Geospatial Multistate Archive and Preservation Partnership

Informational Partners Meeting:
Geospatial File Formats

| May 12 2011 |

Meeting Initiation/ Session Overview

Jeff Brown, Mark Myers

Agenda

- Roll call and introductions
- Introducing File Formats and Data Transfer Documentation Efforts
- Geospatial File Formats
  - **Goal:** Share with Info Partners and get your input on some outstanding questions
  - File Formats document Purpose and Background
  - Introduce the File Formats document
    - Organization, Information
  - Review Survey Monkey file formats survey results
  - Solicit partners' feedback on outstanding questions in File Formats
- Best Practices for Geospatial Data Transfer
  - Introduce the document to the Informational Partners
- Future Informational Partners meetings

Geospatial File Formats Quick Reference

Kelly Eubank, Alec Bethune

Geospatial File Formats Quick Ref

- Goals
  - To provide a useful quick reference guide for GIS file formats encountered in state government
  - Identify some sustainability issues/things to consider for long term preservation
- Background
  - Initial effort was to discuss preservation challenges of emerging GIS formats (kml, pdf, etc)
  - The partnership saw a number of files that were not "standard", started putting together an internal cheat sheet

Introduction

GeoMAPP Geospatial File Formats Spreadsheet - Introduction (4/1/2011)

Electronic file format support is a fundamental challenge in the long-term preservation of digital materials. This issue is especially relevant for geospatial datasets that are created, shared, and stored in many different formats, many of which are proprietary to a specific vendor and/or software application. Geospatial data is primarily comprised of raster data, made up of a two dimensional array of equally sized cells where each cell holds a single attribute value and location coordinates, or vector data, represented as points, lines, and polygons. In addition to the fundamental geographic information, many datasets also include underlying tables of data that further elaborate the geographic elements, such as population counts, income values, average property values, demographic information, or identifying information for individual features such as name, address, etc. Then all of this data may further be wrapped within an encompassing database that enables relationships and analysis across the datasets. Many of the geospatial formats are manifested as numerous files in a variety of file types. All of these aspects pose challenges to GIS professionals and to archivists both today and in the future.

This spreadsheet provides a quick reference of some of the common geospatial raster and vector dataset types, and can serve as a tool to identify geospatial format types based on file extensions. The spreadsheet is organized into tabs with a **Raster** formats tab, a **Vector** formats tab, a **Standards** tab, and an additional **Resources** tab. For the raster and vector format tabs, the spreadsheet provides:

- the file extension, and for multi-file formats a description of the format's file composition,
- a brief description,
- a link to the format specification or standard - if available/applicable,
- an assessment of the format's *currency* - meaning is this a format this is generally in use today,
- an assessment of the format's *prevalence* - meaning how common is it to see this format,
- a sampling of tools available to access the file format
- whether the format is supported by GDAL (for raster files) /OGR (for vector files) or Safe Software's Feature Manipulation Engine (FME). These are tools that can read a large number of file formats, and may become more valuable to archivists especially as data formats obsolesce due to vendor viability and/or market demand.
- sustainability issues
- additional comments

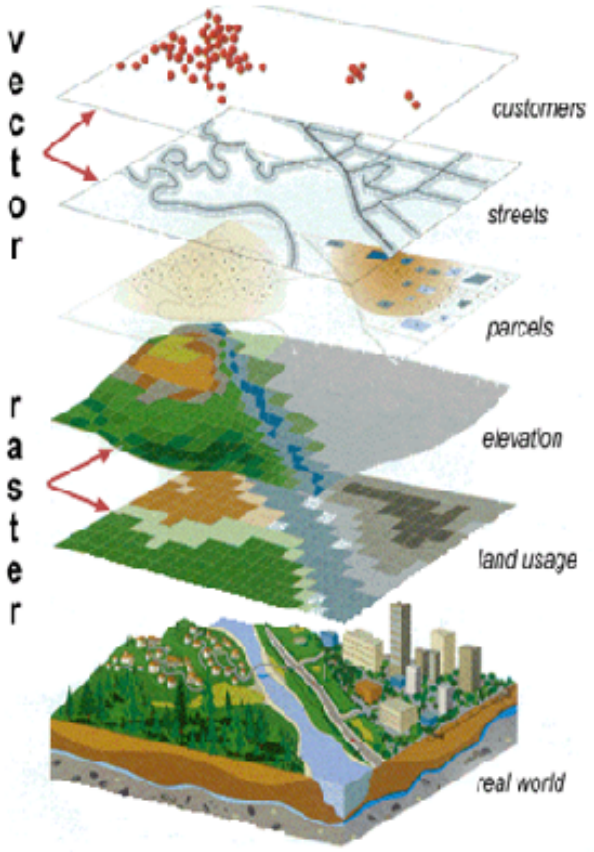


Image Source: NOAA National Coastal Development Center. "Geographic Information Systems (GIS)." Retrieved 3/29/2011 from: http://www.ncddc.noaa.gov/technology/gis/view

Raster Data Formats

	A	B	C	D	E	F	G	H	I	J	K
1	**Raster Formats**	**File Composition**	**Description**	**Link to Format Standard/ Specification**	**Format Currency**	**Format Prevalence**	**Tools for Viewing Format**	**Supported GDAL** http://www.gdal.org/formats_list.html	**Supported by Safe Software Feature Manipulation Engine (FME)** http://www.safe.com/data-	**Sustainability Issues**	**Comments**
2	Raster files generally are used to store image information, such as scanned paper maps or aerial photographs. They are used for data captured by satellite and other airborne imaging systems. Images from these systems are often referred to as remote-sensing data. Unlike other raster files, which express resolution in terms of cell size and dots per inch (dpi), resolution in remotely sensed images is expressed in meters, which indicates the size of the ground area covered by each cell. (http://data.geocomm.com/helpdesk/formats.html)										
3	Arc Digitized Raster Graphic (ADRG)	**File Extension:** *.img or *.gen [old: *.ovr for data, *.lgg for legend file] **3 Files:** TIFF Image (.tif) Metadata file (.txt) ArcInfo world file (.tfw) (optional)	A digital raster graphic (DRG) is a scanned image of a U.S. Geological Survey (USGS) standard series topographic map, scanned, digitized and georeferenced. DRGs are in TIFF format (using the GeoTIFF specifications) and are scanned at a minimum resolution of 250 dots per inch. The digital image is georeferenced to the true ground coordinates of the 2.5-minute grid ticks and projected to the Universal Transverse Mercator (UTM) for projection consistency with USGS Digital Orthophoto Quadrangles (DOQs) and DLGs. The USGS DRG consists of three physical files: the TIFF image, a text file of metadata along with a non-required third file, an ArcInfo world file. The DRG image (*.tif), world (*.tfw), and metadata (*.fgd) files incorporate an intelligent data set name (DSN) consisting of descriptive metadata wrapped around a standardized kernel that describes the spatial location of the file. MIL-A-89007 states ADRG is inactive for new design, and is no longer used, except to support existing systems.	U.S. Military Specification MIL-A-89007 http://earth-info.nga.mil/publications/specs/printed/89007/89007_ADRG.pdf	Legacy	Largely in the Federal government in the military domain.	ESRI Products such as the ArcGIS Military Analyst Suite of Tools (Raster Map Tool, etc.)	Creation-Yes Georeferencing-Yes Compiled by default-Yes	Yes	CADRG (Compressed ADRG file) achieves a nominal compression ratio of 55:1. Enhanced Compressed Raster Graphic (ECRG) uses JPEG 2000 compression using a compression ratio of 20:1 Commonly distributed on CD's, which would require a migration to new media for preservation. Library of Congress Sustainability Page: http://www.digitalpreservation.gov/formats/fdd/fdd000282.shtml	FMEpedia guidance: http://www.fmepedia.com/p/ARC_Digitized_Raster_%28ADRG%29. EROS guidance at http://eros.usgs.gov/#/Products_and_Data_ADRGs

Vector Data Formats

	A	B	C	D	E	F	G	H	I	J	K
1	**Vector Formats**	**File Composition**	**Description**	**Link to Format Standard/ Specification**	**Format Currency**	**Format Prevalence**	**Tools for Viewing Format**	**Supported by OGR** http://www.gdal.org/ogr/ogr_formats.html	**Supported by Safe Software Feature Manipulation Engine (FME)** http://www.safe.com/data-formats/	**Sustainability Issues**	**Comments**
2	Vector datasets are based on the interaction between arcs and nodes, represented as *points* , *lines* , and *polygons*. A point is a single node, a line is two nodes with an arc between them, and a polygon is a closed group of three or more arcs. With these three elements, it is possible to record most all necessary information. Vector data is useful for modeling discrete physical features (http://www.umich.edu/~ipcaa/GIS/General%20GIS%20Concepts.htm) (http://www.mass.gov/mgis/GIS_Glossary.pdf)										
3	Autodesk Drawing eXchange Format	**File Extension:** *.dxf	AutoCAD DXF (Drawing Interchange Format, or) is a CAD data file format developed by Autodesk for enabling data interoperability between AutoCAD and other programs. Versions of AutoCAD from Release 10 (October 1988) and up support both ASCII and binary forms of DXF. Earlier versions	Proprietary, Published AutoDesk DXF Reference http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=12272454&linkID=10809853	Waning	Niche use	A long list of tools that can view DXF files can be found here: http://en.wikipedia.org/wiki/AutoCAD_DXF#Software_that_supports_DXF	Creation-No Georeferencing-Yes Compiled by default-Yes	Yes	As AutoCAD has become more powerful, supporting more complex object types, DXF has become less useful.	
4	Autodesk Drawing File	**File Extension:** *.dwg	DWG ("drawing") is a file format used for storing two and three dimensional design data and metadata. It is the native format for several CAD packages including AutoCAD, IntelliCAD (and its variants) and Caddie. In addition, DWG is supported non-natively by many other CAD applications	Proprietary, Not Available	Current	Widely used in CAD(Architecture, Engineering, Design) Industries. Minimal use in traditional geospatial shops	Autodesk offers a free DWG viewer: DWG TrueView http://usa.autodesk.com/adsk/servlet/pc/index?siteID=123112&id=15314320 ArcGIS Products ; AutoCad ; Other Viewers: http://usa.autodesk.com/adsk/servlet/pc/index?id=15409188&siteID=123112 ; http://usa.autodesk.com/adsk/servlet/pc/index?id=6703438&siteID=123112; http://www.autodesk.co.uk/adsk/servlet/index?siteID=452932&id=8992673 ;	Not listed	Yes		Third party viewers not AutoCad should underst design specifications th limit viewing of certain d elements such as unexp blocks, custom fonts, et

Standards and Other

	A	B	C	D
1	**Standard/Specification Names**	**Link**	**Description**	**Comments**
2	Data Model for Managing and Preserving Geospatial Electronic Records	Center for International Earth Science Information Network (CIESIN) http://ciesin.columbia.edu/ger/DataModelV1_20050620.pdf	Paper prepared by CIESIN describing a comprehensive data model and data dictionary for managing and preserving geospatial data records.	
3	FGDC Content Standard for Digital Geospatial Metadata	Federal Geographic Data Committee http://www.fgdc.gov/metadata/geospatial-metadata-standards	Content Standard for Digital Geospatial Metadata. Prepared by: the Federal Geographic Data Committee. FGDC-STD-001-1998 (http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf). Foundation of ISO 19115. ESRI profile of the Content Standard for digital geospatial metadata can be found at the following website: http://www.esri.com/metadata/esriprof80.html	
4	Geography Markup Language (GML)	OGC Geography Markup Language - Open GIS Geography Markup Language (GML) Encoding Standard: http://www.opengeospatial.org/standards/gml	The Geography Markup Language (GML) is the XML grammar defined by the Open Geospatial Consortium (OGC) to express geographical features. GML serves as a modeling language for geographic systems as well as an open interchange format for geographic transactions on the Internet.	
5	GMLJP2, GML in JPEG 2000 for Geographic Imagery	GML in JPEG 2000 for Geographic Imagery (GMLJP2) Encoding Specification http://portal.opengeospatial.org/files/?artifact_id=13252 Available at OGC GML page: http://www.opengeospatial.org/standards/gml	The OpenGIS encoding specification defines the means by which the OpenGIS GML is to be used within JPEG 2000 images for geographic imagery. Note: The ISO JPEG 2000 standard (http://www.jpeg.org/jpeg2000) is a wavelet based encoding for imagery that provides the ability to include XML data for description of the image within the JPEG 2000 data file.	

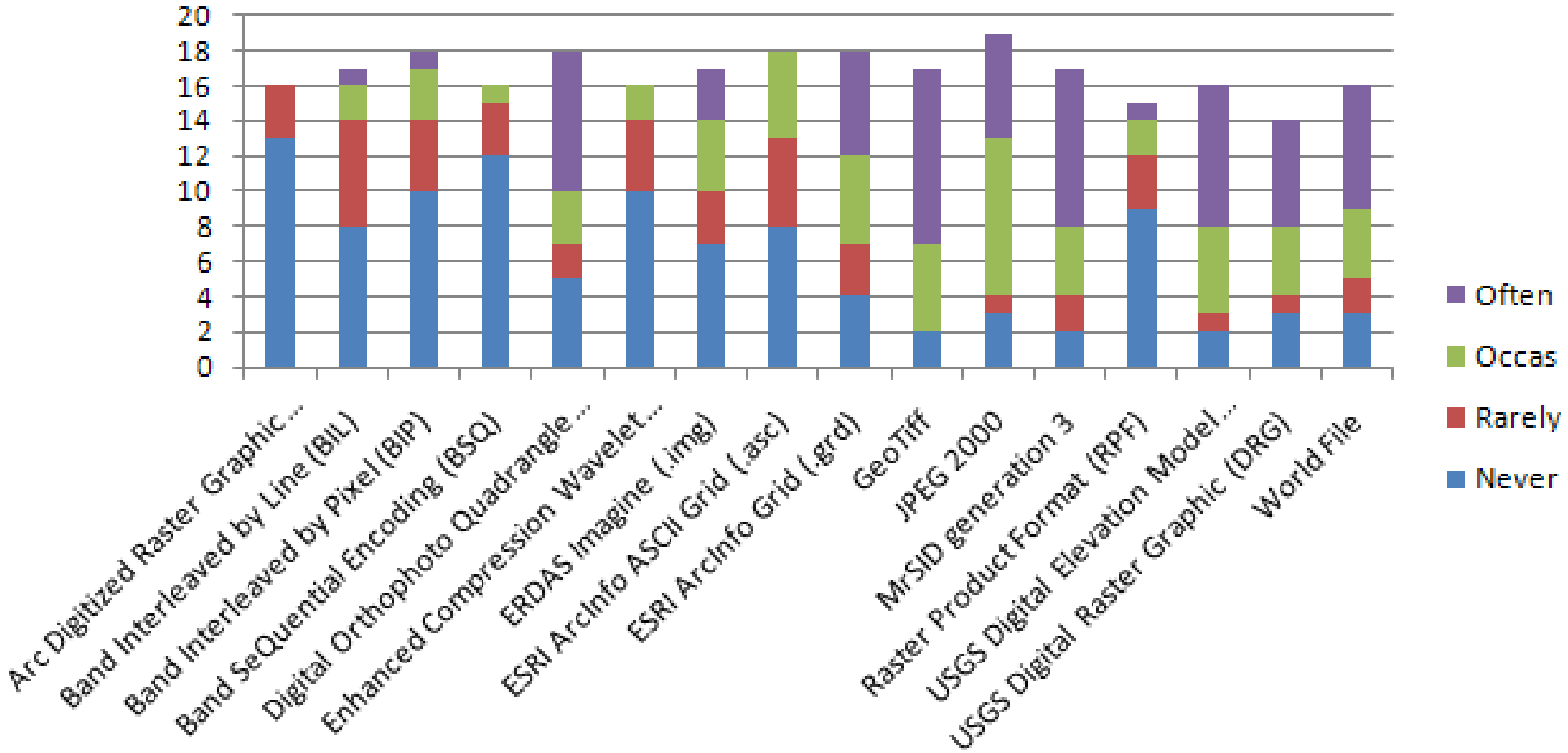
Resources

	A
1	**Additional Resources related to geospatial data formats:**
2	**Digital Image Data Formats**
3	http://www.scs.gmu.edu/~rgomez/EOS%20Lectures/7Lecture%2014%20Oct%2003/Formats/Digital%20Image%20Data%20Formats.doc
4	Provides a nice summary of the 4 major formats used for remotely sensed data (BIP, BIL, BSQ, Run-Length Encoding)
5	
6	**Federal Geospatial Data Systems Engineer Manual [EM 1110-1-2909] (2005)**
7	**Chapter 7: Geospatial Data Overview and Standards**
8	http://140.194.76.129/publications/eng-manuals/em1110-1-2909/c-7.pdf
9	
10	**FGDC Endorsed External Standards**
11	http://www.fgdc.gov/standards/fgdc-endorsed-external-standards/index_html
12	Standards endorsed by the FGDC Steering Committee to promote interoperability
13	
14	**GeoCommunity: GISDataDepot - GIS Data Formats**
15	http://data.geocomm.com/helpdesk/formats.html
16	Offers a brief description of several Vector and Raster data formats
17	
18	**Grime: GIS Formats: File Formats in GIS, 3D & Imagery and What to Treat them with …**
19	http://www.grime.net/gistools/a-z.htm
20	Offers and A-Z listing of numerous file formats, and summarizes a list of tools that can read them and write them.
21	
22	**Grime: GIS Tools: Useful 3rd party Data Conversion Tools for the GIS, 3D & Imagery Entrepreneur**
23	http://www.grime.net/gistools/index.htm
24	Offers an extensive list of utilities and summarizes what file formats they read and write.
25	
26	**Library of Congress: Sustainability of Digital Formats- Planning for Library of Congress Collections**
27	http://www.digitalpreservation.gov/formats/
28	
29	**McGarva, Guy, Steve Morris, and Greg Janee. "Technology Watch Report: Preserving Geospatial Data." May 2009.**
30	http://www.dpconline.org/component/docman/doc_download/363-preserving-geospatial-data-by-guy-mcgarva-steve-morris-and-gred-greg-janee
31	
32	**(NARA) U.S. National Archives and Records Administration: Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files - Raster Images.**
33	http://www.archives.gov/preservation/technical/guidelines.pdf
34	
35	**North Carolina State University Library: Geospatial Data Formats**
36	http://www.lib.ncsu.edu/gis/formats.html
37	Offers a brief description of some of the most prevalent Vector and Raster Data file formats

Survey Results

- Number of Respondents
  - 19 Respondents from 11 states (AZ, IL, KS, KY, ME, MS, MT, NY, NC, WI, WY)
  - 13 GIS Professionals : 6 Preservationists
- Most prevalent datasets noted
  - Vector: Esri File GDB, Esri Shapefile, Esri Personal GDB
  - Raster: GeoTIFF, MrSID, Raster Product Format, DEM, DOQ, Worldfile (*supplements several image file formats*), DRG
- Emerging formats??
  - Geo/Geospatial PDF?
  - KML?

Survey Results – Raster Formats

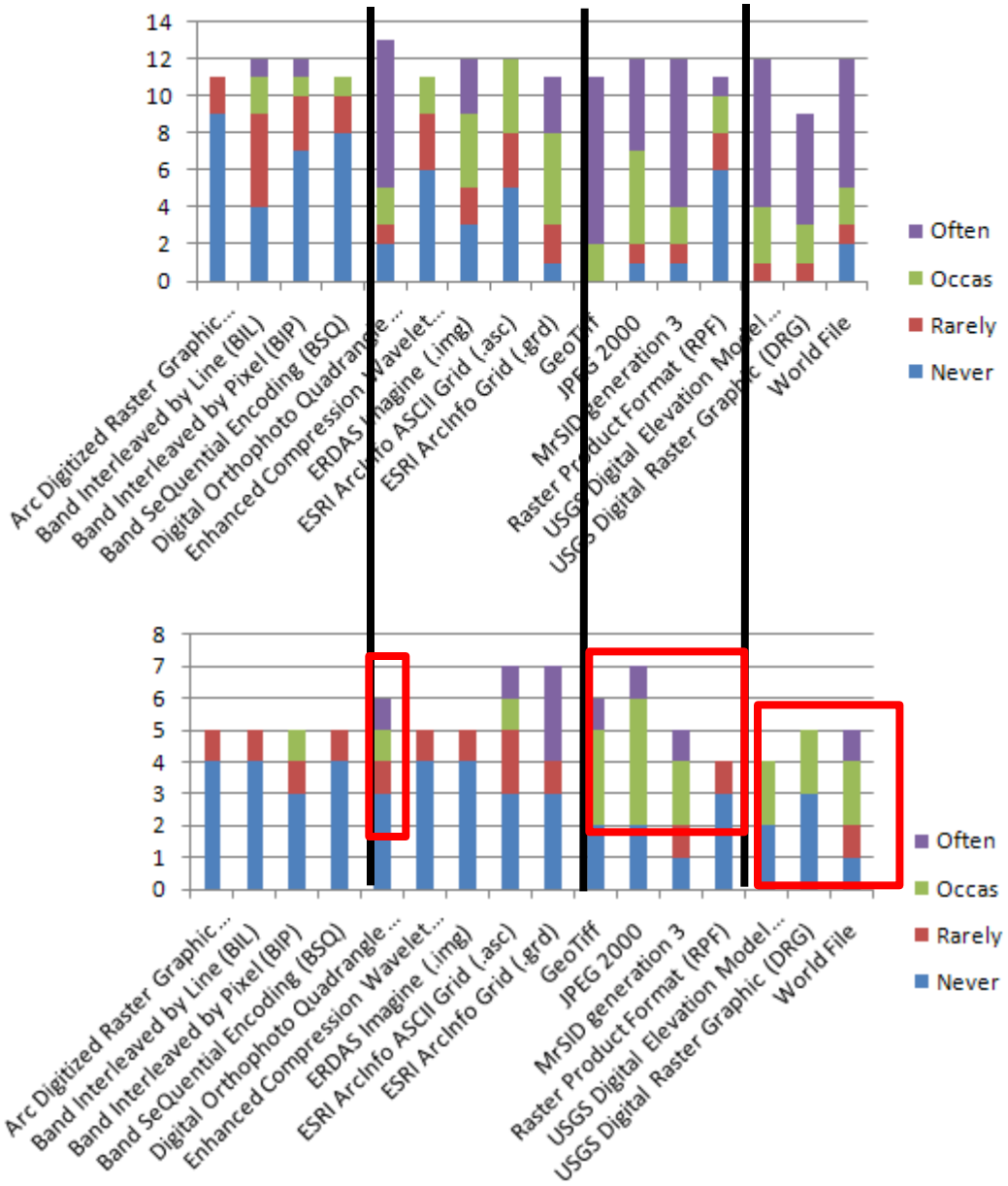


Other: JPEG (KYP), MrSID gen2(MT-G), FLT & HDR – ESRI Grid? (NY-P)

Survey Results – Raster Formats GIS v. Prsvtn

GIS

Preservation



Raster Outstanding Questions

Often Occasionally Rarely

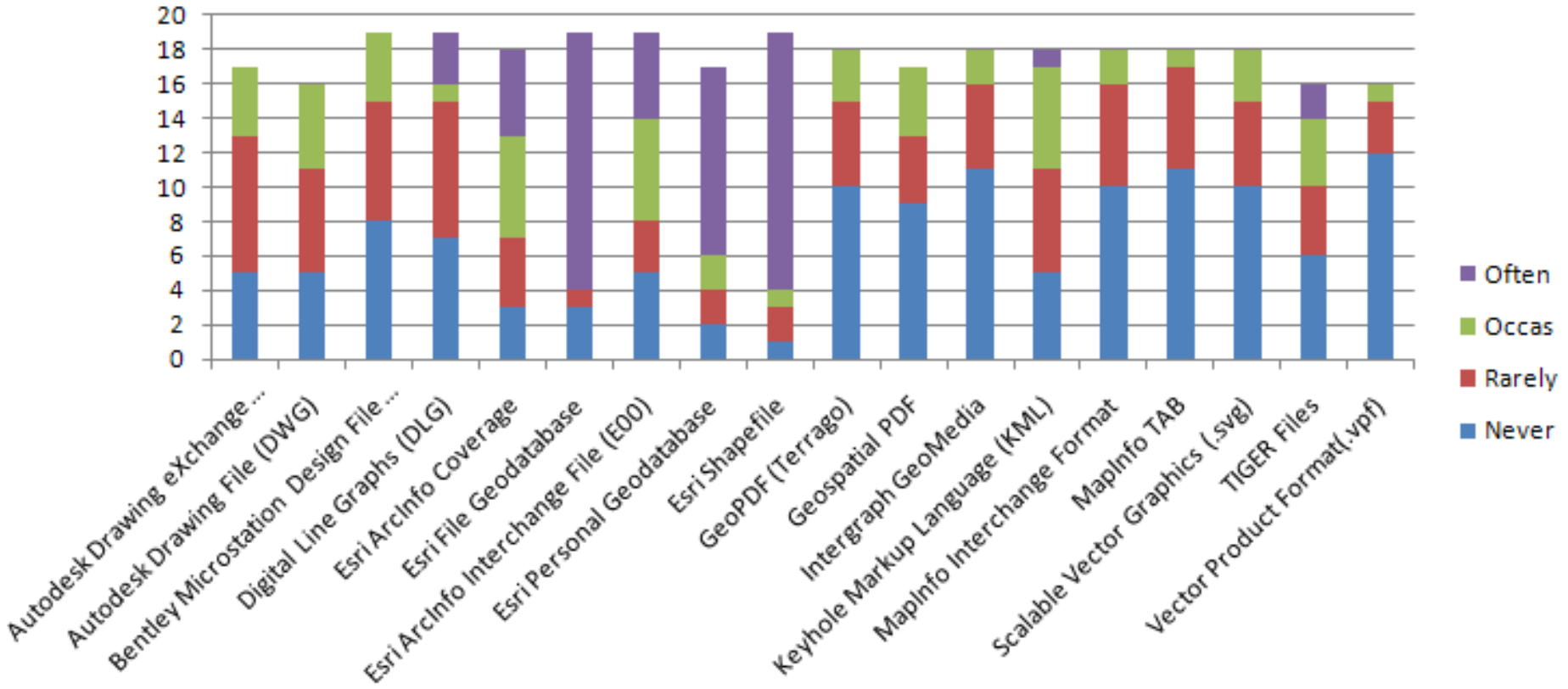
- Enhanced Compression Wavelet (.ecw) [AZ-G, KY-G, KS-G, ME-P, MT-G, NC-G]
  - Geospatial Imagery – aerial and satellite
  - Format Prevalence?
  - Survey Responses: 0 Often; 0 Occasional; 3 Rarely; 13 Never
- ERDAS Imagine (.img) [KS-G, MT-T, WY-G, KY-G, MT-G, NC-G, AZ-G, IL-G, ME-P,]
  - May contain statistics, ground control points, map information, image metadata
  - Format Currency?
  - Format Prevalence?
  - Survey Responses: 3 Often; 4 Occasional; 3 Rarely; 7 Never
- ESRI Arcinfo ASCII Grid (.asc)[IL-G, KS-G, KY-P, MT-G, WY-G, AZ-G, KY-GP, ME-P, MT-G]
  - GRID exchange file
  - Format Prevalence?
  - Survey Responses: 0 Often; 5 Occasional; 5 Rarely; 8 Never

Raster Outstanding Questions

Often **Occasionally** **Rarely**

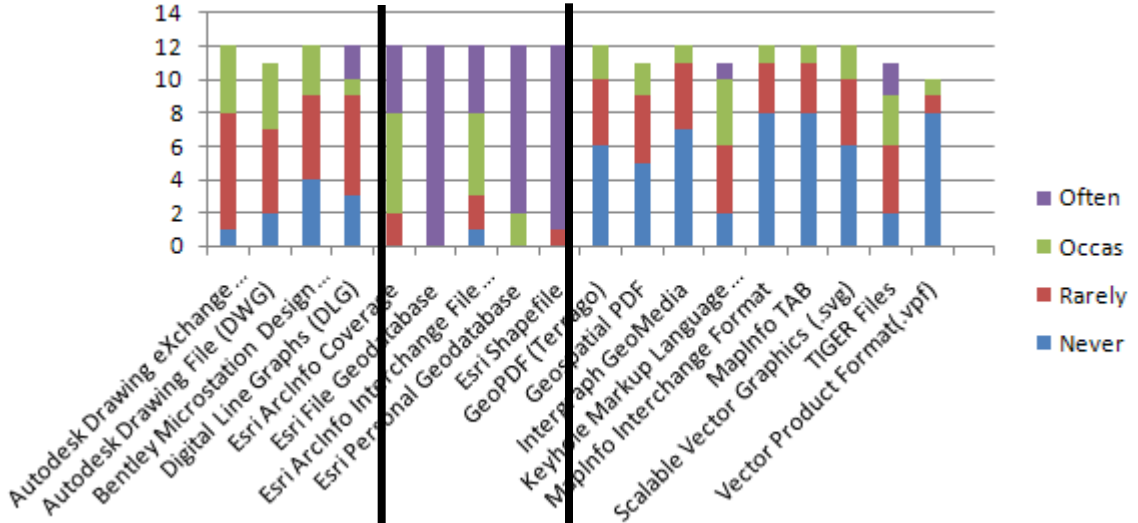
- USGS Digital Raster Graphic(.tif ?.drg) [AZ-G, KS-G, KY-G, MS-G, MT-G, WY-G, IL-G, KY-GP, MT-G, NC-G, WI-G]
  - Scanned, digitized, georeferenced USGS topo maps
  - Only rendered in TIFF? Is there a .drg file format?
  - Survey Responses: 7 Often; 5 Occasional; 1 Rarely; 2 Never
- World Files (*.xxw*) [AZ-G, KY-G, MT-G, NC-GP, WY-G, KS-G, KY-P, MT-G, NY-P, ME-P, WI-G]
  - Plain text data file used by GIS to georeference raster map images
  - Are world files interchangeable across GIS applications?
  - Survey Responses: 8 Often; 4 Occasional; 2 Rarely; 3 Never

Survey Results – Vector Formats

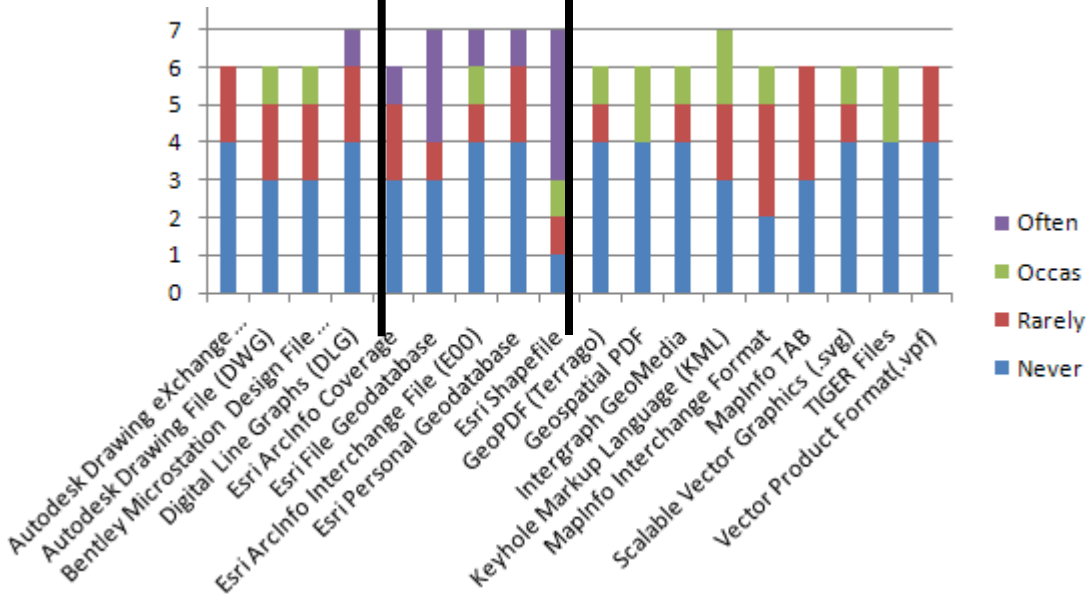


Survey Results – Vector Formats GIS v. Prsvtn

GIS



Preservation



Vector Outstanding Questions

- Digital Line Graphics(?.dlg?)) [AZ-P, KS-G, MT-G, AZ-G, IL-G, KY-GP, ME-P, MT-G, NC-G]
  - Digital vector data representing cartographic information derived from USGS maps and related sources. Contain info depicting geographic features (e.g. hypsography, hydrography, boundaries, roads, utility lines).
  - Is there a .dlg format?
  - Survey Responses: 3 Often; 1 Occasional; 8 Rarely; 7 Never
- Esri Archinfo Coverage (org'ed as a topical directory of feature classes) [AZ-GP, KY-G, MT-G, WY-G, IL-G, KS-G, KY-G, MS-G, MT-G, KY-P, NC-G]
  - A coverage is Esri's core data model – a georelational data model that stores vector data
  - Per ArcGIS Desktop 10 help: format has been superseded by geodatabase
  - Esri offers tools to migrate coverages to geodatabase (since v. 8.1)
  - Esri offers conversion tools to convert to other non-Esri formats
  - Any partners with experience converting coverages?
  - Survey Responses: 5 Often; 6 Occasional; 4 Rarely; 3 Never
- Intergraph GeoMedia (.mdb, mge) [AZ-G, ME-P, KS-G, KY-P, MS-G, MT-G, NC-G]
  - Intergraph suite of GIS editing tools
  - How many files comprise a dataset?
  - Survey Responses: 0 Often; 2 Occasional; 5 Rarely; 11 Never

Vector Outstanding Questions

- MapInfo Interchange Format(.mif, .mid) (KY-G, WY-G, AZ-G, KY-P, ME-P, MT-G, NY-P, NC-G)
  - ASCII text file format that fully describes contents of a MapInfo table
  - How many files comprise a dataset?
  - Format prevalence – ?niche use? Is this seen in the business community?
  - Survey Responses: 0 Often; 2 Occasional; 6 Rarely; 10 Never
- MapInfo TAB (?.tab?) (WY-G, AZ-G, KY-P, MT-G, NY-P, NC-G
  - MapInfo proprietary geospatial vector format
  - How many file comprise a dataset? Dataset composition?
  - Survey Responses: 0 Often; 1 Occasional; 6 Rarely; 11 Never
- Vector Product Format (.vpf) (KY-G, KY-P, ME-P, MT-G)
  - Military standard for vector-based digital map products produced by DOD
  - Format currency? Anybody see this format
  - Survey Responses: 0 Often; 1 Occasional; 3 Rarely; 12 Never

Standards Questions

- GML Simple Features (OGC specification for vector data)
  - GML Simple Features are an open OpenGIS standard which specifies digital storage of geographical data (point, line, polygon, etc.) with both spatial and non-spatial attributes.
  - The GML Simple Features Profile is a more complete profile of GML than the Point Profile and supports a wide range of vector feature objects.
  - Are there "GML Simple Features" files? Or does this describe how "simple feature" can be described in xml?
  - Is this a 'format' or a 'standard'?

Discussion Questions

Often **Occasionally** **Rarely**

- Any other file formats that we overlooked?
  - MrSID – did people respond to this based on the version or just MrSID?
- Have any info partners attempted any format migrations or conversions? What has been your experience with those conversions?
  - Conversions to enable access?
  - Legacy datasets to contemporary datasets?
  - Handing of ESRI Coverage data
- What is the nature of your KML usage [MT-G, WY-G, AZ-GP, KS-G, KY-P, MT-G, IL-G, ME-P, MS-G, MT-G, NC-G]
- How are you using PDF-based geospatial formats (Geospatial PDFs (Esri ArcMap) and GeoPDFs (TerraGo))?
  - Are you seeing *PDFs as a source for datasets or more as a derivative?
    - GeoPDF (TerraGo) [ME-P, MT-G, WY-G, MT-G, AZ-G, KS-G, KY-PG]
    - Geospatial PDF [AZ-GP, KY-P, ME-P, KS-G, KY-G, MT-G]
  - Use case for when using Geospatial PDFS and GeoPDFs (other than display)? Data capture?

More Discussion Topics

Often **Occasionally** **Rarely**

- Is an electronic map considered a legally recognized format of record within your state government?
  - What electronic formats have been approved for electronic maps?
- TIGER formats – other uses beyond Census-originated [MS-G, MT-G, AZ-P, KY-P, ME-P, MT-G, IL-G, NC-G]

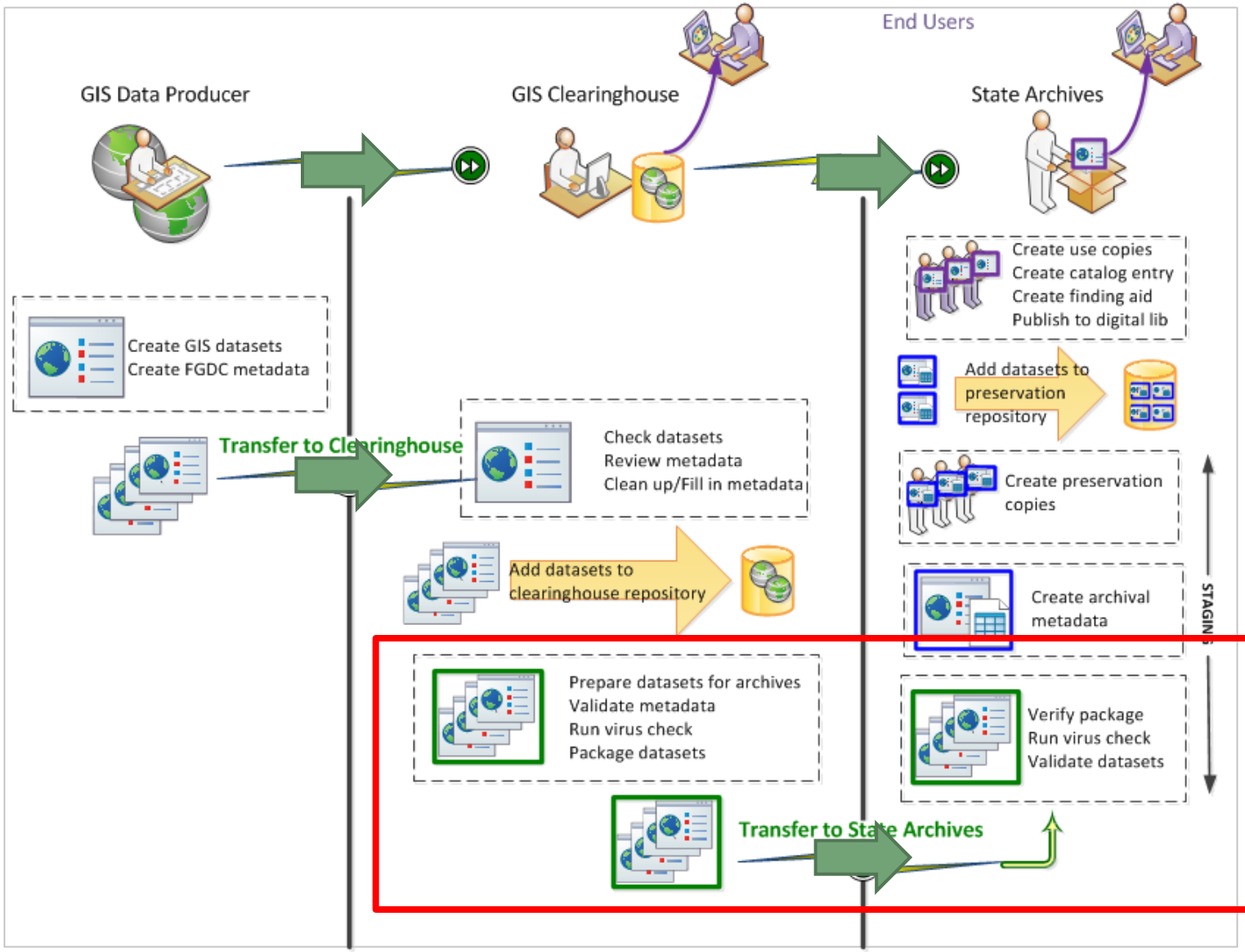
Next Steps

- Future feedback and observations on the File Formats spreadsheet …..
  - Send feedback to Alec Bethune (alec.bethune@nc.gov)

Best Practices for Geospatial Data Transfer for Digital Preservation

Lisa Speaker

Geospatial Data Life Cycle (w/clearinghouse)



GeoArchiving Solution Project Roadmap

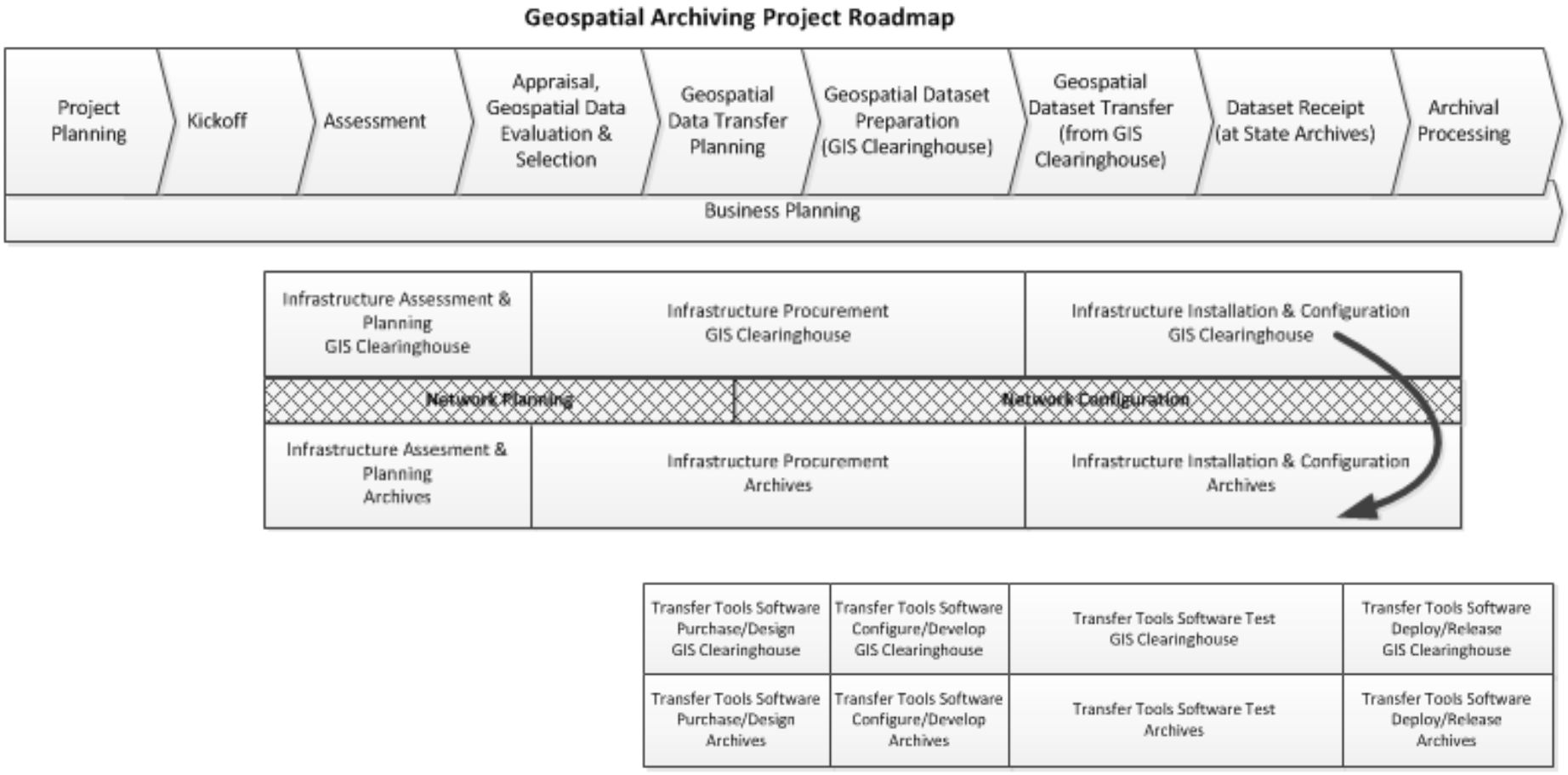


Table of Contents

1. Introduction 1
2. General Planning and Preparations 3
2.1. Getting Started with the Project – The Kick-off Meeting 3
2.2. Inventories, Assessments and Surveys 4
2.3. Appraisal 4
2.4. Geospatial Data Transfer Planning 5
2.4.1 Engage the Information Technology Organization 6
2.4.2. Digital Object Considerations for Data Transfer and Preservation 7
2.4.3 Naming Conventions 7
2.4.4 Data Organization 8
2.4.5 Dataset Verification 11
2.4.6 Dataset Packaging Mechanism for Transfer and Post-transfer Validation 13
2.4.7 Identify File Transfer Size Requirements 13
2.4.8 Identify File Transfer Mechanism 14
2.4.9 Identify Archival Submission Policies 16
2.4.10 Create a Written Document for Data Transfer 17
3. Role-based Tasks for Geospatial Data Transfer 17
3.1 Geospatial Data Creators 17
3.2. GIS Clearinghouse Geospatial Data Transfer Best Practices (or Submitter of geospatial data to the State Archives) 17
3.3. State Archives Geospatial Data Transfer Best Practices 19
4. Conclusions 21
Appendix A – Geospatial Archiving Roadmap A-1
Appendix B: Process Checklist A-2
Appendix C - Resources A-4

Cross-organizational Planning & Preparation for Data Transfer

Role-based activities in conducting Data Transfer

Checklist

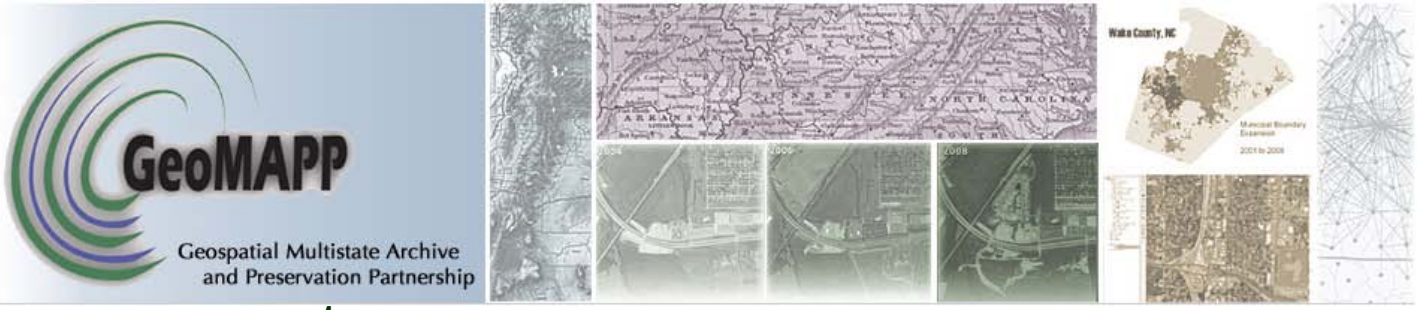
Future Informational Partner Meetings

Kelly Eubank

Upcoming Info Partner Meetings for 2011

- Hold at least 3 more meetings (Jul, Sept, Nov)
- Preliminary date for next meeting: July 28
- Seeking Informational Partner Feedback for topics
  - Will send out a Survey Monkey to get input
  - Some examples include: [use/not use following items to seed]
    - Appraisal and Selection for archiving and long-term preservation
      - Process/Approach to update legislative records schedules to include GIS
    - Data Transfer Best Practices
    - Next Steps/Lessons Learned—GeoMAPP and beyond
    - Other Informational Partner suggestions? What do you want to know more about and discuss?

Thanks!

http://imagecache2.allposters.com/

- Mark Myers, KDLA
  - mark.myers@ky.gov
- Alec Bethune, NC CGIA
  - alec.bethune@nc.gov
- Kelly Eubank, NC Archives
  - kelly.eubank@ncdcr.gov
- Lisa Speaker, NC Archives
  - Lisa.speaker@ncdcr.gov