Dust Free Data:
Developing Dynamic Geospatial Data Archives

Butch Lazorchak- Library of Congress
Joe Sewash- NC Center for Geographic Information & Analysis

NSGIC Midyear Meeting | 28 February 2011 | Annapolis, MD

Project Overview

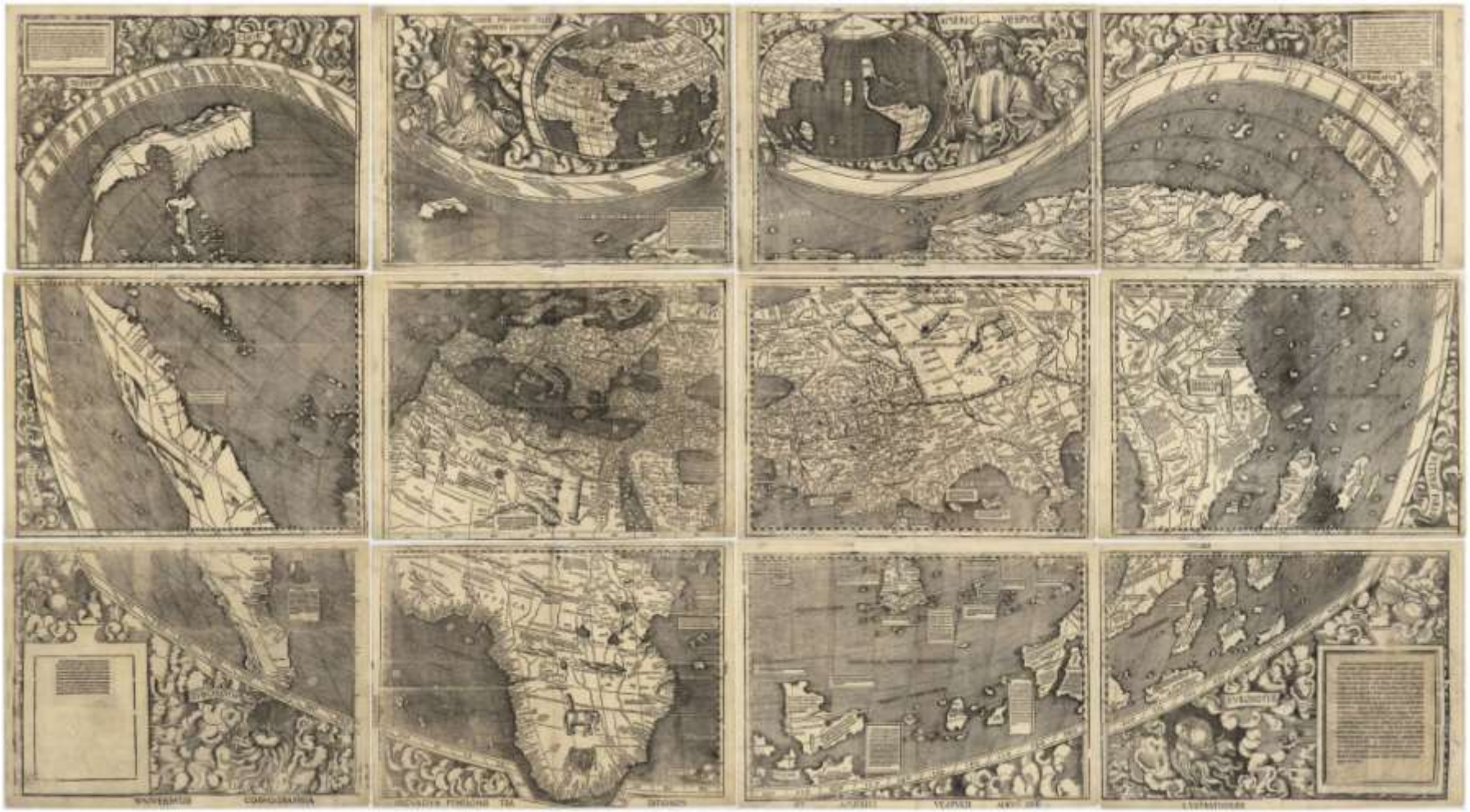
Butch Lazorchak
Library of Congress National Digital Information Infrastructure and Preservation Program

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

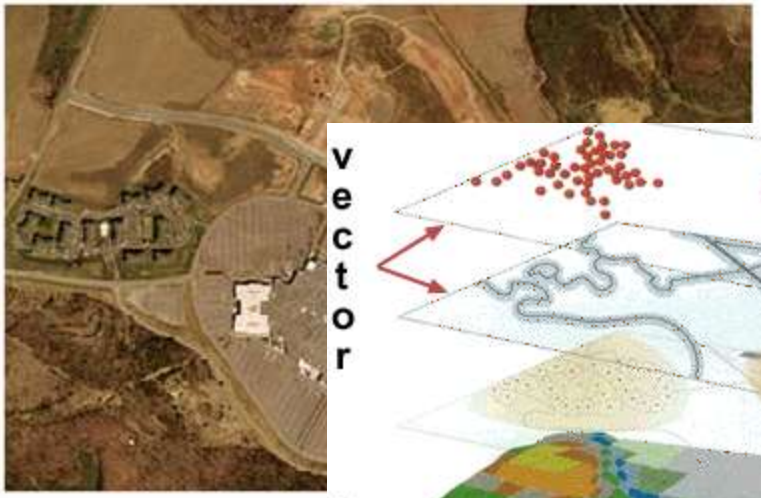
Mission: To ensure access over time to a rich body of digital content through the establishment of a national network of partners committed to selecting, collecting and preserving at-risk digital information

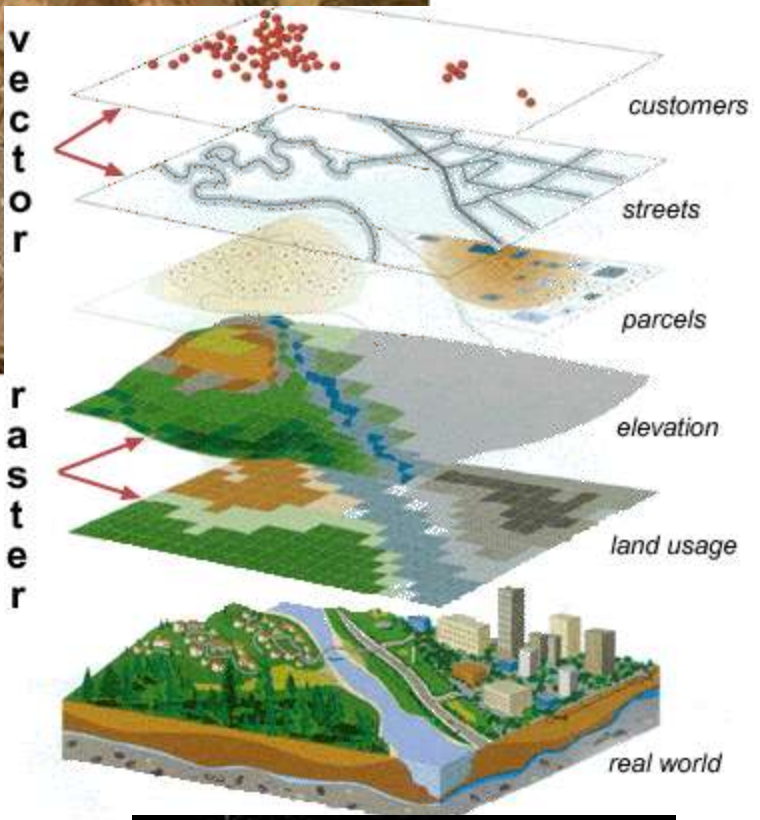
- Learn By Doing
- Catalyze Activity
- Support Collaboration
- Break Down Boundaries

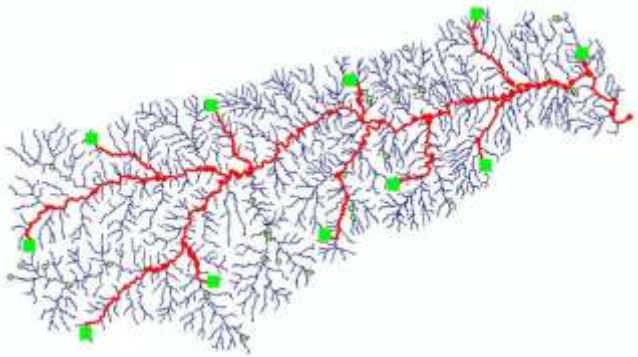
Digitalpreservation.gov

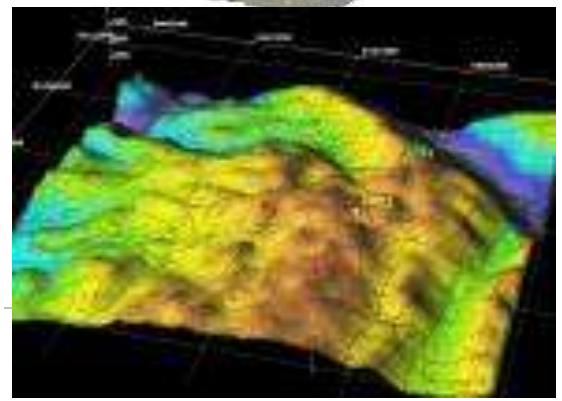


www.geomapp.net




vector
customers
streets
parcels
raster
elevation
land usage
real world

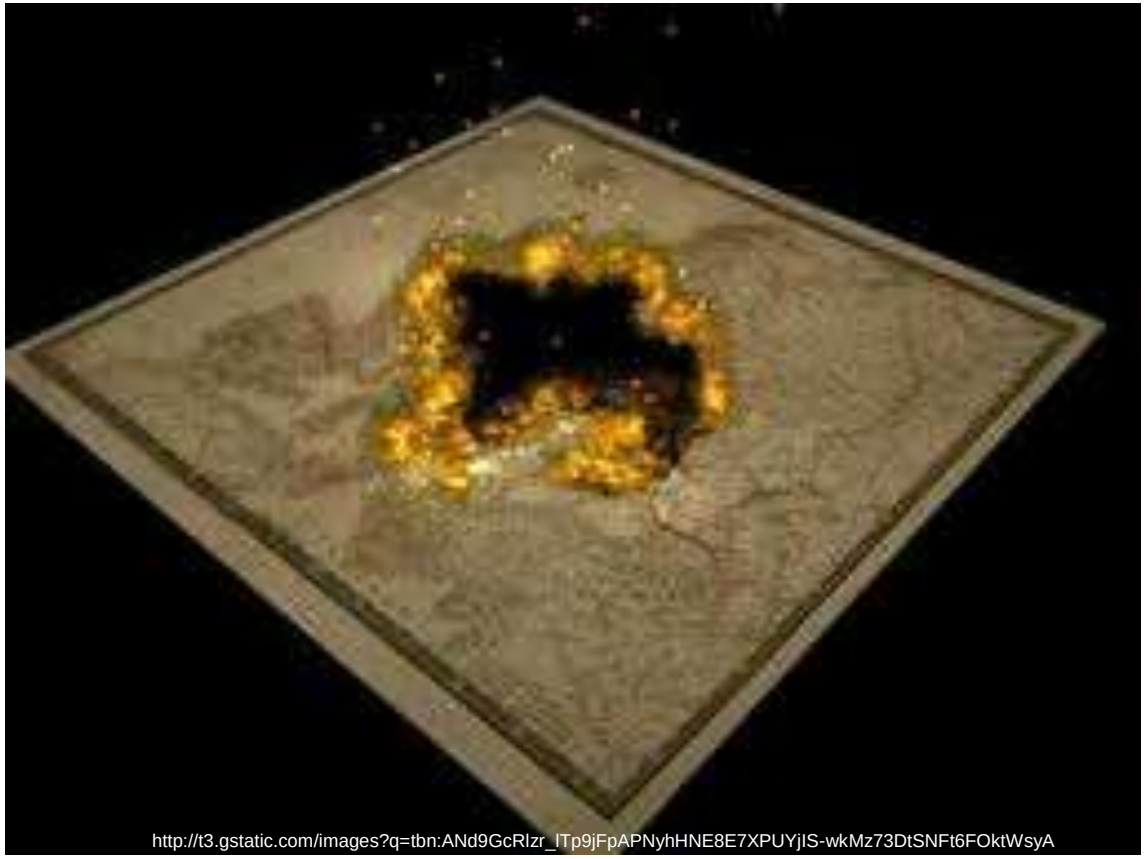




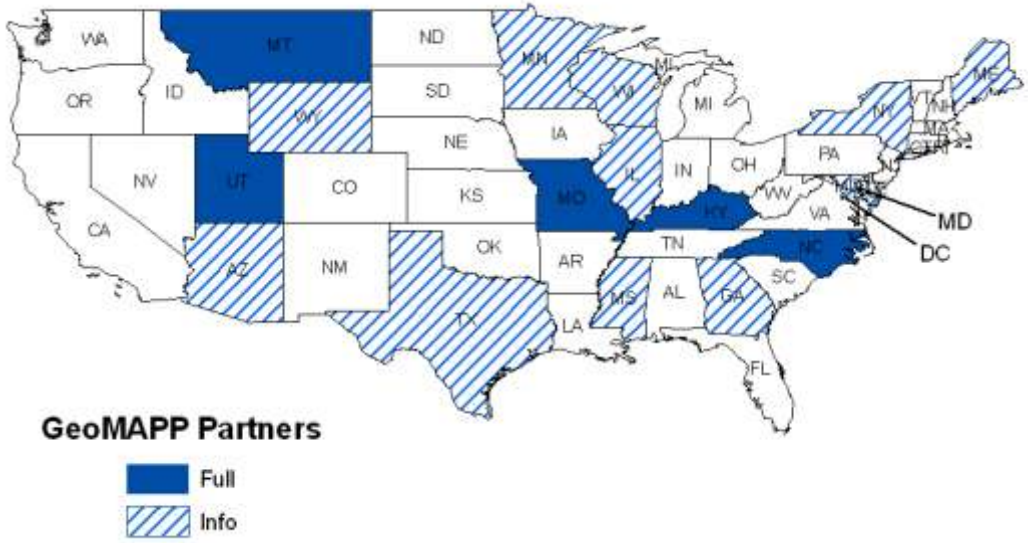
What are the Special Risks to Geospatial Information?

- Unique geospatial data formats
- Spatial database complexity
- Fragility and uncertainty surrounding digital cartographic representation
- Issues related to time-versioned content
- Metadata unavailability or inconsistency
- No generally supported content packaging design for complex geospatial data

GeoMAPP: Saving “at risk” State Geospatial Data



Who is GeoMAPP?



- **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

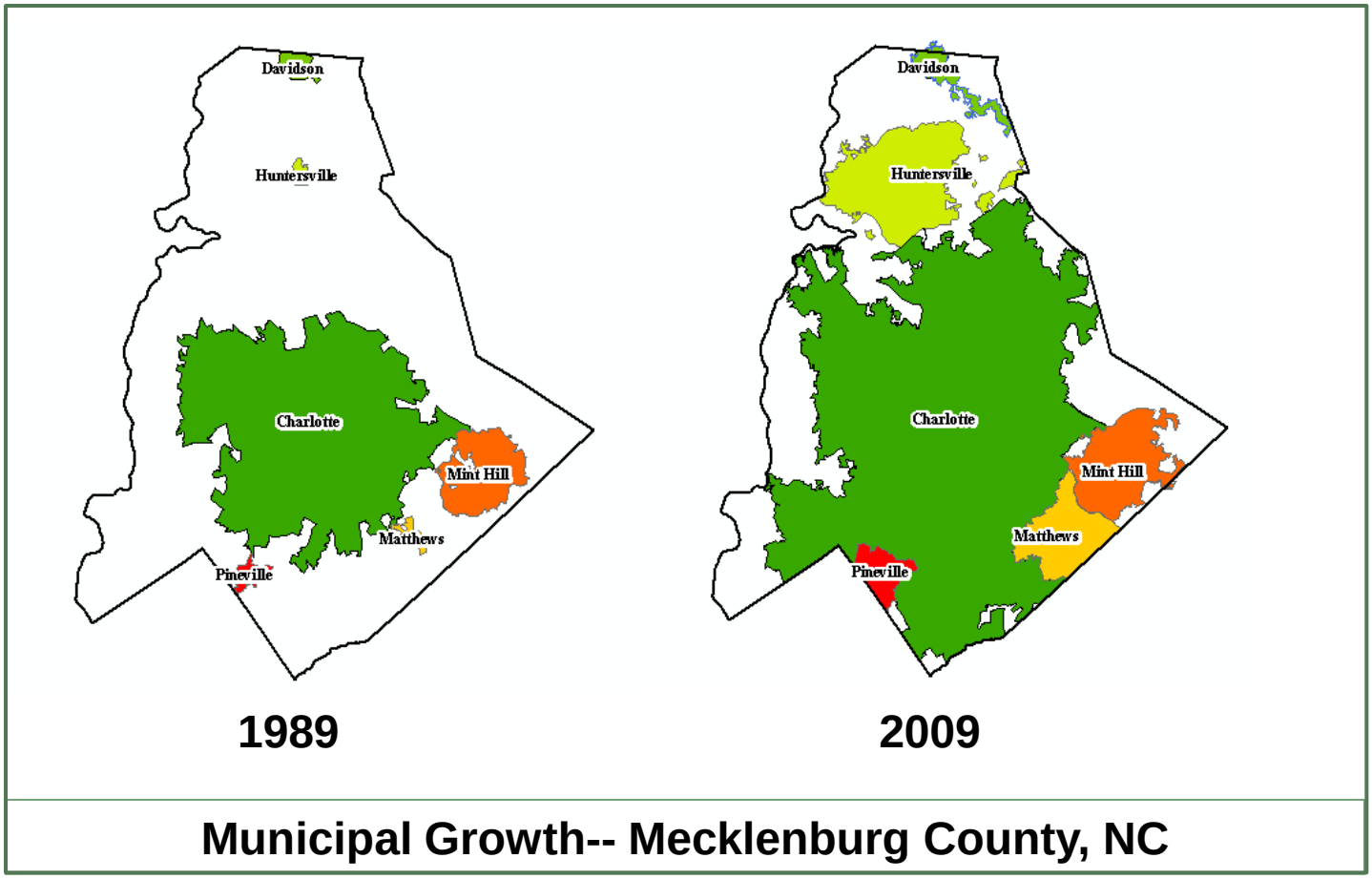
- **12 Informational Partners**: AZ, DC, GA, IL, ME, MD, MS, MN, NY, TX, WI, WY

www.geomapp.net

Why Preserve Geospatial Data?

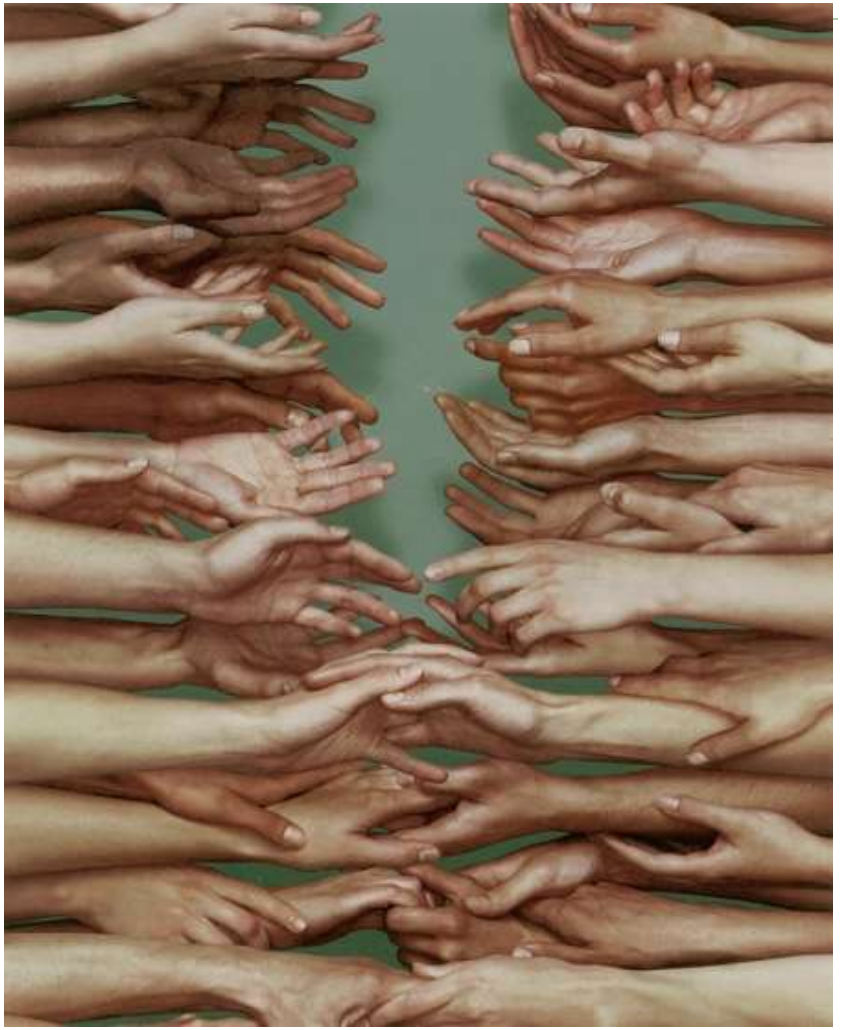
★ ***Modeling Change*** ★ ***Legal Mandate***

★ ***Historic Mapping*** ★ ***Cost of Reproducing*** ★ ***Disaster Recovery/COOP***



Municipal Growth-- Mecklenburg County, NC

Preservation is a Series of Handoffs



- Technological
- Social

http://michael-connor.com/wordpress/wp-content/uploads/2008/10/keegan-web.jpg

GeoMAPP Focus Areas

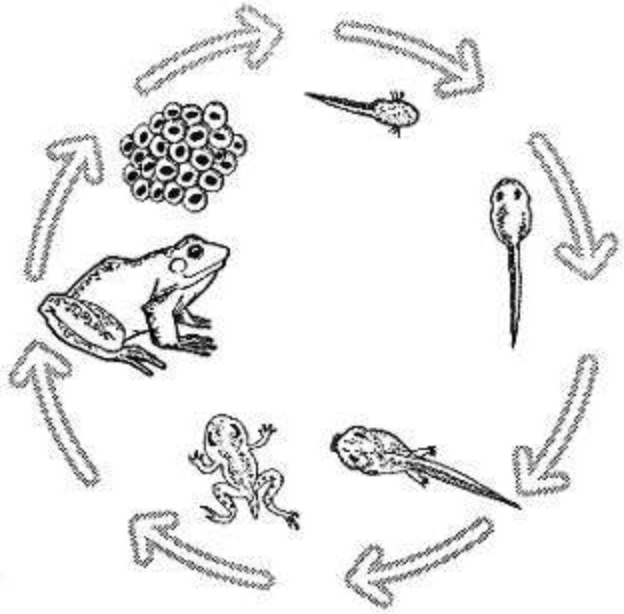
- Technical explorations:
  - file formats
  - metadata
  - data packaging
  - storage solutions
  - long term preservation techniques
- Improving access to archived data
- Business planning for sustainable archives
- GIS Clearinghouses, Archivists and the data creator community

What We've Learned

Joe Sewash
NC Center For Geographic Information and Analysis

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

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 1.1.8 Citation Information		<citation> <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' *e.g. North Carolina Shellfish Growing Areas 2010*
*	**Geospatial Data Presentation Form**	<geoform>	Describes type and format of the dataset *e.g. vector digital data*

- 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

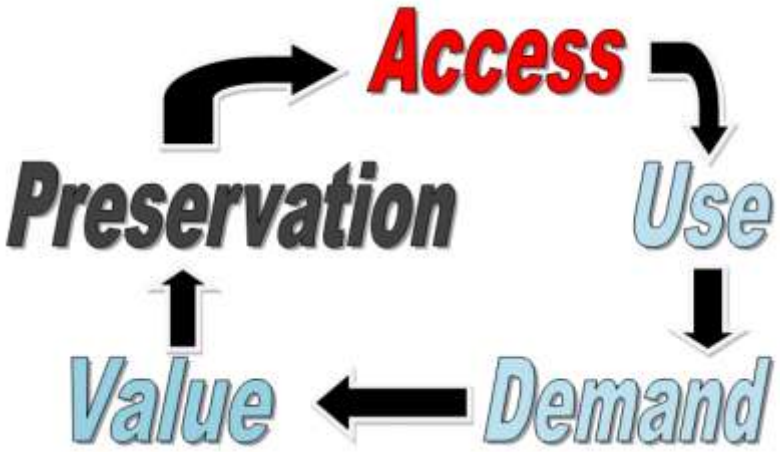
- Sustainability factors
  - Is the data format open or proprietary?
  - Vendor support lifecycle and documentation
  - Data packaging
- Selecting the right format for long term use/ preservation

- 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 pre-dates 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 ARCINFO coverage format through ArcGIS Toolbox or command line ARCINFO.	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.

Leveraging Existing Workflows

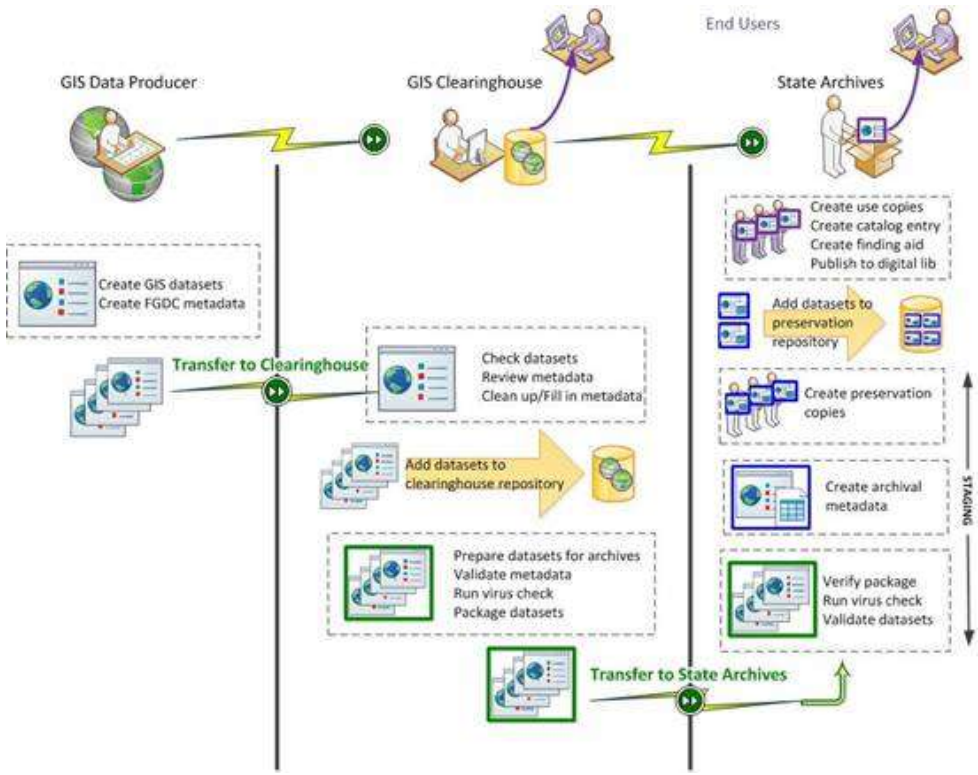
- Existing or Emerging Systems to Leverage
  - Geospatial Platform
  - GIS Inventory powered by Ramona
  - GOS/Geodata.gov/Data.gov
- Focusing on Access



  - Leverage existing GIS or Archives access solutions
    - GeoMAPP Examples: GeoPortal Toolkit Research; Use of ContentDM, Dspace and AXAEM digital repositories;
    - Investigation of free Data consolidation points ArcGIS.com, GeoCommons, etc

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



Ongoing Challenges

Ongoing Preservation Challenges

- Capturing Dynamic Data
  - Database versioning/ Archiving
  - Periodic snapshots
- Dealing with Project files
  - Managing the “package”
  - Interrelationships
  - Diverse files
  - Documentation

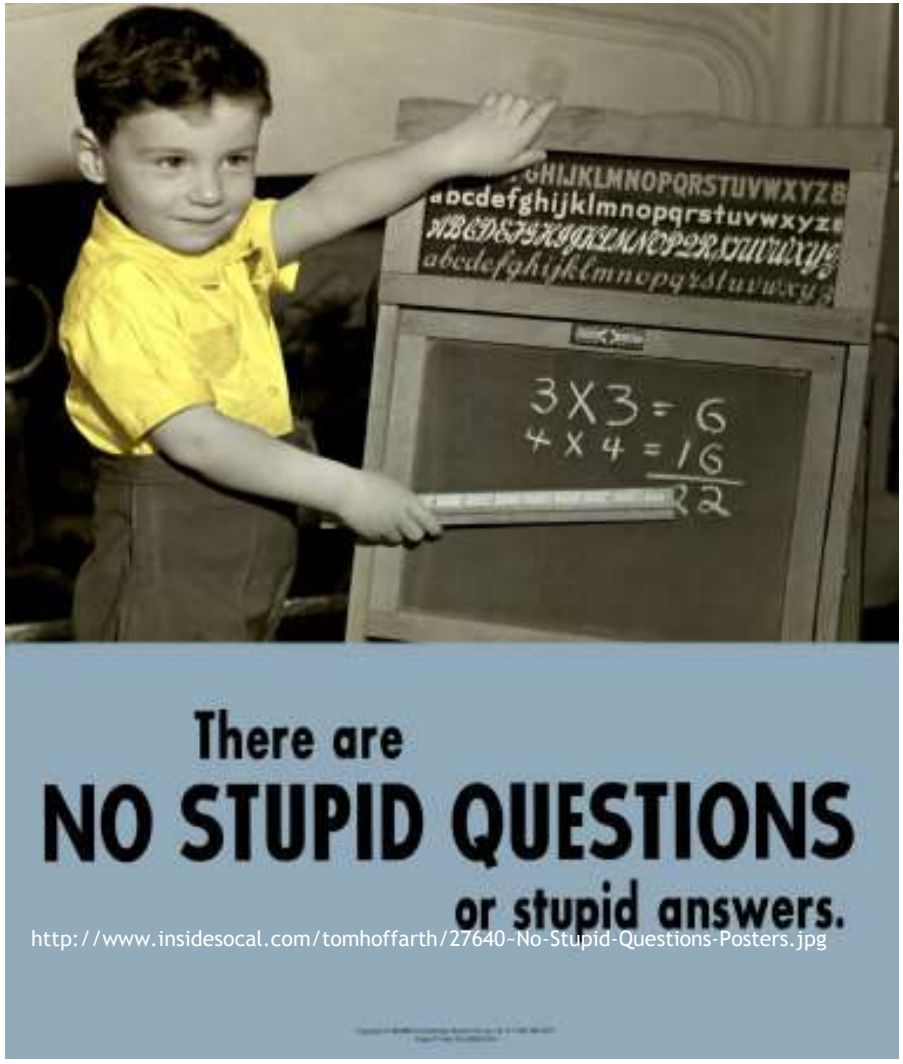
- Web Apps
  - Ephemeral data
  - Web services
  - Capture at the source or tackling the mash up?

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.**

Questions?



www.geomapp.net

Thanks!

http://imagecache2.allposters.com/

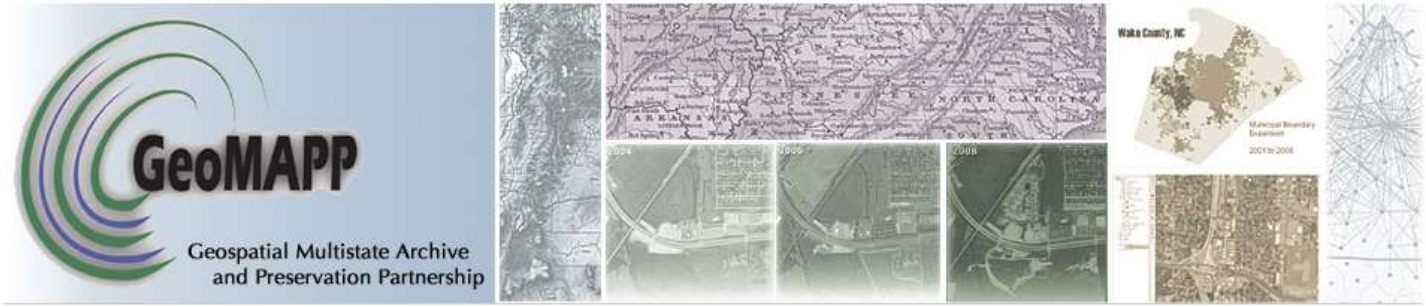
- Butch Lazorchak (Library of Congress)

wlaz@loc.gov

- Joe Sewash (North Carolina CGIA)

joe.sewash@nc.gov

Follow us on Facebook and Twitter!



- **www.geomapp.net**