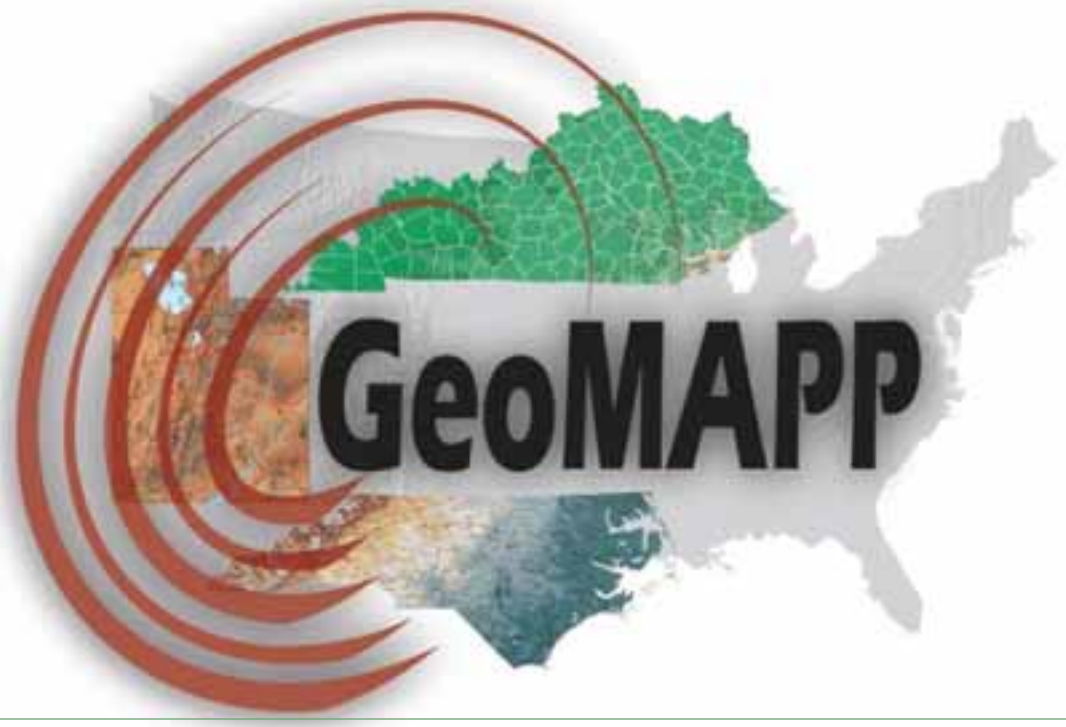


Mapped Today; Zapped Tomorrow? Preserving Government Digital Geospatial Data

Butch Lazorchak, Library of Congress
Alec Bethune, North Carolina Center for Geographic Information and Analysis
Mark Myers, Kentucky Department for Libraries and Archives

ESRI International Users Conference | 14 July 2010 | San Diego, CA

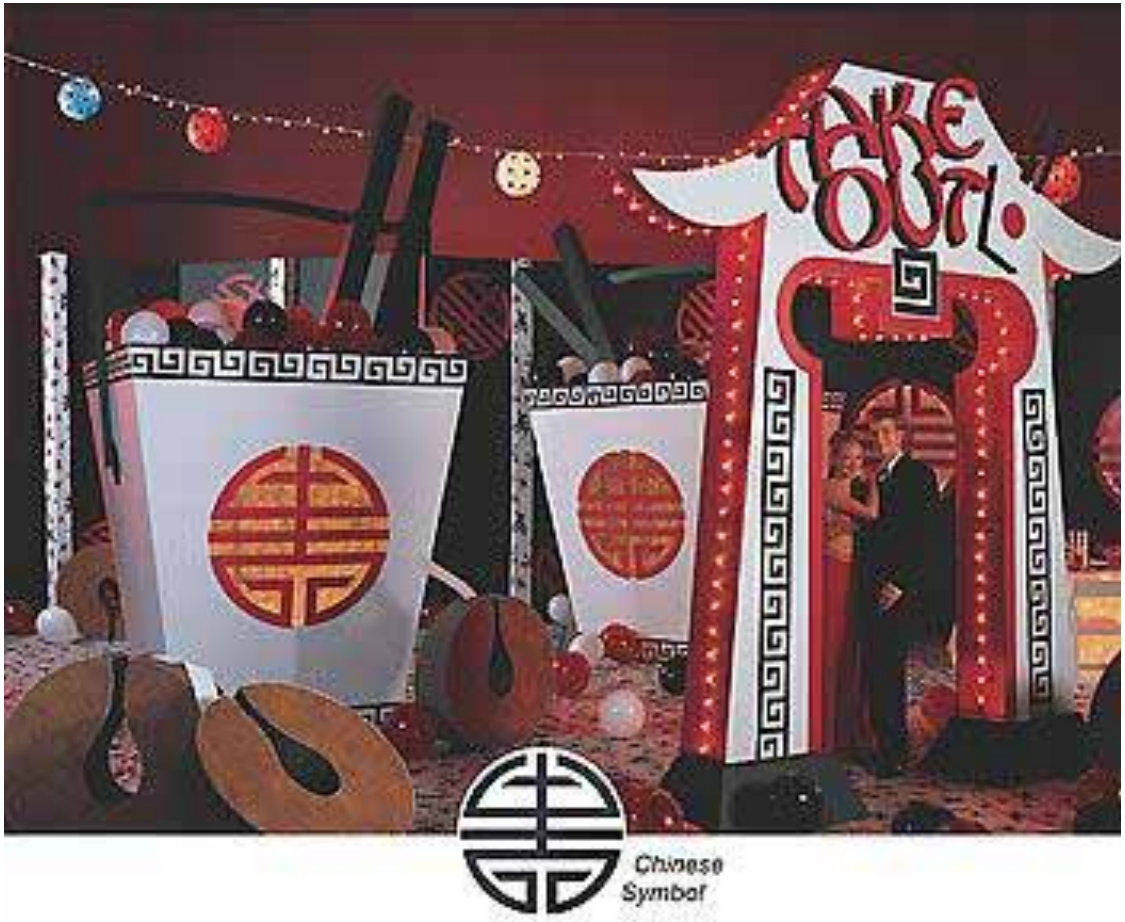
Why is the Library of Congress Interested in Preserving Digital Geospatial Data?

Butch Lazorchak

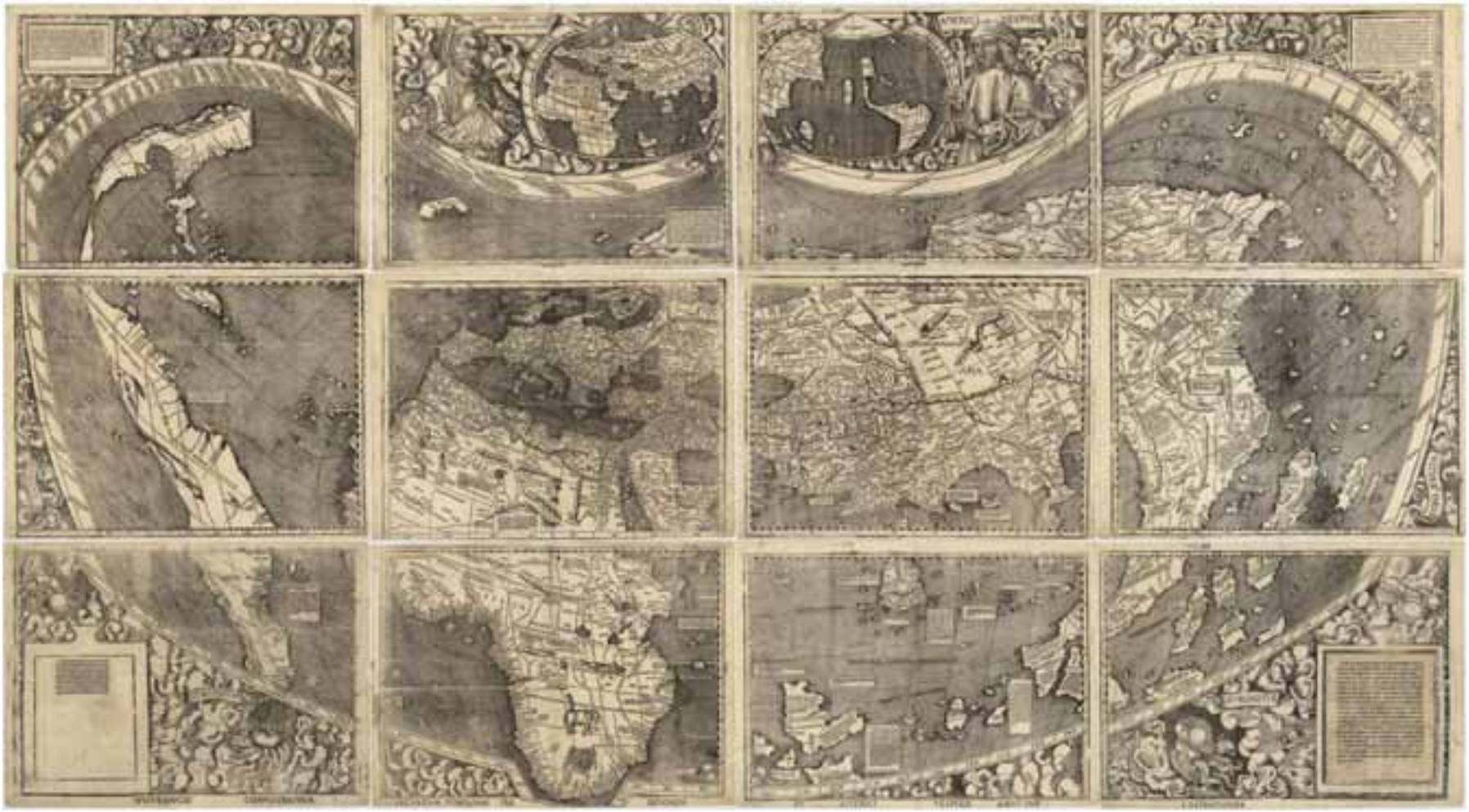
Library of Congress

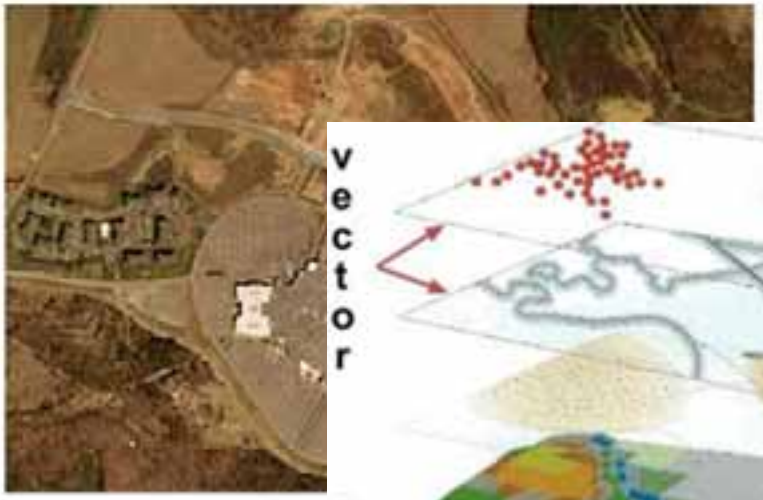
Key Takeaways

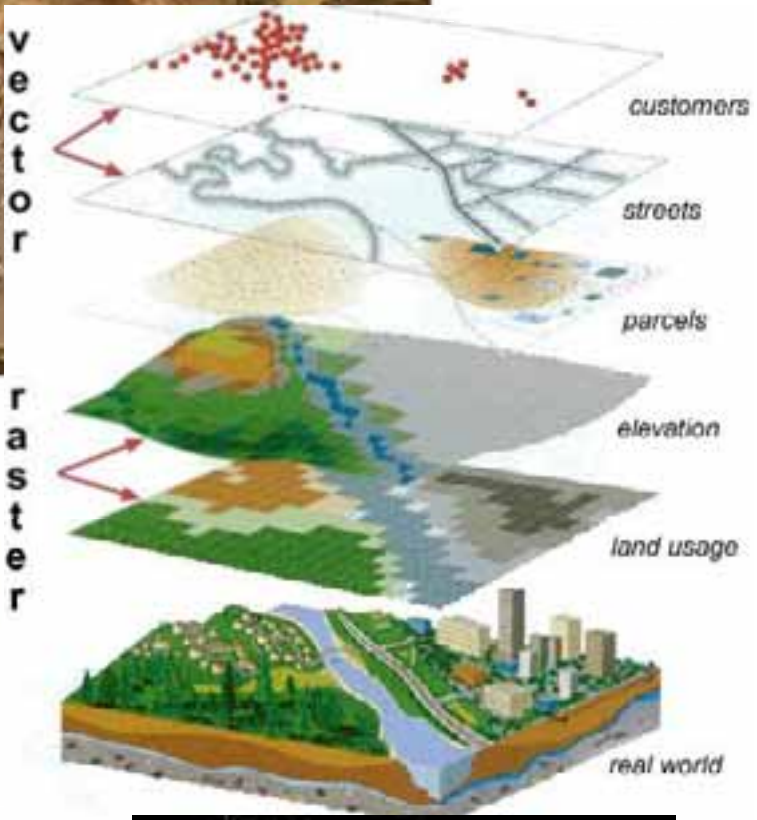
- There are costs and benefits to preserving digital geospatial information
- The Library of Congress and its network are thinking about this issue and can be trusted partners
- The GeoMAPP project is taking a leading role
- We need your help to spread the word

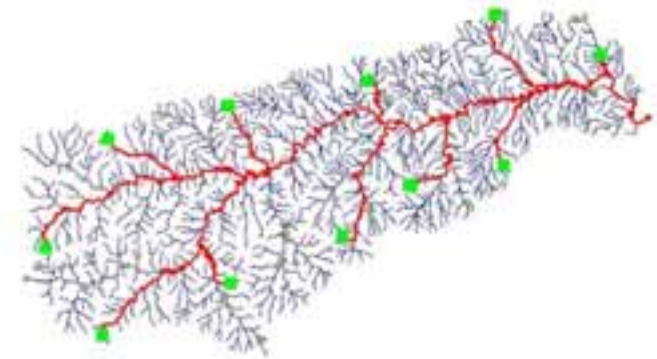


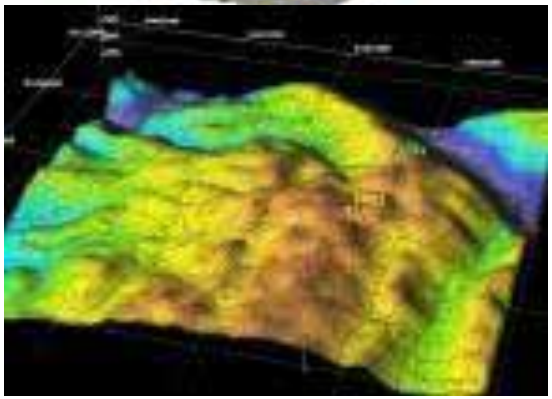
http://shindigzparty.files.wordpress.com/2008/03/4s058a.jpg






vector
raster
customers
streets
parcels
elevation
land usage
real world





Why Preserve? Is there a special value to older materials? Or, is “Preservation” a Dirty Word?

Is Digital Preservation a “cost center” or a “benefits center”?

http://cleantechnica.com/files/2008/05/353493661_0151e8185f.jpg

http://www.afm.ars.usda.gov/hrd/eNeo/images/big-benefits.gif

Current Digital Geospatial Materials: The Benefits are Obvious!

- Useable
- Distributable
- Searchable
- Remix-able
- Accessible

Users Love this!

Demonstrate the VALUE of Preserved / Historic / Superseded Digital Materials by Leveraging Other Use Cases

Generate Revenue

Save Money

But Also…

Disaster Recovery and COOP

Not just "Cultural Heritage"…

Document Business Processes for Improved Decision-making

Enhanced Access Drives Incentives to Preserve

Benefits

- Accessibility creates Use
- Use creates Demand
- Demand creates Value
- Value creates Benefits

There are **significant benefits** to preserving digital geospatial data. What are you going to do with all those benefits?

http://www.artquest.org.uk/valueadded/images/Value-Added-FP-new.gif

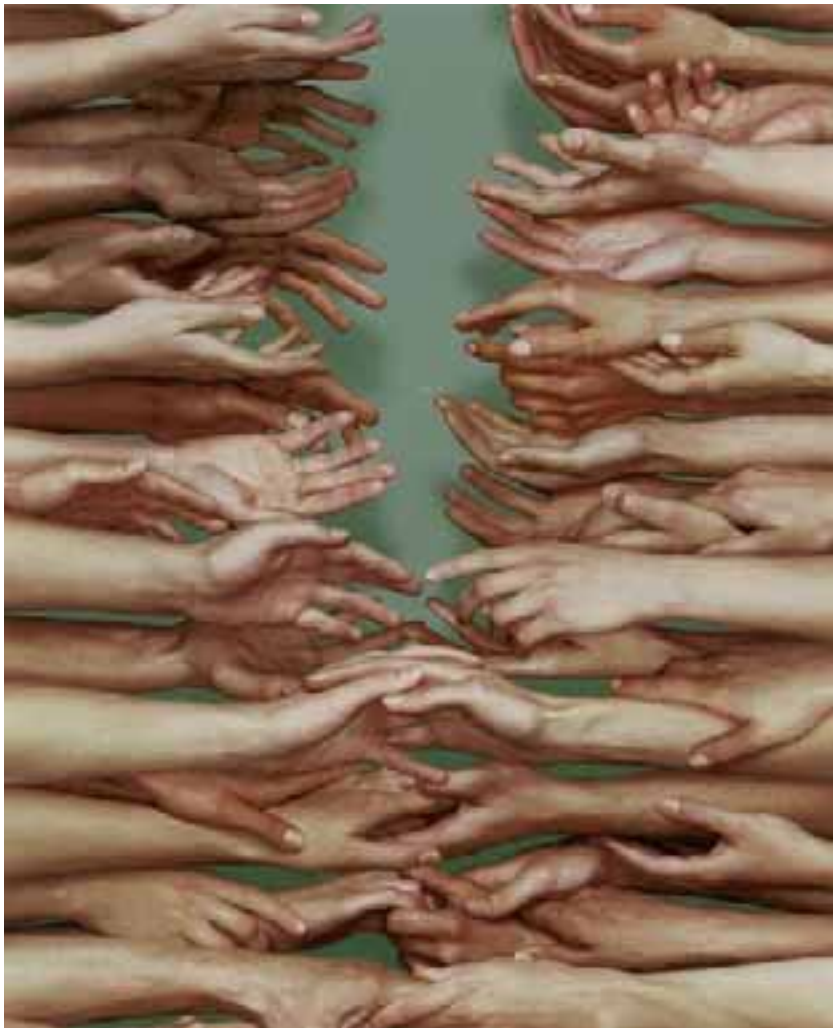
How to Mitigate Costs

Benefits are Decentralized; Costs are Localized

- Leverage Existing Infrastructures
- Take Advantage of Economies of Scale
- Capture “Data in Motion”
- Readdress the lifecycle of information
- Forever or 5 years, whichever comes first!

http://tabootrinity.files.wordpress.com/2009/03/no-money1.jpg

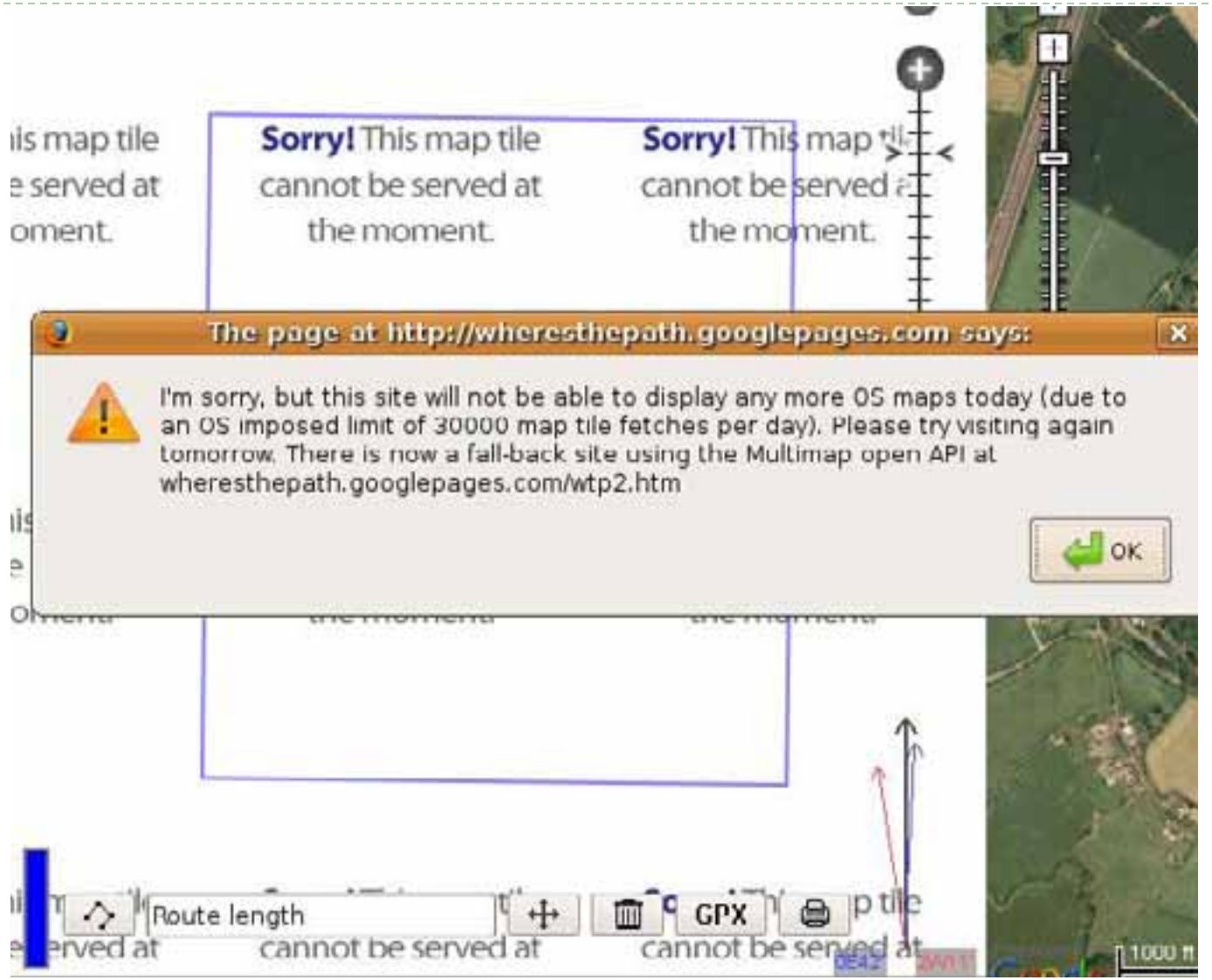
Preservation is a Series of Handoffs



- Technological
- Social

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

Digital: More Fragile Than Analog



http://www.failmaps.com/

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

The Long-term Digital Public Record

What about collections?
What about permanence?
What about eternity?!?!?!?!

http://www.markstivers.com/wordpress/comics/2006-09-10%20Eternity-store.gif

Somebody Has to Be Thinking about the Future

http://img.perezhilton.com/wp-content/uploads/2008/03/foreheadtatsresize__oPt.jpg

Repository of the Past?

Predictor of the future?

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

Taking it to the States: Geoarchiving with GeoMAPP

Alec Bethune

North Carolina Center for Geographic Information and Analysis

How would you describe your geospatial archive?

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

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

http://www.honeysucklecreek.net

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

GIS from an Archivist's Perspective

Mark Myers

Kentucky Department for Libraries and Archives

What is a Record?

North Carolina General Statute 132-1

Utah Code 63G-2-103

Kentucky Revised Statutes 171.410

"all books, papers, maps, photographs, cards, tapes, disks, diskettes, recordings and other documentary materials, regardless of physical form or characteristics, which are prepared, owned, used, in the possession of or retained by a public agency."

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

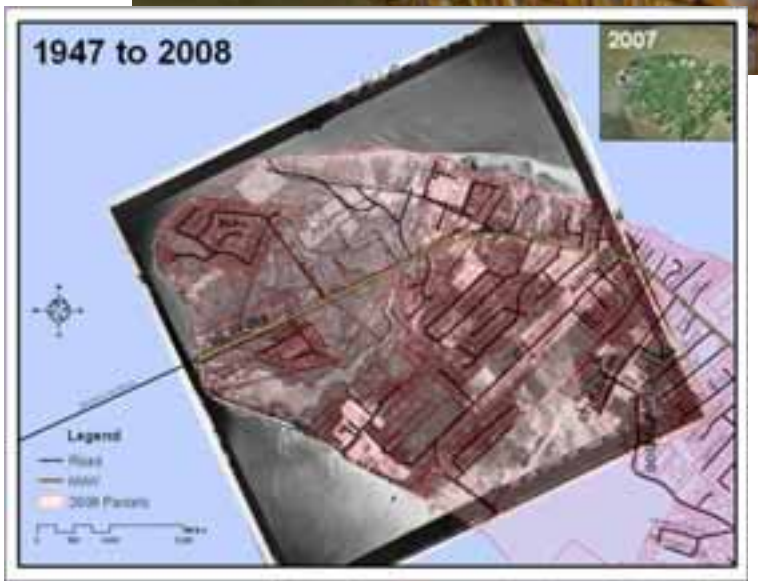
Utah State Archives and Records Service

Kentucky Department for Libraries and Archives

Thinking of GIS Data as Records

- Philosophical shift
  - Paper v/s Digital
  - Present value v/s historic use

- What's currently being preserved?
  - Maps, photo imagery, paper, and administrative records well represented
  - Digital geospatial data…not so much…..

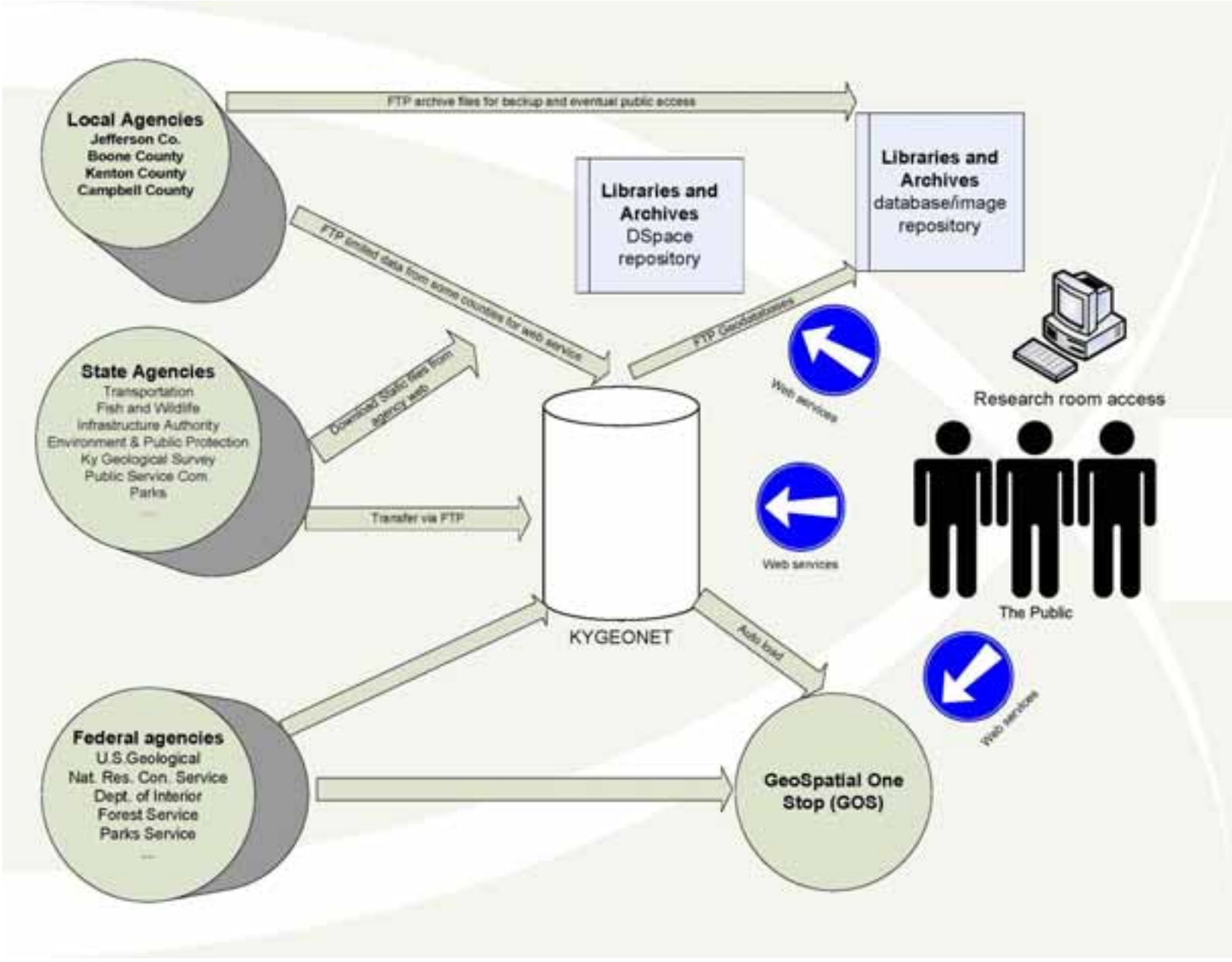


Technical Challenges with Geospatial Data

- 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

Challenges

- Geospatial Data sets—What format do we use to archive?\
  - Shape File
  - Geodatabase
  - Geopdf
- State approaches:
  - NC—transferring shape files
  - UT—transferring shape and geodatabases and creating geoPDF
  - KY—transferring geodatabases
- Took over an hour to transfer 1.5 GB of data from CGIA to State Archives SAN. Anyway to increase speed of transfer?
- Where does GIS data fit? How is it organized in a repository folder structure.


Local Agencies
Jefferson Co.
Boone County
Kenton County
Campbell County
FTP archive files for backup and eventual public access
Libraries and Archives
DSpace repository
Libraries and Archives
database/image repository
FTP limited data from some counties for web service
FTP Geodatabases
State Agencies
Transportation
Fish and Wildlife
Infrastructure Authority
Environment & Public Protection
Ky Geological Survey
Public Service Com.
Parks
Download static files from agency web
Web services
Research room access
Transfer via FTP
Web services
The Public
KYGEONET
Auto load
Federal agencies
U.S.Geological
Nat. Res. Con. Service
Dept. of Interior
Forest Service
Parks Service
GeoSpatial One Stop (GOS)

Current GeoMAPP Focus Areas

- Business planning for sustainable archives
- Improving access to archived data
- Technical explorations:
  - file formats
  - metadata
  - data packaging
  - storage solutions
  - long term preservation techniques
- Outreach, Outreach, Outreach (and mentoring)


Frankfort, KY

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?

Alec Bethune

North Carolina Center for Geographic Information and Analysis

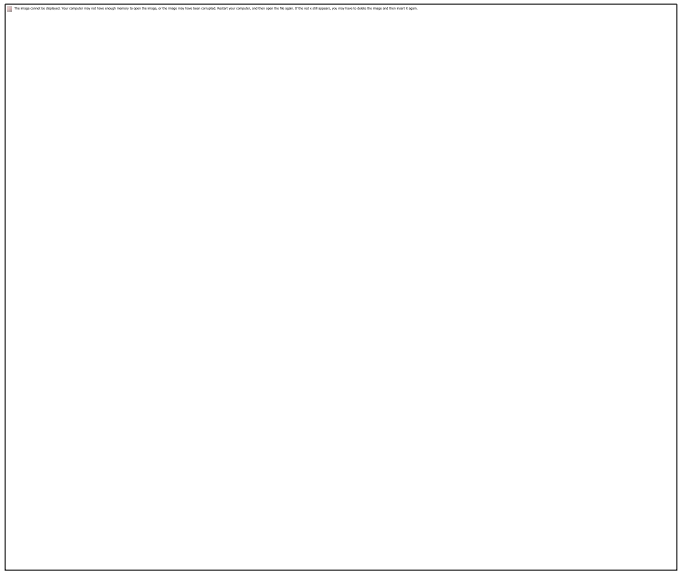
What's in a name?

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	60 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?

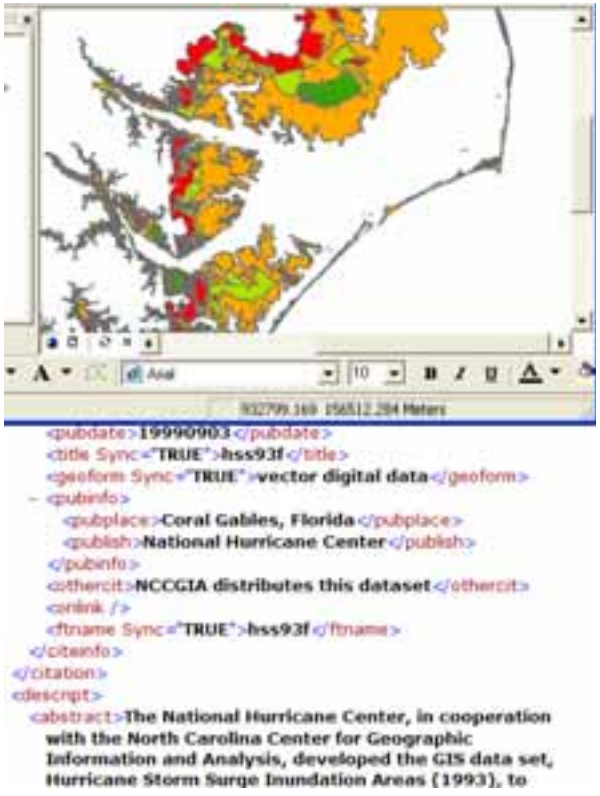
SGA

Attributes of sga

FID	Shape	DSHA_NAME	DSHA_NAME1	HA_CLASS	HA_CLASSID	HA_NAME	HA_STATUS	COUNTY	SURFACE	DATE_CREAT	DATE_UPDAT	ACRES	SQ_MILES
0	Polygon	Atlantic Ocean						nc	water	02-01-04		726986.340798	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		80405.80145	125.634065
3	Polygon							Currituck	land	02-01-04		0.715843	0.001119
4	Polygon							Currituck	land	02-01-04		56.945307	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		1706.572668	2.66652
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			7126.132999	11.134583
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		10.352888	0.016176
19	Polygon							Currituck	land	02-01-04		3.437776	0.005372

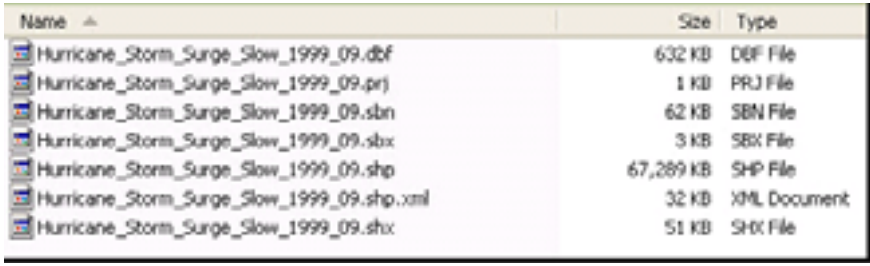
Why Do Metadata?

- 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



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

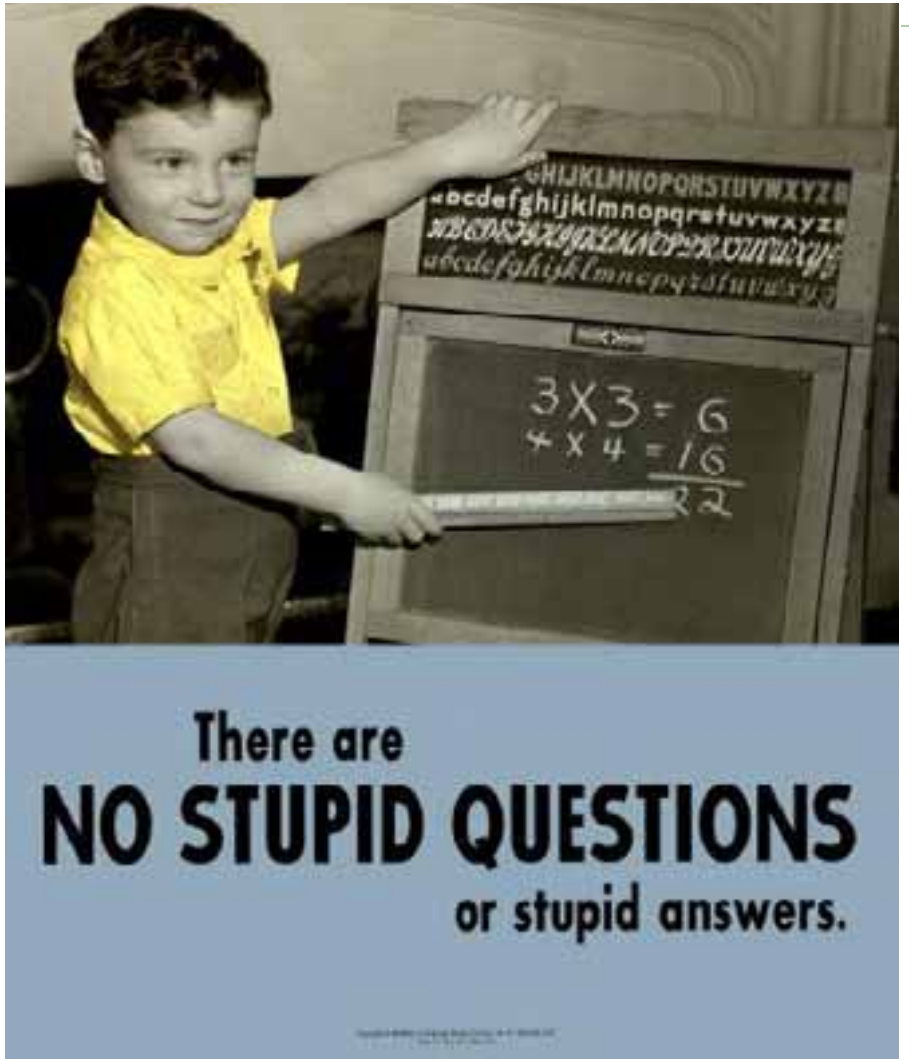


What can you do?

- Acknowledge the costs but recognize the many benefits to preserving digital geospatial information
- Consider the Library of Congress network as trusted partners: How can we work together?
- Tell EVERYBODY!
- Oh yeah, don't forget about the metadata!

http://cnx.org/content/m14808/latest/YouCanDoIt.jpg

Questions?



http://www.insidesocal.com/tomhoffarth/27640~No-Stupid-Questions-Posters.jpg

Thanks!

- Butch Lazorchak (Library of Congress)
  wlaz@loc.gov
  - Alec Bethune (North Carolina CGIA)
    alec.bethune@its.nc.gov
- Mark Myers (Kentucky KDLA)
  mark.myers@ky.gov

http://www.geomapp.net

http://imagecache2.allposters.com/images/pic/Matted_Prints/mp_814104_b~Man-on-Phone-Thanks-Posters.jpg