



# The Disappearing Data Problem:

**Preserving Today's Geospatial Data to  
Meet Tomorrow's Temporal Analysis  
Needs**

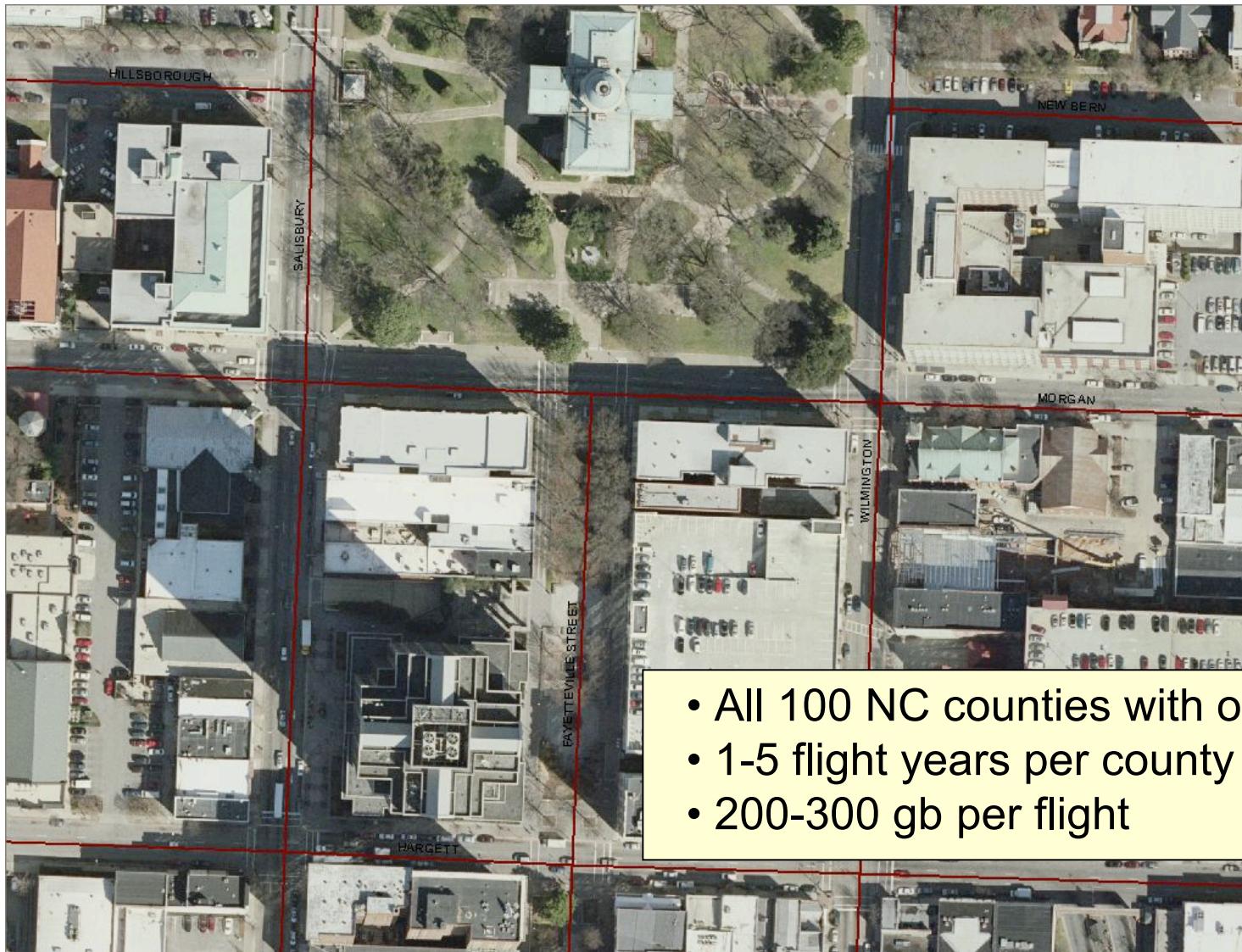
**Steve Morris**  
**Head of Digital Library Initiatives**  
*North Carolina State University Libraries*



## ***Outline***

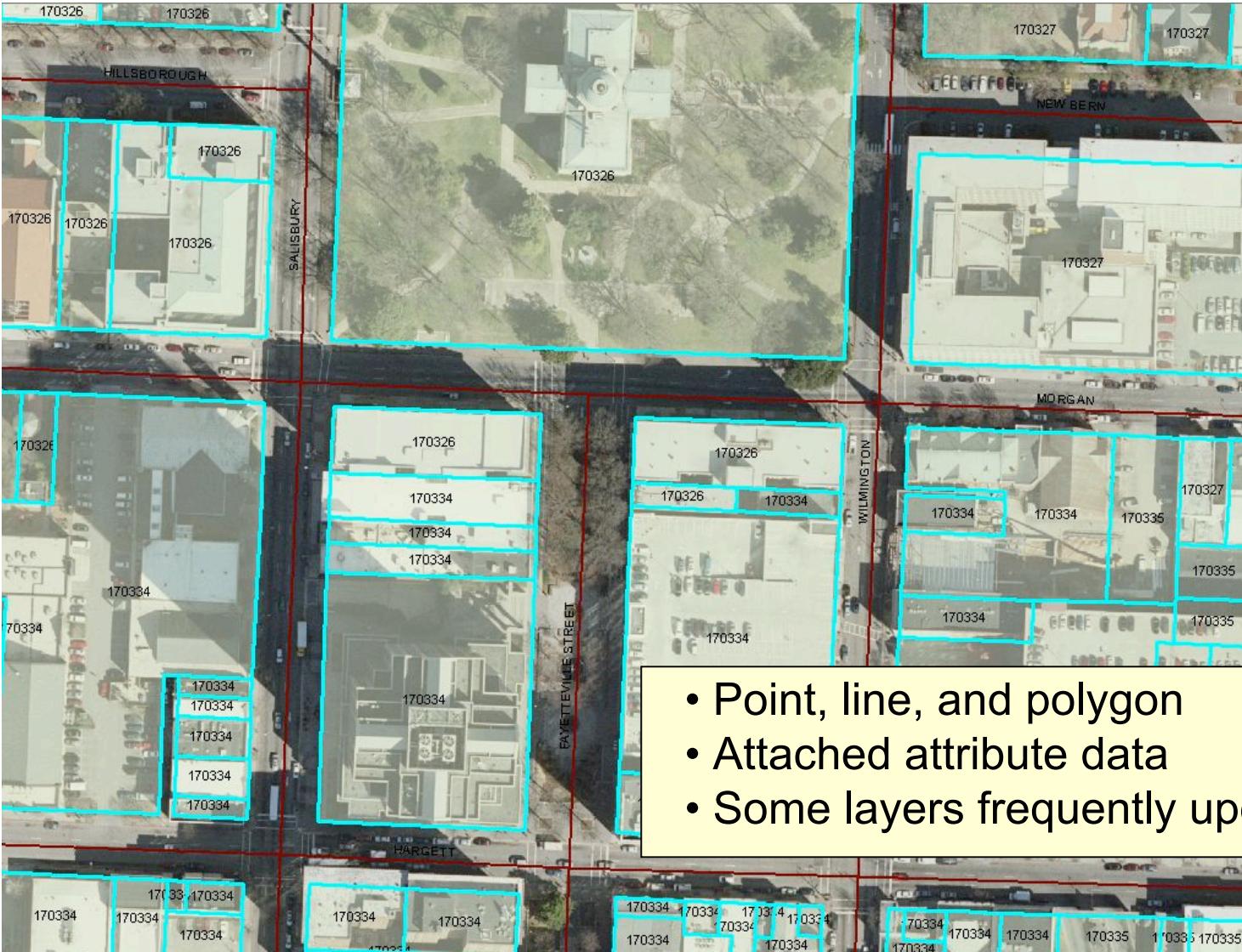
- Background to the geospatial content domain
- Overview of the NDIIIPP project
- Preservation challenges and solutions (?)
- Changes in the content domain
- Moving forward: New initiatives

## *Geospatial Data Types – Digital Orthophotography*



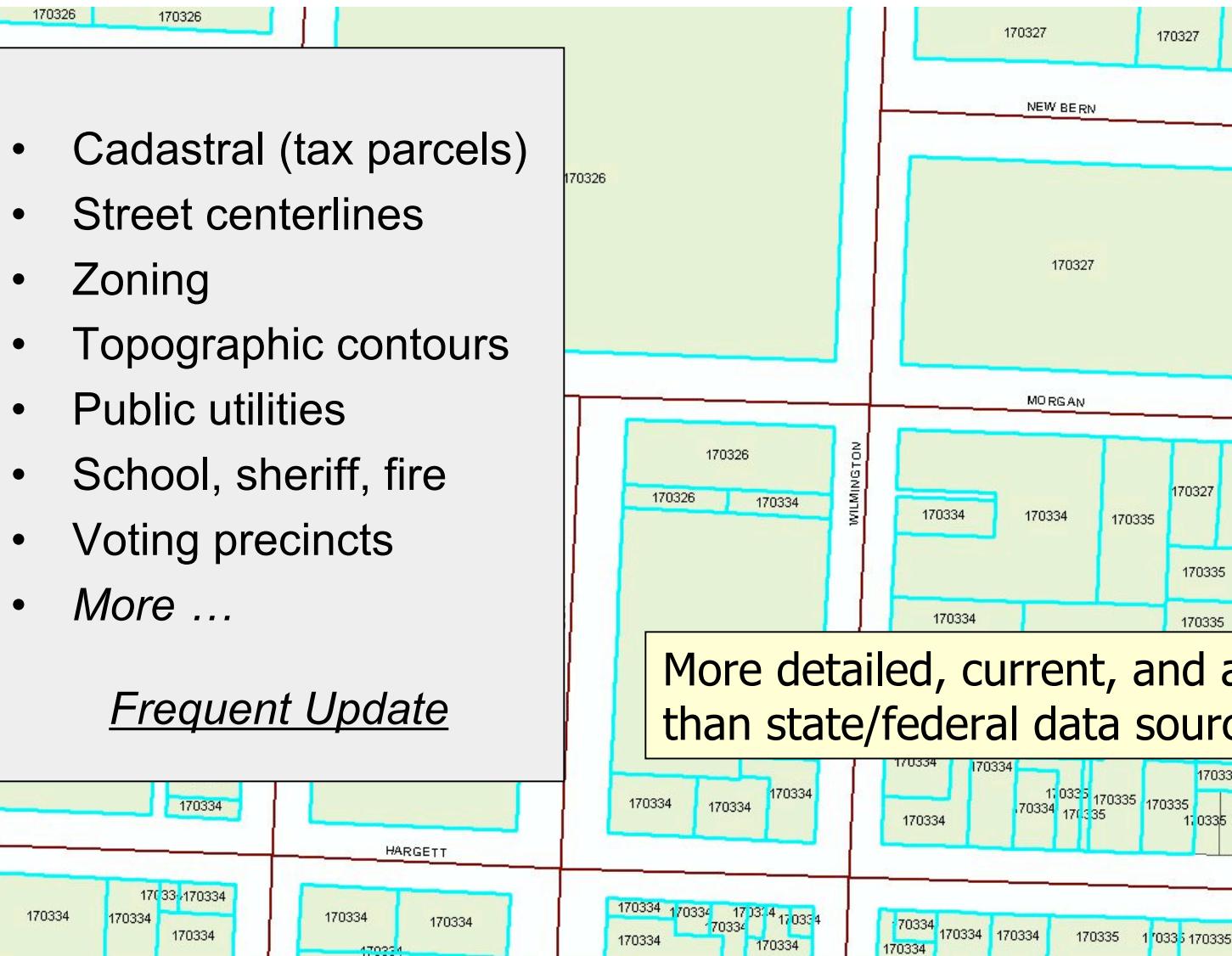
- All 100 NC counties with orthos
- 1-5 flight years per county
- 200-300 gb per flight

# Geospatial Data Types – Vector Data



- Point, line, and polygon
- Attached attribute data
- Some layers frequently updated

# Geospatial Data Types – Vector Data





## ***Background: GIS at NCSU Libraries***

- GIS services program since 1992
- Focus on campus-wide infrastructure, not a lab
  - Data, software, support, evangelism
- Roughly 35 academic departments with GIS activity
- History of close collaboration with state agencies
- Heavy reliance on state/local agency geospatial data
- Data discovery tool development
- *Problem: Access to temporal data*

# ***Example: County and City GIS Data Directories***

Tracking data, map servers, and web services since 2000

Ranked 3<sup>rd</sup> in traffic among entry points to library website

Persistent identifiers

- usage tracking
- IDs used in other sites

Community help in site maintenance

County	Libraries Webpage	GIS Webpage	Online Mapping	Public Downld	NC OneMap	IMS File	WMS Link
Alamance							
Burlington	Lib Web	GIS Web	Web Map	Dwnld	OneMap	.ims File	WMS Link
Graham							
Alexander	Lib Web	GIS Web	Web Map		OneMap	.ims File	WMS Link
Alleghany	Lib Web	GIS Web	Web Map				
Anson	Lib Web	GIS Web	Web Map	Dwnld	OneMap	.ims File	WMS Link
Ashe	Lib Web	GIS Web	Web Map				
Avery	Lib Web	GIS Web	Web Map				
Beaufort	Lib Web		Web Map	Dwnld	OneMap		WMS Link
Bertie	Lib Web	GIS Web	Web Map	Dwnld			
Bladen							
Elizabethtown	Lib Web	GIS Web	Web Map	Dwnld		.ims File	
Brunswick	Lib Web	GIS Web	Web Map		OneMap	.ims File	WMS Link
Buncombe							
Asheville Black Mountain	Lib Web	GIS Web	Web Map	Dwnld	OneMap	.ims File	WMS Link
Burke							
Morganton	Lib Web	GIS Web	Web Map	Dwnld	OneMap		WMS Link
Valdese							
Cabarrus							
Concord	Lib Web	GIS Web	Web Map	Dwnld	OneMap	.ims File	WMS Link
Caldwell							
Lenoir	Lib Web	GIS Web	Web Map			.ims File	
Camden	Lib Web	GIS Web	Web Map		OneMap		WMS Link
Carteret							
Atantic Beach	Lib Web	GIS Web	Web Map		OneMap	.ims File	WMS Link
Emerald Isle							
Caswell	Lib Web	GIS Web	Web Map				
Catawba							
Catawba	Lib Web	GIS Web	Web Map		OneMap	.ims File	WMS Link
Conover Hickory							
Maiden Newton							

# Carrboro, NC : Population 17,797 (2005 est.)

Geographic Information Systems

Download Data

**24 downloadable GIS data layers**

Shapefiles can be imported into your favorite mapping program or viewed in ArcMap or a free viewer (ArcExplorer) that can be downloaded from ESRI at [www.esri.com](http://www.esri.com).

**Boundaries** Municipal and planning boundaries for Carrboro, municipal boundaries for Chapel Hill, and Orange County township boundaries.

**Carrboro City Limits** Carrboro city limits.

**Orange County Index Tiles** Orange County index tiles for the 1998 orthophotography.

**Town of Carrboro**

**6 web mapping applications**

**4 OGC WMS services (web services)**

**Town of Carrboro ZONING Land Use Ordinance Section 15-142**

**9 downloadable PDF map layers**

This map is not a certified survey and no reliance may be placed on it as such.  
Source information used for this map may have been collected at different times and under different conditions, resulting in inconsistencies among features represented together on the map. The map is provided "as is" without warranty of any kind by the map's creator in that map is supporting data and makes no warranty, expressed or implied, as to the accuracy of the information presented. The fact of distribution does not constitute such a warranty.

1600 Feet 500 0  
N.C. State Plane Coordinate System (NAD 83)

Layers

- Carrboro Places
- Carrboro City Limit
- Streams
- Railroad in Carrboro
- Streets
- Carrboro Bikeways
- Wireless Network
- Access
- Properties
- Roads
- Water Bodies
- Downtown Parking
- 2003 Aerial Photography
- Apartments and Condos
- Subdivisions
- Zip Codes
- Boundaries

Refresh Map



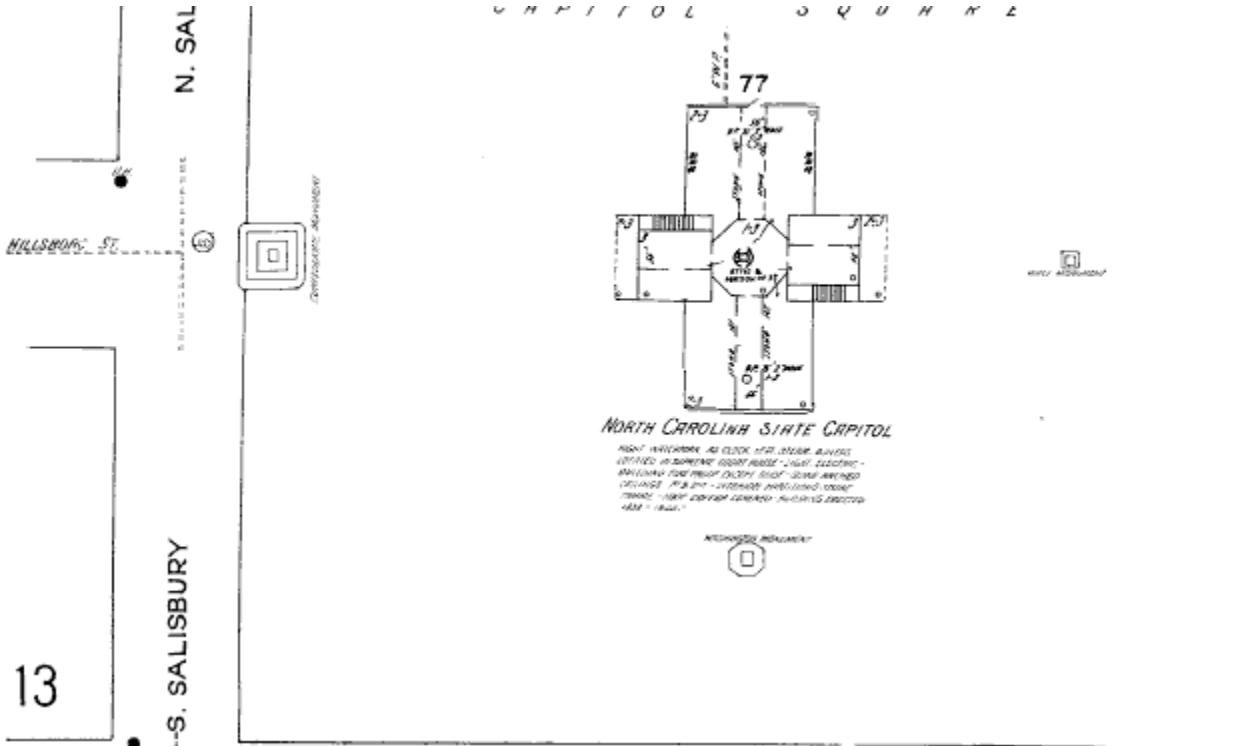
## ***Problem: Lack of Temporal Data***

- Industry focus on “latest and greatest” data
- Industry *temporally-impaired* from the point of view of data availability, software support, etc.
- “Kill and fill” as a common approach to data management (past versions of vector data lost)

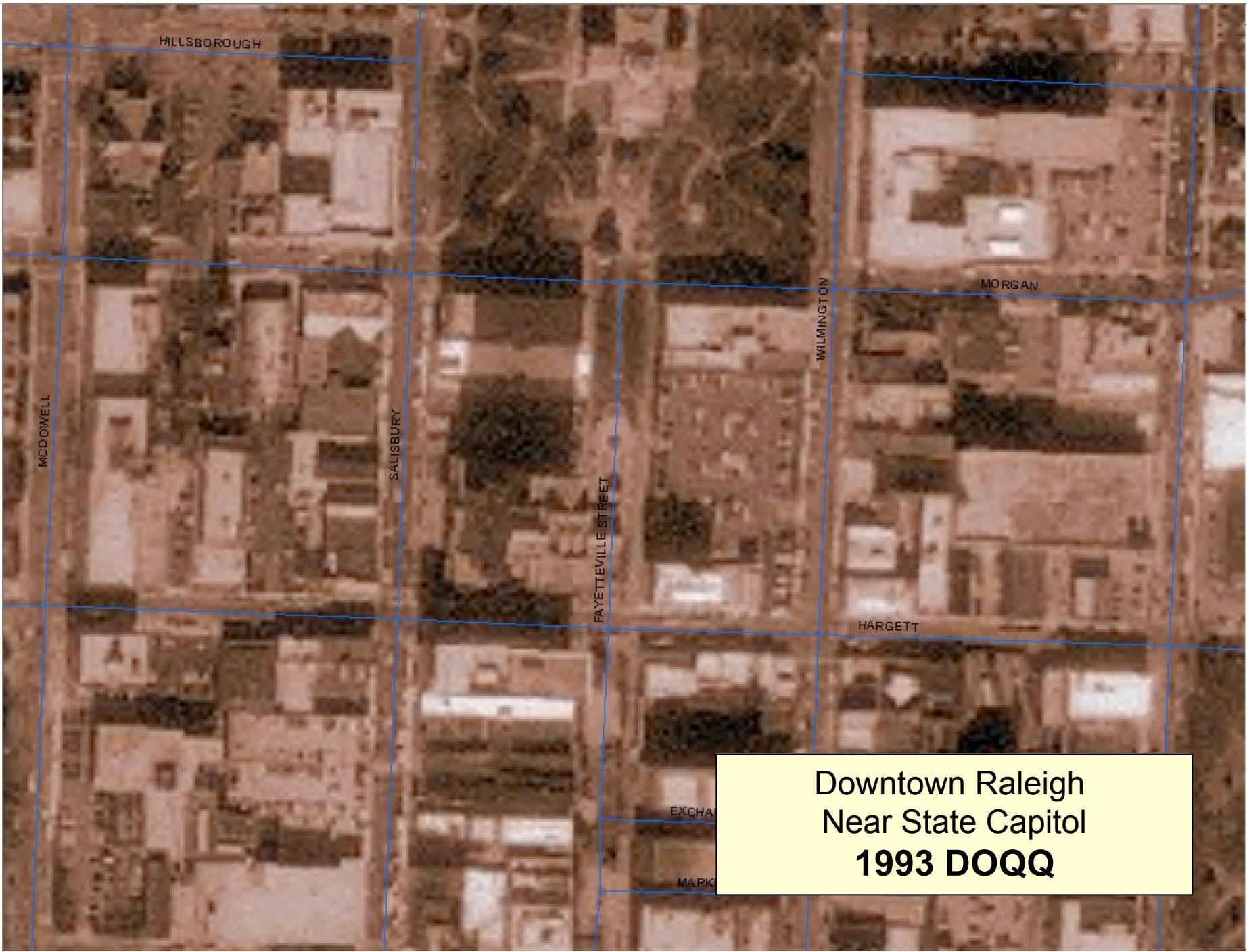
### Loss of memory about the data

- Of superceded county orthophoto flights in NC:
  - Only 22% recorded in the state’s GIS inventory
  - Only 30% accessible through county map servers

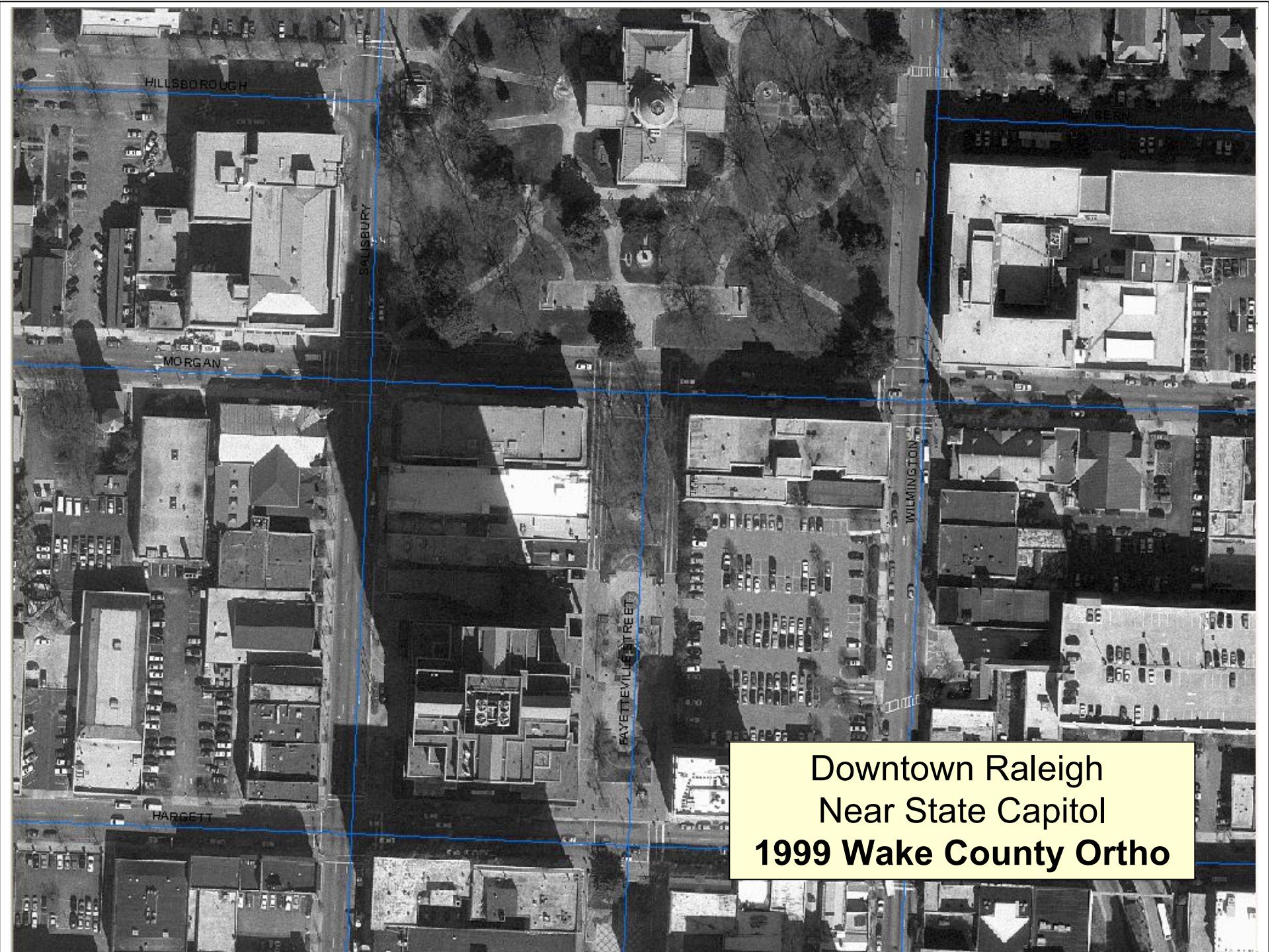
*Some older inventories only available through Internet Archive*



Downtown Raleigh  
Near State Capitol  
**1914 Sanborn Map**



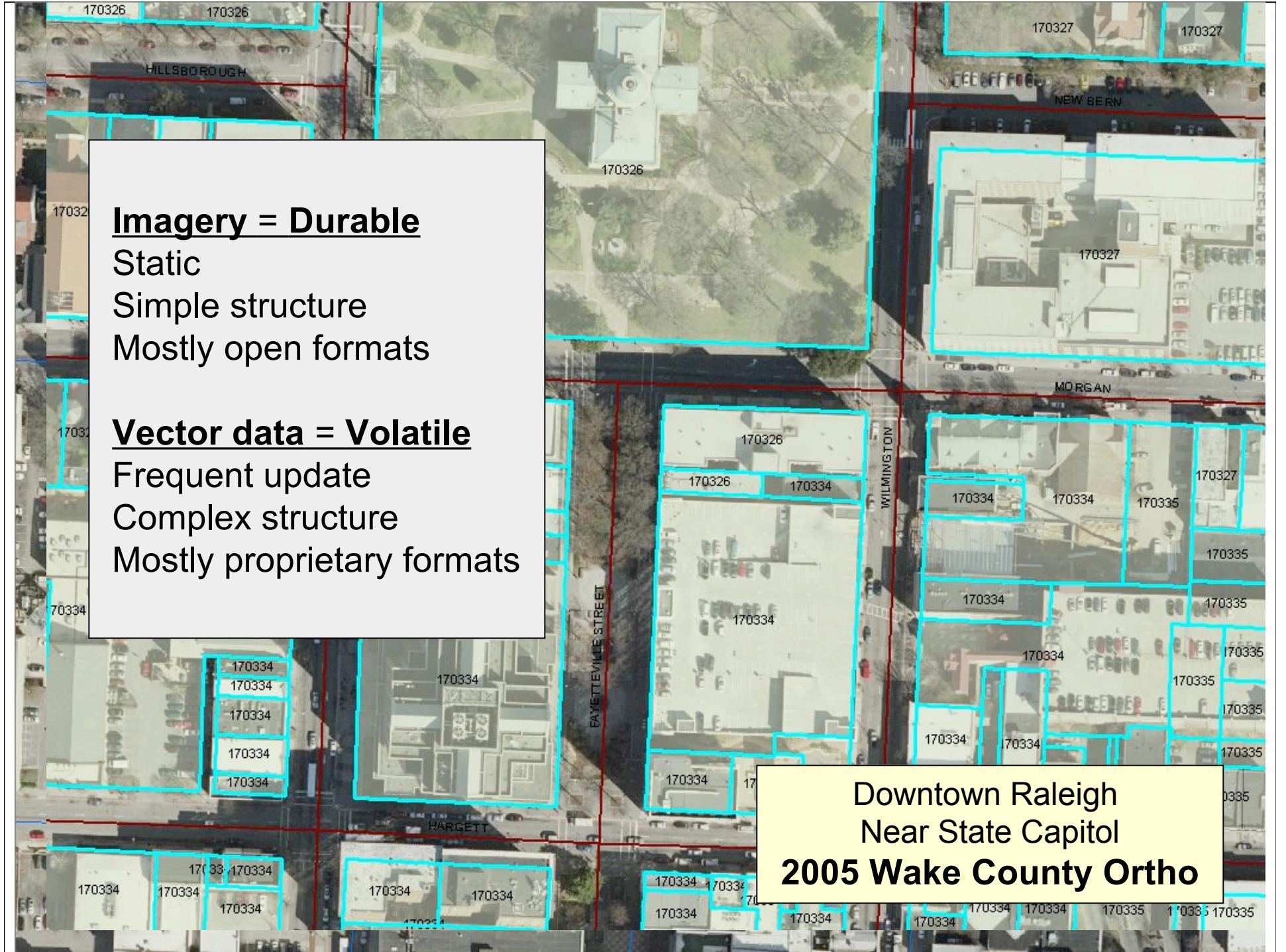
Downtown Raleigh  
Near State Capitol  
**1993 DOQQ**



Downtown Raleigh  
Near State Capitol  
**1999 Wake County Ortho**



Downtown Raleigh  
Near State Capitol  
**2005 Wake County Ortho**



# ***NDIIPP Project Overview***



## ***NC Geospatial Data Archiving Project***

- Partnership between university library (NCSU) and NC Center for Geographic Information & Analysis
- Part of the Library of Congress National Digital Information Infrastructure and Preservation Program (NDIIPP)
- Focus on state and local geospatial content in North Carolina (*state demonstration*)
- Tied to NC OneMap initiative, which provides for seamless access to data, metadata, and inventories
- Objective: engage existing state/federal geospatial data infrastructures in preservation

*Serve as catalyst for discussion within industry*



## ***Background to Spatial Data Infrastructure***

- Ca. 1990: Response to high costs of recreating data
  - Produced data not discoverable or not reusable
- 1<sup>st</sup>: Metadata standard: 1994 (FGDC)
  - Enable data discovery and evaluation for use
- 2<sup>nd</sup>: Data clearinghouse network: 1996 (using Z39.50:)
  - Search metadata encoded in SGML (later XML)
- 3<sup>rd</sup>: Cultivate content standards: late 1990's -
  - Enhance reusability, compatibility, semantic consistency
- 4<sup>th</sup>: Develop web services specifications: 2000 - (OGC)
  - Specs facilitate interoperability of data/services (e.g., WMS)

*Temporal aspects of SDI not well developed*



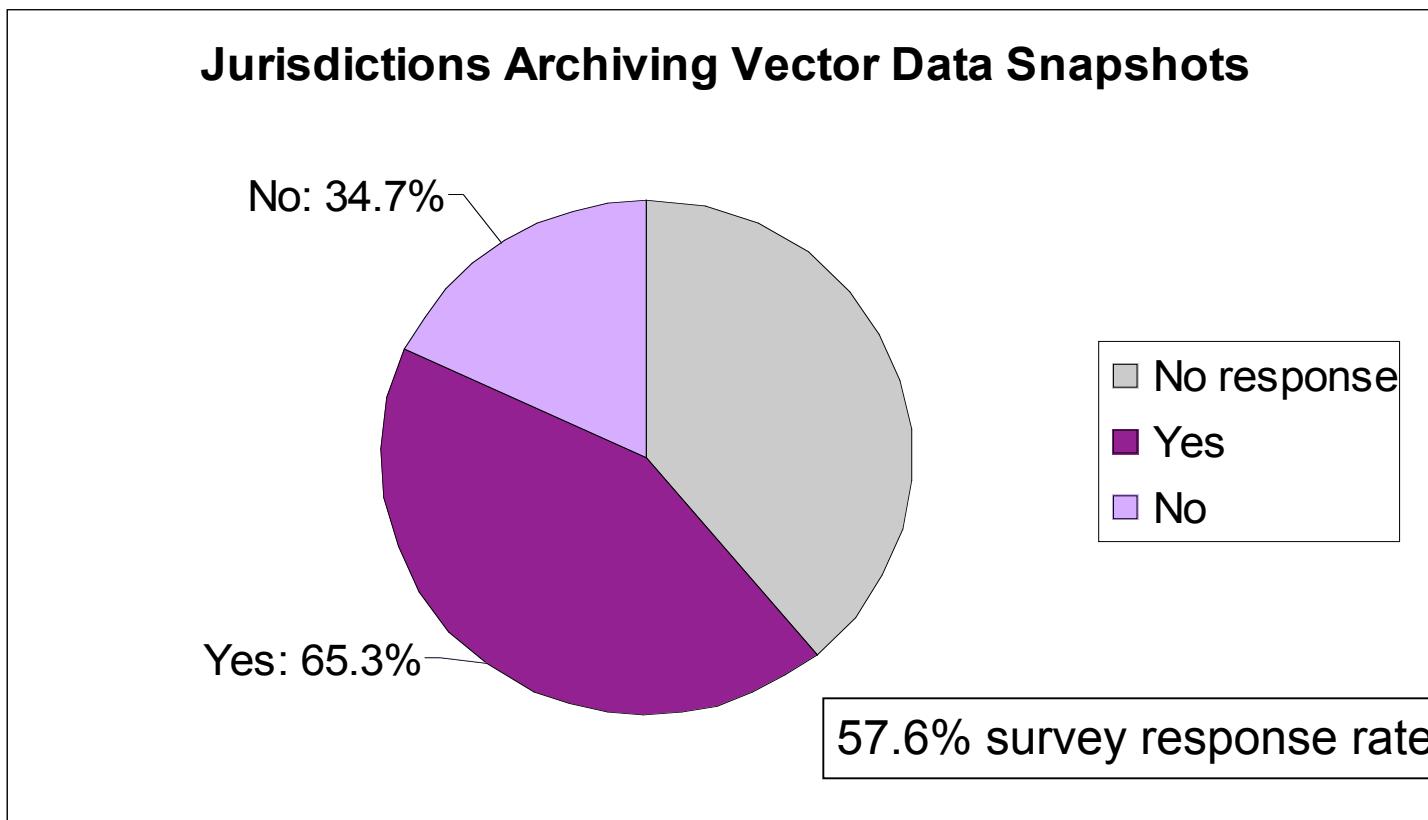
## ***Technical Challenges with Geospatial Data***

- Complex vector formats: multi-file, multi-format
  - No non-proprietary, well-supported format for vector data
- Shift to web services-based access
  - Data becoming more ephemeral
- Often: Inadequate or nonexistent metadata
  - Impedes discovery and use
- Increasing use of spatial databases for data management
  - The whole is greater than the sum of the parts but the whole is very hard to preserve

## ***Problems and (Elusive) Solutions***

## ***Problem: Data Loss***

### **Survey of current archiving practice among NC counties and municipalities**





## **Survey of current archiving practice among NC counties and municipalities**

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

“... I do a weekly backup of existing data but it is overwriting the previously saved data.”

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

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



## **Survey of current archiving practice among NC counties and municipalities**

“We are only an emerging GIS. But it is my intention that ALL data will be archived.”

“Getting ready to implement this type of archiving of data.”

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

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

*Tremendous data producer interest in digitizing and georeferencing old analog imagery and maps*



## ***Data Loss: Looking for Solutions***

- Sept. 2006: Survey of current archiving practice among NC county and municipal agencies
- Nov. 2007: NC Geographic Information Coordinating Council (GICC):  
*Ten Recommendations in Support of Geospatial Data Sharing* released
  - Recommendation: “Establish archive and long term data access strategies”
  - Suggested best practices include: “Establish a policy and procedure for the provision of access to historic data, especially for framework data layers.”
- Feb. 2008: NC GICC Archival and Long-Term Access Working Group formed

# ***Problem: Making the Business Case for Data Archiving***



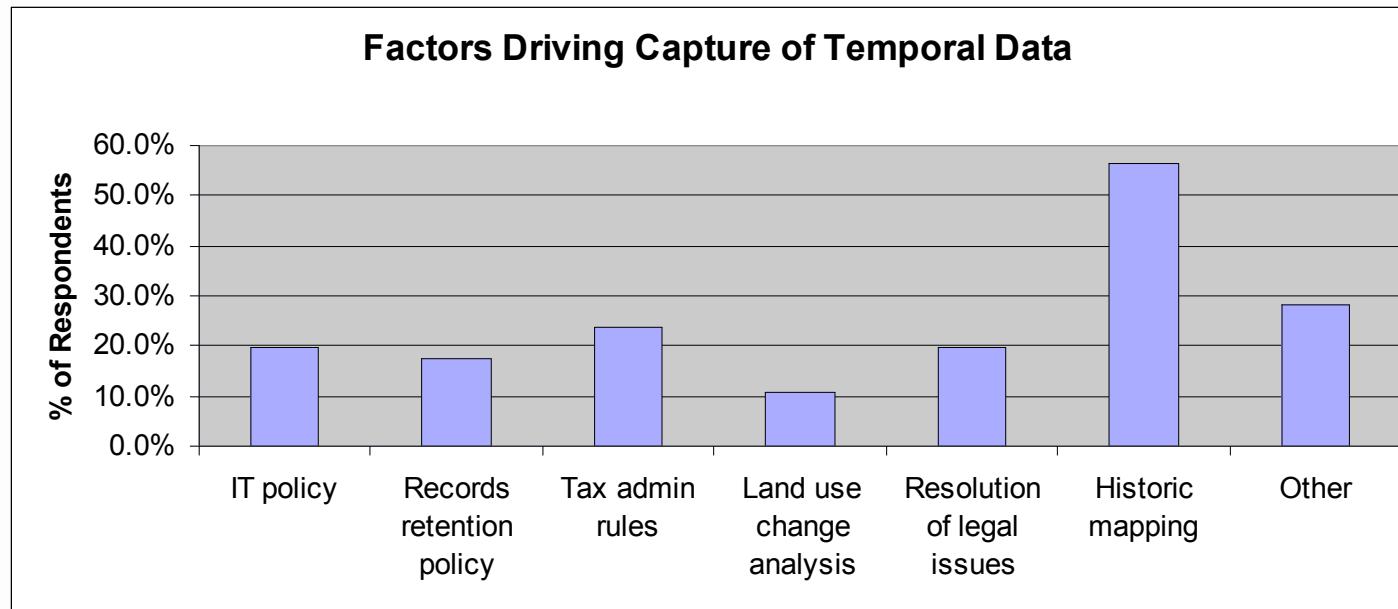
**Use case:**  
Land use and impervious  
surface change analysis



## ***Business Case: Looking for Solutions***

- Harvesting use cases for older data as part of outreach
- Formal surveys of current archiving practice and business drivers

### **Survey of current archiving practice among NC counties and municipalities**



## ***Problem: Putting the Data in Motion***

Most costly part of archive development is identifying, negotiating acquisition, and then transferring data

Local agency “contact fatigue” resulting from repeated state, federal, and university requests for data

Archive development is low priority  
– leverage other business uses that can put the data in motion

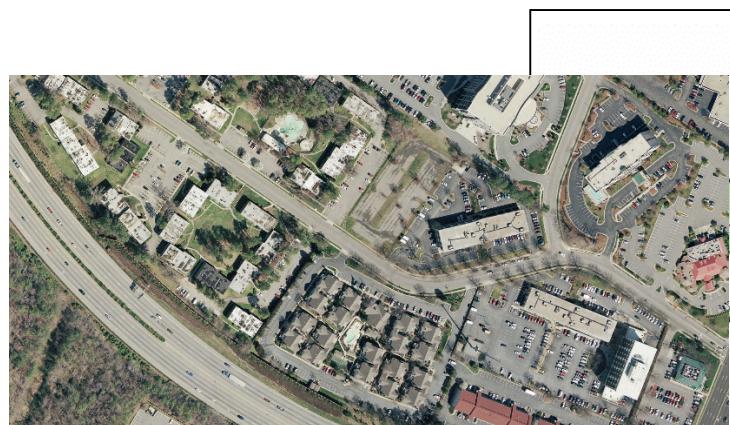
- ***Continuity of operations***
- ***Highway planning***
- ***Floodplain mapping***



### Objective

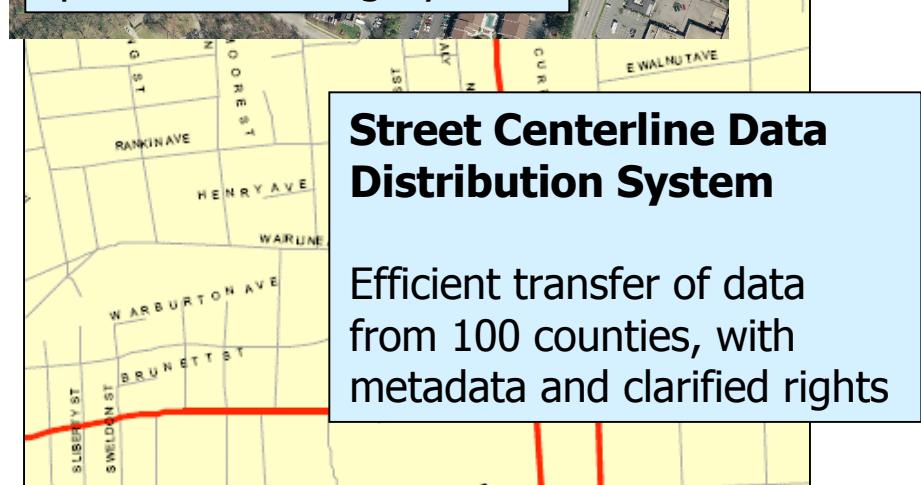
- Minimize direct contacts
- Document data
- Clarify rights
- Routinize transfer

# *Putting the Data in Motion: Looking for Solutions*



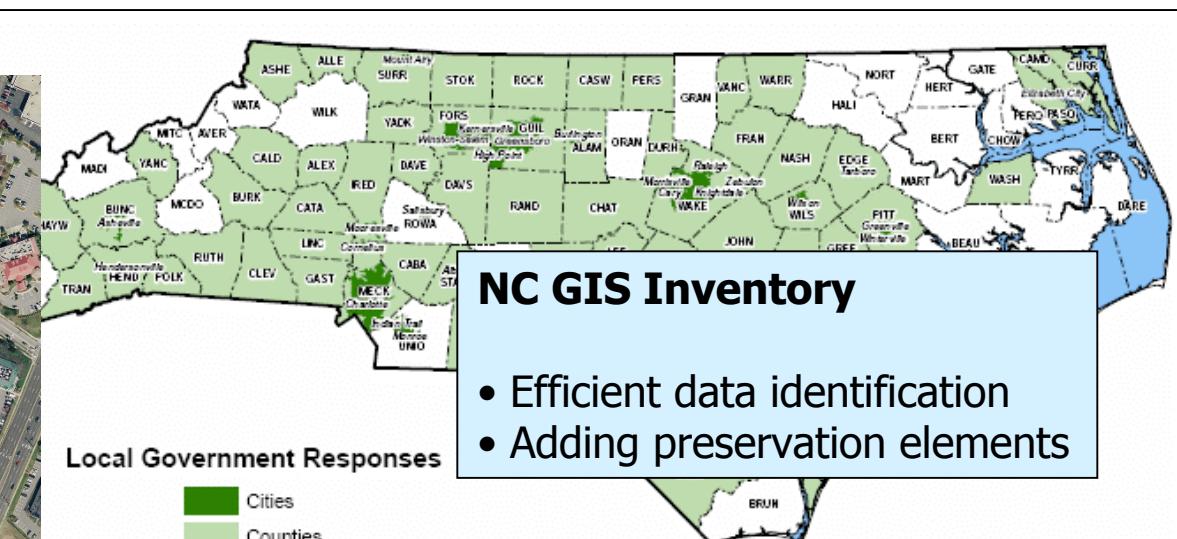
## Orthophoto Data Distribution System

Transfer of large quantities of imagery



## Street Centerline Data Distribution System

Efficient transfer of data from 100 counties, with metadata and clarified rights



**Local Government Responses**

- Cities (dark green)
- Counties (light green)

**NC GIS Inventory**

- Efficient data identification
- Adding preservation elements

## NC OneMap Data Discovery and Download

Data can be downloaded from the NC OneMap FTP site at no cost. [Vector](#) and [raster](#) data are available. If you do not find what you're looking for

Below is the list of available data. You can download using an FTP client. To download the ZIP file to your computer, click the column to download the ZIP file to your computer otherwise noted. The projection is NC StatePlane Feet.

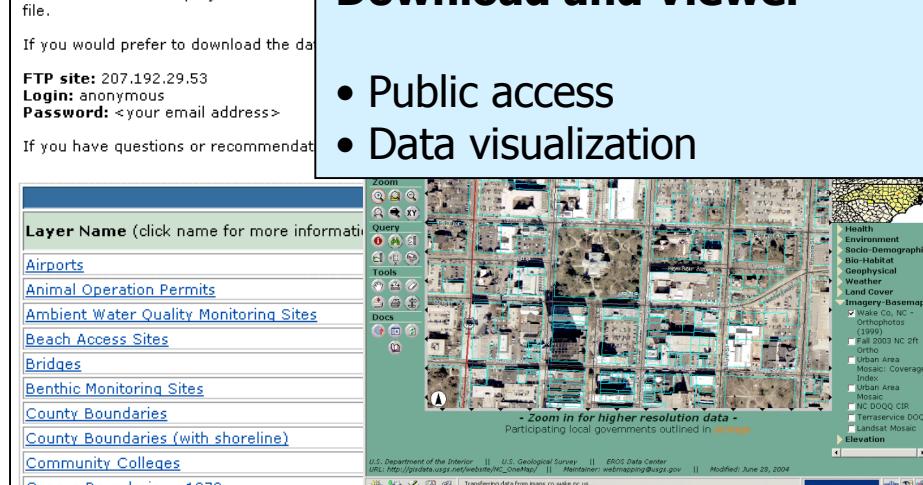
If you would prefer to download the data

**FTP site:** 207.192.29.53  
**Login:** anonymous  
**Password:** <your email address>

If you have questions or recommendations, please contact us.

## NC OneMap Data Download and Viewer

- Public access
- Data visualization



The interface includes a sidebar with various data categories and a main map area showing a detailed view of a geographic area with red outlines and labels like "SILVER ST", "SWEDEN ST", "RANKIN AVE", "HENRY AVE", "WARLINE", "E WALNUT TAVE", "W ARBURTON AVE", and "BRUNETTE ST".

Layer Name (click name for more information)

- Airports
- Animal Operation Permits
- Ambient Water Quality Monitoring Sites
- Beach Access Sites
- Bridges
- Benthic Monitoring Sites
- County Boundaries
- Community Colleges
- Census Boundaries - 1970

Zoom in for higher resolution data. Participating local governments outlined in orange.

U.S. Department of the Interior || U.S. Geological Survey || EROS Data Center URL: [http://gisdata.usgs.gov/websvc/NC\\_OneMap/](http://gisdata.usgs.gov/websvc/NC_OneMap/) || Maintainer: webmapping@usgs.gov || Modified: June 28, 2004

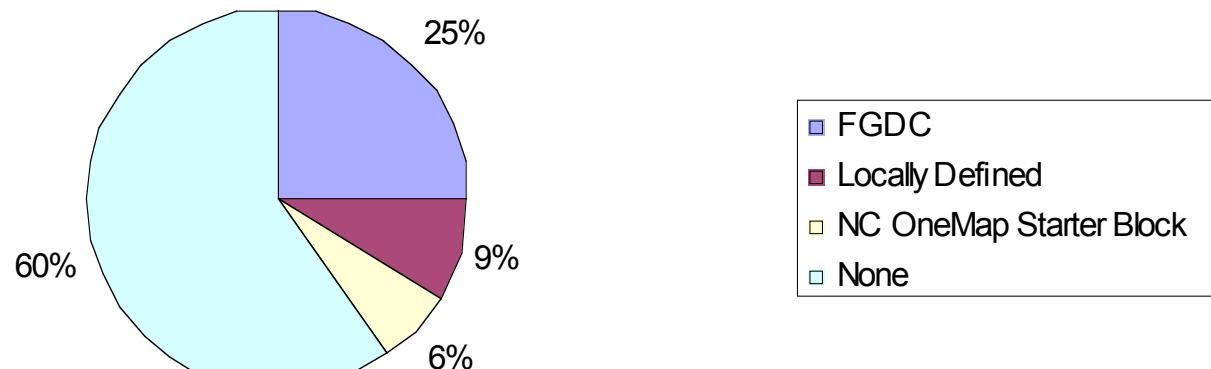
Healthcare Environment Socio-Demographic Bio-Habitat Geospatial Weather Land Cover Imagery-Basemap Wake Co, NC - Orthophotos (1999) Fall 2003 NC 2ft DEM Urban Area Mosaic Coverage Index Urban Area Mosaic ID: 0000 CIR TerrainService DDO Landsat Mosaic Elevation

## ***Problem: Metadata***

Metadata is often asynchronous, inconsistently structured, incomplete, or missing.

### **Survey of current archiving practice among NC counties and municipalities**

Metadata archived with data?





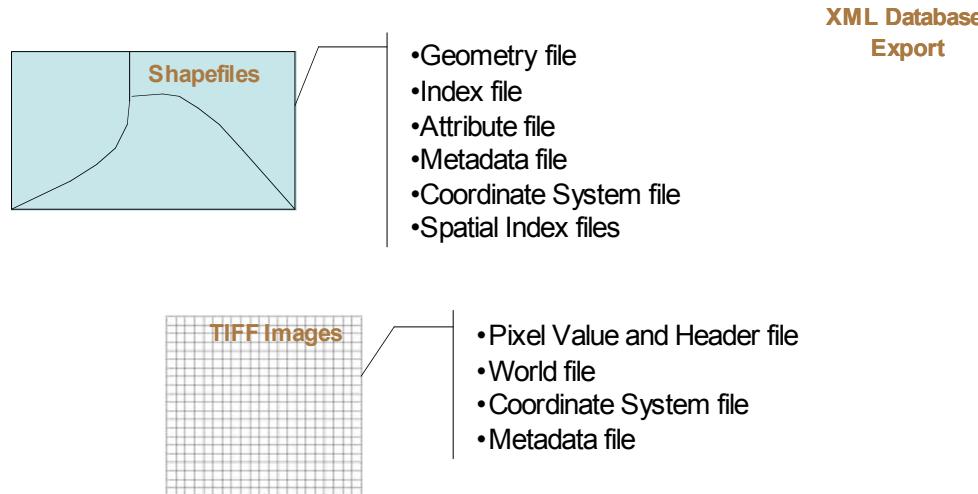
## ***Metadata: Looking for Solutions***

- NC OneMap Metadata Outreach
  - Workshops, support
- NC OneMap Metadata Starter Block
  - Starter templates for key data layers
- NC GIS Inventory
  - Builds minimal metadata
- Emerging content exchange networks
  - e.g., NC StreetMap.com
  - Accrete metadata as part of submission and transfer process

## ***Problem: Content Packaging***

- Complex multi-file, multi-format objects
- Shared ancillary components
- Need to add administrative & technical metadata beyond FGDC

### **Potential Ingest Objects**





## ***Content Packaging: Looking for Solutions***

- Open Geospatial Consortium (OGC) Data Preservation Working Group formed
  - Content packaging now a topic of discussion
- Emerging content exchange networks
  - .e.g., NCStreetMap.com

### Objective

- Automated processing of received data
- Reduce costly and error-prone human intervention
- Capture additional technical and administrative metadata

## ***Changes in the Domain***

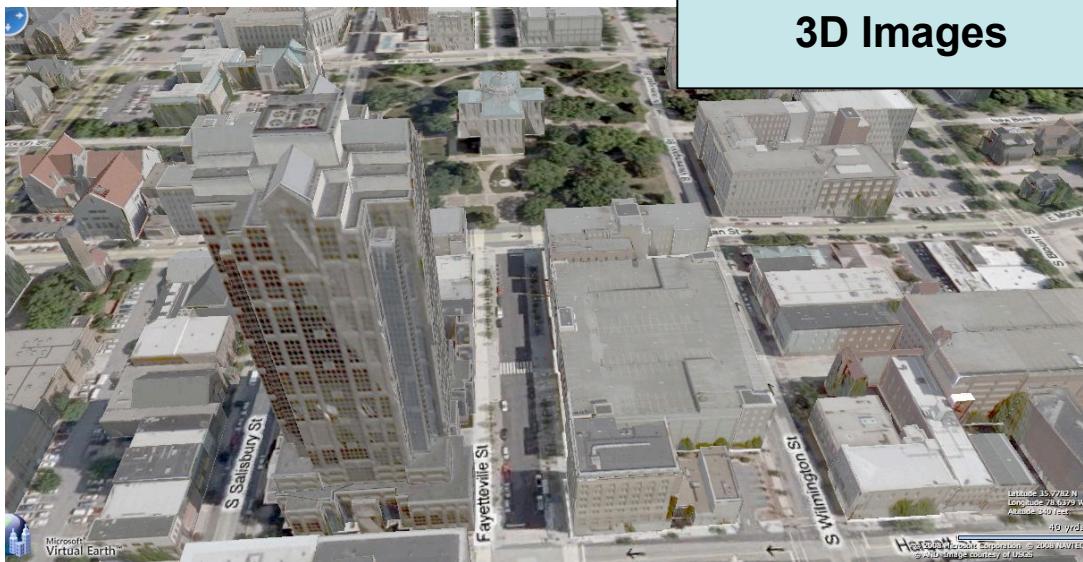
# ***Changes in the Domain: New Location-Based Content***



**Street Views**



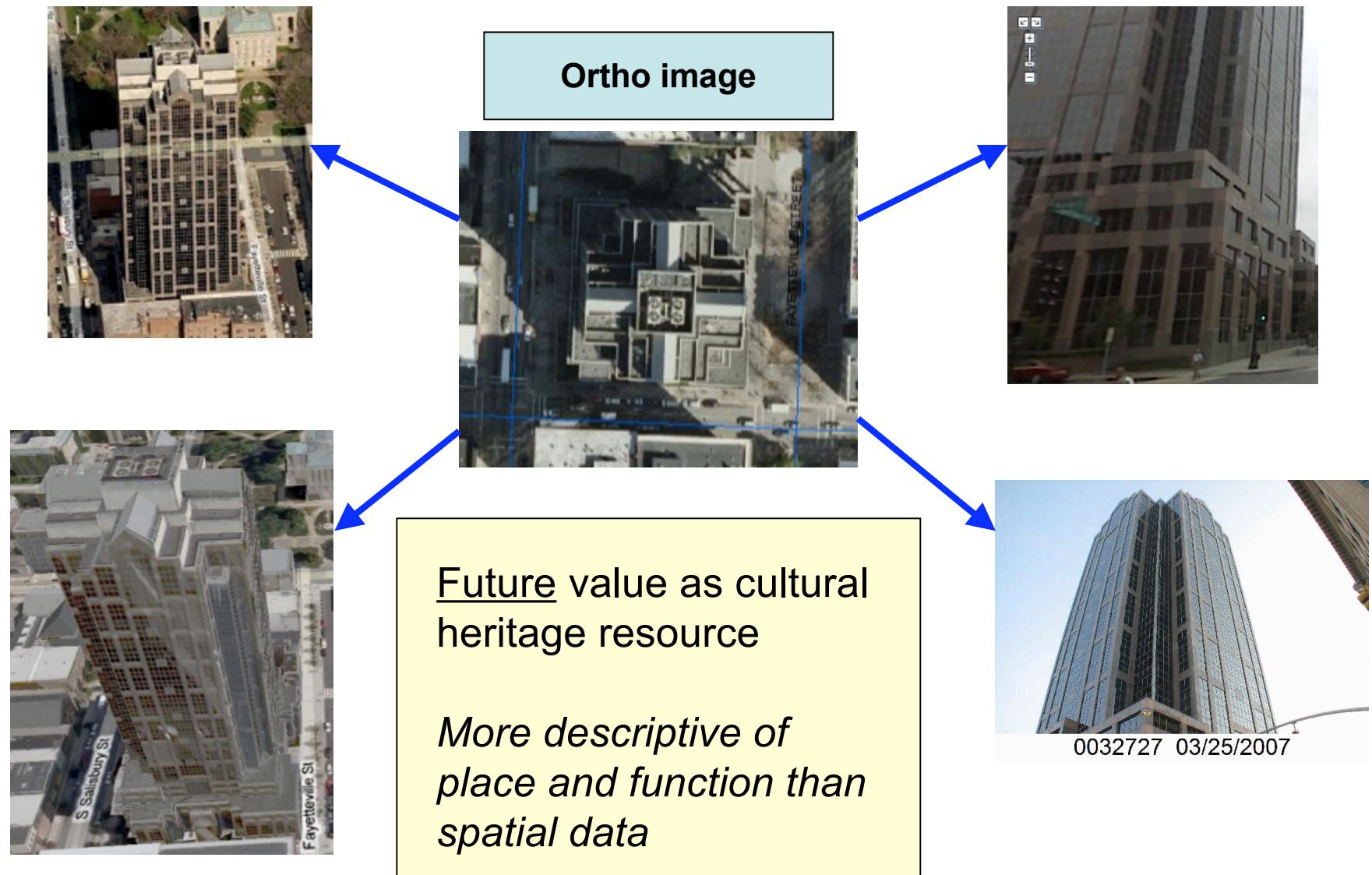
**Oblique Imagery**



**3D Images**

Present-day value in  
**location-based  
services and mobile  
applications**

# ***Changes in the Domain: New Location-Based Content***



# *Changes in the Domain: Geospatial PDF*

General Maps - Microsoft Internet Explorer provided by NCSU Libraries

File Edit View Favorites Tools Help

Address http://www.co.gaston.nc.us/Planning/maps/GeneralMaps.htm

Search Our Site! Home | Departments | Events Calendar | Maps/GIS | Document Center

Land Use  
Long Range Planning  
Environmental Planning  
Planning GIS  
G-CaMP  
Downloads  
Ordinances  
Board of Adjustment  
Planning Board  
UDO  
Historic Preservation  
Calendars  
FAQ  
Comp Plan  
Census

**Planning & Development Services**

**General Maps**

- [Census Tracts \(PDF, 677K, 6/29/2006\)](#)
- [Census Tract Boundary with Election Precinct Districts Overlaid \(PDF, 890K, 7/25/2006\)](#)
- [Communication Tower Sites \(PDF, 808K, 7/3/2006\)](#)
- [Election Precincts \(PDF, 710K, 7/25/2006\)](#)
- [Election Precinct Districts with Township Boundary Overlay \(PDF, 7/25/2006\)](#)
- [Extra-Territorial Jurisdictions \(PDF, 1463K, 7/10/2006\)](#)
- [Fire Districts \(PDF, 1014K, 7/13/2006\)](#)
- [Health Department \(PDF, 618K, 7/21/2006\)](#)
- [Hydrography \(PDF, 2195K, 7/13/2006\)](#)
- [Municipal Areas \(PDF, 4316K, 7/13/2006\)](#)

The screenshot shows a Microsoft Internet Explorer window displaying the Gaston County Planning & Development Services website. The left sidebar contains a navigation menu with links to Land Use, Long Range Planning, Environmental Planning, Planning GIS, G-CaMP, Downloads, Ordinances, Board of Adjustment, Planning Board, UDO, Historic Preservation, Calendars, FAQ, Comp Plan, and Census. The main content area features a large map of Gaston County with various municipal areas shaded in different colors. A legend on the right identifies the city areas: BELMONT (purple), BESSERME CITY (green), CHERRYVILLE (orange), CRAMERTON (blue), DALLAS (pink), DELIVIEW (yellow), GASTONIA (light blue), HIGH SHOALS (light green), KINGS MTN (dark blue), LOWELL (light purple), MCADENVILLE (light pink), MOUNT HOLLY (light green), RANLO (light blue), SPENCER MTN (light green), and STANLEY (light blue). Above the map is a section titled "GASTON COUNTY MUNICIPAL AREAS" with a detailed map showing the boundaries of these areas. The bottom of the page has a yellow footer box containing text about the counterpart to analog maps.

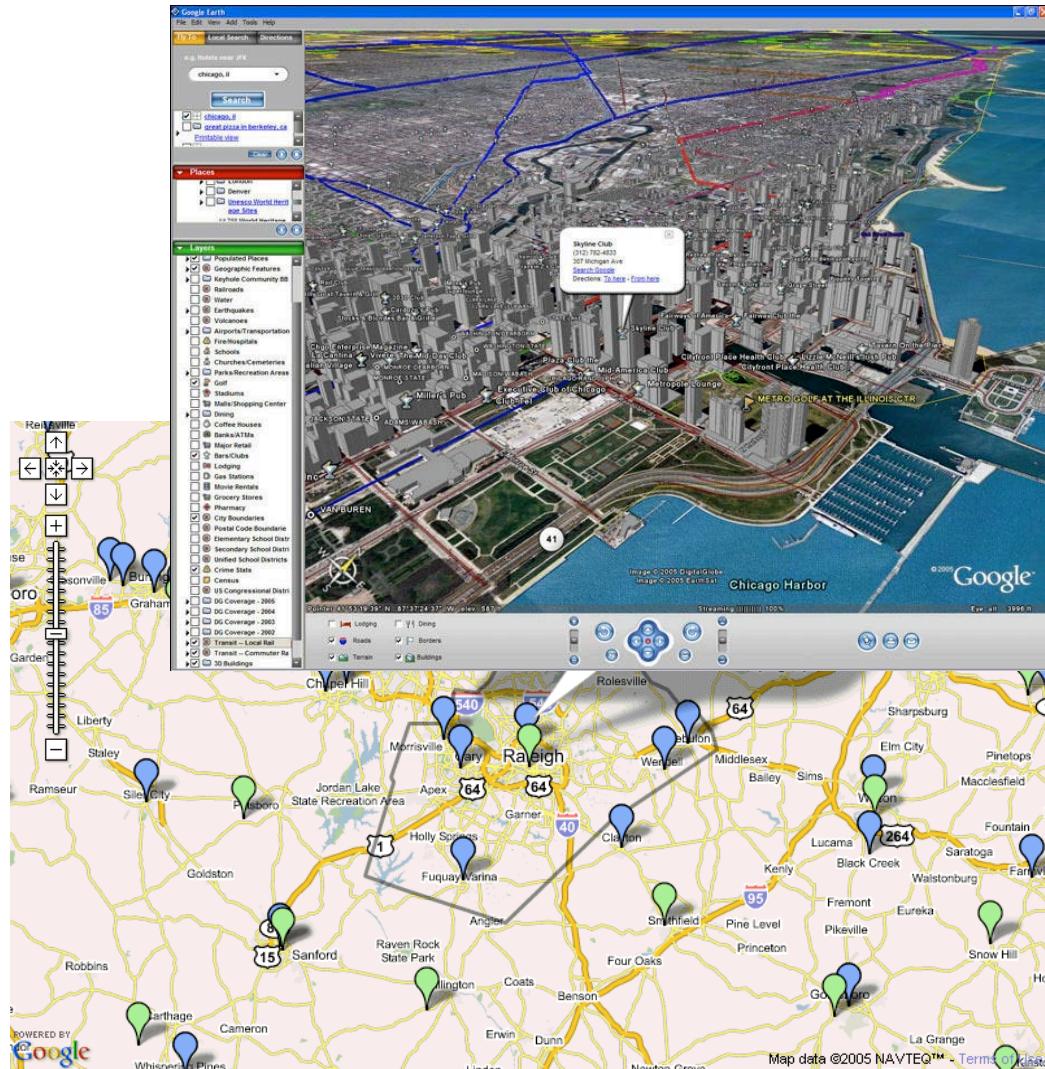
## PDF and GeoPDF

Counterpart to analog map = datasets *plus* data models, symbolization, classification, annotation, etc.

More data intelligence survives in PDF documents than survives in most other “baked” formats

# *Changes in the Domain: New Network Payloads*

KML  
GeoRSS  
GeoJSON  
Tile Map Service  
More ....



- Lightweight
- AJAX-friendly
- Often ephemeral

## ***Moving Forward: New Initiatives***



## ***NC GICC Archival and Long-Term Access Committee***

- Initiated by NC Geographic Information Coordinating Council in 2008 to address growing concerns of state and local agencies about long-term access to data
- Federal, state, regional, and local agency representation
- Key focus
  - Best practices for data snapshots and retention
  - State Archives processes: appraisal, selection, retention schedules, etc.
  - Who, What, Why, When, Where, How



## ***NDIIPP Multi-State Geospatial Project***

- Lead organizations: North Carolina Center for Geographic Information & Analysis (NCCGIA) and State Archives of NC
- Partners:
  - Leading state geospatial organizations of Kentucky and Utah
  - State Archives of Kentucky and Utah
  - NCSU Libraries in catalytic/advisory role
- State-to-state and geo-to-Archives collaboration
- 2 year project: Nov. 2007-Dec. 2009
- Archives as part of Spatial Data Infrastructure



## ***Conclusion***

- “Supporting temporal analysis requirements” gets more attention than “archiving and preservation”
- Leverage existing infrastructure
- Current data sharing needs drive infrastructure improvements that help archiving
- Leverage business needs that are more compelling than preservation (e.g., continuity of operations)
- Facilitate stakeholder ownership of the solutions
- Mine state and local archiving innovations

***Thanks to Library of Congress and the NDIIPP Partners!***



## *Questions?*

Steve Morris  
Head, Digital Library Initiatives  
NCSU Libraries  
ph: (919) 515-1361  
[Steven\\_Morris@ncsu.edu](mailto:Steven_Morris@ncsu.edu)

<http://www.lib.ncsu.edu/ngdap>