



Geospatial Multistate Archive and Preservation Partnership

Moving from 'Keeping Stuff' to Geoarchiving: Developing Dynamic Geospatial Data Archives

Alec Bethune- NC Center for Geographic Information and Analysis
Kelly Eubank- North Carolina State Archives

North Carolina GIS Conference | 17 February 2011 | Raleigh, NC





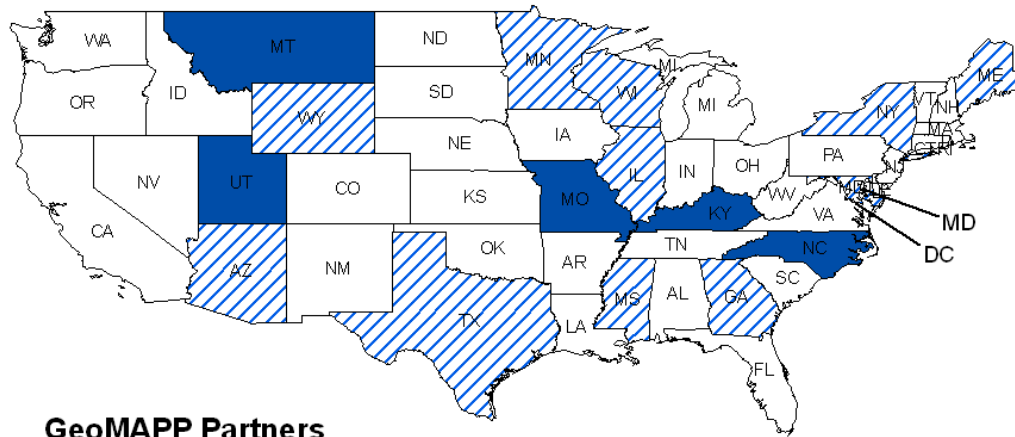
Project Overview

Alec Bethune

NC Center For Geographic Information and Analysis



Who is GeoMAPP?



GeoMAPP Partners

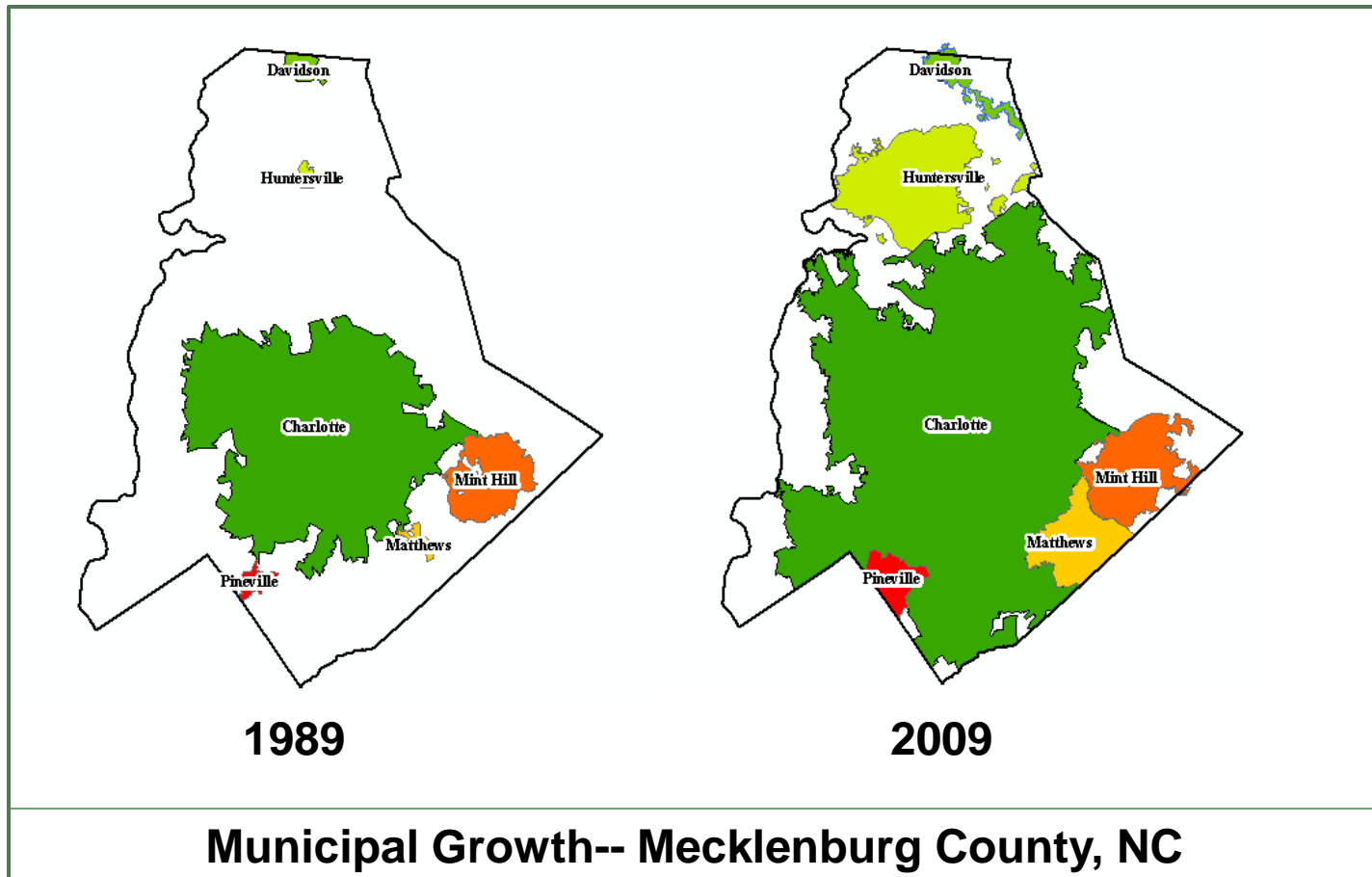


- ▶ **North Carolina** Center for Geographic Information and Analysis (CGIA)
 - ▶ North Carolina State Archives
 - ▶ NC State University Libraries
 - ▶ **Kentucky** Department of Libraries and Archives (KDLA)
 - ▶ Kentucky Division of Geographic Information (DGI)
 - ▶ Kentucky State University
 - ▶ **Utah** Automated Geographic Reference Center (AGRC)
 - ▶ Utah State Archives
 - ▶ **Missouri** IT Services Division
 - ▶ Missouri State Archives
 - ▶ Missouri Spatial Data Information Service
 - ▶ **Montana** State Library
- ▶ **12 Informational Partners:** AZ, DC, GA, IL, ME, MD, MS, MN, NY, TX, WI, WY

▶ www.geomapp.net

Why Preserve Geospatial Data?

- ★ *Modeling Change*
- ★ *Legal Mandate*
- ★ *Historic Mapping*
- ★ *Cost of Reproducing*
- ★ *Disaster Recovery/COOP*

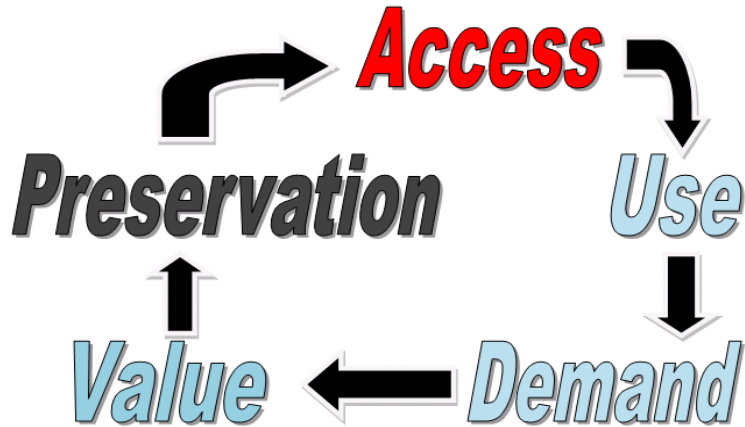


GeoMAPP Focus Areas

- Technical explorations:
 - file formats
 - metadata
 - data packaging
 - storage solutions
 - long term preservation techniques
- Improving access to archived data
- Business planning for sustainable archives
- Outreach: National and State GIS & Archives groups and with the data creator community




A Strong Focus on Access

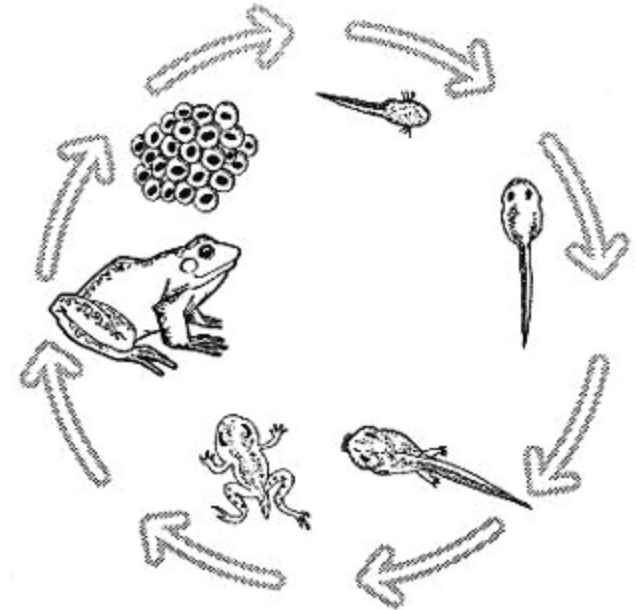


NOT



The Geoarchiving Process Lifecycle

- 
1. Establishing key relationships
 2. Inventory
 3. Appraise
 4. Data Preparation
 5. Transfer
 6. Ingest
 7. Preservation
 8. Access
 9. Business planning for sustainability



<http://www.ipcc.ie/lifecycle.html>

GeoMAPP Deliverables

Also available or coming soon at:

<http://www.geomapp.net/documents.htm>

- ▶ Interim Report: (GeoMAPP 2007-2010)
- ▶ Geoarchiving self-assessment
- ▶ Business Case Documentation
- ▶ Utilizing Geospatial Metadata in Preservation Practices
- ▶ Geodatabase Investigation
- ▶ Data Transfer Documentation
- ▶ State record retention schedules for geospatial data
- ▶ FGDC and Dublin Core metadata comparison
- ▶ NSGIC/CoSA/NAGARA/State/Local government Geo-Archival Practices Surveys
- ▶ **...and much much more.**





Project Progress/ What We've Learned

Kelly Eubank
North Carolina State Archives



We're doing backups, isn't that archiving?

- ▶ Backups – a means to save and recover current records
- ▶ Personal Archives- “Keeping stuff” on external media or on hard drives or SAN
- ▶ True Archiving – formally preserving important data **permanently** in a trusted digital repository



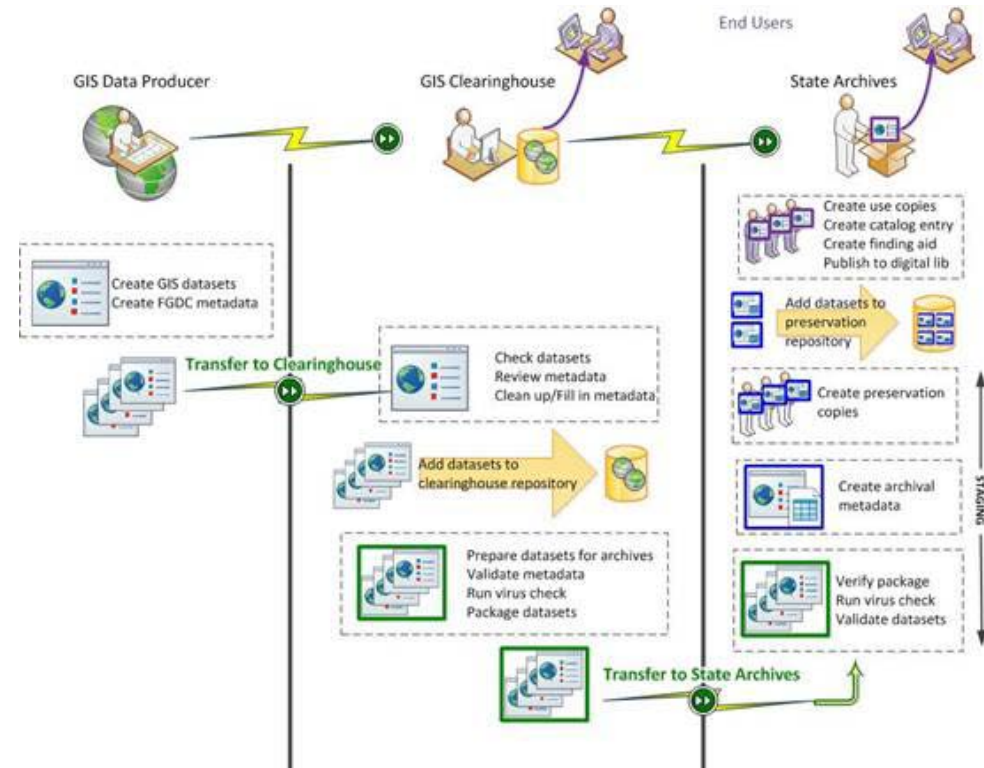
Statewide Geoarchiving Awareness :

NC Geographic Information Coordination Council

- ▶ Archival/ Long-term Access AdHoc Committee proposed:
 - ▶ “Develop a plan and implement steps at the Archives and Records Section to receive geospatial content as prescribed in emerging schedules.”
 - ▶ “...leveraging of the NC OneMap clearinghouse and data work flow as a content transfer point between data stewards and the NC State Archives.”
 - ▶ Team has also investigated direct transfers to the Archives
 - ▶ “...leveraging the recommendations from the Ten Recommendations for Geospatial Data Sharing.” (GICC 11/07/07)

Geoarchiving In NC

- ▶ Data Transferred to Date
 - ▶ CGIA current and historic datasets
 - ▶ Local Government Data
 - ▶ State Interstate Data Transfer Activities
- ▶ Current SAN Capacity
 - ▶ Assistance from ARM
 - ▶ Recording Fee



Finding Archived Data

- ▶ MARS Catalog (<http://mars.archives.ncdcr.gov>)
 - ▶ Database of archival holdings\ Dataset request to DCR

The screenshot shows the MARS website interface. On the left, there is a search results table with columns for Mars Id, Class, and Title. The first row is highlighted. Below the table are navigation arrows and the text 'North Carolina State Archives'. On the right, a 'Details for - GIS Data' window is open, displaying an abstract and a contents section. The abstract describes the GIS Data collection spanning 1947-2009, created by the GeoMAPP project. The contents section mentions 'demonstration' data and provides a link to 'View Full Scope / Contents'.

- ▶ NC Digital Collections (<http://digital.ncdcr.gov>)
 - ▶ Limited datasets available for online download

The screenshot shows the North Carolina Digital Collections website. The header includes navigation links: Home, Advanced Search, All Collections, Other Resources, Browse By..., and Preferences. Below the header, the search results for 'Carolina' are displayed. On the left, there is a 'Refine your search' sidebar with 'Subjects' listed: Geographic Information Systems -- North Carolina (20), North Carolina -- Maps (20), Boundaries (6), Transportation (5), Structures (4), and Show more... On the right, the search results are shown as a table with columns: Image, Title, Date Published, and Subjects. The first result is '1992 NC Congressional Districts', published in July 1998, with subjects: Geographic Information Systems--North Carolina; North Carolina--Maps; Boundaries.

Outreach

- ▶ Data Transfer demonstrations
- ▶ Sharing findings within NC
 - ▶ Local Government
 - ▶ State Agencies
 - ▶ GICC
 - ▶ ArcUsers Group
- ▶ National Discussions
 - ▶ ESRI
 - ▶ SAA/ CoSA/NAGARA
 - ▶ Library of Congress
 - ▶ National Archives (NARA)
 - ▶ FGDC



Making your data “preservation ready”
for the future and more useful now



Lots of Copies Keeps Your Data Safe



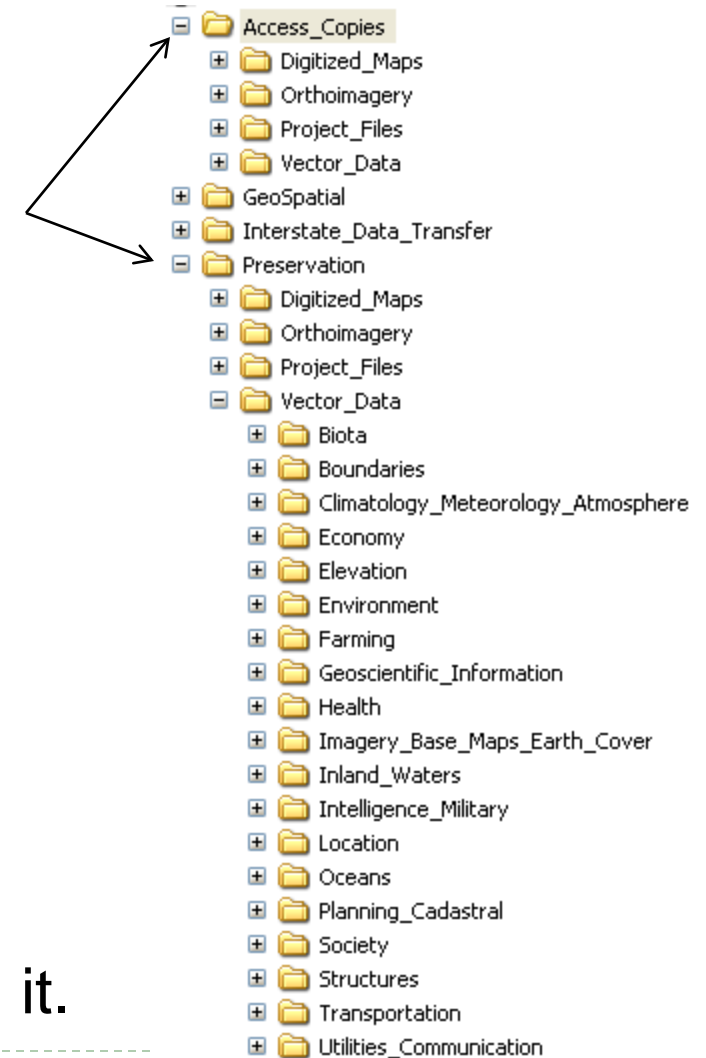
magento.siteground.com



barry.wordpress.com



krunker.com



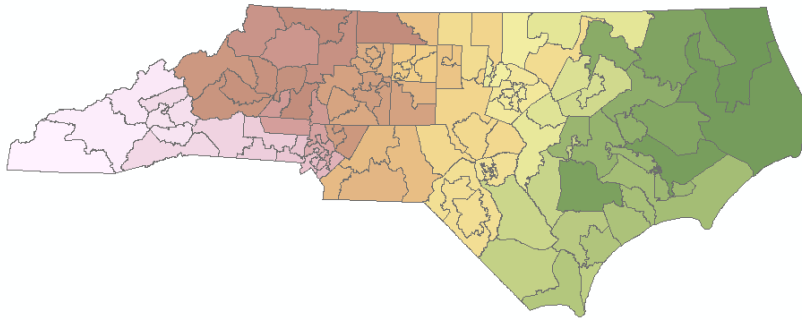
...but don't forget where you stashed it.

Why worry???



Dataset Names without Shame

dnchou



NC House Districts 2003

3036_01

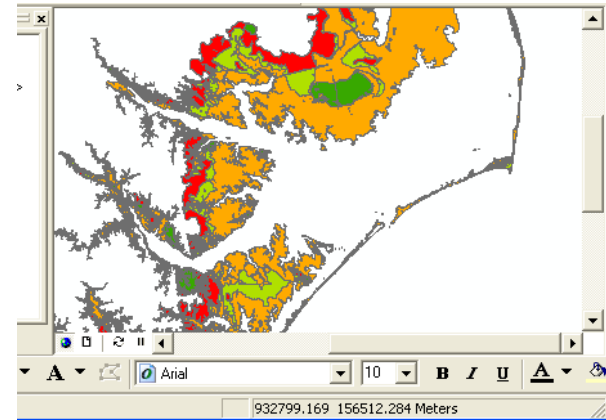


OR

OC6i0 37 000 30053604 20100411R0

Don't Duck the Metadata!

- Dual value for data management/ usability and preservation
- GeoMAPP Products:
 - FGDC / Dublin Core Crosswalk
 - Preservation and Administrative Elements for the Archives
 - GeoMAPP Metadata Whitepaper
 - Background on Metadata research
 - Quick metadata key element reference guide



```
<pubdate>19990903</pubdate>
<title Sync="TRUE">hss93f</title>
<geoform Sync="TRUE">vector digital data</geoform>
- <pubinfo>
  <pubplace>Coral Gables, Florida</pubplace>
  <publish>National Hurricane Center</publish>
</pubinfo>
<othercit>NCCGIA distributes this dataset</othercit>
<onlink />
<ftname Sync="TRUE">hss93f</ftname>
</citeinfo>
</citation>
<descript>
<abstract>The National Hurricane Center, in cooperation
with the North Carolina Center for Geographic
Information and Analysis, developed the GIS data set,
Hurricane Storm Surge Inundation Areas (1993), to
```

Metadata Makes it Last Longer

Important CSDGM Fields for Preservation:

	Friendly Field Name	XML Tag	Description / Example Value
1. Identification Information Section		<idinfo>	
1.1 Citation		<citation>	
1.1.8 Citation Information		<citeinfo>	
	Originators	<origin>	The party responsible for the dataset. This is often the dataset creator except in cases where the dataset's creation was contracted out to a third party, but is 'owned' and maintained by another party.
	Publication Date	<pubdate>	The date the dataset was completed and was made ready for use.
	Title	<title>	Title of dataset, ideally including 'where', 'what', 'when' <i>e.g. North Carolina Shellfish Growing Areas 2010</i>
*	Geospatial Data Presentation Form	<geoform>	Describes type and format of the dataset <i>e.g. vector digital data</i>

➤ Key FGDC CSDGM fields for preservation

➤ Citation Info, Abstract, Purpose, Time Period, Theme & Place
Keywords, Access/Use, Contact Info, Source/ Process Step Info,
Projection/ Coordinate Info

Format Matters

- Sustainability factors
 - Is the data format open or proprietary?
 - Vendor support lifecycle and documentation
 - Data packaging
- Selecting the right format for long term use/preservation
- GeoMAPP Format Efforts
 - Deliverable: Data Formats Index

Vector Formats	File extensions	Description	Link to Format Standard/ Specification	Format Currency	Format Prevalence	Tools for Viewing Format	Convertible with FME	Sustainability Issues
ESRI ArcInfo Interchange File (E00)	*.e00	ArcInfo Interchange File Format is a vector file format which is produced by ESRI and predates the shapefile. It is as an export file used to enable coverage, grid or TIN and an associated INFO table to be transferred between different machines, which are not connected by any type of file sharing network. A volume has the extension .E00 through .E99.	http://avce00.maptools.org/docs/v7_e00_cover.html	Legacy	Infrequent use	Not a viewable format. Exported files (*.e00) are imported to ARC/INFO coverage format through ArcGIS Toolbox or command line ARC/INFO.	Yes	Preservation Strategy: There are interior tables within the coverage and the *.e00 files. If these files are not worked with in the Arc environment the data will be corrupted and unusable. Any *.e00 files or coverages should be converted to either feature classes or shapefiles.

Where We Go From Here

- ▶ NC OneMap records retention schedule
- ▶ Continued outreach with local and state government
 - ▶ NC Commerce/ DOT
 - ▶ City of Charlotte
- ▶ Investigating preservation and access storage options for geospatial data
- ▶ Investigation of file transformations
- ▶ Preservation of geodatabases?
- ▶ Providing greater access to archived content



Questions?



Thanks!

▶ Alec Bethune (CGIA)

alec.bethune@nc.gov

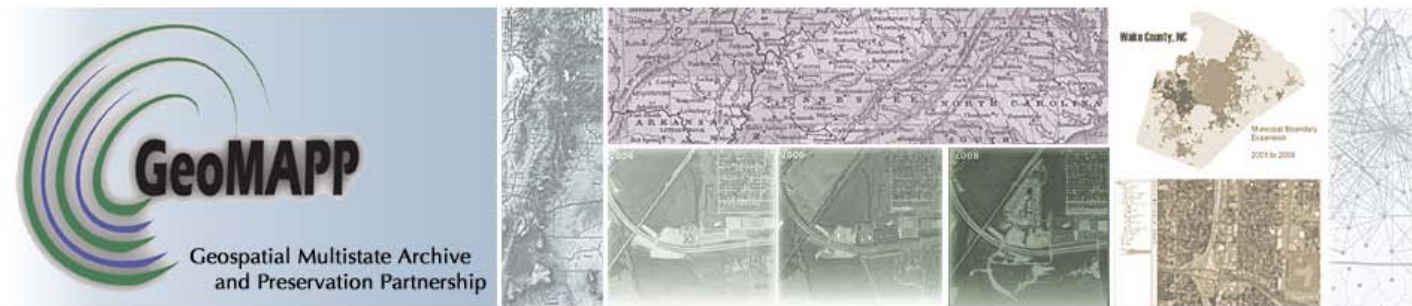
▶ Kelly Eubank (NC Archives)

kelly.eubank@ncdcr.gov



<http://imagecache2.allposters.com/>

Follow us on Facebook and Twitter!



▶ www.geomapp.net