North Carolina Geospatial Data Archiving Project

Frequency of Geospatial Data Capture

Survey conducted by

NC Center for Geographic Information & Analysis

Under a Partnership with

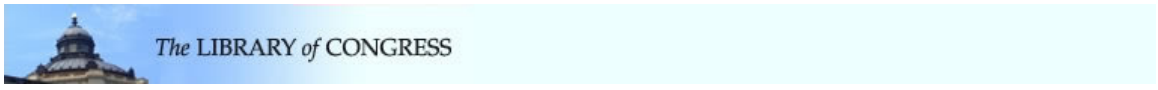
The Library of Congress

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	NAME	TYPE
Alexander	County	Laurinburg	Municipal
Ashe	County	Macon	County
Beaufort	County	Martin	County
Boone	Municipal	Mitchell	County
Burke	County	Nash	County
Burlington	Municipal	New Bern	Municipal
Carrboro	Municipal	Onslow	County
Caswell	County	Richmond	County
Charlotte	Municipal	Robeson	County
Dare	County	Salisbury	Municipal
Graham	County	Transylvania	County
Granville	County	Vance	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).

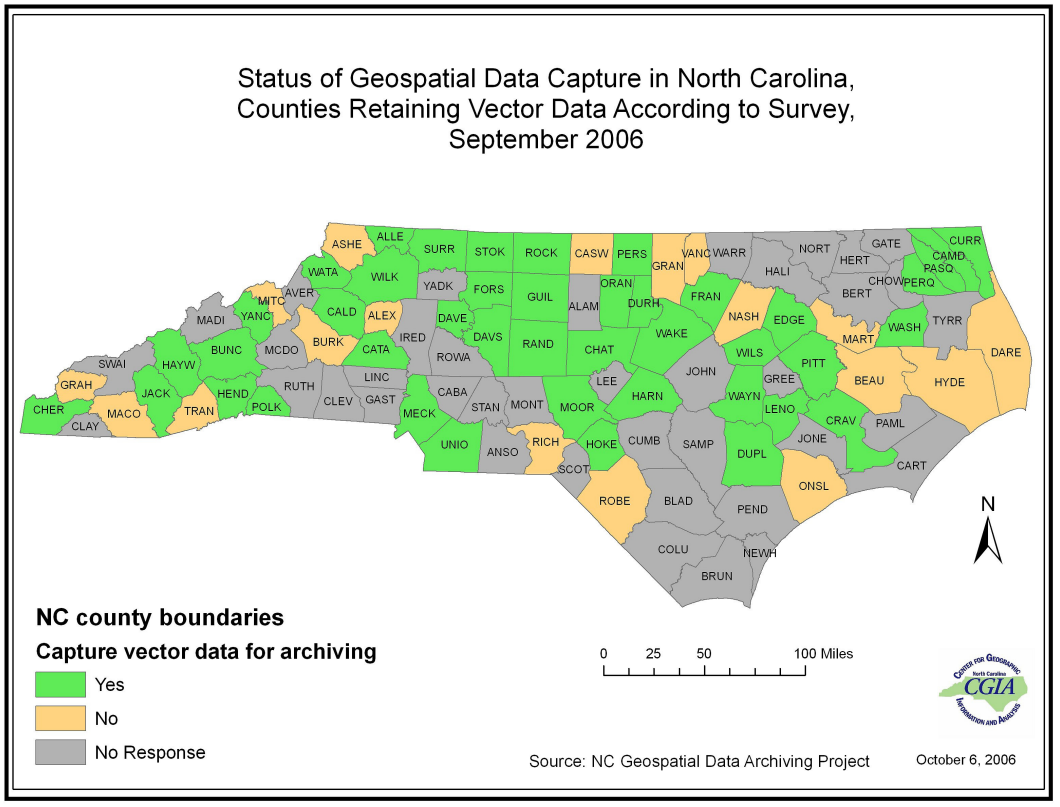
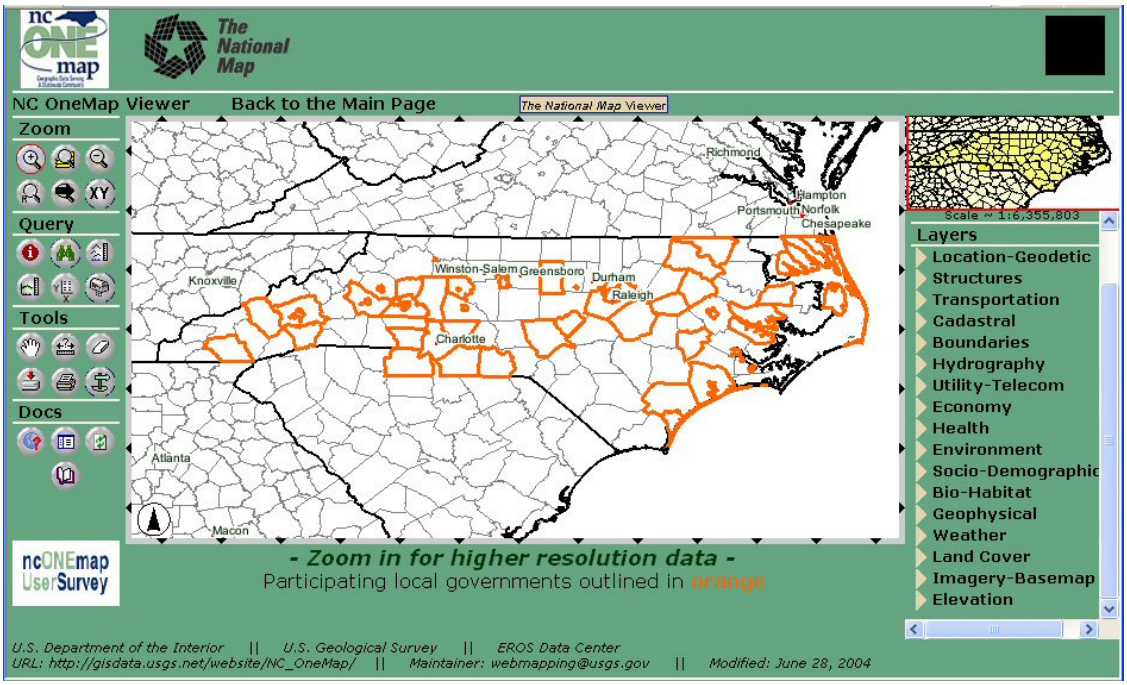


Figure 1: Data Capture Status by County



• 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.