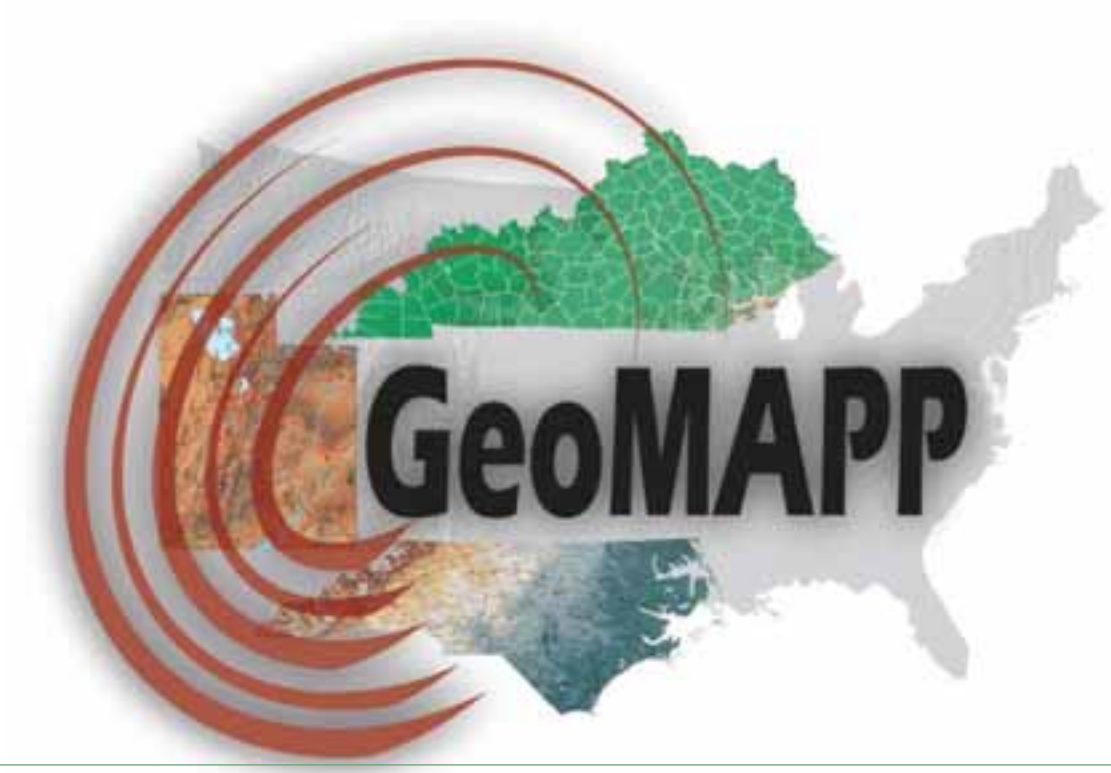


Mapped Today; Zapped Tomorrow? Preserving Government Digital Geospatial Data

Butch Lazorchak, Library of Congress
Alec Bethune, North Carolina Center for Geographic Information and Analysis
Kelly Eubank, North Carolina State Archives

ESRI Federal Users Conference | 19 February 2010 | Washington, DC

Why is the Library of Congress Interested in Preserving Digital Geospatial Data?

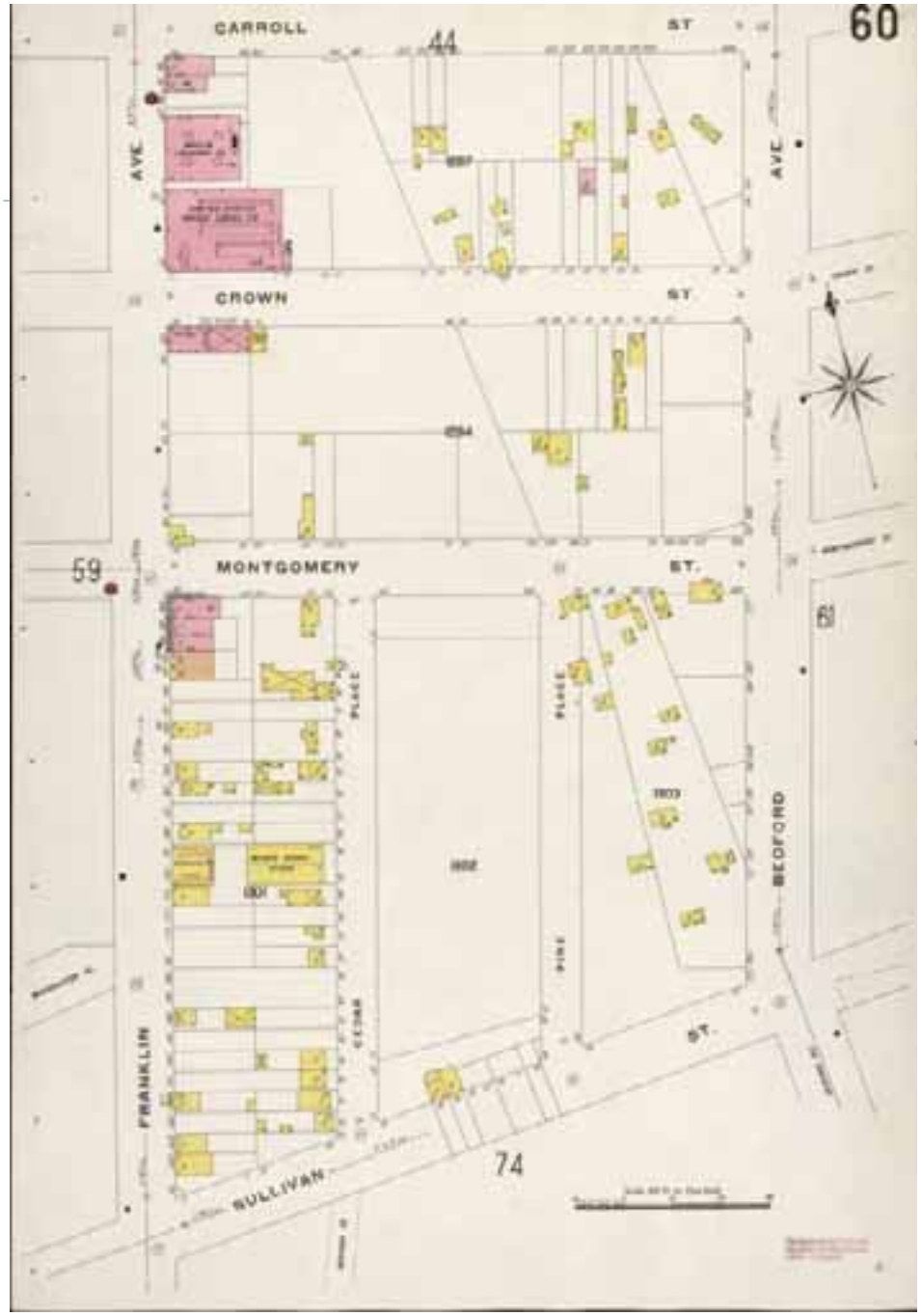
Butch Lazorchak

Library of Congress

- The Library's mission is to make its resources available and useful to the Congress and the American people and to sustain and preserve a universal collection of knowledge and creativity for future generations.
- Geography and Maps Division features the largest and most comprehensive cartographic collection in the world

- Of course, the mission to collect doesn't prescribe exactly those future generations will do with all that knowledge and creativity. In fact, we can't know.

- We do know that serendipity's record is pretty good


CARROLL
ST
44
60
AVE
AVE
CROWN
ST
59
MONTGOMERY
ST.
61
PLACE
PLACE
BEDFORD
FRANKLIN
PINE
ST.
SULLIVAN
74

http://sidesalad.net/archives/ApartmentsAtTheSiteOfTheFormerEbbetsField.JPG

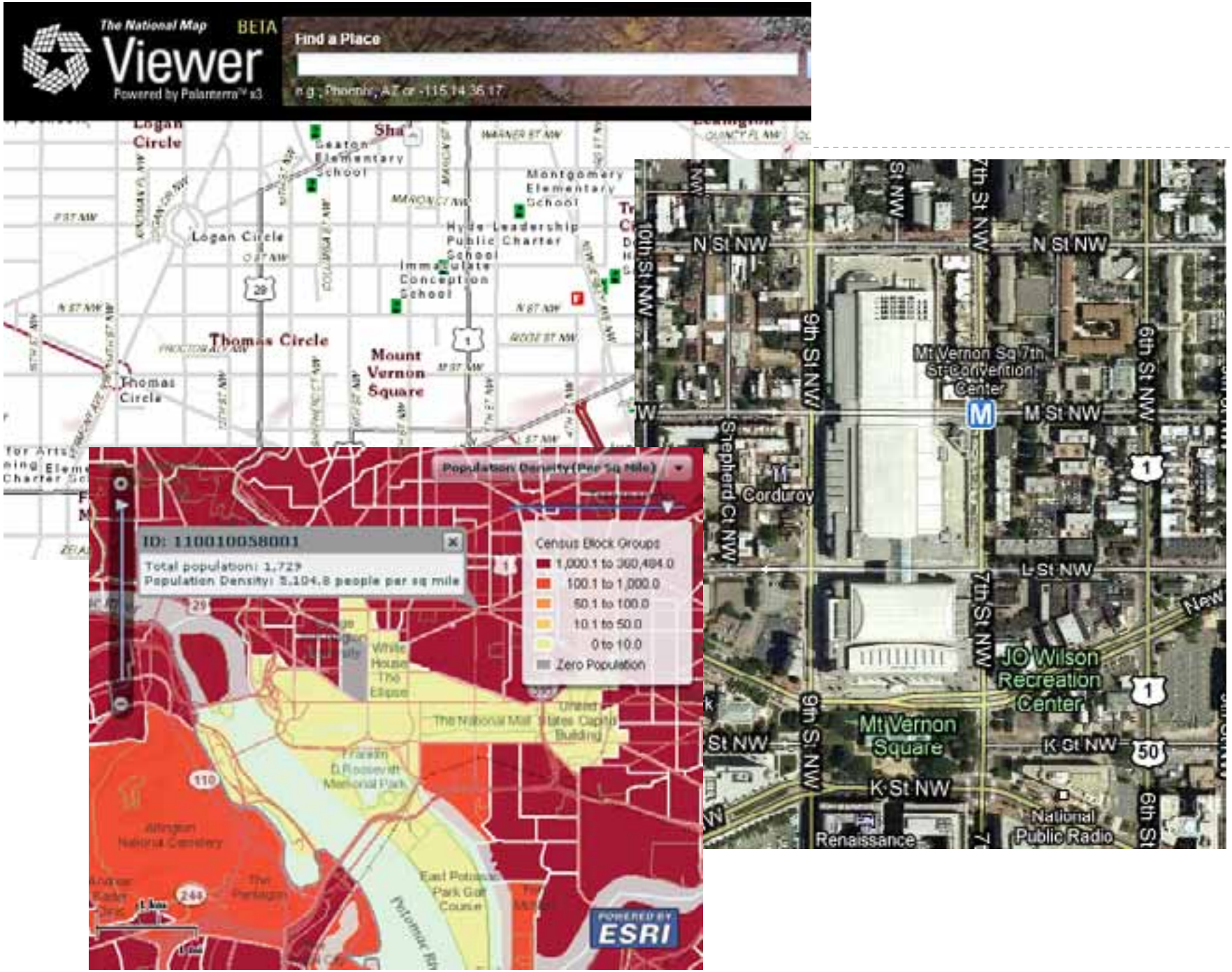
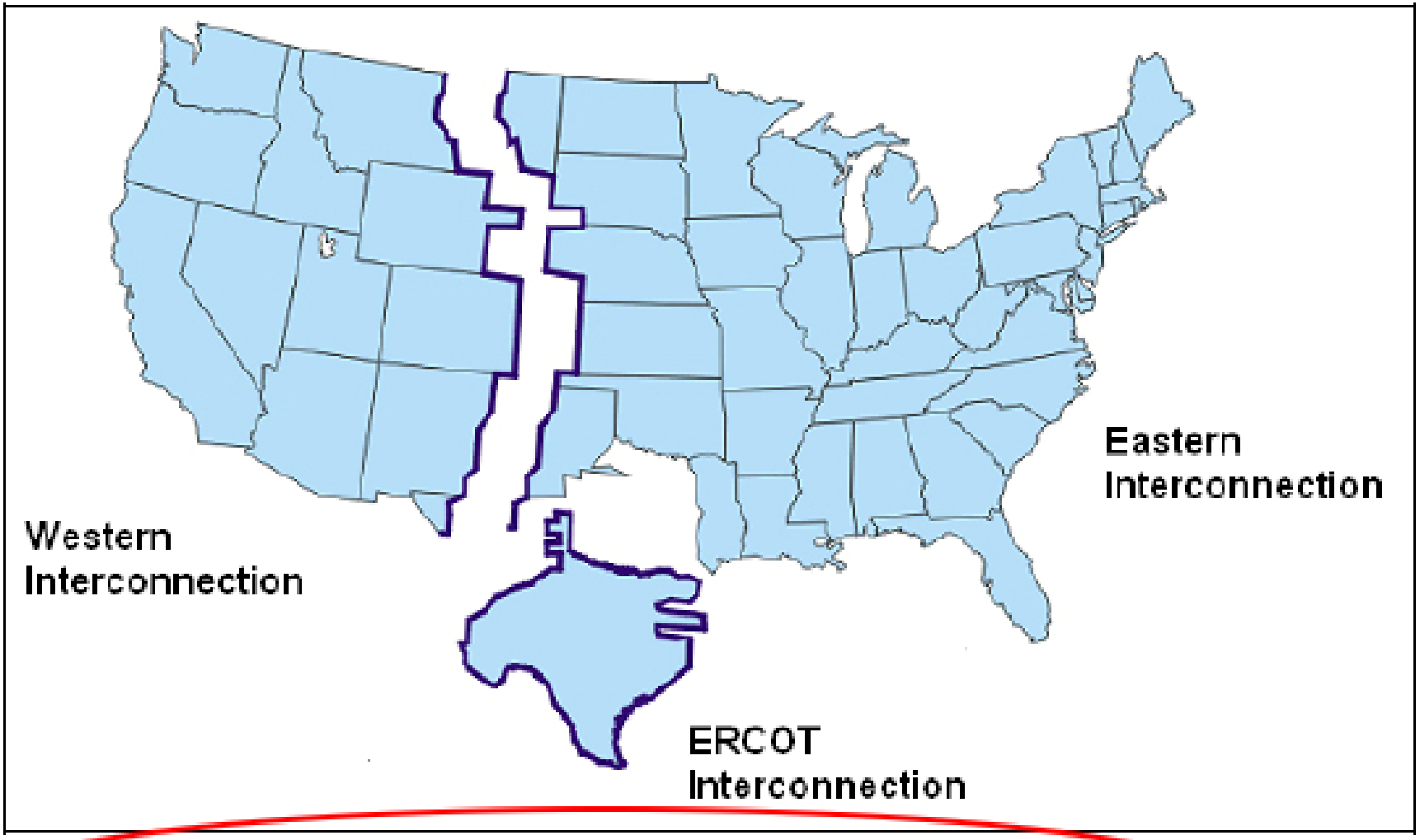

The National Map BETA
Viewer
Powered by Polaris
Find a Place
e.g.: Phoenix, AZ or -115.14 36.17
Logan Circle
Seaton Elementary School
Montgomery Elementary School
Hyde Leadership Public Charter School
Immaculate Conception School
Thomas Circle
Mount Vernon Square
Population Density (Per Sq Mile)
ID: 110010058001
Total population: 1,729
Population Density: 5,104.8 people per sq mile
Census Block Groups
1,000.1 to 360,494.0
100.1 to 1,000.0
50.1 to 100.0
10.1 to 50.0
0 to 10.0
Zero Population
White House / The Ellipse
The National Mall
Arlington National Cemetery
The Pentagon
East Potomac Park Golf Course
POWERED BY ESRI
N St NW
7th St NW
9th St NW
10th St NW
6th St NW
M St NW
Mt Vernon Sq 7th St Convention Center
Shepherd Ct NW
Corduroy
L St NW
JO Wilson Recreation Center
Mt Vernon Square
K St NW
National Public Radio
Renaissance

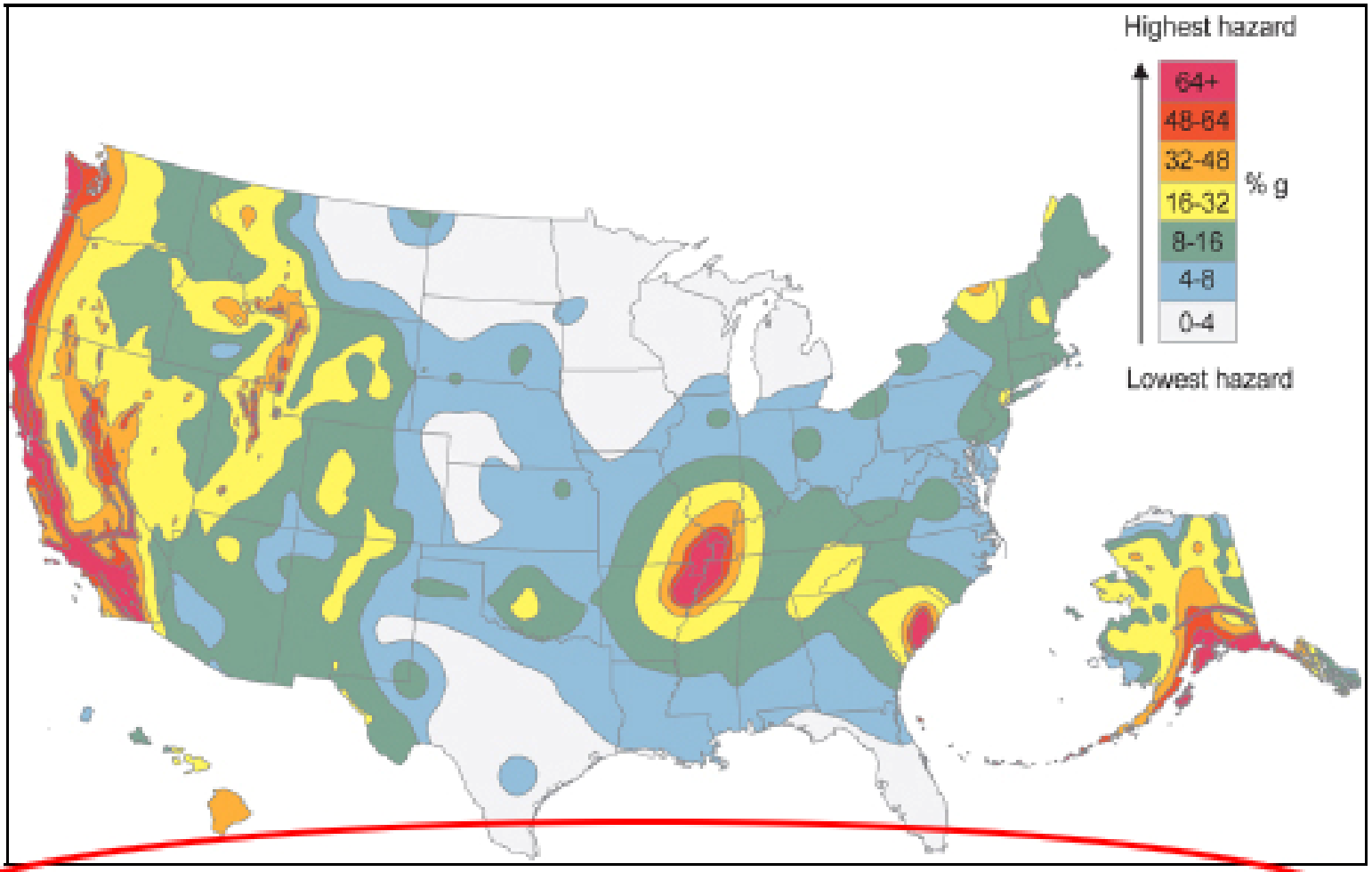
Figure 6. United States Power System Interconnections



Source: adapted from a map located on the Energy Information Administration website at http://www.eia.doe.gov/cneaf/electricity/page/fact_sheets/transmission.html.

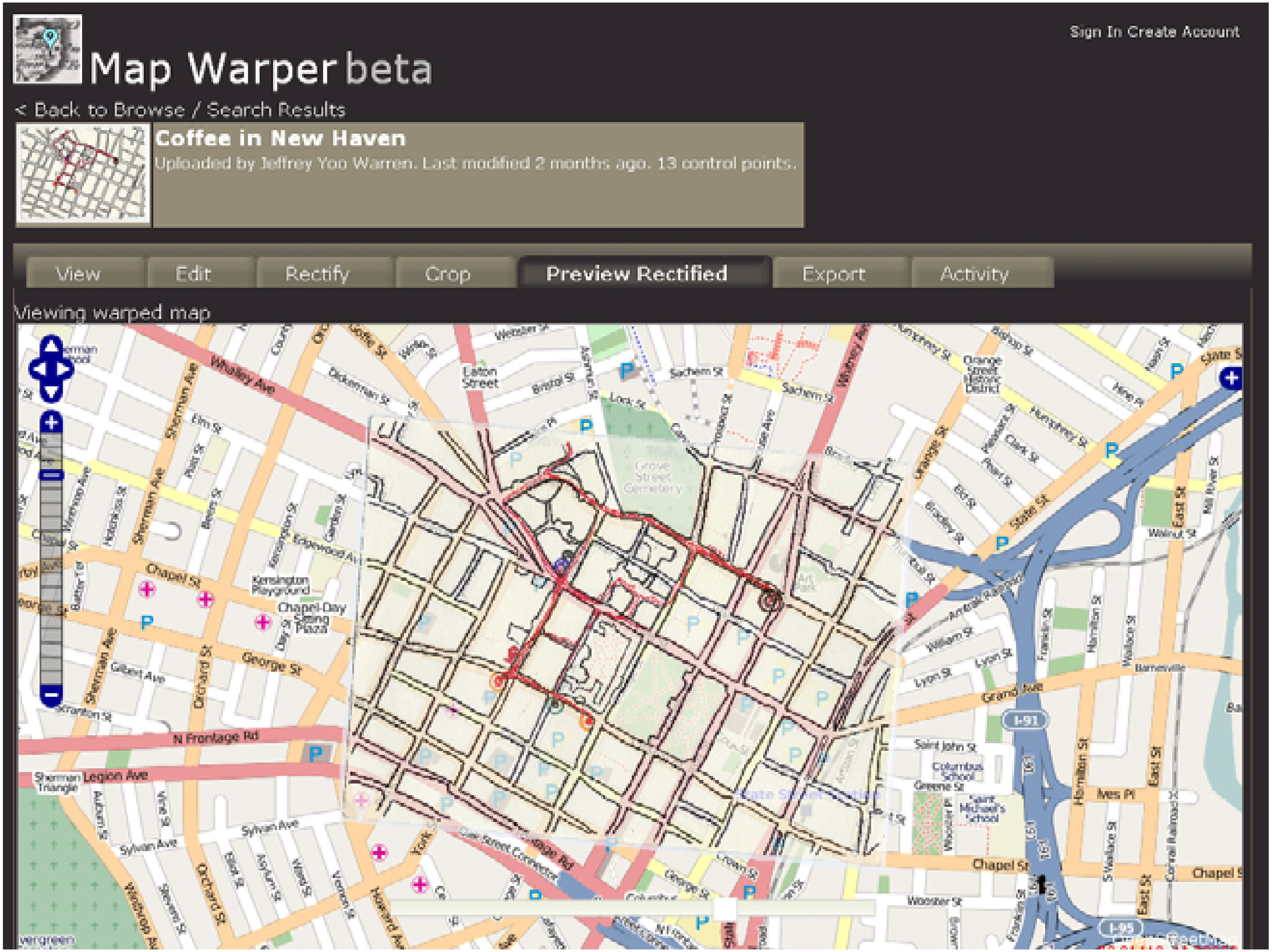
Notes: ERCOT = Electric Reliability Council of Texas.

Figure 2. Earthquake Hazard in the United States



Source: USGS Fact Sheet 2008-3018 (April 2008), at http://earthquake.usgs.gov/research/hazmaps/products_data/images/nshm_us02.gif. Modified by CRS.

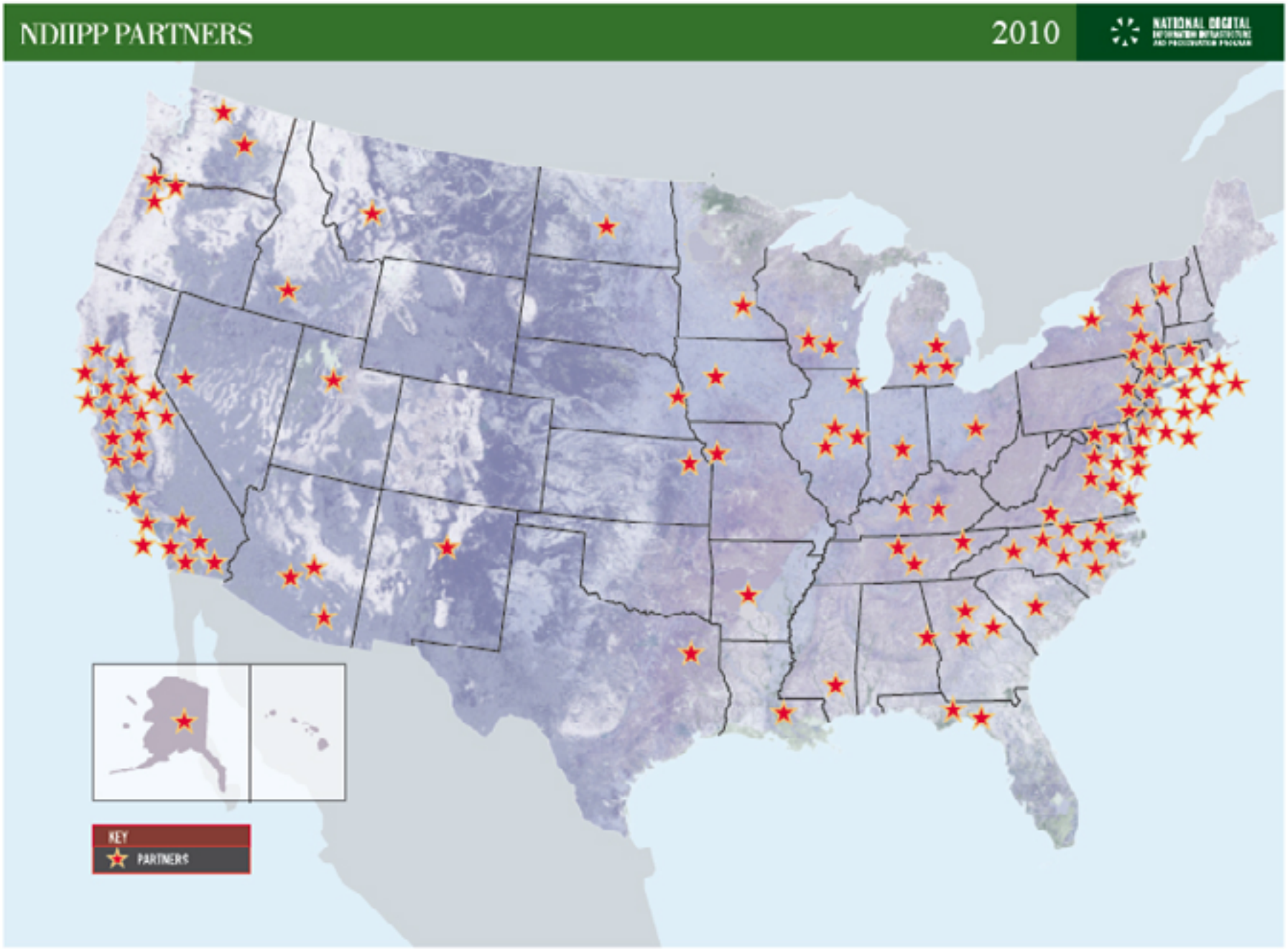
Note: The bar in the upper right shows the potential ground motion—expressed as a percentage of the acceleration due to gravity (g)—with up to a 1 in 10 chance of being exceeded over a 50-year period.


Sign In Create Account
Map Warper beta
< Back to Browse / Search Results
Coffee in New Haven
Uploaded by Jeffrey Yoo Warren. Last modified 2 months ago. 13 control points.
View
Edit
Rectify
Crop
Preview Rectified
Export
Activity
Viewing warped map

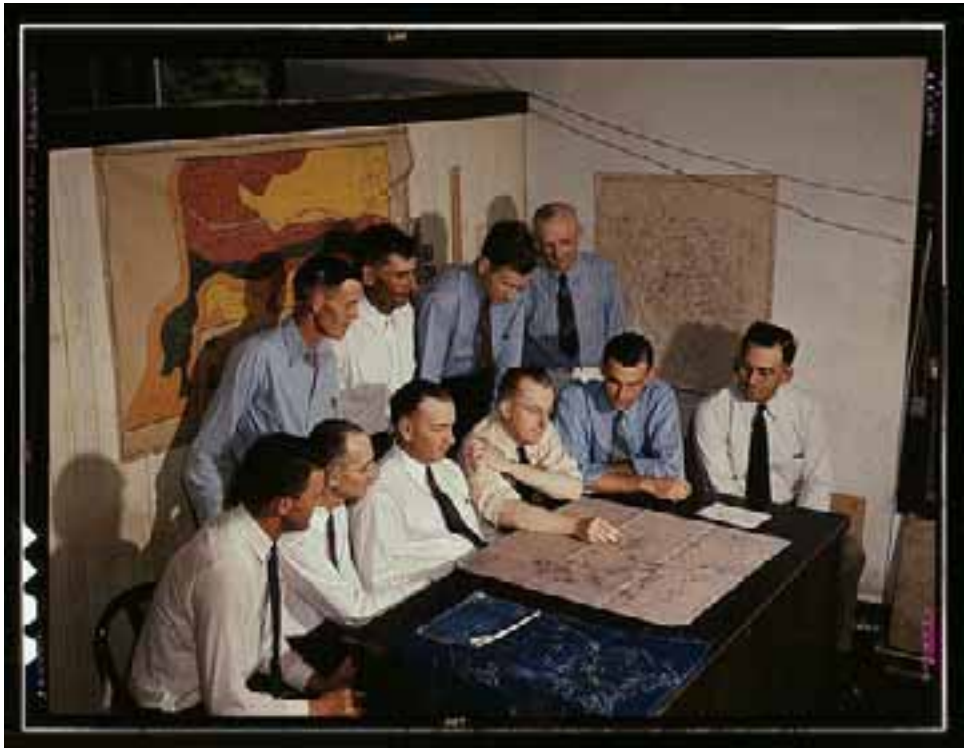
Library of Congress National Digital Information Infrastructure and Preservation Program (NDIIPP)

Mission: To ensure access over time to a rich body of digital content through the establishment of a national network of partners committed to selecting, collecting and preserving at-risk digital information

- Learn By Doing
- Catalyze Activity
- Support Collaboration
- Break Down Boundaries

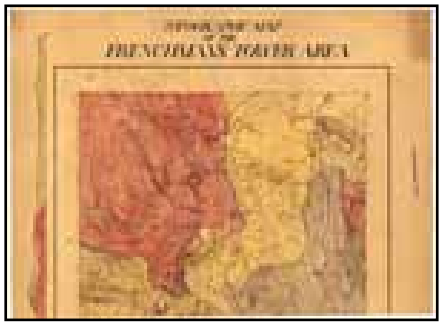

NDIIPP PARTNERS
2010
NATIONAL DIGITAL
KEY
PARTNERS

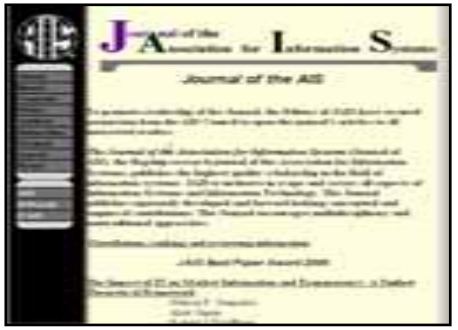
NDIIPP Focus Areas

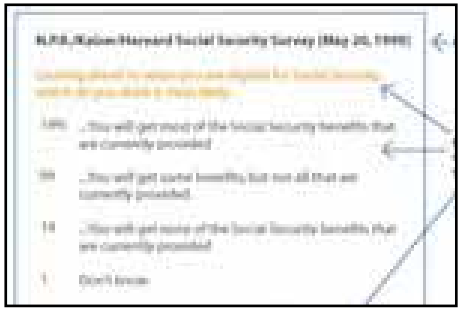


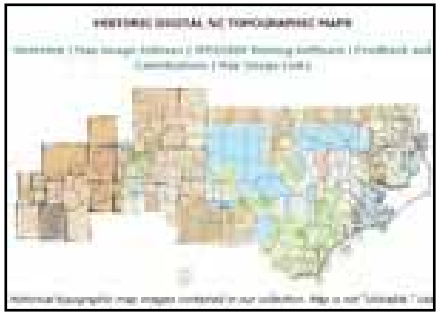
- National Digital Collection
- Partnerships: Government, Industry, Academia
- Technical Infrastructure
- Sustainability
- Public Policy

NDIIPP Content: Collections









Text & Images

Audiovisual

Geospatial

Web sites

Leverage! Leverage! Leverage!

- Cultural Heritage
- Generate Revenue
- Save Money
- Legal Mandate
- Maximize Investment
- Disaster Recovery and Coop
- Document Business Processes for Improved Decision-making
- Ensure Enhanced Access to Support a Variety of Unforeseen Uses

Framing a National Preservation and Access Strategy for Geospatial Data

- November 2009 Summit
  - (Info at http://www.digitalpreservation.gov:8081/news/2009/20091120news_article_geospatial_meeting.html)
- The Library is working with a variety of participants to develop a national strategy for preserving and supporting enhanced access to geospatial data.
- Participants included: NARA, NOAA, USGS, NSGIC, OGC, Census, EPA, FGDC, academia, others
- Effort is aligned with existing mandates and programs in connection with geospatial information.
- Issues of interest include:
  - Identifying geospatial data types or categories that are most important to preserve
  - Scoping preservation and access models
  - Exploring options for expanding educational and community building activities

Possible Next Steps?

- Working Groups
- Explore Licensing Issues
- Information Clearinghouse/Outreach
- Format Registry Information
- Study of Appraisal and Selection Issues
- Convene Commercial Data Providers
- Data Management Policies
- Template for Cooperative Agreements
- Build on existing NDIIPP Projects

Taking it to the States: Geoarchving with GeoMAPP

Alec Bethune

North Carolina Center for Geographic Information and Analysis

Preserving State and Local Government Data

- Initial challenges faced
  - Determining what content to preserve
  - Creating relationships
  - Identifying technical challenges

- What data is important
  - Dynamic “At Risk” datasets
  - Framework data with temporal value

Who is GeoMAPP?

- NC 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

- 10 Informational Partners: DC, GA, ME, MD, MN, MT, NY, TX, WI, WY

Digital Preservation Points of Failure

- Data is not saved, or …
- can't be found, or …
- media is obsolete, or …
- media is corrupt, or …
- format is obsolete, or …
- file is corrupt, or …
- meaning is lost

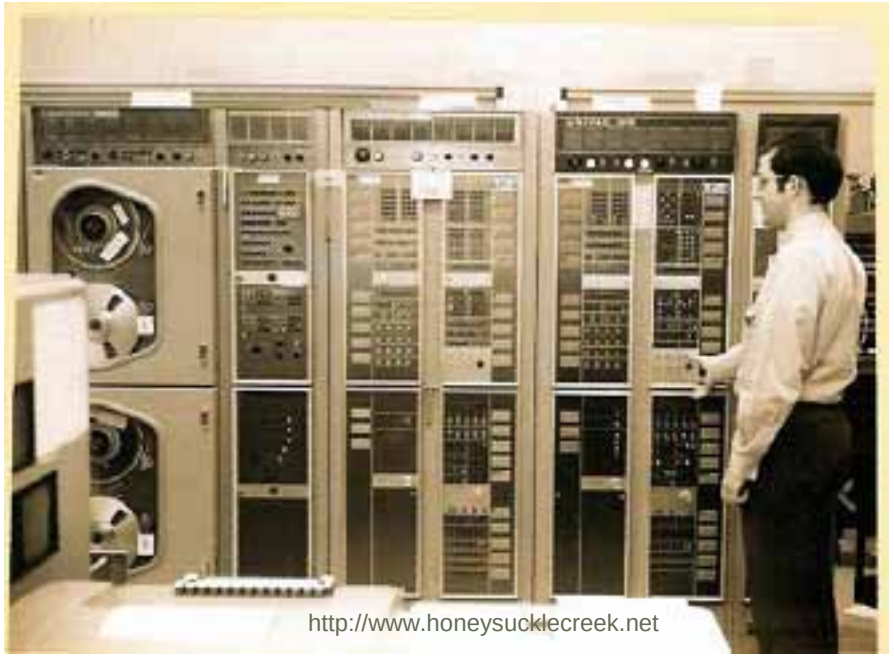


Photo Courtesy
Steve Morris
NCSU Libraries

How would you describe your geospatial archive?

Bob's hard drive

Last week's set of nightly tape backups

Several boxes of CD's and DVD's

The data back-end for our internet mapping application

A collection of files in our "GIS Folder"

A stand-alone spatial database

An enterprise GIS

Graphic Courtesy:
Steve Morris
NCSU Libraries

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

Archiving 101

Kelly Eubank
North Carolina State Archives

Archivists Make it Last Longer

- Transfer of responsibly for records maintenance and access
- Trusted source for legal matters
- Policies to address the long term access and utility of the data

NC Government Records Center

North Carolina Public Records Law

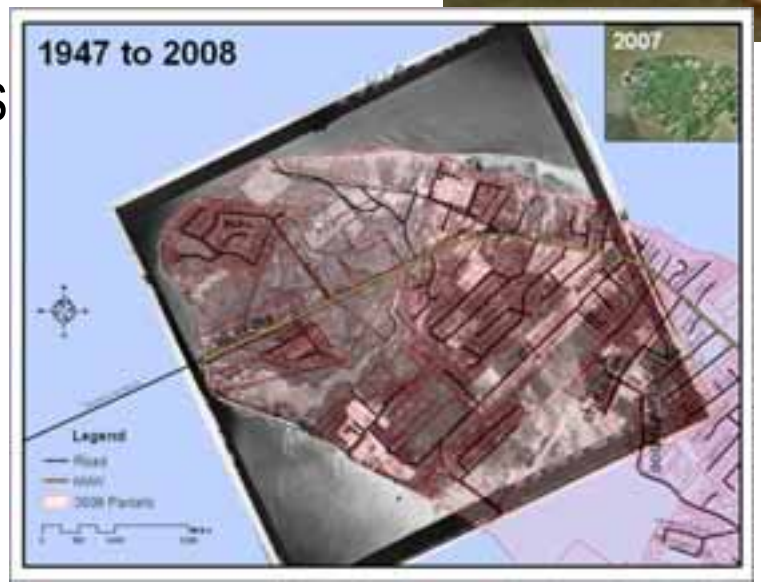
General Statute § 132-1 Public Records Act

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

Geospatial Data are Public Records

Thinking of GIS Data as Records

- Vector data
  - Shapefiles (3-6 separate files)
  - Geodatabases
- Raster data
  - TIFF images
- Metadata
  - XML files
- GIS Project files
- GIS Output
  - GeoPDF files
  - Printed maps



Records Maintenance Principles

- How do records end up at the State Archives?
- Under what authority does DCR advise agencies?
- How do employees determine how long to maintain a record prior to disposal?
- Who makes the decision?
- How should electronic records be maintained?

  (Hint: Back-up is not an archive.)

Records Retention Schedules 101

- Agreement between government (local/state) and State Archives
- Lists common types of records (record series) found in government offices
- Provides instructions for the retention and disposition of the records series – Uniform descriptions.
- List minimum records retention periods
- Identifies permanent, confidential or restricted records

- http://www.records.ncdcr.gov/local/default.htm

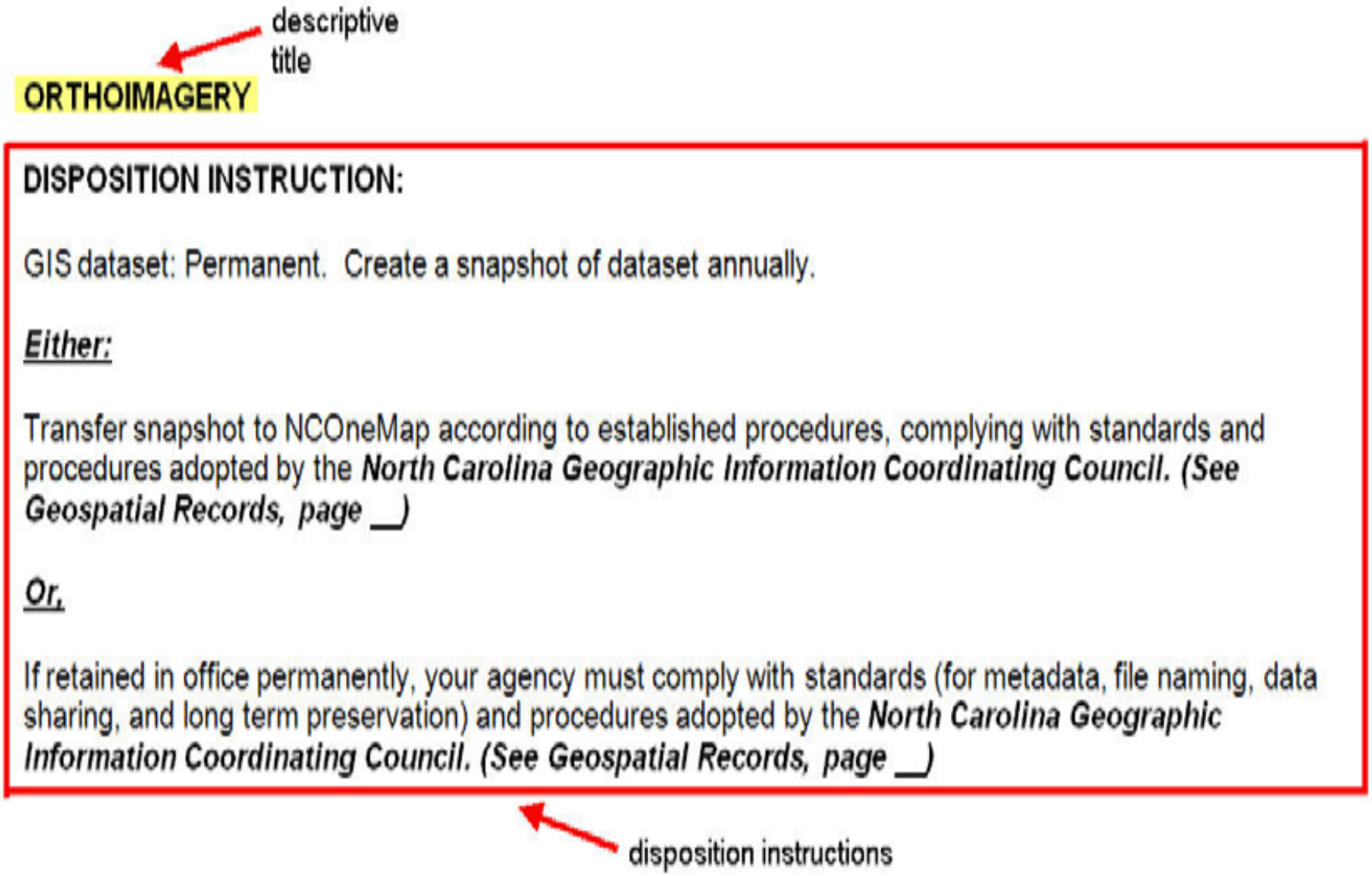
Sample Schedule Page

ITEM #	STANDARD-: PROGRAM RECORDS: GIS RECORDS		
	RECORD SERIES TITLE	DISPOSITION INSTRUCTIONS	CITATION
	GEOGRAPHIC INFORMATION SYSTEM (GIS): OPERATIONAL RECORDS Includes user guides, system flowcharts, job or workflow records, system specifications, and similar documentation	Destroy in office system operational records when the system is discontinued or when system data has been transferred to a new operating environment (platform).	
	GEOGRAPHIC INFORMATION SYSTEM (GIS): BACKUP FILE	Destroy in office system backup files when superseded, obsolete, or administrative value ends, whichever occurs first.	
	GEOGRAPHIC INFORMATION SYSTEM (GIS): PROJECT RECORDS	Retain GIS datasets and accompanying documentation (metadata) with historical and/or legal value permanently. Destroy other items when administrative value ends.	
	GEOGRAPHIC INFORMATION SYSTEM (GIS): INTERNAL STANDARDS AND PROCEDURES Include requirements that are intended to make hardware, software and data compatible and that cover data capture, accuracy, sources, base categories, output, and data element dictionaries.	Permanent.	
	Individual Data Sets:		
	ADDRESS POINTS MAPS	Paper: Retain in office permanently. GIS dataset: Permanent. Create a snapshot of dataset annually. ***Either:*** Transfer mapdata to NCOneMap according to established procedures, complying with standards and procedures adopted by the ***North Carolina Geographic Information Coordinating Council. (See Geospatial Records, page ___)*** ***Or,*** If retained in office permanently, your agency must comply with standards (for metadata, file naming, data sharing, and long term preservation) and procedures adopted by the ***North Carolina Geographic Information Coordinating Council. (See Geospatial Records, page ___)***	

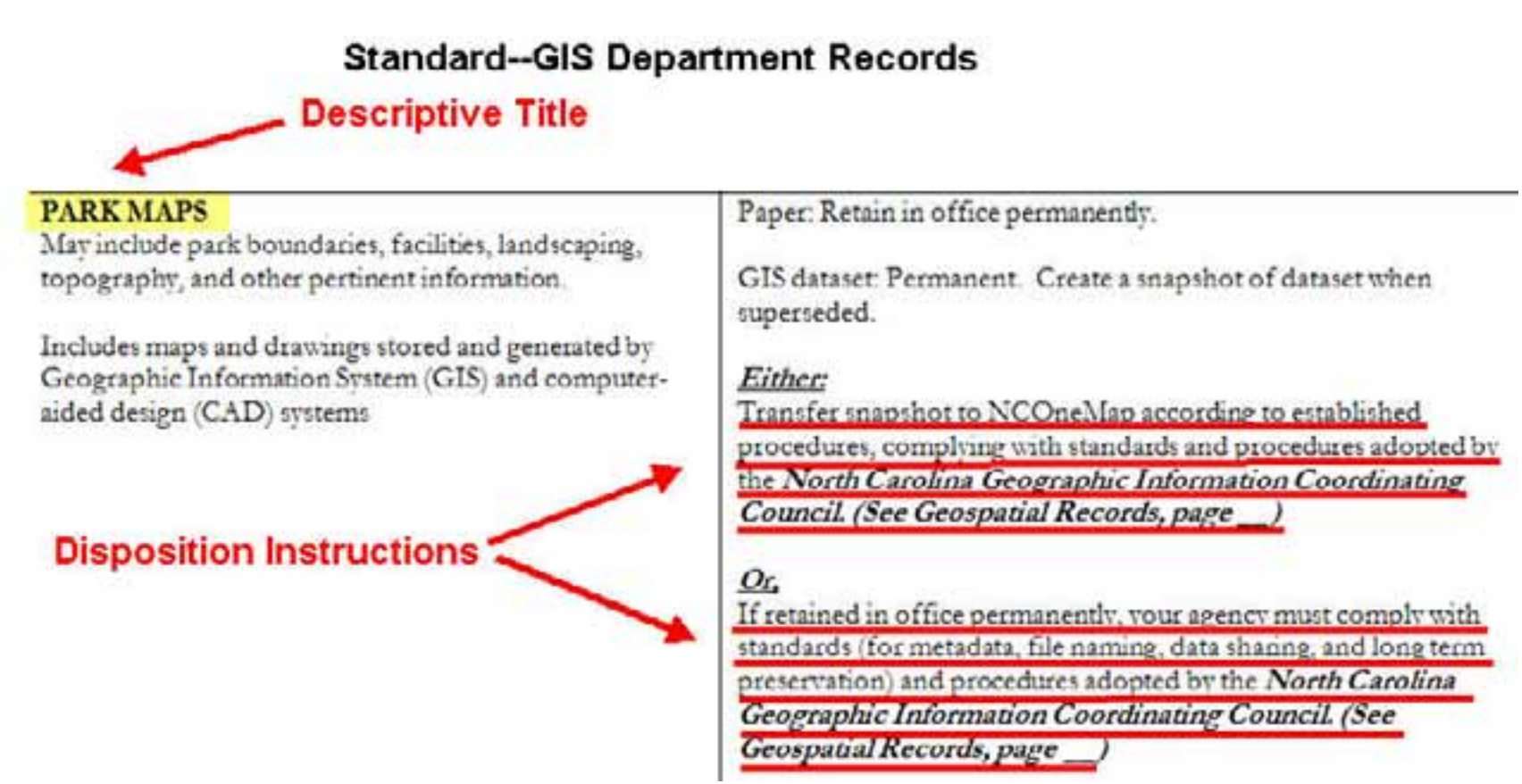
Current State of Geo-Centric Records Schedules in NC

- Maps, photo imagery, paper, and administrative records well represented
- Digital geospatial data…not so much…..
  - CGIA's Records Schedule: last updated in 1997
  - Similar challenges in other state agencies and local governments
- Where to go from here?

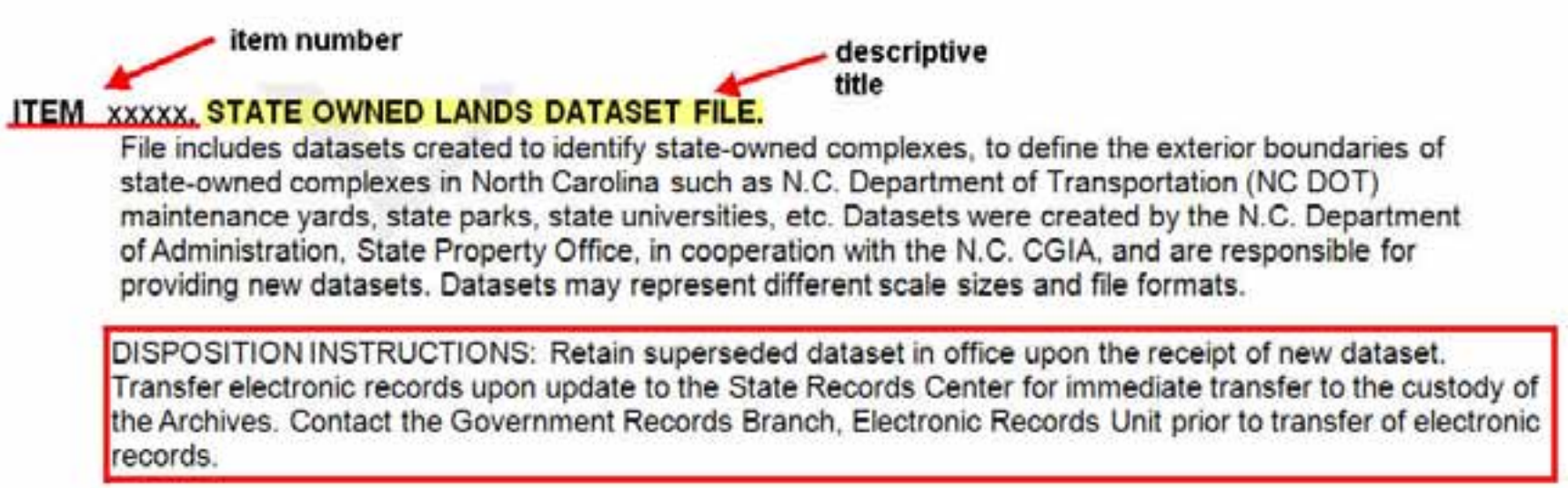
Draft – Local Government Schedule Item



Draft - Schedule Items (Vector)

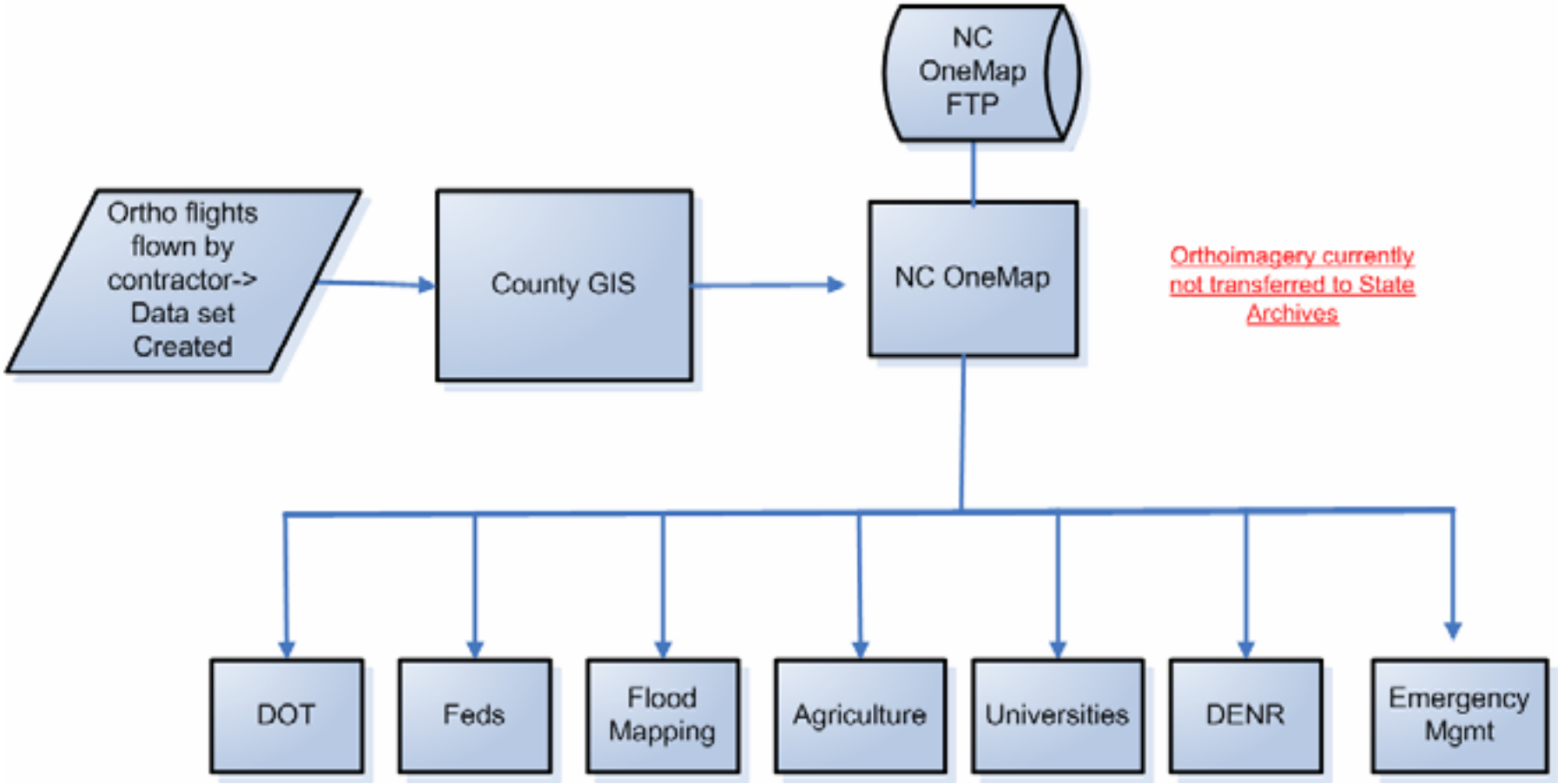


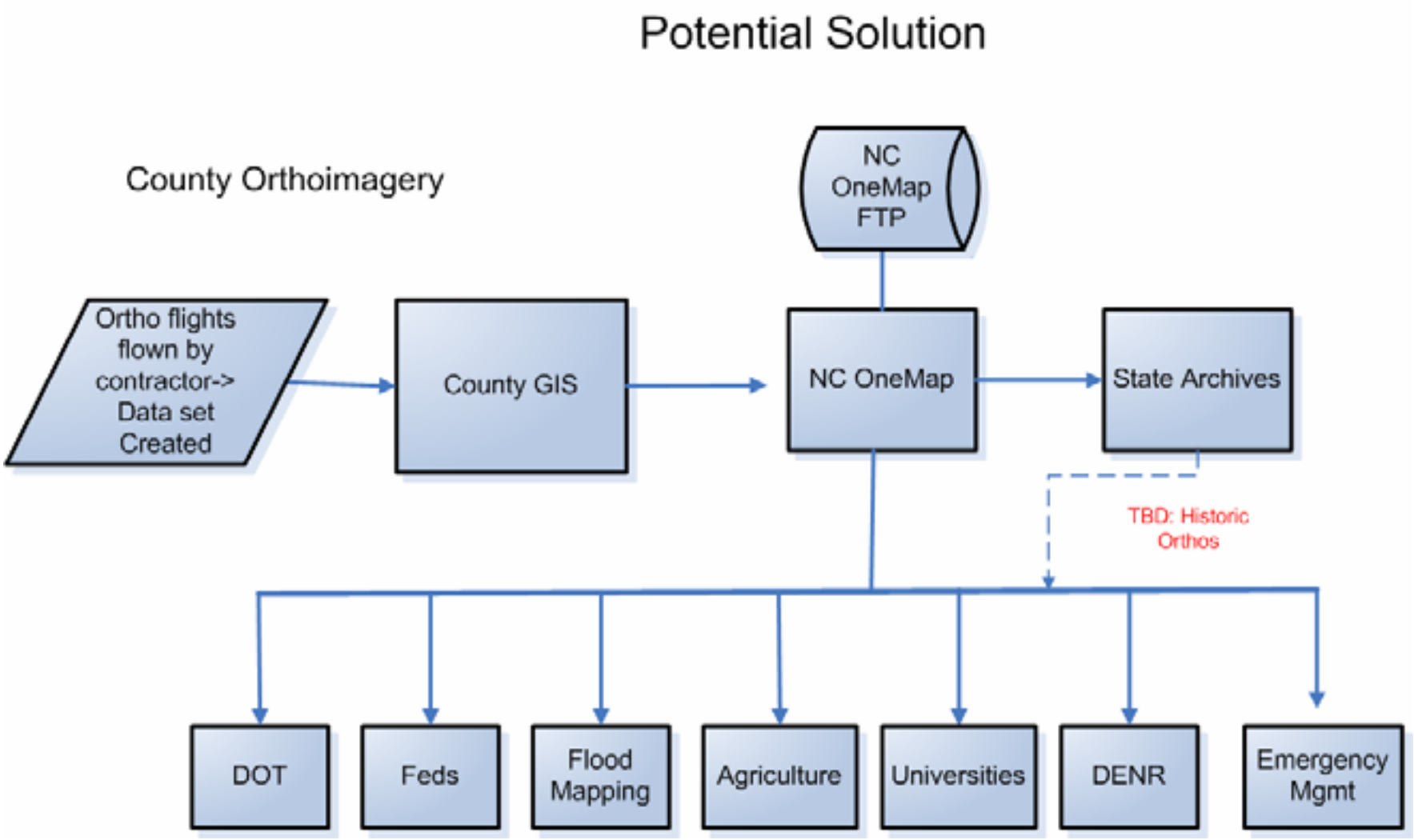
Sample Schedule Item from NC OneMap



Archives Challenge...

County Orthoimagery




Potential Solution
County Orthoimagery
NC OneMap FTP
Ortho flights flown by contractor-> Data set Created
County GIS
NC OneMap
State Archives
TBD: Historic Orthos
DOT
Feds
Flood Mapping
Agriculture
Universities
DENR
Emergency Mgmt

GeoMAPP Key Findings and how they can be applied to archiving within your agency

Alec Bethune

North Carolina Center for Geographic Information and Analysis

GeoMAPP's Approach to Geoarchiving

- Surveys to capture the archival "lay of the land"
- Inventory and comparison of each state's data
- Appraisal of GIS data and creating records schedules
- Business planning for sustainable archives
- Data Transfer
  - Intrastate: GIS→ Archives (NC 15 TB SAN)
  - Interstate: State Archives→ State Archives
- Outreach, Outreach, Outreach
  - Within each state
  - Nationally

Surveying the Geoarchiving Landscape:
National, State and Local Surveys

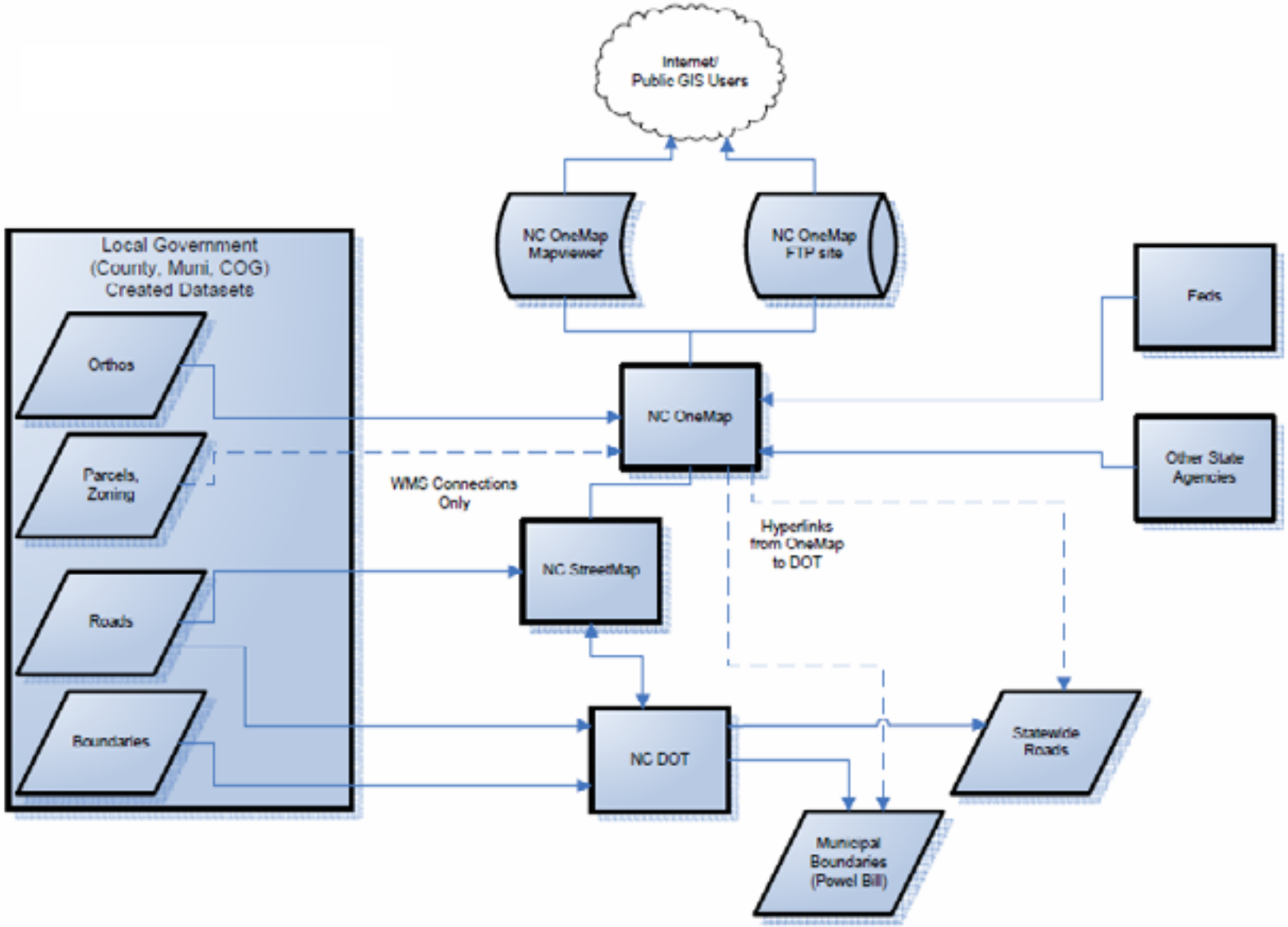
- Targeted Communities:
  - Statewide GIS and Archives Leaders
  - State Agency and Local Government GIS data creators

- Findings:
  - Who is archiving?
  - What are the archiving catalysts?
  - What data are being preserved?

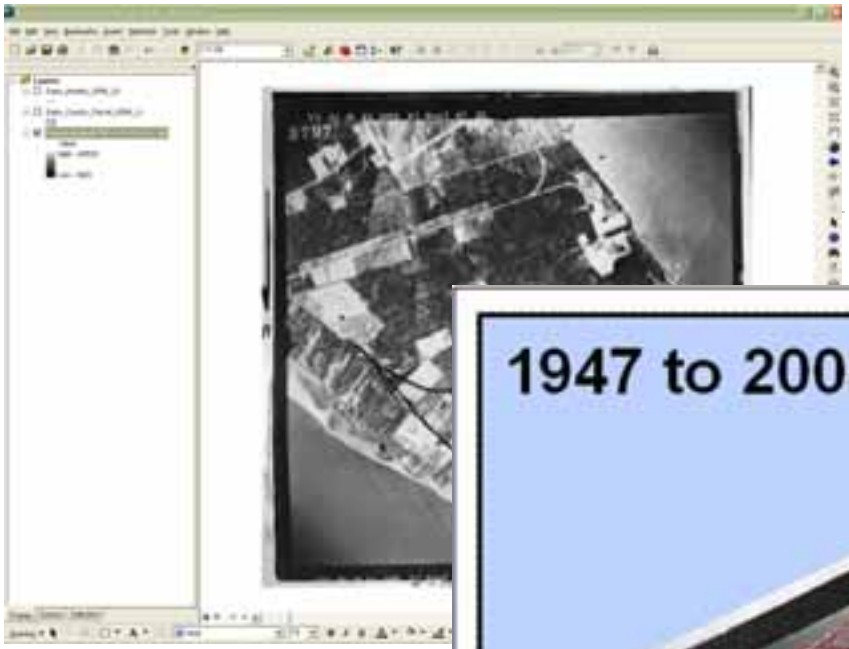
Justifying the Investment:
Developing Business Case Documentation

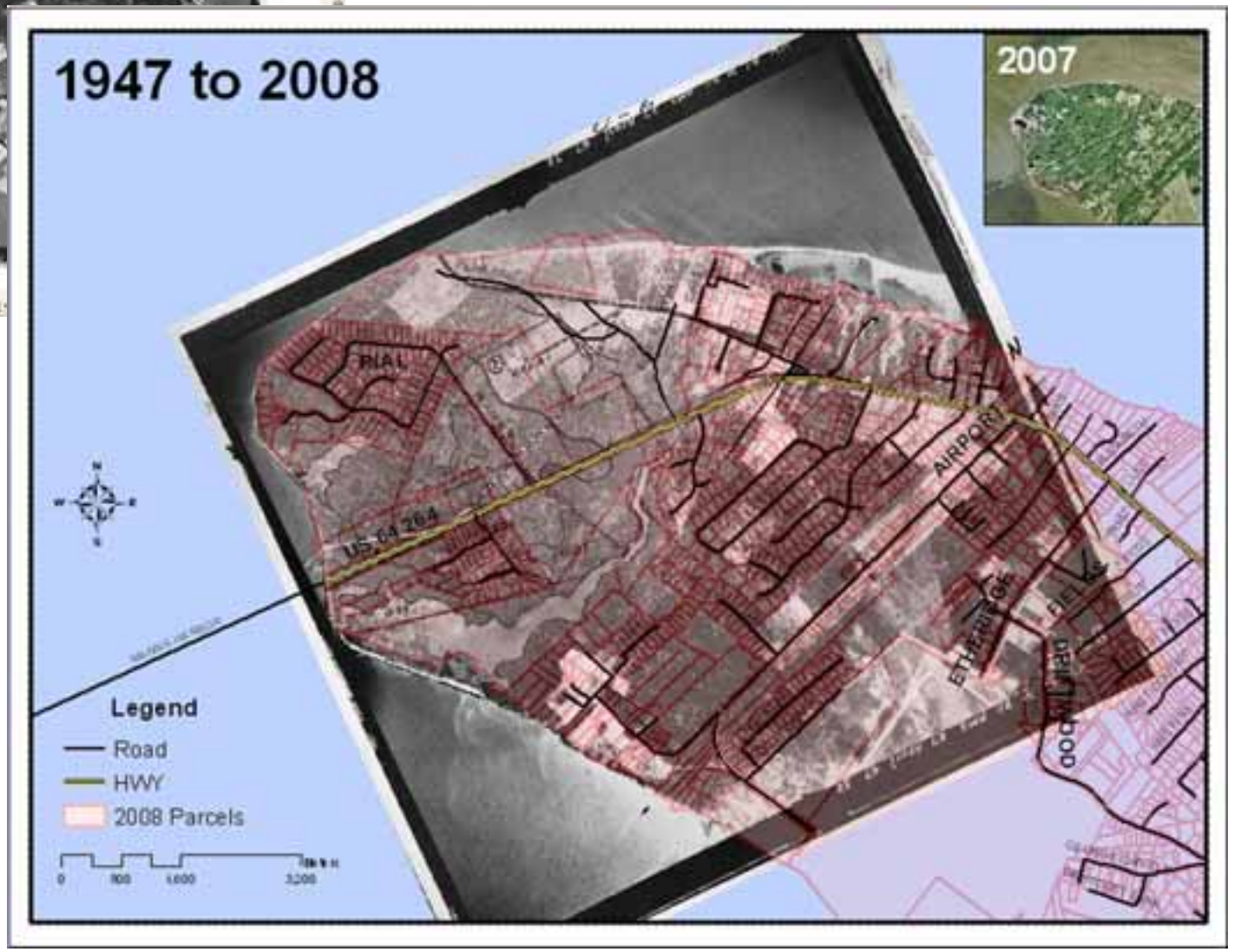
- Develop business case materials to justify the human and technology resources to archive data
- Important themes
  - Value of the data
  - Projecting near and long term costs
  - Cost benefit analysis

Data Transfer is Complicated



Using Historic Data to Model Changing Natural Environments





What's Next

- GeoMAPP 2010
  - A Growing Partnership
    - New and Informational Partners
  - Focus Areas:
    - Business planning for sustainable archives
    - Improving access to archived data
    - Technical exploration
    - Outreach, Outreach, Outreach

Big Picture GeoMAPP Takeaways

- Collaboration is a key
  - GIS & Archives staff
  - Get to know your data producers
- Know what you have (data inventory)
- Make it official (data appraisal/ records scheduling)
- Leverage existing workflows and investigate new sustainable processes to make the data last
  - Don’t re-invent the wheel
  - Keep data discoverable/ accessible/ usable for future use
  - Justify the investment (business case)

What can I do to make my data more useful for others now

…and more “preservation ready” for the future?

Preparing your data for future use

- Naming Convention
  - Both Filename and Attribute names

- Metadata, metadata, metadata!

- Format awareness
  - Versioning and file type

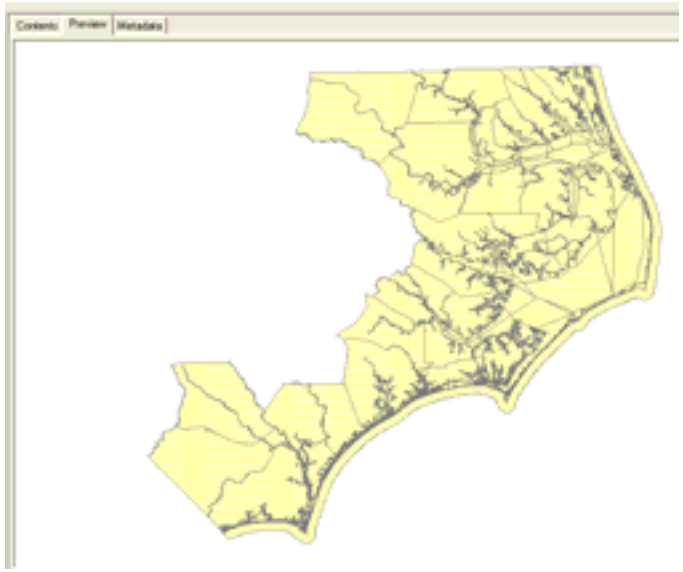
What's in a name?

What is SGA?

Name	Size	Type
sga.dbf	2,502 KB	DBF File
sga.prj	1 KB	PRJ File
sga.sbn	86 KB	SBN File
sga.sbx	4 KB	SBX File
sga.shp	49,668 KB	SHP File
sga.shx	68 KB	SHX File
sga.txt	23 KB	Text Document

How about PES?

Name	Size	Type
pes.dbf	620 KB	DBF File
pes.prj	1 KB	PRJ File
pes.shp	68 KB	SHP File
pes.shx	20 KB	SHX File
pes.txt	26 KB	Text Document



Can the attributes tell us what we're looking at?

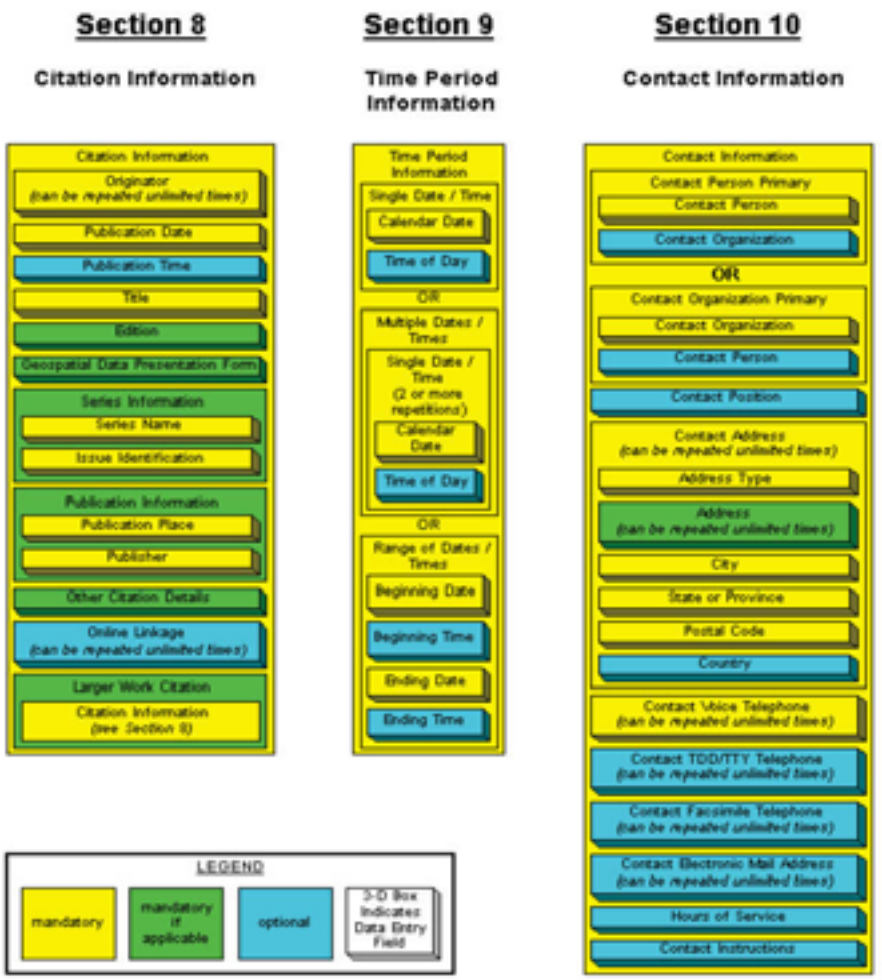
SGA

Attributes of sga

FID	Shape	DSHA_NAME	DSHA_NAME1	HA_CLASS	HA_CLASS1D	HA_NAME	HA_STATUS	COUNTY	SURFACE	DATE_CREAT	DATE_UPDAT	ACRES	SQ_MILES
0	Polygon	Atlantic Ocean						nc	water	02-01-04		726895.340798	1135.93022
1	Polygon							Currituck	land	02-01-04		11868.185721	18.54404
2	Polygon	Currituck Sound Area	I-16	CSHA - Prohibited	CSHA-P	Currituck Sound	Closed	Currituck	water	02-01-04		80405.80146	125.634065
3	Polygon							Currituck	land	02-01-04		0.715843	0.001119
4	Polygon							Currituck	land	02-01-04		56.946307	0.088977
5	Polygon							Currituck	land	02-01-04		9288.470022	14.513234
6	Polygon	Currituck Sound Area	I-16	CSHA - Prohibited	CSHA-P	Currituck Sound	Closed	Currituck	water	02-01-04		153.970266	0.240579
7	Polygon							Currituck	land	02-01-04		0.110463	0.000173
8	Polygon							Currituck	land	02-01-04		1706.572686	2.666652
9	Polygon							Currituck	land	02-01-04		20.35279	0.031801
10	Polygon							Currituck	land	02-01-04		0.38658	0.000604
11	Polygon							Currituck	land			420.888478	0.657638
12	Polygon	Currituck Sound Area	I-16	CSHA - Prohibited	CSHA-P	Currituck Sound	Closed	Currituck	water	02-01-04		16.863203	0.026349
13	Polygon							Currituck	land			7120.132999	11.124583
14	Polygon							Currituck	land	02-01-04		40.51575	0.063306
15	Polygon	Currituck Sound Area	I-16	CSHA - Prohibited	CSHA-P	Currituck Sound	Closed	Currituck	water	02-01-04		497.105567	0.776727
16	Polygon							Currituck	land			0.07873	0.000123
17	Polygon							Currituck	land	02-01-04		90003.122131	140.629878
18	Polygon							Currituck	land	02-01-04		10.352886	0.016176
19	Polygon							Currituck	land	02-01-04		3.437779	0.005372

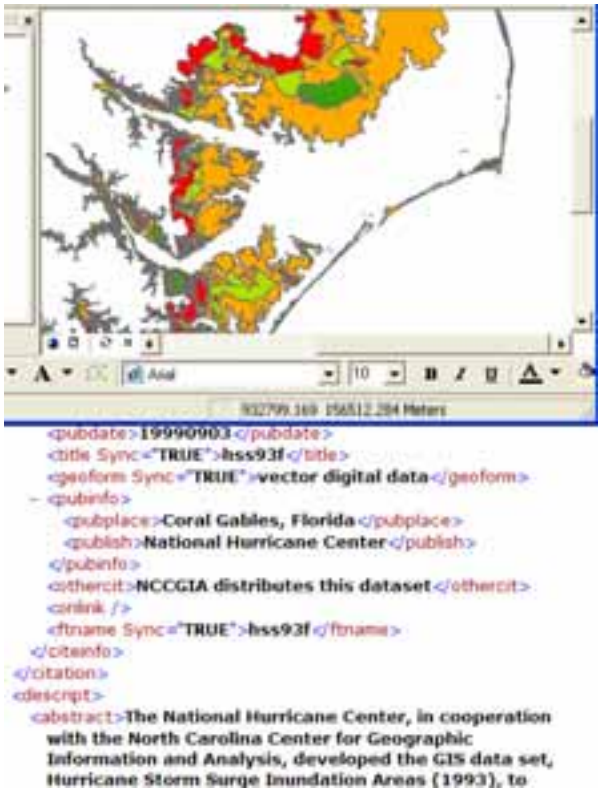
GIS Metadata

```
- <citeinfo>
    <origin>NC, DENR-Division of Marine Fisheries; NC, DENR, DEH, Shellfish
      Sanitation and Recreational Water Quality Seciton</origin>
    <pubdate>20090415</pubdate>
    <title Sync="TRUE">Shellfish_Growing_Areas_2009_04</title>
    <geoform Sync="TRUE">vector digital data</geoform>
    <ftname Sync="TRUE">Shellfish_Growing_Areas_2009_04</ftname>
    <onlink>www.nconemap.com</onlink>
  </citeinfo>
</citation>
- <descript>
  <abstract>The North Carolina (NC) Department of Environment and Natural
    Resources (DENR) Division of Environmental Health (DEH) - Shellfish
    Sanitation and Recreational Water Quality Section (SS+RWQS) identifies
    Shellfish Growing Areas (SGA). The SS+RWQS performs water sampling
    throughout the year to determine the extent of contamination or
    cleanliness of SGA, and submits recommendations to the Division of Marine
    Fisheries (DMF) to close or open waters for shellfish harvesting. The DMF
    disseminates a proclamation with the SS+RWQS's recommendation. The
    SGA contain waters that are permanently closed (prohibited), open
    (approved), or subject to being opened or closed (conditionally approved -
    open or conditionally approved - closed).</abstract>
  <purpose>The SGA contain waters that are permanently closed (prohibited),
    open (approved), or subject to being opened or closed (conditionally
    approved - open or conditionally approved - closed).</purpose>
  <supplinf>For online maps of growing areas and more information, see
    http://www.deh.enr.state.nc.us/shellfish/maps.htm</supplinf>
  <langdata Sync="TRUE">en</langdata>
</descript>
- <timeperd>
  - <timeinfo>
    - <sngdate>
        <caldate>20090415</caldate>
```



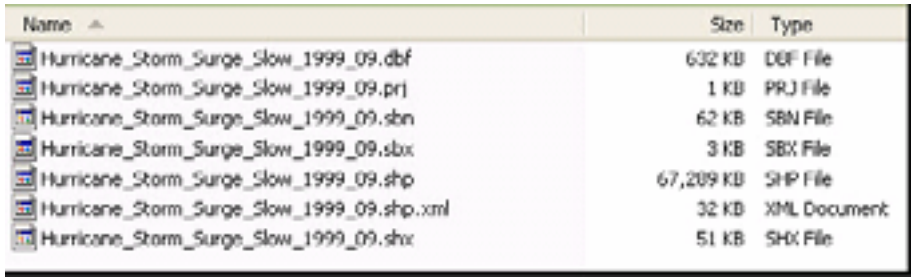
Why Do Metadata?

- Document the specifics about your data:
  - The purpose of the dataset
  - When it was created
  - How it was created
  - Information about the attributes
  - Technical specs
  - ..and much, much more
- Increases data utility
- Improves data management/ discovery
- Makes data more transferrable



ESRI Format Awareness

- Software Versioning
- Data Formats
  - Shapefile
    - Open format/ complex object
    - Older standard

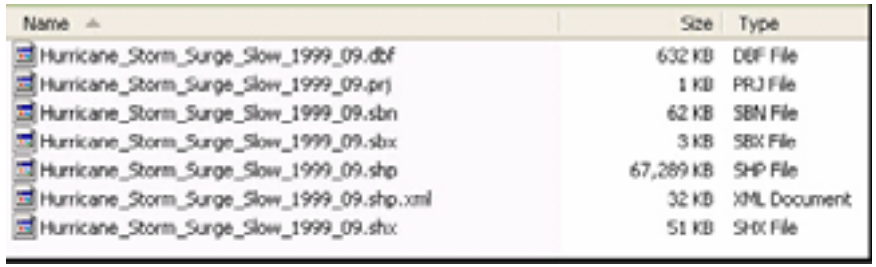


  - Geodatabases
    - Personal

      - Proprietary/ consolidated object/
      - Replaced by FGdB
    - File
      - Proprietary/ complex object
      - Current ESRI “standard”

Best Practices Recap

- File naming
  - Descriptive title
    - Wake_Parcels_2006
    - Shellfish_Growing_Areas_2009
- Attributes
  - Logical Name
  - Explanation in metadata record
- Metadata
  - Do IT!
  - Ideally FGDC compliant
  - Important fields: Title, Abstract, Publication Date, Contact Info, Process steps, Attributes description



Name	Size	Type
Hurricane_Storm_Surge_Slow_1999_09.dbf	632 KB	DBF File
Hurricane_Storm_Surge_Slow_1999_09.prj	1 KB	PRJ File
Hurricane_Storm_Surge_Slow_1999_09.sbn	62 KB	SBN File
Hurricane_Storm_Surge_Slow_1999_09.sbx	3 KB	SBX File
Hurricane_Storm_Surge_Slow_1999_09.shp	67,289 KB	SHP File
Hurricane_Storm_Surge_Slow_1999_09.shp.xml	32 KB	XML Document
Hurricane_Storm_Surge_Slow_1999_09.shx	51 KB	SHX File

Contacts

- Butch Lazorchak (Library of Congress)
  wlaz@loc.gov
  - Alec Bethune (North Carolina CGIA)
    alec.bethune@its.nc.gov
- Kelly Eubank (North Carolina Archives)
  kelly.eubank@ncdcr.gov

http://www.geomapp.net