



# Preservation of Digital Geospatial Resources: A Team Climb

Geospatial Multistate Archive and Preservation Partnership

NSGIC Annual Conference | September 9, 2008 | Keystone, CO

# Library of Congress, NDIIPP, and Fifty States

Butch Lazorchak  
Library of Congress

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

---

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



## E-JOURNALS

- Scholarly electronic journals
- Small-circulation on-line journals

## STATE LEGISLATION & AGENCIES

- Laws, bills, proceedings, committee reports
- Rule-making and regulatory documentation
- Judicial decision, opinions, reports, & rules

## GEOSPATIAL

- At-risk state, regional, and local Gov't data (e.g., Infrastructure maps, Jurisdictional boundaries)
- Aerial and satellite imagery
- Atlantic and Pacific coastal imagery

## WEB SITES & BLOGS

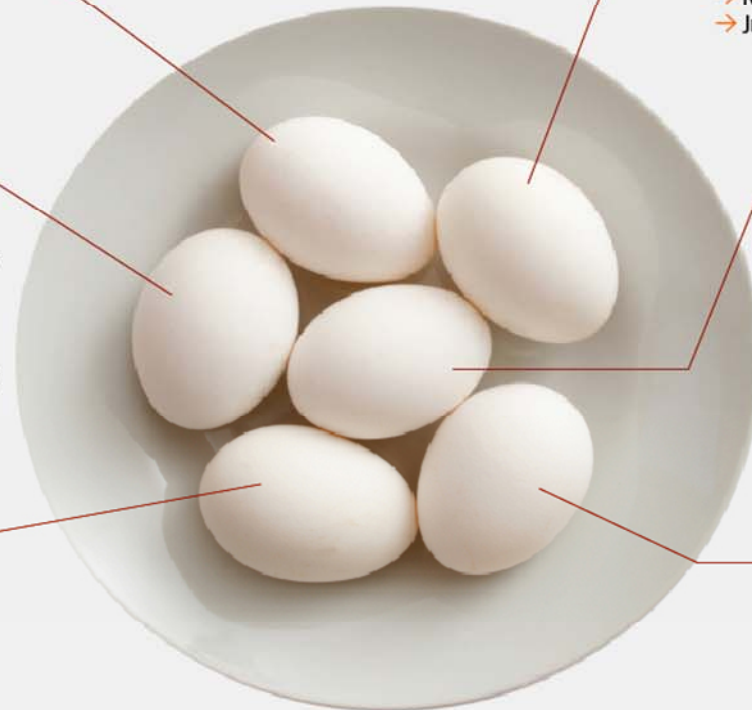
- International political movements
- Elections, Congressional Confirmations, Legal blogs
- Politics & social movements in the Western US
- Hurricane Katrina aftermath

## SOCIAL SCIENCE DATASETS

- Regional and national polling data
- Social and economic surveys
- Output of government-funded research (e.g., NSF, NIH)
- Surveys of international opinion (e.g., USIA datasets 1952-1999)

## VIDEO

- Foreign news programming (SCOLA)
- U.S. television news
- Public television (e.g., Frontline, Nature)



**BORN DIGITAL CONTENT**

# North Carolina Geospatial Digital Archiving Project (NCGDAP): 2003

---

**Lead Partner:** North Carolina State University Libraries

**Additional Partner:** North Carolina Center for Geographic Information and Analysis

## **Objectives:**

- ▶ Intrastate partnerships
- ▶ Identify and acquire at-risk geospatial data



# State-Level Preservation and Access Challenges

---



- ▶ Capability
- ▶ Authority
- ▶ Funding and Staffing

# NDIIPP Work with State and Local Governments

---

- ▶ Workshops
- ▶ Reports
- ▶ Requests for Interest
- ▶ Funding



# Preserving State Government Information Initiative: 2008

---

- ▶ AZ State Library, Archives and Public Records
  - ▶ FL, NY, SC, WI
- ▶ MN Historical Society
  - ▶ CA, IL, KS, MS, TN, VT
- ▶ NC Center for Geographic Information and Analysis/NC Office of Archives and History
  - ▶ KY, UT
- ▶ WA State Archives
  - ▶ AK, CA, CO, ID, IN, LA, MT, OR





# Desired Outcomes of State Initiatives

---



- ▶ Acquire and provide access to digital content
- ▶ Model best practices
- ▶ Catalyze collaboration
- ▶ Demonstrate concrete results
- ▶ Share lessons learned

# Geospatial Multistate Archive and Preservation Partnership (GeoMAPP)

---

**Lead Partner:** North Carolina Center for Geographic Information and Analysis

**States:** NC, KY, UT

## **Objectives:**

- ▶ Interstate partnerships
- ▶ Implement a geographically dispersed content-exchange network
- ▶ Explore data replication among several states





# What is at risk?

Steve Morris  
North Carolina State University

# How Would You Describe Your Current 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

# 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

## Solutions:

Migration  
Encapsulation

Emulation  
XML



## Problem: Temporal Data Unavailability

---

- ▶ Industry focus on “latest and greatest” data
- ▶ “Kill and fill” as a common approach to data management (past versions of vector data lost)

### Not just data loss, also: Loss of memory about data

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

*Some older inventories only available through Internet Archive*

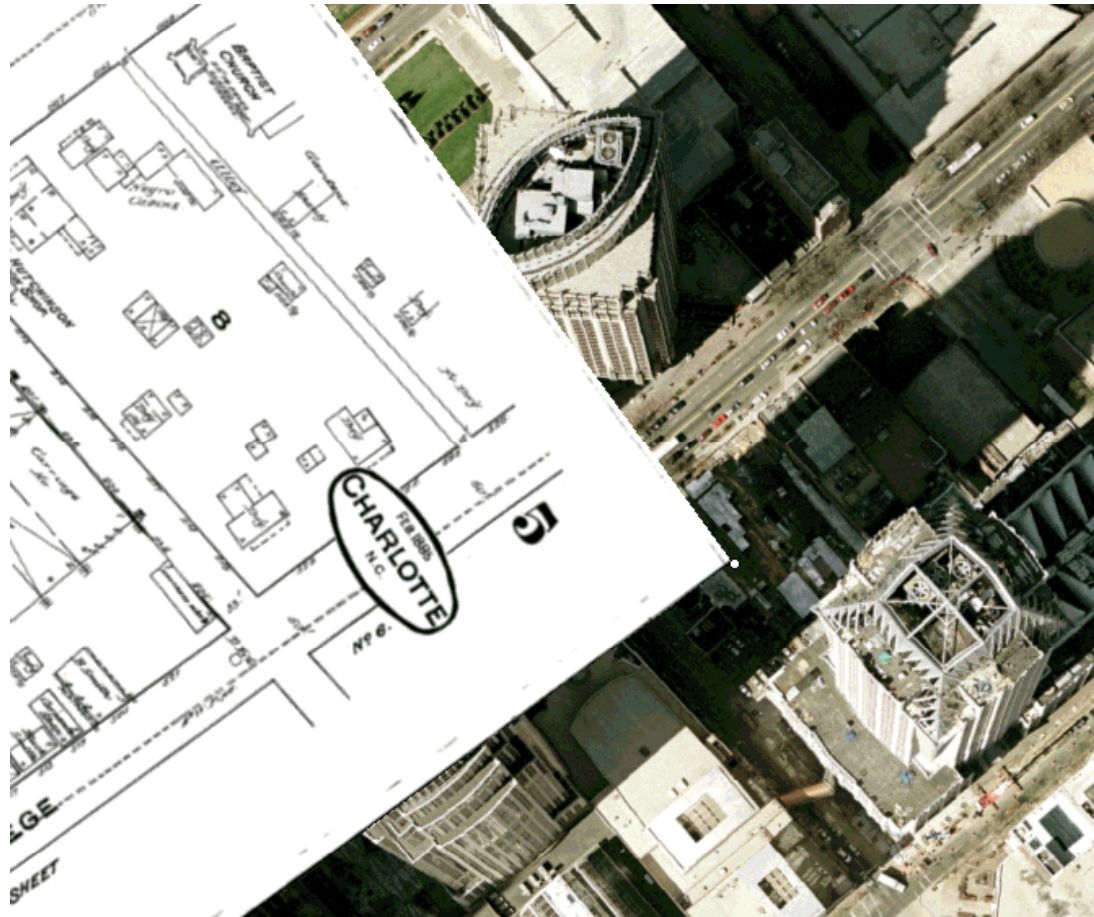
# Findings from Survey of Archiving Practice in NC Local Agencies

---

- ▶ “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.”

# Value in Older Data: Cultural Heritage

Future uses of data are difficult to anticipate (as with Sanborn Maps)





# Value in Older Data: Solving Business Problems

---

**Land use change analysis**

**Site location analysis**

**Real estate trends analysis**

**Disaster response**

**Resolution of legal challenges**

**Impervious surface maps**



**Suburban Development 1993/2002**

Near Mecklenburg County-Cabarrus County NC border

## Different Ways to Approach Preservation

---

- ▶ Technical solutions: How do we preserve acquired content over the long term?
- ▶ Cultural/Organizational solutions: How do we make the data more preservable—and more prone to be preserved—from point of production?

*Current use and data sharing requirements – not archiving needs – are most likely to drive improved preservability of content and improvement of metadata*

# Spatial Data Infrastructure: Where Does Archiving Fit?

---

- ▶ **Metadata standards and outreach**
  - ▶ metadata quality, best practices
- ▶ **Inventories**
  - ▶ Reduce “contact fatigue”, shareable info store
- ▶ **Content exchange networks**
  - ▶ Leverage more compelling business reasons to put data in motion
  - ▶ Automate process, add technical & administrative metadata
- ▶ **Framework data communities**
  - ▶ Snapshot frequency, schemas, format strategies

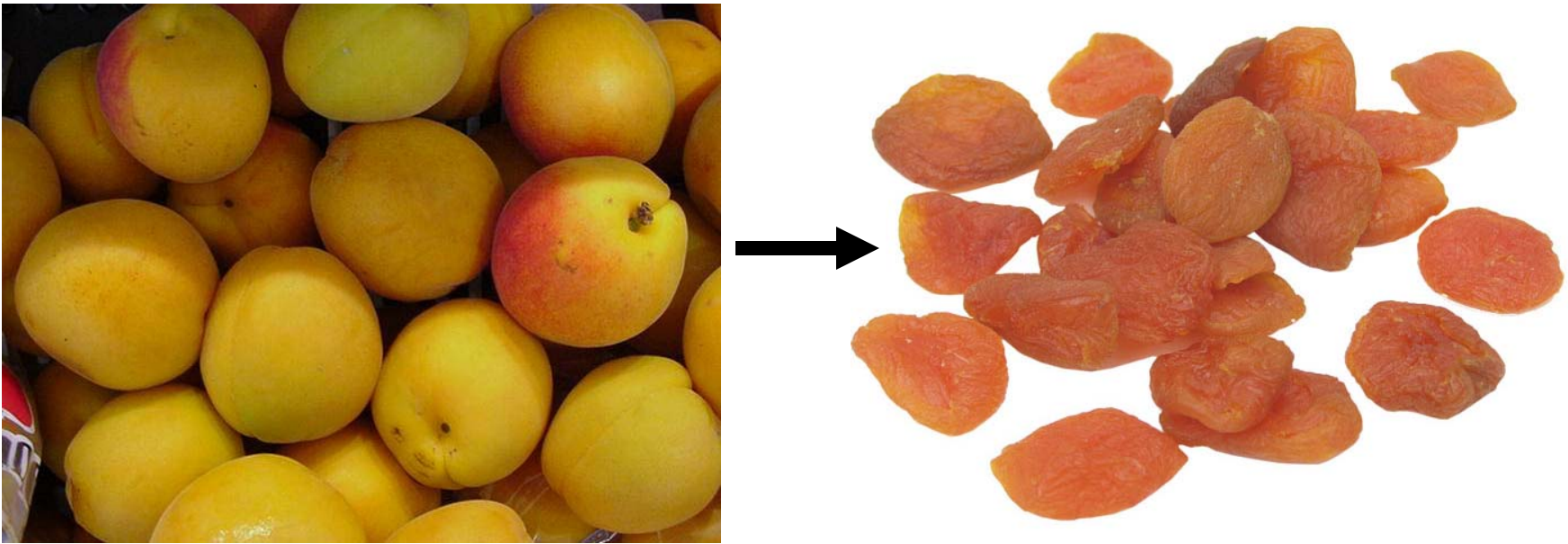
# Technical Challenges with Geospatial Data

---

- ▶ **Complex vector formats: multi-file, multi-format**
  - ▶ No non-commercial, well-supported format
- ▶ **Shift to web services-based access**
  - ▶ Data ephemeral, how to record decisions?
- ▶ **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
- ▶ **Content packaging**
  - ▶ No geospatial industry standard

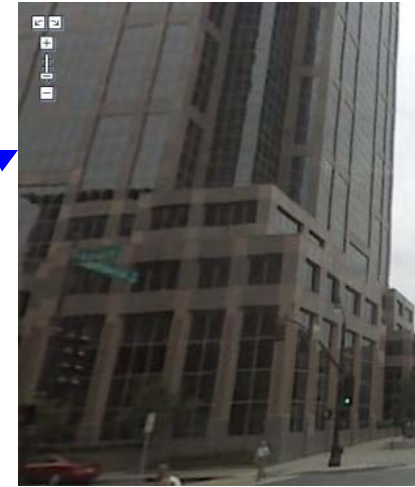
# Preservation Approaches: Original Data vs. “Dessicated” Data

---



Complex data representations can be made more preservable (and less useful) through simplification

# Changes in the Domain: New Location-based Content



Future value as cultural heritage resource:  
*More descriptive of place and function than spatial data*



0032727 03/25/2007

# Changes in the Domain: Geospatial PDF

## PDF and GeoPDF

General Maps - Microsoft Internet Explorer provided by NCSU Libraries

File Edit View Favorites Tools Help

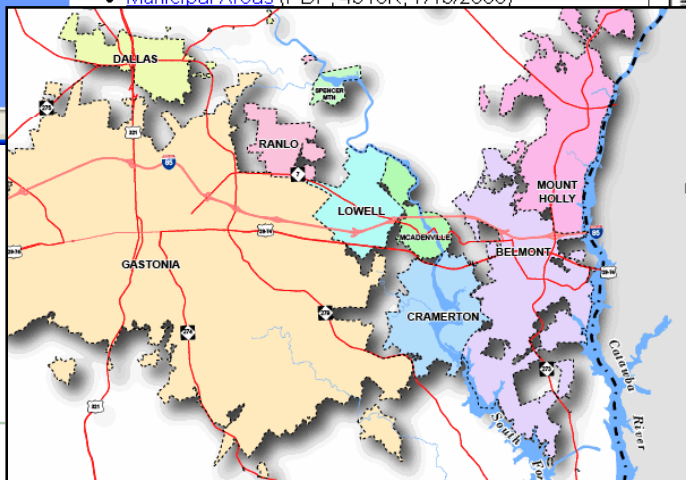
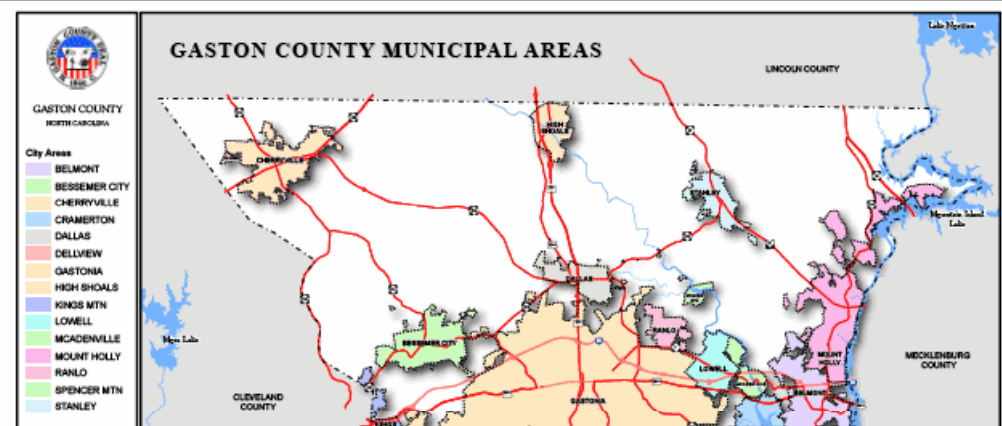
Address <http://www.co.gaston.nc.us/Planning/maps/GeneralMaps.htm> Go Links SnagIt

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](#) (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 Overlap](#) (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)



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

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



# Technology

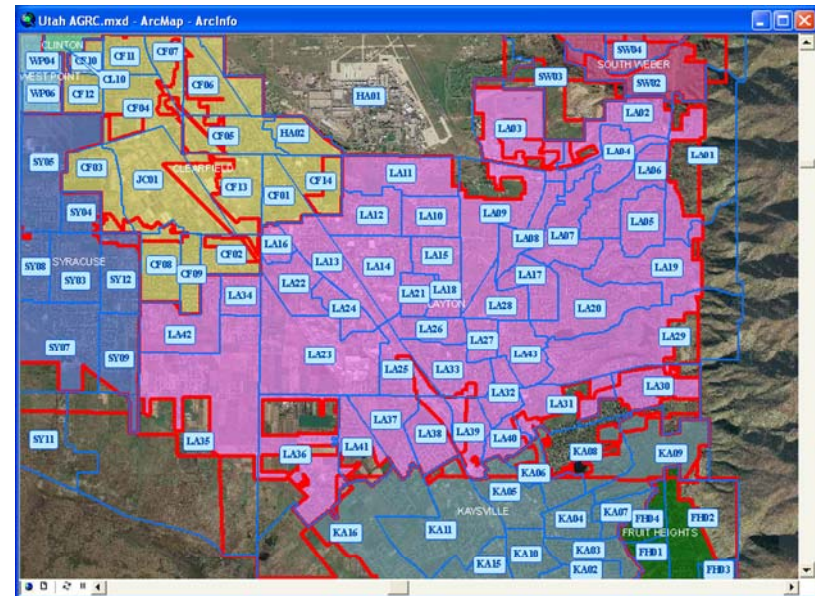
Matt Peters

Utah Automated Geographic Reference Center



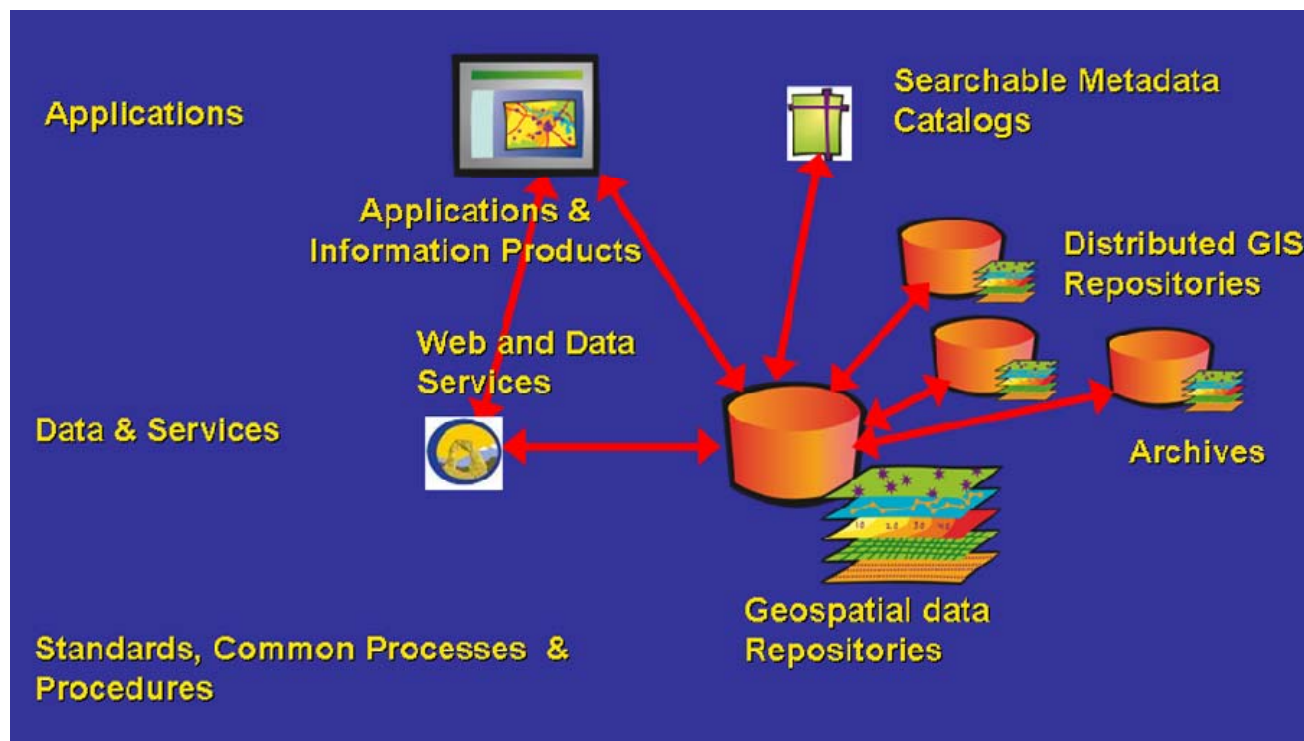
# State of Utah

- ▶ **Governance of the Utah Geospatial Infrastructure**
  - ▶ Department of Technology Services / State CIO
  - ▶ Automated Geographic Reference Center
  - ▶ Utah GIS Advisory Committee
  - ▶ Supported by dozens of “responsible” partners
- ▶ **State Geographic Information Database (SGID):**
  - ▶ Centralized shared database (350 data themes)
  - ▶ Internet portal viewer (GIS.UTAH.GOV)
  - ▶ Web Services and applications (MAPSERV.UTAH.GOV)



# State of Utah

## Components of a State SDI





# Business Plan

Elizabeth Perkes  
Utah Archives

# Geospatial Data Preservation Business Case Drivers

---

- ▶ Loss of investment if records are mismanaged or stored in vulnerable ways
- ▶ Legal requirements for preservation of data to ensure decisions and analysis are documented and repeatable
- ▶ Government Records Access and Management Act requires information is available to public
- ▶ Management of records enables less duplication of storage by multiple agencies
- ▶ Data must comply with Archival Professional Guidelines

# Geospatial Data Preservation Business Case Objectives

---

- ▶ Support State Archivist's responsibility to preserve digital geospatial data
- ▶ Provide assistance to the geospatial community
- ▶ Establish procedure and mechanism for inventory of state and local data
- ▶ Enable compliance with GRAMA
- ▶ Increase awareness about preservation and archives

# Inventory Process

---

- ▶ Deciding series boundaries: individual layers vs. layer groups
- ▶ Structuring retention schedules with growth in mind
- ▶ Relationships between state and local government retention schedules
- ▶ Naming conventions: ISO vs. local

# Retention Schedules

SCHEDULE 1  
AUTOMATED GEOGRAPHIC REFERENCE  
Geospatial Data Sets

---

1-17.

**UTILITIES AND COMMUNICATION RECORDS**

These geospatial records pertain to utility routes and distribution facilities located in Utah. These include coal seams, mine locations, oil and gas well locations, oil, gas, and water pipelines, telephone service areas, water distribution facilities, electrical lines, and communication towers.

**RETENTION:**

Permanent. Retain by State Archives.

**PRIMARY CLASSIFICATION**

Public.

---



# Creating Finding Aids

---

- ▶ Finding aids help end-users gain access to data
- ▶ Produced by database from retention schedule
- ▶ Repurposing metadata: details included in finding aid
- ▶ Searching issues (GOS, Ramona, other)



# Geospatial Data Sets

**Geo-Spatial Data Sets** [?] [v] [x]

**Series:** 26771 - 1      **Created:** 01/DD/2000 - 02/DD/2007      **File Size:** 3.00 MB  
**Accessioned:** MM/DD/CCYY      **Published:** 07/01/2007      **File Type:** Shapefile

Provide resolution or scale, whichever is most accurate

**Resolution: 1 inch =** 1.50 Meters      **Scale:** 1:24,000      **Accuracy:** [ ]

**Projection:** UTM, Zone 12, meters      **Datum:** NAD83

**Created by Agency:** 1142 [H] Utah Transit Authority

**Created by Other:** [ ]

**GIS Title of Data Set:** BusStops\_UTA

**Descriptive Title:** Utah Transit Authority bus stops

**Local Path to Data:** \\AC02\Data\Users\Archives\GISData\2000\BusStops\_UTA.zip

**URL to Online Data:** http://gis.utah.gov/index.php?option=com\_dbquery&Itemid=87

**URL to MetaData:** ftp://ftp.agrc.state.ut.us/SGID\_Vector/MetadataHTML/SGID\_U024\_BusStops\_UTA.html

[ < Exit ]      [ OK ]      [ Cancel ]      [ Next > ]

# Container List

## Container List

ITEM	DATES	TECHNICAL SPECS	DESCRIPTION
1	2000 Jan - 2007 Feb	<b>Scale:</b> 1:24,000 <b>Resolution:</b> One inch = 1.50 Meters <b>Datum:</b> NAD83 <b>Projection:</b> UTM, Zone 12, meters <b>File Name:</b> BusStops_UTA <b>File Type:</b> Shapefile <b>File Size:</b> 3.00 MB	<u>Utah Transit Authority bus stops</u>





# Inventory, appraisal, and selection

Kelly Eubank  
North Carolina State Archives

# Speaking the same language

---

- ▶ GIS practitioners organize their own data—GeoOne Stop, National Inventory powered by Ramona, ISO
- ▶ Example: Trails
  - ▶ Tourism?
  - ▶ Parks and Recreation?
  - ▶ Natural Resources?
  - ▶ Transportation?
- ▶ Translating GIS speak to Archives speak
  - ▶ Determining layers
  - ▶ What to keep?

# Appraisal

---

- ▶ Making sense of archivists
- ▶ What do we do? How do we do it?
- ▶ How does digital content alter what we do?
  - ▶ Digitized content
  - ▶ Born Digital
- ▶ Traditional work versus work in the digital age
  - ▶ Get it near the beginning. Get it now.
- ▶ Do we keep everything?
  - ▶ House appraisals
  - ▶ Legal, fiscal, historical, evidential values

# Retention Schedule

STATE ARCHIVES AND RECORDS COMMISSION  
 Public Records Division  
 Kentucky Department for Libraries and Archives

Schedule Date: December 09, 1999

## STATE AGENCY RECORDS RETENTION SCHEDULE

Finance and Administration  
 Commonwealth Office of Technology  
 Geographic Information

Series	Records Title and Description	Function and Use	Contents	Retention		
				Disposition Instruction		
06481	Geographic Information System (C) KRS 61.878, amended by HB 59 for Homeland Security Change Date: 9/9/2005 (V)	The GIS database is a series of map layers. These layers store graphic information about map features (parcels, streets, utility lines) and geographic areas (tax neighborhoods, subdivisions, service districts) and are linked to detailed attribute information or descriptions stored in a database. GIS software accesses this integrated graphic and attribute information to support all mapping, geographic query, and analysis applications. Data layers in a GIS database may be categorized as "base map" or "thematic." Base map layers are those that contain map features that are commonly needed for most applications and which serve as a foundation and reference for other map layers. The Commonwealth Map is a base map made up of 12 layers (see Contents). Other thematic data layers contain other map features, normally displayed with base maps, that are needed for specific applications. Spatial features in the GIS database are stored in either vector or raster form. A vector format represents the location and shape of features and boundaries precisely as a string of x,y coordinates. In contrast, the grid-based or raster format generalizes map features as cells or pixels in a grid matrix. Map layers that need to accurately depict map features, such as manholes, street centerlines, or parcel boundaries, most often use a vector format. Map layers that are stored as images, such as digital aerial orthophotographs, use a raster format. Vector and raster map layers may be overlaid and displayed together in a GIS.	Raster graphic image layers and Vector data layers superimposed individually or cumulatively. Organized into categories: Agriculture, Boundaries, Environment, Fish & Wildlife, Geodesy, Geophysical, Historic, Imagery, Institutional, Misc., Soils, Standards, Transportation, Utilities, Water Resources.	Agency: P	Records Center:	Archives Center:
				Retain in agency. Replace information as updated; move data offline as necessary. Provide periodic data snapshots to KDLA.		

## The Commonwealth Map Appraisal Report

+	<b>1.) Category:</b>	Transportation
	<b>2.) Layer:</b>	KYTC State Maintained Roads
	<b>3.) Agency Records Officer Name:</b>	Ann Stansel & Todd Shipp
	<b>4.) Technical Contact Name:</b>	KYTC Office of Technology, GIS & Engineering Branch, Branch Manager
	<b>5.) Business Use:</b>	
	<b>a.) Legal, Administrative, Fiscal Value: Why does the agency maintain these records?</b>	Administrative Value
	<b>b.) Purpose: How are the records used?</b>	Provides statewide coverage of state maintained roads. Provides a base network for evaluating transportation access and a base reference layer for other mapping. Coverage built on county based data collection originally using USGS 1:24000 topographic quadrangles. The data is in the process of being replaced by GPS data. See Standard for KYTC Road Centerlines: <a href="http://giac.state.ky.us/giac_standards_trans.htm">http://giac.state.ky.us/giac_standards_trans.htm</a> .
	<b>c.) Support: What program or programs do they support?</b>	Kentucky Transportation Cabinet
	<b>d.) User: Who uses the records?</b>	Kentucky Transportation Cabinet engineers, other state government employees, general public
	<b>e.) Update Frequency:</b>	Monthly





# The Challenge is On!!

---

- ▶ Link to NSGIC Survey
- ▶ [http://www.surveymonkey.com/s.aspx?sm=mNR5P0TSFyxJbNAPI8pQcA\\_3d\\_3d](http://www.surveymonkey.com/s.aspx?sm=mNR5P0TSFyxJbNAPI8pQcA_3d_3d)



# Kentucky's Perspectives

Ken Bates  
Kentucky GIS  
Kentucky State University

---

▶ Archivists and GISers...

Different Perspectives but not different

GIS....locomotive breath.....faster newer...etc

Archives.. Hey what about the stuff you're throwing out?

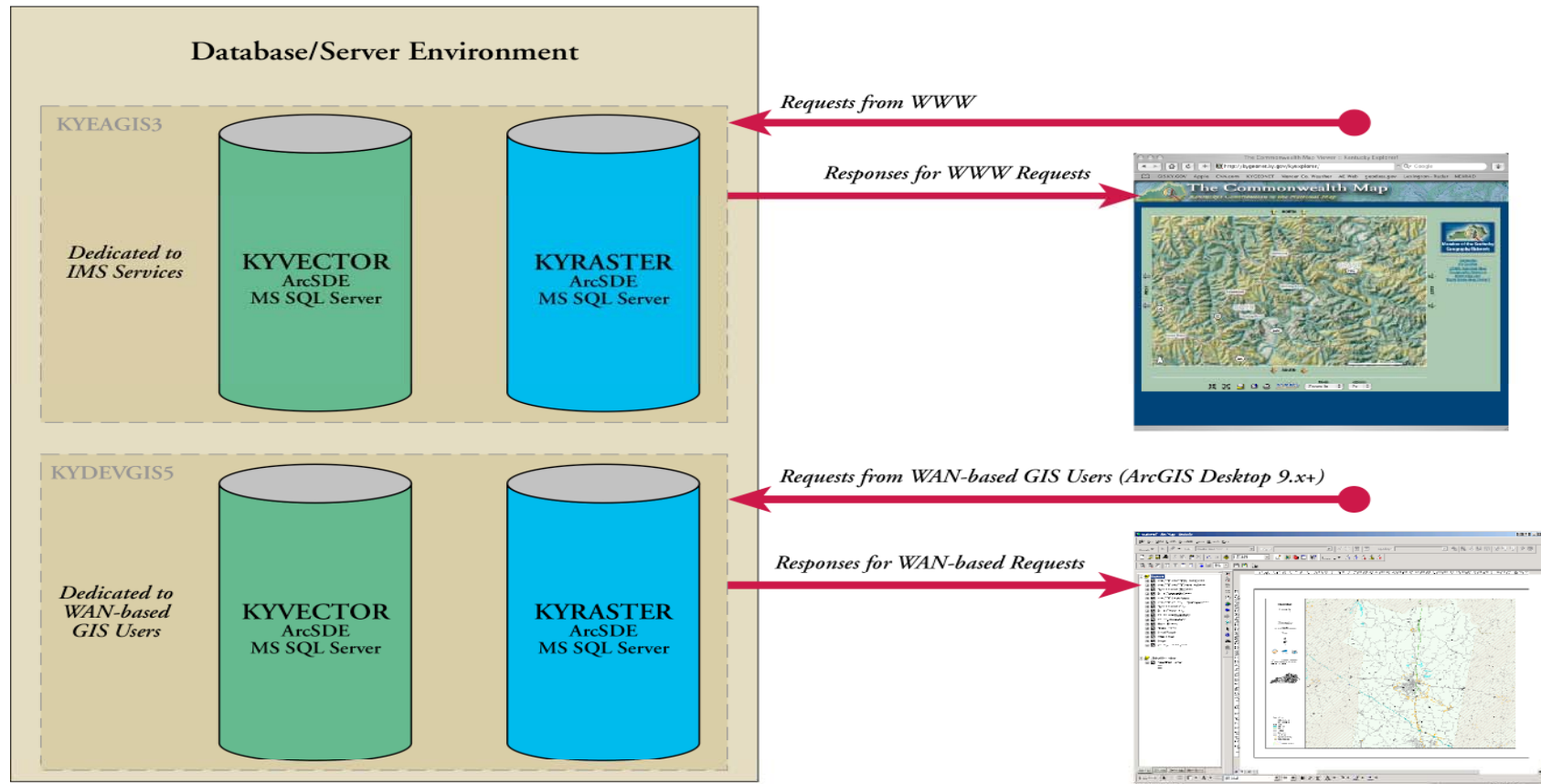
WHAT

Ironically....GISists often talk about change detection

Similar....in the response to questions about other's data..

# Commonwealth of Kentucky

## KYRASTER & KYVECTOR *The Commonwealth's GIS Utility*



KyRaster and KyVector are based on ArcSDE 9.x and Microsoft SQL Server 2000 and run on HP DL560 Proliant Rack Servers with data storage on the Commonwealth's SAN. The services are housed at the Commonwealth's world-class Cold Harbor Computing Center in Frankfort, Kentucky.



# Commonwealth of Kentucky

---

- ▶ ***Things that leverage KyRaster & KyVector ...***
- ▶ KYGEONET - (<http://kygeonet.ky.gov/>)
- ▶ IMS Sites - TCM, TNM, KyParks, KDFWR, KyHydro, KyExplorer, ...
- ▶ GIS Applications - Mine Mapping Application, KEMAP
- ▶ Desktop Users – Most GIS users within the WAN use this resource on a daily basis
- ▶ ....and the GeoSpatial One-Stop
- ▶ Completed on snapshot Archive and transferred to Kentucky State Archives

What do you want to know?

---

▶ Panel and Audience Discussion/Questions

# For More Information

---

[www.geomapp.com](http://www.geomapp.com)



Zsolt Nagy | North Carolina  
CGIA

- ▶ 919-733-2090
- ▶ [zsolt.nagy@ncmail.net](mailto:zsolt.nagy@ncmail.net)



Elizabeth Perkes | Utah Division  
of Archives

- ▶ 801-531-3852
- ▶ [eperkes@utah.gov](mailto:eperkes@utah.gov)



Butch Lazorchak | Library of  
Congress

- ▶ 202-707-2603
- ▶ [wlaz@loc.gov](mailto:wlaz@loc.gov)



Kelly Eubank | NC State  
Archives

- ▶ 919-807-7355
- ▶ [kelly.eubank@ncmail.net](mailto:kelly.eubank@ncmail.net)



Steve Morris | NCSU Libraries

- ▶ 919-515-1361
- ▶ [steven\\_morris@ncsu.edu](mailto:steven_morris@ncsu.edu)



Ken Bates | Kentucky State  
University

- ▶ 502-597-7016
- ▶ [ken.bates@kysu.edu](mailto:ken.bates@kysu.edu)



Matt Peters | Utah AGRC

- ▶ 801-538-3168
- ▶ [mpeters@utah.gov](mailto:mpeters@utah.gov)