisive response to animal disease outbreaks, it is assumed that the benefits are substantial given the level of commerce that exists in the state for animals and animal products.

Case #5: The NC Department of Agriculture and Consumer Services pre-loaded statewide parcel data in cooperation with local governments and as a result, FEMA was able to utilize the data to begin Hurricane Isabel recovery efforts in a timelier manner. Another case study conducted after Hurricane Isabel identified the benefits of having parcel data in place and coordinated statewide in advance of events, and having ‘core’ parcel data published on a regular basis. The report highlights five specific findings for sharing of parcel data for emergency response, including the savings of time to assessors and adjusters for purposes of insurance claims and federal disaster loans, among other activities. See ATTACHMENT D2.

The examples above are only a few of the numerous business cases that could be documented to demonstrate the benefits of data sharing. Other significant cases could

include the NC Floodplain Mapping Program in the development of local flood insurance rate maps and the Department of Commerce for industry recruitment tools, where the combined data resources of state and local government yield meaningful and powerful results to all participants.

Members of the Local/State/Regional/Federal Data Sharing *ad hoc* Committee

| | |
|------------------|---|
| Bill Holman | Committee Chair, Duke University |
| Mary Combs | U.S. Dept of Agriculture – Natural Resources Conservation Service |
| Jim Dolan | North Carolina Office of State Budget and Management |
| John Farley | North Carolina Department of Transportation |
| Tim Johnson | North Carolina Center for Geographic Information and Analysis |
| Chris Koltyk | Moore County |
| Steve Morris | North Carolina State University Libraries |
| Zsolt Nagy | North Carolina Center for Geographic Information and Analysis |
| Doug Newcomb | U.S. Fish and Wildlife Service |
| Anne Payne | Wake County |
| Jake Petrosky | City of Raleigh/Capital Area Metropolitan Planning Organization |
| Allan Sandoval | North Carolina Department of Commerce |
| Colleen Sharpe | City of Raleigh |
| John Spurrell | North Carolina League of Municipalities |
| Steve Strader | U.S. Geological Survey – National Geospatial Programs Office |
| Richard Taylor | North Carolina Wireless 911 Board |
| Rebecca Troutman | North Carolina Association of County Commissioners |
| David Lawrence | UNC School of Government, Invited Speaker (March 15, 2007) |
| DeWayne Branch | NCSU Graduate Student, Observer (March 15, 2007) |

Meetings of the Committee

The Committee held all of its meetings at the Albert Coates Local Government Center, Raleigh NC. The meetings were held on the following dates: March 15, 2007; April 19, 2007; May 17, 2007; June 21, 2007; and October 10, 2007.

Attachments

(Available: http://www.ncgicc.com/Portals/3/documents/GICC_presentations_080807.zip)

| | |
|---------------|---|
| 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 <i>Calculating the Value of Nature</i> |
| ATTACHMENT D2 | Parcel Data and Hurricane Isabel, A Case Study |

Appendix G: NCGDAP Outreach and Engagement Events

| Event Name | Location | Dates | Scope | Presentations |
|---|-------------------|-------------------------|---------------|--|
| All Hands Kickoff Meeting | Washington D.C. | Jan. 12-13, 2005 | National | NDIIPP Project: Collection and Preservation of At-Risk Digital Geospatial Data |
| 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 | Long Term Preservation of Geospatial Data (poster) & Long-Term Preservation of At-Risk Digital Geospatial Data: A Cooperative Agreement with Library of Congress |
| UCSB Workshop | Santa Barbara, CA | March 7-8, 2005 | National | Content and Practice: Background to the NC Geospatial Data Archiving Project |
| 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 | NDIIPP Project: North Carolina Geospatial Data Archiving Project |
| 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 | North Carolina Geospatial Data Archiving Project/NDIIPP: Collection and preservation of at-risk digital geospatial data |
| 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 | Collection and Preservation of At-Risk Digital Geospatial Data: the North Carolina NDIIPP Project |

| | | | | |
|---------------------------------------|------------------------|------------------|---------------|---|
| Partners Meeting | Warrenton, VA | July 11-13, 2005 | National | Key Issues and Findings in Work to Date |
| NDSAB Meeting | Washington D.C. | July 14, 2005 | National | North Carolina GDAP Project Overview |
| 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 | Map Portals and Geochiving: New Opportunities in Geospatial Information Service |
| NC ArcGIS Users Group | Wrightsville Beach, NC | Oct. 26-28, 2005 | State | Cooperative Project with Library of Congress on Preservation of Digital Geospatial Data |
| 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 | Long-term preservation of digital geospatial data: challenges for ensuring access and encouraging reuse |
| Charlotte/Mecklenburg GIS Day Exhibit | Charlotte, NC | Nov. 16, 2005 | State | GIS Day 2005 (poster) |
| CNI Task Force Panel | Phoenix, AZ | Dec. 5-6, 2005 | National | NDIIPP Project Briefing |
| NARA Technology Team Meeting | Washington, DC | Dec. 14, 2005 | National | Preservation of Digital Geospatial Data: Challenges and Opportunities |
| LC Brown Bag and G&M Visit | Washington, DC | Dec. 15, 2005 | National | Collection and Preservation of At-Risk Digital Geospatial Data: North Carolina Geospatial Data Archiving Project (NDIIPP Partnership) |
| Partners Meeting | Berkeley, CA | Jan. 9-11, 2006 | National | NCGDAP Update (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 | Collection and preservation of at-risk digital geospatial data |
| NSDI Partnership Office Webex Meeting | | March 1, 2006 | National | Collection and preservation of at-risk digital geospatial data |
| 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 | Workflow: Tools and Resources ; Collection and preservation of at-risk digital geospatial data ; Identification, Selection, and Appraisal within the North Carolina Geospatial Data Archiving Project (NCGDAP) ; Metadata Handling in the North Carolina Geospatial Data Project (NCGDAP) ; Collection Building Processes within the North Carolina Geospatial Data Archiving Project (NCGDAP) ; Preservation Strategies in the North Carolina Geospatial Data Archiving Project (NCGDAP) |
| GITA | Tampa, FL | April 23-26, 2006 | National | GITA 2006 poster presentation |
| 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 | NCGDAP Project Overview |
| NCPMA NC OneMap Outreach – Jacksonville | Jacksonville, NC | May 12, 2006 | Local | Project presentation and discussion |
| Archiving 2006 | Ottawa, CA | May 23-26, 2006 | International | Preservation of State and Local Government Digital Geospatial Data |
| JCDL | Chapel Hill, NC | June 11-15, 2006 | International | NDIIPP Preservation Network: Progress, Problems, and Promise |
| Where 2.0 | San Jose, CA | June 13-14, 2006 | International | The Disappearing Data Problem |
| July Partners Meeting | Washington D.C. | July 19-20, 2006 | National | Preserving North Carolina Legacy Geologic and Topographic Maps |

| | | | | |
|---|-------------------|------------------------|---------------|--|
| 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 | Spatial Data Infrastructure and Data Preservation in North Carolina |
| NC Property Mappers Assn. Fall Meeting | Raleigh, NC | Oct. 11-13, 2006 | State | GIS Data Preservation: Partnership with Library of Congress |
| DCC Workshop on Maintaining Long-term Access to Geospatial Data | Edinburgh, UK | Oct. 27, 2006 | International | Maintaining Long-Term Access to Geospatial Data |
| GRADE 2nd Project Meeting | Edinburgh, UK | Oct. 30, 2006 | International | |
| DLF 2006 Fall Forum | Boston, MA | Nov. 8, 2006 | National | Collection and Preservation of At-Risk Digital Geospatial Data: NDIIPP Project Update on the NC Geospatial Data Archiving Project |
| CRADLE | Chapel Hill, NC | Nov. 17, 2006 | State | Preserving Digital Geospatial Data: The NC Geospatial Data Archiving Project |
| 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 | NDIIPP Project Update: NC Geospatial Data Archiving Project (NGDAP) |
| NCGIS 2007 | Winston-Salem, NC | March 1-2, 2007 | State | Twenty Years of Spatial Vision, But What Does 1987 Look Like in Your GIS? Emerging Issues, Hindsight and Insights from the NC Preservation Partnership |
| 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 | Curation and Preservation of Complex Data: North Carolina Geospatial Data Archiving Project |
| ESRI Annual Users Conference | San Diego, CA | June 18-22, 2007 | International | Data Snapshot Archiving: A Frequency of Capture Survey |
| American Library Annual Meeting NDIIPP Symposium | Washington, DC | June 25, 2007 | National | Preserving State and Local Government Digital Geospatial Data |
| NDIIPP Partners Meeting | College Park, MD | June 26-27, 2007 | National | Preserved Digital Content: Collections, Value, and Stewardship |

| | | | | |
|---|--------------------|--------------------|---------------|--|
| OGC Technical Committee Meeting | Paris, France | July 9-12, 2007 | International | Co-chaired Working Group session |
| Cambridge Conference | Cambridge, UK | July 18, 2007 | International | Geospatial Data Preservation Challenges at the Sub-National Level: The North Carolina Experience |
| National Digital Strategic Advisory Board Meeting | Washington, DC | July 24, 2007 | National | Preserved Digital Content: Value to Public Policy Decision Making Now and in the Future |
| URISA Annual Meeting | Washington, DC | August 21-22, 2007 | International | Preserving State and Local Agency Digital Geospatial Data |
| 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 | Library of Congress Partnerships for Managing Geospatial Data |
| 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 | Archiving State and Local Agency Digital Geospatial Data: Overview of the Problem Area ; Archiving State and Local Agency Digital Geospatial Data: Looking for Solutions |
| NC GICC Archival and Long-Term Access Committee Kickoff | Raleigh, NC | Feb. 29, 2008 | State | Archiving State and Local Agency Digital Geospatial Data: An Overview of the Problem Area |
| NCCGIA/State Archives Metadata Meeting | Raleigh, NC | Mar. 4, 2008 | State | Metadata Handling in the North Carolina Geospatial Data Project (NCGDAP) |
| 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 |

Appendix H: Points of Engagement with Standards Efforts

| Topic | Possible Points of Intersection with OGC Working Groups | Issues |
|---|---|---|
| Geography Markup Language (GML) for Archiving | GML Working Group | <ul style="list-style-type: none"> Quality and functionality tradeoffs that would have to be made in favor of sustainability Learning from the PDF/A experience in terms of creating an archival profile for complex content and engaging software vendors in the process |
| Content Packaging and Metadata Wrappers | Metadata Working Group Geo RM Working Group | <ul style="list-style-type: none"> Bundling of data, metadata, ancillary components with data Learning from examples in other industries, such as METS, MPEG 21 DIDL, XFDU, and IMS-CP Studying FAO's Metadata Exchange Format as a lightweight wrapper currently in use |
| Data Versioning | Metadata Working Group Catalog Working Group | <ul style="list-style-type: none"> Support for versioned data (e.g. feature data subject to ongoing change) |
| Persistent Identifiers | Metadata Working Group | <ul style="list-style-type: none"> Durable or permanent links to resources, services, schemas, etc. Learning from solution in other industries: Handle, DOI, ARK, etc. |
| Archive Rights | GeoRM Working Group | <ul style="list-style-type: none"> Archival rights and static file rights use cases Addressing derivative data use cases |
| Temporal WMS | WMS Revision Working Group Mass Market Geo Working Group Decision Support Working Group Context Revision Working Group | <ul style="list-style-type: none"> Saving data state and not just service state in service interactions. Temporal component in Tiled WMS implementations |
| Graphic representation | <i>TBD</i> | <ul style="list-style-type: none"> Addressing standards issues for geospatial PDF content |
| Content Replication/ Distributed Storage | <i>In connection with the OGC/OGF (Open Grid Forum) Memorandum of Understanding for collaboration</i> | <ul style="list-style-type: none"> Large scale data replication or transfer Scenario: replication of content for a redundant WMS that is both geographically- and organizationally-remote from the original WMS |

Appendix I: Storage and Backup System

Primary Storage

Primary data storage is provided by two Nexsan ATABeast storage arrays. Each array contains 42 400GB Parallel ATA drives arranged in 4 Raid 5 stripes of 10 drives per stripe. RAID 5 was chosen for block-level striping and distributed parity to improve read speed and fault tolerance. The two arrays are mirrored externally by a Sun Enterprise 450 connected via 1GB Storage Area Network (SAN) switches. Observed read/write speed is ~70 MB/second.

Each ATABeast is partitioned into 1TB slices and are made available with several networking protocols including NFS, Samba, and SSH. The ATABeast storage is mounted via Network File System (NFS) over a 2GB fibre channel. Storage is then shared via Samba with mount points on the NDIIPP Solaris virtual zone.

Backup

The current backup schedule calls for monthly full backups and nightly incremental backups. Backup images are retained for 3 months. The current network hardware provides a 1GB network connection through the Libraries' Cisco 4600 1GB switch to the ITD (Campus Information Technology Division) Cisco 6500 switch.

Off-site backups are maintained at Poe Hall on the campus of NC State. The two-drive, C4-LT03 tape library is controlled by a Sun V210. The tape library holds 38 400 GB tapes and has a write speed of ~80MB/second. Analysis of data in the collection demonstrated that compression of data did not significantly decrease tape usage or increase time to completion due to the high volume of already-compressed images in the collection.

Appendix J: Repository Ingest Workflow

The repository ingest workflow, as planned, is described in charts on the subsequent pages. Major workflow components include:

- Receipt of data
- Threat scanning
- Formation of transfer sets (including metadata “seed file” and file manifest)
- Format scanning and migration
- Metadata preparation and remediation
- Ingest item migration
- Repository ingest

The charts on the next six pages outline the overall repository ingest process, including pre-processing steps. Charts include:

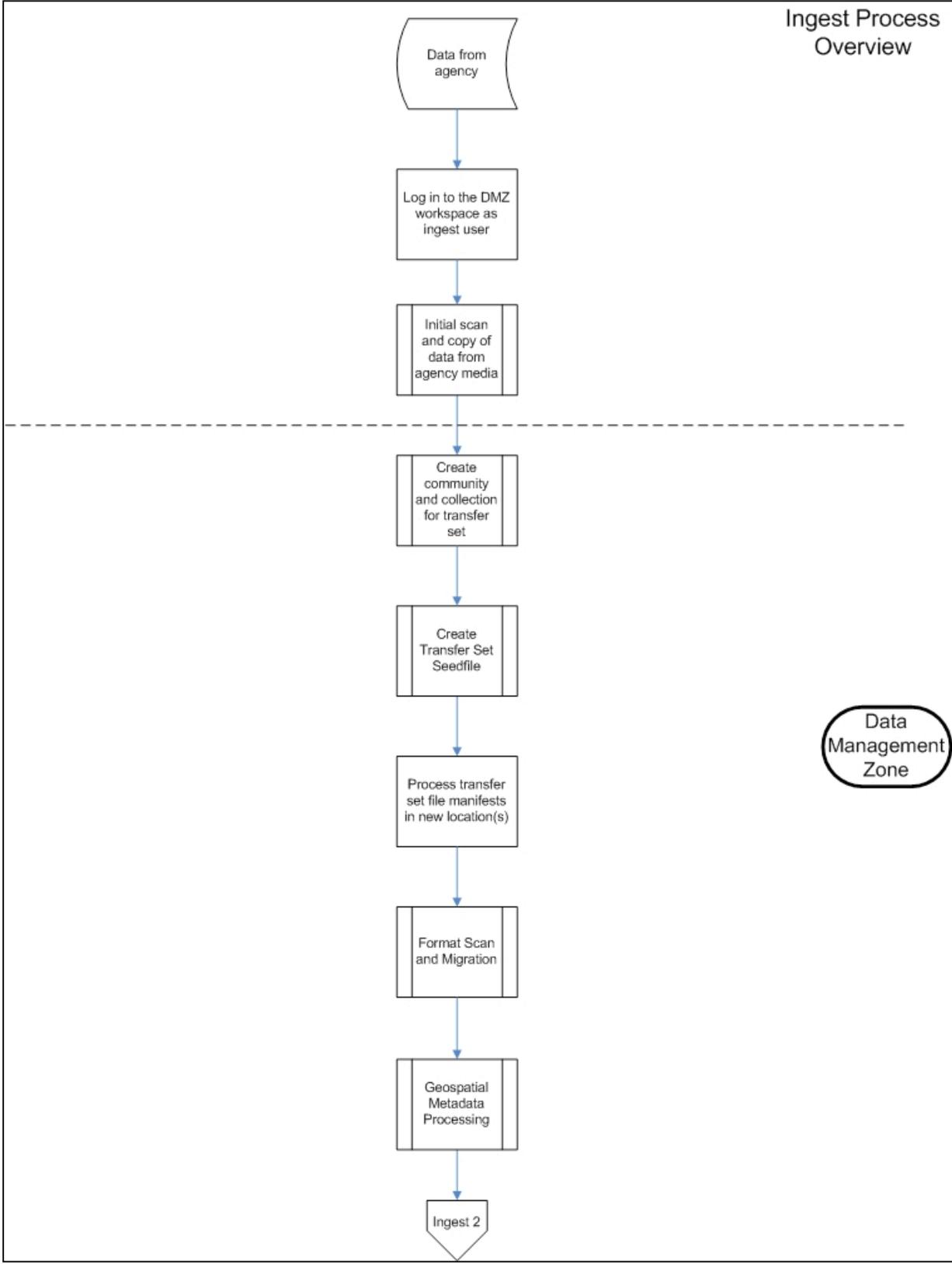
- Charts 1-2: Ingest Process Overview and Ingest Process Overview (Continued), which outline the overall repository ingest process.
- Chart 3: Format Scan, which illustrates the types of checks done for special format types.
- Chart 4: Geodatabase Handling, which outlines the specific processing steps to be taken with Geodatabases.
- Chart 5: Metadata Process Overview, which describes preparation of individual metadata records and metadata templates.
- Chart 6: Metadata Template Process, which outlines in detail the creation of metadata templates

Some components of the workflow have not yet been implemented, including

- METS record construction
- Synchronization of existing metadata
- Automation of metadata processes for vector data

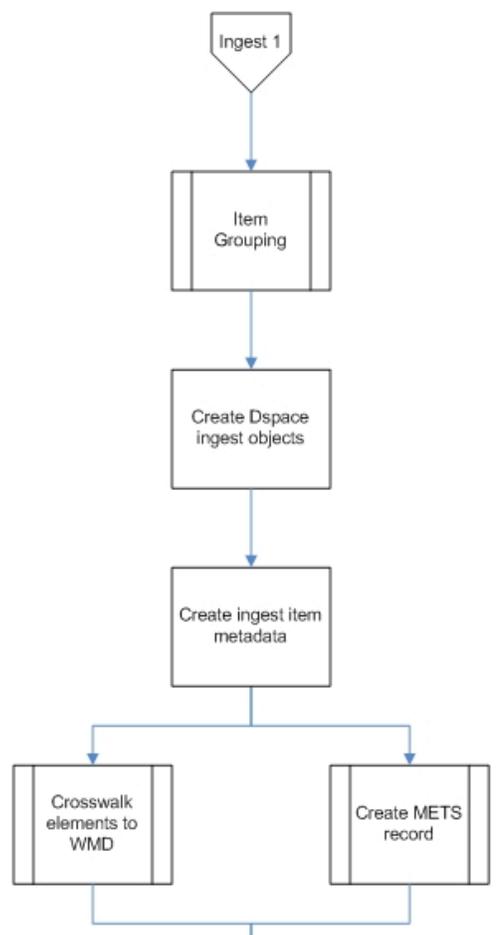
Due to processing costs, format migrations are only occurring selectively, with an emphasis on test cases that enhance the project learning experience.

Ingest Process Overview

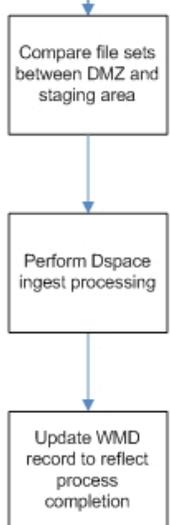


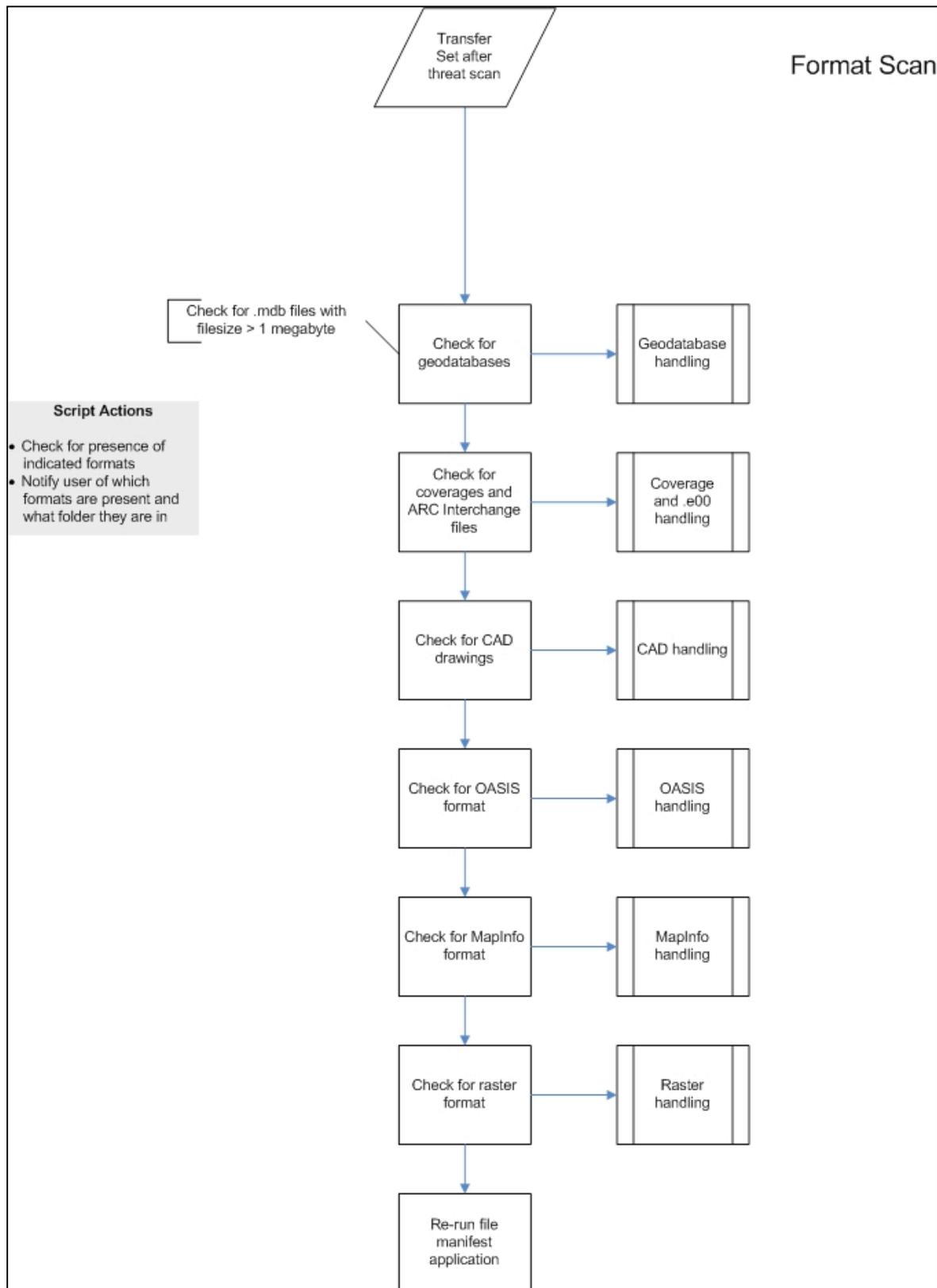
Data Management Zone

Ingest Process Overview (Continued)

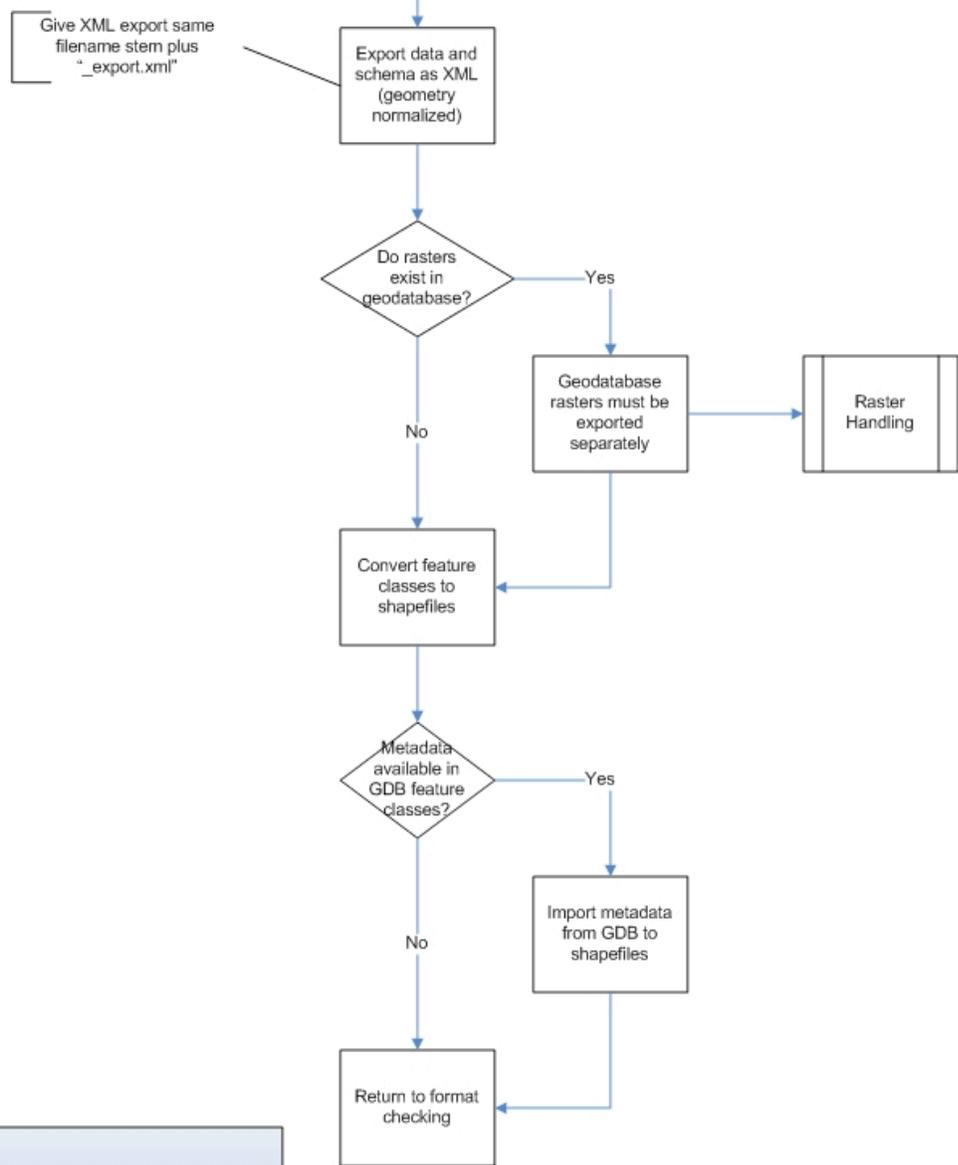


Ingest Staging (ATA Beasts)





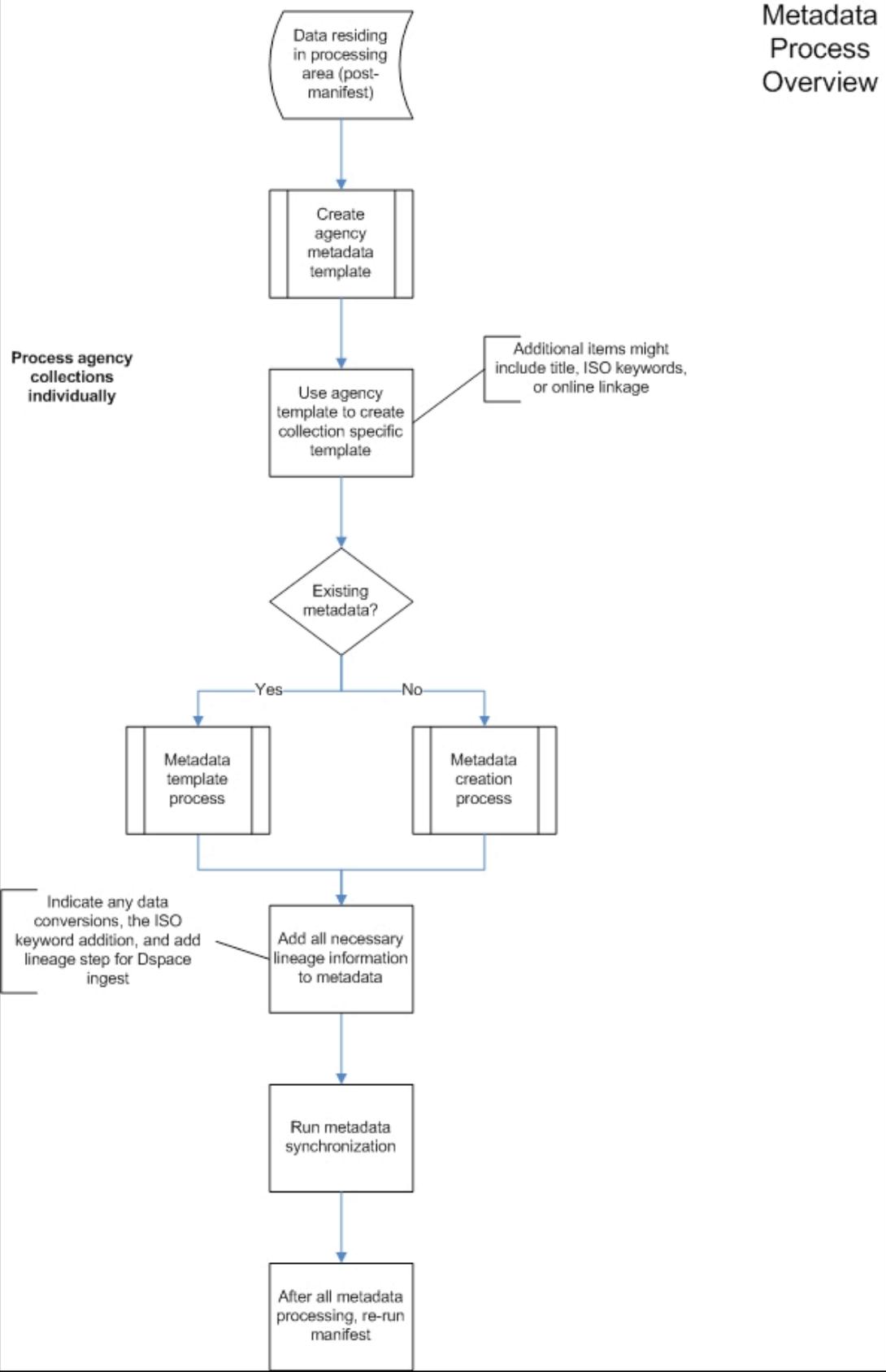
Geodatabase Handling



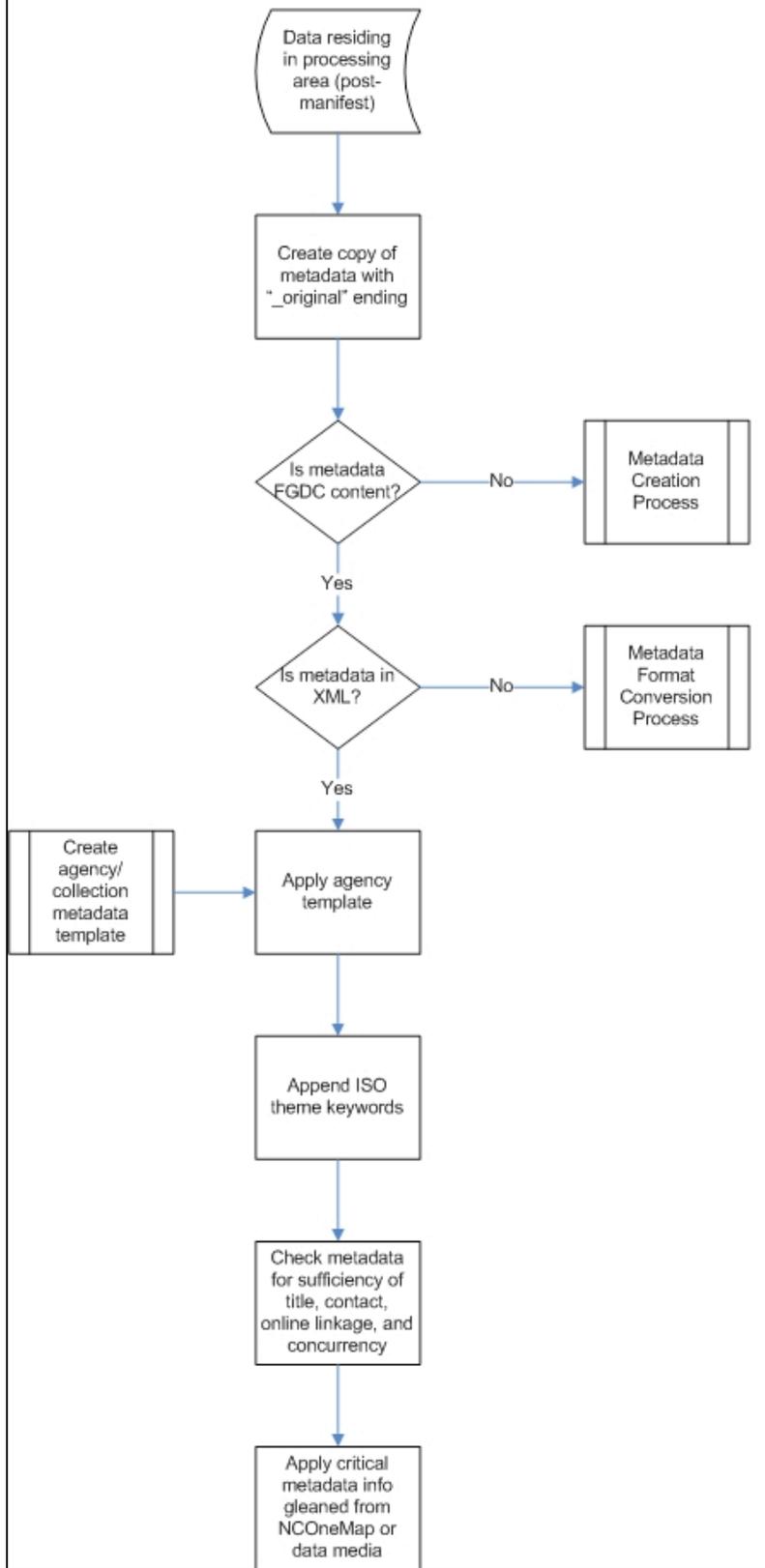
- Items Produced
- 1) Geodatabase paired with XML export and the ERDAS images linked to the XML schema.
 - 2) Individual TIFF images with world files
 - 3) Individual shapefiles with metadata and projection files (when provided)

Metadata Process Overview

Process agency collections individually



Metadata Template Process



Appendix K: Transfer Set Seed Files

The Transfer Set Seed File provides a means to capture descriptive, technical and administrative metadata which applies to entire received collections. Metadata elements related to manner of acquisition are included. The acquired metadata propagates to the individual items created during the ingest workflow. The seed file form and sample output are shown below.

Seedfile XML Form

| | |
|--|---|
| Community & Collection Wake County ▾ Orthophotos ▾ create | |
| transfer info | Received by JE as part of CGIAsneakernet. Return to CGIAsneakernet. |
| media label | 2005 Mecklenburg Orthos 2005 Catawba Orthos |
| notes | includes tiff and sid |
| date received (yyyy-mm-dd) | 2007 ▾ 10 ▾ 16 ▾ |
| method of access | mediated: <input type="radio"/> unmediated: <input type="radio"/> third party: <input checked="" type="radio"/> |
| signed agreement: | present: <input type="radio"/> absent: <input checked="" type="radio"/> |
| disclaimer: | present: <input type="radio"/> absent: <input checked="" type="radio"/> |
| NCSU faculty, staff, and students may ... | may use, redistribute, and use for commercial purposes. ▾ |
| The general public may ... | may use, redistribute, and use for commercial purposes. ▾ |
| Library of Congress may ... | may use, redistribute, and use for commercial purposes. ▾ |
| <input type="button" value="Create XML"/> | |

Sample Transfer Set Seed File output:

```
- <seed_file>
  <community_name>Wake County</community_name>
  <collection_id>3</collection_id>
  <collection_name>Orthophotos</collection_name>
  - <rights_verbose>
    NCSU faculty, staff, and students may use, redistribute, and use for commercial purposes. The general public may use, redistribute, and use for commercial purposes. Library of
    Congress may use, redistribute, and use for commercial purposes. Acquired from a third party.
  </rights_verbose>
  <rights_code>T00N111P111L111</rights_code>
  - <transfer_info>
    Received by JE as part of CGIA sneakernet. Return to CGIA.
  </transfer_info>
  <media_label>2005 Mecklenburg Orthos 2005 Catawba Orthos</media_label>
  <notes>includes tiff and sid</notes>
  <date_received>2007-10-16</date_received>
</seed_file>
```

Appendix L: Project Extension Work Plan

NDIIPP Project Extension Work Plan: North Carolina Geospatial Data Archiving Project

Lead Institution: North Carolina State University Libraries

Partner Institution: North Carolina Center for Geographic Information & Analysis

Period: Project extension October 2007-March 2009

Overview

The North Carolina Geospatial Data Archiving Project (NGDAP), one of the eight initial NDIIPP Preservation Partners projects, is focused on the preservation of state and local agency digital geospatial data. The project also focuses on the engagement of existing spatial data infrastructure in the challenge of preserving digitally-born content. In addition to the development of a demonstration data archive, NCGDAP has focused on outreach and engagement activities involving statewide coordination initiatives, government agencies, archives, commercial vendors, and standards organizations. This work plan outlines continued work in three key areas: 1) outreach and, engagement with federal, state, and local government agencies, relevant national associations, State Archives, and standards organizations; 2) content exchange tests with involving exchange of digital content with other NDIIPP partners; and 3) continued participation in NDIIPP network development.

Work Plan Elements by Phase

| Content Identification and Selection | |
|---|---------------------------------|
| Action | Timeline |
| <p>Deliverable:</p> <p>Task/Objective: Refine targeted list of state agency data resources to be acquired. (NCSU)</p> <p>Detail: Follow up on initial selective acquisition of state agency data with a prioritized list of new data acquisition targets. Target agencies include Dept. of Transportation, Dept. of Environment and Natural Resources and others. Specific data collections to be targeted include:</p> <ul style="list-style-type: none"> • NC OneMap Database: Miscellaneous statewide data resources • NC Dept. of Transportation GIS and Map Products: Road data and maps,; elevation data • NC Dept. of Environmental and Natural Resources: Geologic maps • NC Flood Mapping Information System: Selected LIDAR data and derivatives; flood maps | <p>Oct. 2007- Dec. 2007</p> |

| | |
|---|---------------------------------|
| <p>Task/Objective: Refresh the list of county and city data resources and map services and implement harvestable KML representations as well as GeoRSS feeds of this meta-information in order to support broader access to and reuse. (NCSU)</p> <p>Detail: In order to increase discoverability of county and municipal data resources and services NCSU has been experimenting with exposure through use of KML surrogates to be indexed for discovery in web indexes as well as specialized geospatial indexes. Initial development of GeoRSS feed of this content has also begun.</p> | <p>Oct. 2007- Dec. 2007</p> |
| <p>Task/Objective: Continue content development, registration of data providers, and reporting in the NC GIS Inventory, part of RAMONA. (CGIA)</p> <p>Detail: CGIA is continuing to promote participation of state, local, regional, and federal agencies in population of the NC GIS Inventory and reporting.</p> | <p>Ongoing</p> |
| <p>Task/Objective: Provide feedback on RAMONA inventory tool with regard to satisfaction of the needs of data preservation, including additional questions, functionality and reporting options (NCSU) (CGIA)</p> <p>Detail: NCGDAP feedback to the NC GIS Inventory development effort will highlight additional requirements for preservation efforts. A list of functions and customizations will be submitted to the RAMONA team. Items of interest include the number, the types of information RAMONA attempts to gather (more limited than the earlier NC One Map Inventory), the extent to which those information fields are completed, and flexibility of access to RAMONA information (e.g., integration of RAMONA data lookups in other data discovery applications).</p> | <p>Nov. 2007</p> |
| <p>Task/Objective: Conduct data analysis of external use of the county and city data and services directories, including KML surrogates and GeoRSS feeds. (NCSU)</p> <p>Detail: The nature of user interaction with resources in emergent data discovery environments is an unknown. Log analysis may be used to better characterize these interactions and guide commitments to and innovation in different methods of cultivating data discovery.</p> | <p>Nov. 2007- Jan. 2008</p> |

| | |
|--|---------------------------------|
| <p>Deliverable:</p> <p>Task/Objective: Formalize collection development strategy</p> <p>Detail: Reassess and formalize collection development strategy with regard to state and local data products, with special attention to the following emergent content forms:</p> <ul style="list-style-type: none"> • Cartographic product forms such as PDF map series and GeoPDF documents • Annotation and syndication forms such as KML and GeoRSS • Place-based data (not spatial) such as tax assessor building images, street view imagery, Dept. of Transportation videologs, and oblique imagery • Scanned and/or georectified map and aerial imagery derived from analog products | <p>Jan. 2008- Dec. 2008</p> |
| <p>Deliverable:</p> <p>Task/Objective: Finalize reports from Frequency of Capture Survey. (CGIA, NCSU)</p> <p>Detail: Reassess county and municipal practice with regard to creating snapshots of geospatial data layers. This is a follow-up survey to the original survey completed in Aug. 2006. Comparison of results with 2006 returns will be assessed Report materials will be posted to the public website.</p> | <p>Aug. 2008- Nov. 2008</p> |

| Content Acquisition | |
|---|-----------------|
| Action | Timeline |
| <p>Task/Objective: Acquire additional county and municipal data resources, including from agencies for which acquisition has already occurred. (NCSU, CGIA)</p> <p>Detail: Local agency data resources will continue to be acquired with an emphasis on acquisition in “low friction” situations (involving fewer negotiations, legal agreements, expenditures). “High friction” situations will be documented to inform the project learning experience, while acquisition efforts will defer to statewide efforts to increase the number of “low friction” situations through cultivation of open data sharing agreements. Off-line transfers will occur, however, on-line applications will be used, enhanced or created to facilitate network acquisition to data</p> | <p>Ongoing</p> |
| <p>Task/Objective: Acquire additional state agency data resources. (NCSU, CGIA)</p> <p>Detail: Targeted agencies include Dept. of Transportation, Dept. of Environment</p> | <p>Ongoing</p> |

| | |
|---|---|
| <p>and Natural Resources, and other agencies, as well as data made available through the NC OneMap data download system initiated in 2006. Off-line transfers will occur, however, on-line applications will be used, enhanced or created to facilitate network acquisition to data</p> | |
| <p>Task/Objective: Complete processing of existing NCSU holdings. (NCSU)</p> <p>Detail: Data acquired several years ago using a more primitive process continues to provide both challenges and learning experiences with regard to data validation and development of minimal metadata for undocumented sources.</p> | <p>Oct. 2007- Dec. 2007</p> |
| <p>Task/Objective: Acquire data through the new Centerline Data Distribution System (CDDS), currently under development by NCCGIA. (NCSU, CGIA)</p> <p>Detail: The Working Group for Roads and Transportation (WGRT) Centerline Data Distribution System (CDDS), developed with USDOT funding (and NCGDAP support) will provide a nexus for the collection, organization, and coordinated distribution of centerline data developed and maintained by public sector organizations in North Carolina. Major features of the CDDS include: upload of centerline data and metadata contributed by public sector agencies; organize and store uploaded centerline data and metadata; and facilitate download access of stored centerline data and metadata.</p> | <p>Ongoing</p> |
| <p>Task/Objective: Target additional university and NGO data resources to be acquired. (NCSU)</p> <p>Detail: Acquisition of university data resources has been highly selective, focusing on specific learning cases such as LIDAR data derivatives produced in research labs. <i>Acquisition of data in this area is expected to be limited.</i></p> | <p>Ongoing</p> |
| <p>Task/Objective: Revisit use of METS records in data ingest objects. (NCSU)</p> <p>Detail: Although an initial mapping to METS was completed for the project, actual implementation was deferred because of competing priorities and in response to a sense that developments in the community might lead to emergence of broadly applicable METS profiles that could be utilized by the project. <i>Actual implementation of METS may not be an outcome.</i></p> | <p>Jan. 2008- June 2008</p> |

| | |
|---|-----------------------------|
| <p>Task/Objective: Reassess mix of preservation metadata elements captured during the acquisition and ingest process. Explore use of PREMIS for encoding of preservation metadata elements. (NCSU)</p> <p>Detail: Initial efforts to define an approach to PREMIS implementation have been made, but actual implementation was deemed to be initially out of scope for cost reasons. In addition, PREMIS was seen to be in a state of flux from the point of best practices. Use of PREMIS will be reconsidered in light of evolving practice in the community, but <i>actual implementation of PREMIS may not be an outcome.</i></p> | <p>Jan. 2008- June 2008</p> |
|---|-----------------------------|

| Partnership Building | |
|---|--|
| Action | Timeline |
| <p>Deliverable:</p> <p>Task/Objective: Further develop and communicate business case (CGIA)</p> <p>Detail: Develop a concise business case for long-term retention and access of geospatial data by describing potential applications of retained data and the related benefits. Describe best practices to serve the business needs. Other activities may include development of promotional materials, participation in workshops, enhancement of web presence, and consultation with the GICC to develop a statement of direction for long-term retention and access of geospatial data. Work to proceed in context and support of NC OneMap business case and planning.</p> | <p>Ongoing</p> <p>Interim report June 2008</p> |
| <p>Deliverable:</p> <p>Task/Objective: Identify and/or develop model long-term access policy, short term best practices and guidelines for use by stakeholders in state and local government. (CGIA, State Archives)</p> <p>Detail: The GICC Ad Hoc Local/State/Federal Data Sharing Working Group is currently developing recommendations regarding data sharing practice at the local agency level. These recommendations are expected to include an encouragement to state and local agencies to develop an archival data management plan. Identify short term retention practices and guidelines that can be implemented for use by local and state agencies. Short term retention ensures at-risk data are available and saved while long term preservation strategies are being developed and implemented.</p> | <p>Oct. 2008</p> |

| | |
|---|------------------------------|
| <p>Task/Objective: Engage existing and new content providers in NC OneMap provider services, through training and technical assistance, including metadata development. (CGIA)</p> <p>Detail: Adding additional state and local partners to the network increases data access and increases opportunities to engage partners on preservation issues. In addition, through the metadata outreach, local agencies are encouraged and enabled to produce their own metadata.</p> | Ongoing |
| <p>Task/Objective: Attend and participate in appropriate outreach and planning, events and activities. (CGIA, NCSU)</p> <p>Detail: State, local, regional, and national level events provide an opportunity to market the preservation problem, disseminate findings, and engage both public and private partners in the preservation efforts. Such meetings include regional GIS user meetings as well as state- and national-level professional meetings.</p> | Ongoing |
| <p>Deliverable:</p> <p>Task/Objective: Inventory existing requirements, administrative rules, laws, policies that may impact plans for preservation of geospatial data and related records, (CGIA, State Archives)</p> <p>Detail: Work with user community and state archives to identify existing polices and rules that apply to geospatial preservation.</p> | Oct. 2007 – March 2008 |
| <p>Deliverable:</p> <p>Task/Objective: Conduct capacity and requirements assessment for state government (CGIA, State Archives)</p> <p>Detail: Assist in the determination of current capacity and potential requirements in state government for the provision of long term access, and preservation</p> | Dec 2007 – May 2008 |
| <p>Task/Objective: Establish a formal project team as point of collaboration with the State Division of Archives and State Library (CGIA, State Archives, NCSU)</p> <p>Detail: Members of State Archives have been participating in NCGDAP meetings and activities since early 2006. This new formal group will provide the framework for conducting an assessment of existing requirements and capacity to ingest digital records.</p> | Jan. 2008 |

| | |
|---|-------------------------|
| <p>Task/Objective: In collaboration with NARA, provide leadership for the Data Preservation Working Group within the Open Geospatial Consortium. (NCSU)</p> <p>Detail: This international venue provides an opportunity to engage software vendors, data vendors, consulting firms, and major national agencies on the issue of preservation in the context of standards development. In parallel, there may be opportunities to engage key open source organizations, in particular the Open Source Geospatial Foundation (OSGeo), on preservation issues that intersect with emerging specification and interoperability efforts in the open source community.</p> | Ongoing |
| <p>Task/Objective: In the geospatial and government agency community, engage relevant professional organizations such as the National States Geographic Information Council (NSGIC), Western Governor's Association, the National Association of Counties (NACO), and the National Association of Chief Information Officers (NASCIO), and potentially others. (CGIA, NCSU)</p> <p>Detail: NSGIC has already hosted two presentation sessions on the preservation topics and there is broad interest among states in this issue. NACO has entered into an agreement with LC regarding to retention of county map information. Outreach efforts will focus on generating awareness among boards, leadership and appropriate subcommittees.</p> | Ongoing |
| <p>Task/Objective: In the archives community, engage relevant professional organizations such as the National Association of Government Archives and Records (NAGARA) and the Council of State Archivists (CoSA). (NCSU, CGIA, State Archives)</p> <p>Detail: There is much activity in the archiving community in the area of data preservation but few projects which focus on geospatial information. Outreach efforts will focus on generating awareness among boards, leadership and appropriate subcommittees.</p> | Ongoing |
| <p>Task/Objective: Participate in Library of Congress report to Congress (NCSU, CGIA)</p> | Oct. 2007- Mar. 2009 |
| <p>Task/Objective: File quarterly reports about project status and activities. (NCSU)</p> | Quarterly |
| <p>Task/Objective: Complete interim project report covering initial 3-year project period and file with Library of Congress (NCSU)</p> | Oct. 2007 |

| Content Retention and Transfer | |
|--|-------------------------|
| Action | Timeline |
| <p>Task/Objective: Continue submission of data resources into the repository and refine ingest workflow for additional data configurations. (NCSU)</p> <p>Detail: Automated processes for repository ingest have been deployed but it is recognized that, given the variety of means by which data is transmitted and the variety of emergent content forms being encountered, that ingest mechanisms will need to continue to evolve. Data ingest will continue at a moderate rate, though it should be noted that the outreach and engagement components of the project will take priority over growth of the repository, the major function of which is to create the learning experience and to catalyze discussion in the geospatial community.</p> | Ongoing |
| <p>Task/Objective: Continue database administration of NC OneMap with refresh and archive of data layers from participating agencies and continue with functional enhancements of NC OneMap, including consideration of temporal content viewing and download. (CGIA)</p> <p>Detail: Ongoing development of the OneMap viewer will include components supporting temporal observation and services based on older versions of framework and imagery data. . Work will aim to support short term (10 year) retention of local and state content and will facilitate the development of a preservation process.</p> | Ongoing |
| <p>Task/Objective: Re-evaluate metadata ingest and authoring processes. (NCSU)</p> <p>Detail: Current processes focus on use of the FGDC version 2 content standard. The new FGDC standard, based on the North American Profile of the ISO 19115 content standard, using the ISO 19139 encoding standard, will be considered for use if tool support and standard adoption have become sufficient.</p> | Oct. 2008- Mar. 2009 |
| <p>Task/Objective: Review digital repository software performance and options. (NCSU)</p> <p>Detail: DSpace will continue to be used for the remainder of the project. Shortcomings and challenges presented by DSpace will be documented.</p> | July 2008- Mar. 2009 |
| <p>Task/Objective: Explore re-ingest of data into a different software environment (possibly FEDORA). (NCSU)</p> <p>Detail: FEDORA is known to offer functionality not offered by DSpace but at higher deployment costs. Initial investigations into the state of FEDORA technology and implementations <u>may</u> lead to a very limited implementation involving ingest tests.</p> | Oct. 2008- Mar. 2009 |

| | |
|---|------------------------------|
| <p>Task/Objective: Cultivate one or more partners with which to collaborate on content exchange tests. (NCSU)</p> <p>Detail: Initial discussions have been carried out with Lockheed Martin (in connection with the NARA ERA program), UCSB National Geospatial Data Archive (NDIIPP) about possible content exchanges. There is also interest in exploring use of generic METS profiles developed by the ECHO DEPOSITORY (NDIIPP) for repository exchange</p> | <p>Nov. 2007- June 2008</p> |
| <p>Task/Objective: Define role in tests, reformulate data objects as needed, and engage in content exchange tests. (NCSU)</p> <p>Detail: Formulate DIPs for repository exchange, participate in METS/PREMIS developments as appropriate and as resources permit, and contribute to repository exchange objects best practices development. Assess impact of ingest processes in remote systems on integrity of data and metadata and impacts on durability.</p> | <p>June 2008 – Mar. 2009</p> |
| <p>Task/Objective: Transfer selected content to Library of Congress. (NCSU)</p> <p>Detail: Under a technical process to be determined, transfer content to Library of Congress as needed and as allowed by any rights limitations that may be imposed on the data.</p> | <p>Oct. 2007- Mar. 2009</p> |
| <p>Task/Objective: Contribute to efforts within the NDIIPP network to cultivate a common set of repository ingest tools and frameworks. (NCSU)</p> <p>Detail: In the course of repository development work, content exchange tests, and participation in NDIIPP discussion exchange forums.</p> | <p>Ongoing</p> |

Appendix M: DSpace Qualified Dublin Core Mapping

DSpace Qualified Dublin Core metadata is utilized to describe the data while in DSpace. The QDC is supplemental to and derivative of descriptive, technical, and administrative metadata already captured in the FGDC metadata or elsewhere as part of the ingest workflow process. In order to support workflow automation, QDC elements have been defined so as to not require human intervention (i.e. only metadata elements that are auto-extractable are utilized). Some FGDC elements are complex and have to be aggregated in QDC elements.

| Destination | Source Schema | Source Element(s) | Notes |
|----------------------------|-----------------------|---|---|
| contributor_ author | FGDC/ literal | //metadata/idinfo/citation/citeinfo/origin | For NCSU created content: "NCSU Libraries, North Carolina Geospatial Data Archiving Project." |
| contributor_ others | FGDC | //metadata/dataqual/lineage/srcinfo/srccite/citeinfo/origin | Often not populated in FGDC. |
| coverage_ spatial | FGDC | //metadata/idinfo/spdom/bounding/westbc, //metadata/idinfo/spdom/bounding/eastbc, //metadata/idinfo/spdom/bounding/northbc, //metadata/idinfo/spdom/bounding/southbc | Coverage_spatial is a combined field containing all 4 bounding coordinates |
| date_issued | Seedfile | //seedfile/date_issued | Leave blank if pubdate field not populated. |
| coverage_ temporal | FGDC | //metadata/idinfo/citation/citeinfo/pubdate | FGDC field untrusted. Source from master spreadsheet. Never blank. May use date range.. |
| description_ none | FGDC | //metadata/idinfo/descript/purpose, //metadata/idinfo/descript/supplinf | Combined field with supplinf. |
| description_ abstract | Seedfile | //metadata/idinfo/descript/abstract, //seedfile/abstract | FGDC when available. Source from master spreadsheet when not present. |
| description_ provenance | FGDC/ literal | //metadata/dataqual/lineage/procstep/procdesc | NCGDAP statement appended to FGDC element. |
| publisher_ none | FGDC | //metadata/idinfo/citation/citeinfo/pubinfo/publish | Often not populated in FGDC. |
| rights_ none | Seedfile | //seedfile/rights_verbose, //seed_file/rights_code | Verbose description followed by code. |
| title_ none | Seedfile | //seedfile/title | Remediate in QDC. County/City/location, State,Year Content Type-Filename. |
| identifier_ other | 3 rd Party | NOID | Unique identifier |
| type_ none | string literal | Geospatial | |