Geospatial Multistate Archive and Preservation Partnership

In Partnership with

The Library of Congress

National Digital Information Infrastructure and Preservation Program (NDIIPP)

Final Report: 2007-2011

December 2011

Prepared by:

North Carolina Center for Geographic Information and Analysis
North Carolina Department of Cultural Resources

In partnership with:

Kentucky Department for Libraries and Archives
Kentucky Division of Geographic Information
Kentucky State University
Montana State Library
North Carolina State University Libraries
Utah Automated Geographic Reference Center
Utah State Archives

Table of Contents

Executive Summary

State governments have long understood the value of using geospatial information in decision making processes and planning efforts. State agencies now embrace the use of GIS information to analyze real world problems, to display and describe the physical world in digital graphical formats in order to provide more efficient and effective services to their citizens. State governments are also beginning to recognize the value of having access to older geospatial data as a resource to explore societal, environmental and economic change over time. Compelling business uses such as tracking changes in population, land, or vegetation over time or providing a cultural record of place are spurring users to seek out and use superseded geospatial content.

State GIS and archival organizations are making efforts to respond to this information need, however, they face serious obstacles. Traditionally, it has not been a priority for data creators to preserve superseded geospatial information. Limited resources, diminishing budgets, and in some cases a lack of understanding by decision-makers or practitioners about the benefits of preserving geospatial data can stifle efforts to implement a formal preservation plan. As a result, older data is often overwritten or lost when more current information is received or as data is updated. As such, geospatial data is extremely susceptible to either temporary or permanent loss.

The Geospatial Multistate Archive and Preservation Partnership (GeoMAPP) was formed in 2007 to address the challenges associated with identifying, preserving and providing long-term access to temporally significant digital geospatial content in state and local governments; dynamic data that is at-risk of being lost when updates are made. The project is one of four initial state government partnerships funded by the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP), and includes representatives from state geospatial and archival organizations from Kentucky, Montana, North Carolina and Utah.

From November 2007 to December 2009, three state partners (KY, NC, UT) worked together to investigate approaches for preserving and providing access to superseded geospatial data. Concurrently the partners engaged GIS data creators and archives leaders from local and state government within each state and nationally to raise awareness about geoarchives issues and solicit feedback.

GeoMAPP Key Findings

1. **Establish a geoarchiving team** with participation from the GIS, archives and IT communities to work with data producers, cross train and tackle the geoarchiving challenge.
2. **Inventory GIS holdings** and document information such as data ownership, theme, age, frequency of update, format and size about data to be considered for preservation.
3. **Appraise**-Develop a formal policy to assess which datasets need to be preserved based on legal, historical, business, and research value.
4. **Data Prep, Transfer and Ingest**- Develop standards for metadata, file formats, file naming and data packaging, and create attainable processes to prepare, transfer, review and ingest geospatial data into a robust archive for long-term preservation
5. **Preserve**- Store multiple copies of archived data on diverse storage systems that track the location and integrity of each file
6. **Provide Access** to your archived holdings to allow the public to take advantage of these resources, to add their own value to the data and to become supports of the geoarchiving process
7. **Justify the Investment** through the development of metrics for measuring costs and benefits derived from specific use cases and the preservation process. Develop a programmatic strategy to track and document benefits over time to demonstrate success.

In the final phase of work from January 2010-December 2011, GeoMAPP added the state of Montana as a fourth full partner. GeoMAPP research during this period focused on detailed technical explorations relating to the long-term preservation, storage and access of archived geospatial content. The project documented iterative findings and best practices in a series of technical tools and whitepapers. Outreach continued with state and national GIS and archives communities, and the team continued to engage with an expanding network of project Informational Partners. A final key project focus was to engage with a contractor team to develop tools and templates for generating business planning and justification documentation to support the funding of geoarchiving activities.

The GeoMAPP partnership brought together GIS practitioners, archives professionals, university researchers and librarians to build awareness that older, archived content needs to be preserved and has great value once made accessible.

The core objectives during the second phase of the project included:

- Integrate Montana as funded full partner and expand the voluntary Informational Partner program;
- Conduct detailed collaborative research on technical matters surrounding the transfer and long-term preservation of geospatial data, and explore advanced access methods to these data while publishing results in the form of project whitepapers and tools;
- Create business planning tools, templates and documentation that will help states develop metrics to support the establishment and ongoing support of sustainable geoarchives;
- Continue the project's outreach and mentoring mission of engaging with GIS data creators and custodians, archives practitioners, and members of industry at local statewide and national events.

Key GeoMAPP Deliverables:

- Project Interim Report (2007-09)
- Geoarchiving Self Assessment Tool
- National, State, and Local Geoarchiving Survey Results
- Data Transfer Best Practice Documentation
- Geospatial Data File Formats Guide
- Geospatial Metadata for Preservation Whitepaper
- Archival Metadata for Geoarchiving Whitepaper
- Archival Processing Guide
- Storage Infrastructure for Geoarchiving Assessment
- Business Planning Toolkit and Documentation

*note: All of these documents can be found on www.geomapp.net

GeoMAPP would like to sincerely thank the Library of Congress' NDIIPP program for its generous funding and support. The breadth of research and the scale of outreach and engagement far exceeded any preconceived notions on the part of the partners. Each state takes with them a much broader depth of knowledge, a greater appreciation for the challenges and opportunities with geospatial information, and a wider network of practitioners with whom to continue to collaborate and pursue answers. Although the partnership accomplished quite a bit, there is still much more to do to ensure that critical geospatial datasets are preserved and made available to the public.

Project Overview

Introducing GeoMAPP

Over the past thirty years the practice of generating maps and other cartographic products within state and local governments has been subject to incredible technological shifts: a migration from entirely paper-based workflows, to computer generation of paper maps, to our current all-digital world where born-digital geospatial datasets are used to create digital maps, new digital geospatial datasets, and dynamic web mapping applications and services. These shifts have necessitated changes in the records management of geographic data, evolving from capturing and preserving paper and Mylar-based geographic products derived manually from sketches, digitizing pucks and scanners, to the need to capture and preserve complex and diverse geospatial databases reliant on various software tools and relational databases.

The quantity, size and dynamic nature of modern GIS technologies and data introduces new challenges to archivists and librarians: how does one capture and preserve snapshots of superseded critical state and local government geospatial data that may be updated on a daily basis; how does one ensure that large archival repositories of geospatial data stay usable and authentic over time; how can the public discover and easily access these collections to benefit research and add value; and what resources are required to develop and maintain an archive for large geospatial data collections?

In November 2007, under the auspices of the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP)[1], state government archives and GIS practitioners from Kentucky, North Carolina, and Utah chartered a partnership with the Library of Congress to investigate these questions and other issues related to the preservation of geospatial content. This effort, the Geospatial Multistate Archive and Preservation Partnership (GeoMAPP)[2], began with four key objectives:

- Identify and inventory geospatial content within each state that is temporally valuable or is "at-risk" of being lost when updates made or are not preserved due large size;
- Explore the challenges of building collaborative relationships across organizational units within each state and across state lines to investigate the technical challenges related to the inventory, appraisal, ingest, storage and preservation processes to ensure the long-term viability and accessibility of valuable digital geospatial data;
- Develop business planning materials and practices that can be used to justify the creation, expansion or maintenance of a sustainable geoarchive;
- Conduct outreach with both the data creator community and with representatives from the geospatial and archives technologies industry to build awareness of the need to address the long-term preservation and access of valuable digital geospatial data.

The initial research efforts spanning 2007- 2009 are documented in the project's interim report.[3].

[1] Library of Congress' NDIIPP: http://www.digitalpreservation.gov/
[2] GeoMAPP: http://www.geomapp.net
[3] GeoMAPP Interim Report: http://www.geomapp.net/docs/GeoMAPP_InterimReport_Final.pdf

In January 2010, the project began a new phase of work that built on the earlier efforts with new work concentrating on:

- Adding new states as funded Full Partners and voluntary Informational Partners;
- Conducting detailed research into technical matters surrounding the transfer, storage and long-term preservation of geospatial data, and providing advanced access methods to these data;
- Engaging an outside contractor to assist with developing tools, templates and documentation to assist in the development of business planning materials to support the establishment and ongoing support of sustainable geoarchives;
- Continuing the project's outreach and mentoring mission of engaging with GIS data creators and custodians, archives practitioners, and the GIS and archives commercial sector at local, statewide and national events.

After conducting a national search and a formal review and assessment process, GeoMAPP selected the state of Montana to join the partnership in January of 2011. Montana provided a new perspective on the geoarchiving challenge and assisted with new technical explorations. Montana benefited from the project's earlier findings and was able to test the initial partners' assumptions from existing best practice documentation.

GeoMAPP's technical explorations for the project's second phase drew heavily on the foundations that were built during the project's initial phase and the focus areas for the new work were meant to address gaps in understanding identified during the project's initial efforts. However, having already learned to speak each other's language and having learned from the experience of developing early data preparation, transfer and ingest workflows the existing partners were well positioned to dig deeper into technical elements to benefit the packaging and transfer, long-term storage and preservation, and discoverability and access to archived geospatial content.

A challenging national economic climate helped catalyze a project focus on the need to investigate tools and techniques for developing business case and business planning materials and to develop a business planning tool suite to make the business case for funding geoarchiving activities. This effort drew heavily from states' previous funding request efforts and other successful business models within the partnership. GeoMAPP also engaged an outside contractor to review existing project business planning materials and help the project develop new tools and templates. The resulting business planning products are available for use by the partners and the community at large to develop business plans to support the establishment and continued support of dynamic geospatial data archives.

GeoMAPP: The Power of Partnership

One of the unique elements of the GeoMAPP partnership has been the distributed nature in which collective project tasks have been completed with direct involvement from each of the project partners. While each partner took a unique "state-centric" approach to implementing geoarchiving systems and workflows, all participants brought their findings and questions back to the partnership for discussion and application.

Collaboration in a multistate consortium is atypical of how state governments customarily address technological challenges. Tight staffing constraints often limit organizations to be focused on managing existing processes and addressing issues only when production challenges occur. Partners worked diligently to share their experiences, to learn from each other and to form project-wide generalized recommendations, best practices and standards.

The GeoMAPP partnership is comprised of the following agencies:

- North Carolina Center for Geographic Information and Analysis (Principal Investigator)
- North Carolina State Archives- Electronic Records Branch (Co- Principal Investigator)
- North Carolina State University Libraries

- Kentucky Department for Libraries and Archives
- Kentucky Division of Geographic Information
- Kentucky State University

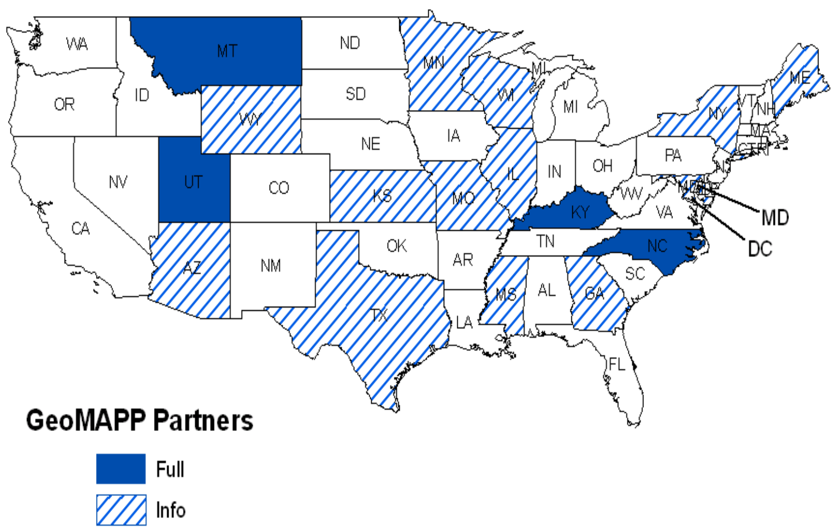
- Montana State Library

- Utah State Archives and Records Service
- Utah Automated Geographic Reference Center

GeoMAPP's pairing of preservation and GIS staff from each state enabled each of the state partners to establish or enhance the relationship between these organizations and to jointly investigate the challenges of preserving geospatial content within their state and among the project. The "getting to know each other" process featured building a familiarization of each discipline's terms and jargon while providing formal cross-training opportunities between groups on both archival and GIS tools and technologies. By understanding each other's language and learning both groups' workflows and responsibilities, the state teams were better prepared to tackle the challenge of preserving geospatial content.

To support the continued relationship building and bolster collaborative efforts among the state partners, the project held three face-to-face partner meetings during the project's second phase: June 2010 in Salt Lake City Utah, September 2010 in Phoenix Arizona, and in June 2011 in Helena Montana. Each of these events allowed the partners to engage directly with each other, to strengthen team efforts and to discuss technical explorations. The meetings also allowed the project to interface with individuals from the host state that were not directly engaged with the project. These interactions with individuals from state government, industry, and the community allowed GeoMAPP partners to engage new communities, share its geoarchiving message, and to learn from the related efforts and areas of interest from the speakers.

The GeoMAPP partnership focused significant effort during the final phase of work to growing the size of the project's Informational Partnership and to increase project engagement with this unique arm of the partnership. Initially composed of both GIS and archives practitioners from state governments when initiated in 2009, the GeoMAPP Informational Partnership grew to include standalone GIS or archival entities and federal partners such as the National Archives and Records Administration as well as unofficial university and local government participants. In late 2011 the Informational Partnership included official engagement from 18 states (including the 4 full partners) representing over one third of the states in the union.



GeoMAPP Partners

GeoMAPP Informational Partners:

Arizona Arizona State Library, Archives and Public Records Arizona State Cartographer's Office	**Kansas** Kansas Historical Society Kansas Information Technology Office	**Mississippi** Mississippi Department of Archives and History Mississippi Geospatial Clearinghouse	**Wisconsin** University of Wisconsin-Madison Wisconsin Department of Administration
The District of Columbia District of Columbia Office of Public Records District of Columbia Office of the Chief Technology Officer	**Maine** Maine State Archives Maine Office of GIS	**Missouri** Missouri State Geographic Information Office Missouri Spatial Data Information Service	**Wyoming** American Heritage Center, University of Wyoming Wyoming Geographic Information Science Center, University of Wyoming
Georgia Records and Information Management Services - Georgia Archives Information Technology Outreach Services Division, CVIOG-UGA	**Maryland** Maryland State Archives Maryland Department of Natural Resources	**New York** New York State Archives New York State Office of Cyber Security and Critical Infrastructure Coordination	
Illinois Illinois State Geological Survey	**Minnesota** Minnesota Historical Society Minnesota Department of Administration, Geospatial Information Office (MnGeo)	**Texas** Texas State Library and Archives Commission Texas Natural Resource Information System	

GeoMAPP Partner Highlights: States' Unique Approaches and Next Steps

A core tenant of the GeoMAPP effort has been cross-state collaboration to jointly investigate, dissect and propose solutions to the challenges surrounding the preservation of geospatial data. While the joint findings are highlighted under the GeoMAPP banner, many of the project's unique findings were discovered within the Petri dishes of discovery within each state. Each state took a unique approach to tackling geoarchiving and implemented unique workflows and technologies that worked within their own state context and leveraged existing systems within their organizations.

The following section provides background on some of the unique approaches and advancements of GeoMAPP's partner states and provides an additional organizational and technical overview for Montana's GIS and preservation programs. For more organizational background detail on Kentucky, North Carolina and Utah please see the GeoMAPP Interim Report.

Kentucky[4]

The Kentucky GeoMAPP team is comprised of staff from the Department for Libraries and Archives (KDLA), the state's primary archival body and the Department of Geographic Information (DGI), which manages the Kentucky Geography Network (KYGEONET)[5] Kentucky's geospatial data clearinghouse. Organizationally, DGI falls under Kentucky's Commonwealth Office of Technology (COT). The team also received technical GIS training, consultation, and project assistance from Kentucky State University.

During the first phase of the project Kentucky concentrated on inventory, scheduling, and physically transferring records to the archives and between GeoMAPP partner archival organizations. During the second phase of the project, Kentucky consolidated what it had learned from the first phase and concentrated on preservation, access and documentation of data transfer.

Electronic Records Program Background

At GeoMAPP's inception KDLA had three staff members who accessioned geospatial data, e-mail, website snapshots, state publications, governor's records and meeting minutes into their archive. Despite the loss of a team member during the project period, GeoMAPP allowed Kentucky to continue to expand its electronic records program through the grant's financial support, sharing of ideas/techniques, and development of best practices. The team developed a DSpace repository application that houses GIS and other electronic records. The Kentucky DSpace repository stores shapefiles, small images and PDFs, and plans are in place to describe and reference Esri File Geodatabases and large image stores that are external to the DSpace instance. Throughout the project, Kentucky's electronic records holdings continued to grow and the team is focusing on accessioning additional records.

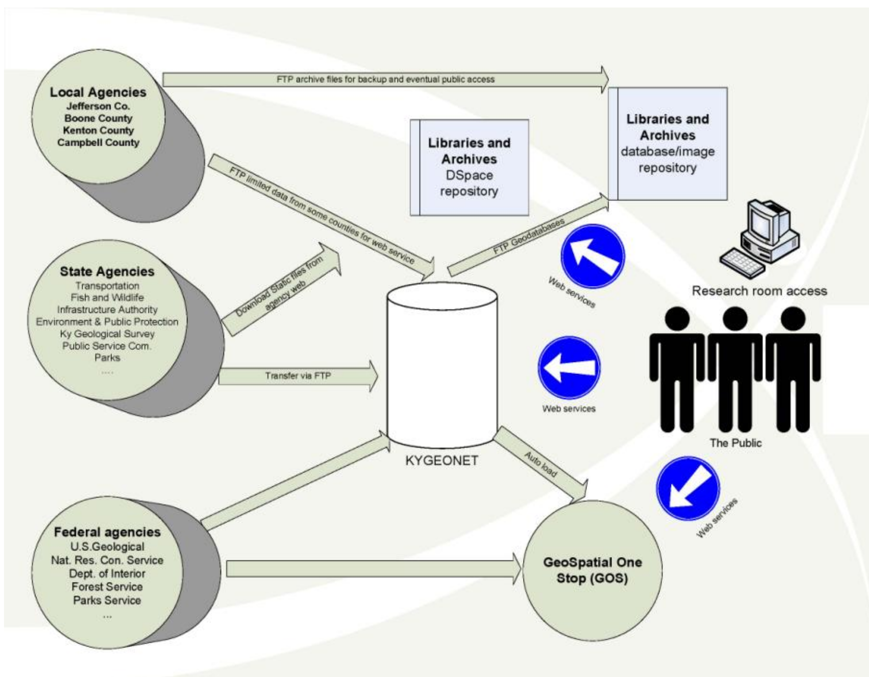
Kentucky's Geospatial Architecture

The Commonwealth of Kentucky takes a fairly centralized approach for their geospatial holdings and hosts data for local, regional, state and federal entities on the Kentucky Geography Network. All of the resources made available via the KYGEONET feed the Commonwealth's Enterprise GIS Databases,

[4] For more background on Kentucky, see the GeoMAPP Interim Report: pp12-14
[5] KYGEONET, http://KYGEONET.ky.gov/

KyRaster and KyVector, which are managed by the Division of Geographic Information (DGI). These databases are accessed by hundreds of GIS users in State Government on a daily basis. There are no formal agreements in place nor do any mandates exist that require data producers to provide their geospatial data resources to the KYGEONET. Participation is voluntary; however, entities have chosen to contribute due to the exposure their data receives and the benefits that are realized from having the data accessible in a "self-serve" manner.

In order for data to be ingested into the KYGEONET, geospatial data resources must include a minimum set of Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM)-compliant metadata[6]. If the required metadata is not present, the data will not be ingested into the KYGEONET or the Enterprise Databases. In most instances the data submitted for distribution is an Esri shapefile, or file and tile-based image datasets. Transfer of this data occurs via network shares, FTP, DVD/CD, and portable hard drives. One of the primary challenges the Kentucky team has faced in data acquisition has been with several regional agencies responsible for hosting local government data that charge for data access. This restricted access has limited the archiving efforts for this data, but participation in GeoMAPP has helped catalyze discussion between KDLA, DGI and the data providers.



How geospatial data moves within the Commonwealth of Kentucky

Project Accomplishments

Kentucky began the GeoMAPP project with a small number of image files transferred to the archives based on a single broad series in the state records retention schedule. Through participation in the GeoMAPP project, the Kentucky team accomplished the following:

[6] Geospatial metadata details can be found on p.27 of this report

- Established an archival collection of geospatial records (File Geodatabase snapshots, geospatial PDF files and shapefiles) transferred from state and local agencies;
- Created a web interface providing description of and access to archived geospatial records and an in-house tool that permits user access to all vector snapshots arranged chronologically, with accompanying image files;
- Participated in conference presentations and Informational Partners presentations that fostered learning about best practices for geospatial record archiving;
- Raised awareness within the geospatial record creating community for the need to manage and preserve records using a revised records retention schedule and archival techniques;
- Executed Memoranda of Agreement with local government GIS record creators to store preservation copies of geospatial records at the archives and make them accessible on a predetermined schedule.

In addition to participating in and leading project working group efforts, the KDLA and DGI teams worked together to develop a map interface to provide access to records in the DSpace e-archives. DGI had limited involvement with GeoMAPP during the final months of the project due to depletion of their budget and limited staff resources, and they were not able to participate in later activities of the project, including the writing of the final report. Despite DGI's withdrawal from the project, DGI and KDLA have continued the archiving process.

Without a formal contract to GeoMAPP, staff from Kentucky State University offered the archivists Esri training, participated in numerous meetings, and advised KDLA in the formation of the geospatial portion of its e-archives.

Introducing- Montana

The Montana State Library was integrated into GeoMAPP as a full partner in February 2011 after participating for a year as a GeoMAPP Informational Partner. As a full partner, Montana provided a unique organizational structure to the GeoMAPP program because the GIS clearinghouse and informal GIS archiving activities have long been a combined and integral responsibility of the Montana State Library (MSL). In 1985, the Montana State Legislature first funded the Natural Resource Information System (NRIS). NRIS has its roots in the environmental impact statement process that placed increased demands on natural resources agencies throughout state government at that time. Agencies and the legislature recognized that duplication of effort could be significantly reduced through an aggregated resource for the natural resource data needed to complete these processes. They also recognized that the home for this resource should be the State Library. In the late 1980's the role of NRIS evolved to include GIS data management because of the value that GIS brings to natural resource information management.

In a more traditional library role, MSL also manages the State Publications Depository Program to provide permanent public access to Montana state publications. This program actively manages both print and digital state publications for Montana. As one agency, the Montana State Library, brings the resources and expertise of these two programs to GeoMAPP.

Electronic records and publications program background

The Secretary of State (through the State Records Manager) is responsible for records management policy setting for the State of Montana, though, as in most states, responsibility for records management remains at the records producer level. Records deemed to be of permanent value are eventually transferred to the State Archives, which is a program of the Montana Historical Society (MHS).

MHS has an Electronic Records Project Archivist who is funded by the Legislative Services Division. The Electronic Records Project Archivist works with Montana Legislative Services to preserve the digital audio and video recordings of legislative standing committee hearings, floor sessions and interim committees. The Electronic Records Project Archivist also manages analog recordings and paper records of the Legislature. The digitized recordings of sessions dating from 2005 have been migrated to an archival format and preserved as part of the NDIIPP sponsored Washington State Digital Archives. Ultimately MHS would like to fully fund this position with a full time permanent staff member within the archives devoted to electronic records throughout state and local government.

As of December 2011, efforts to implement a statewide Montana electronic records management system have been slow moving. A Montana Electronic Records Initiative was established in 2008, which published a strategic plan.[7] Leadership from across the executive branch participated in the initiative. In the coming years, the goals of the strategic plan will be addressed through legislation and cooperative agreements between agencies. In the meantime the State Records Committee and Local Government Records Committee address electronic records retention issues via retention schedules and records management guidelines including mandatory migration plans for records with retentions longer than 10 years (per statute). These committees also work to educate agencies about the importance of, and legal requirements to, define and protect essential records in electronic formats, both born-digital and digitized, through workshops, webinars and guidelines.

As stated, MSL is responsible for permanent public access for state publications, a subset of state records. This responsibility is defined broadly in statute to mean any information produced by the government of Montana that is intended for public distribution. To this end, MSL actively digitizes print publications and collects born digital state publications. Since 2007 these publications have been managed, made accessible and preserved through a partnership with the Internet Archive. As of March, 2011 MSL had digitized 1 million pages of state publications, comprising more than 15,000 publications made available online through the Internet Archive. MSL also subscribes to Archive-it, the Internet Archives web archive service. By archiving state agency websites, MSL captures state publications published to those websites that do not make it to MSL through the state depository system and preserves the context of those publications on the web.

Although responsibility for electronic government information in Montana is distributed in this way, the Secretary of State, the Historical Society, and the State Library work closely with one another. The three programs joined together to create a Permanent Public Access Committee which was responsible for updating the Montana state publications statute.

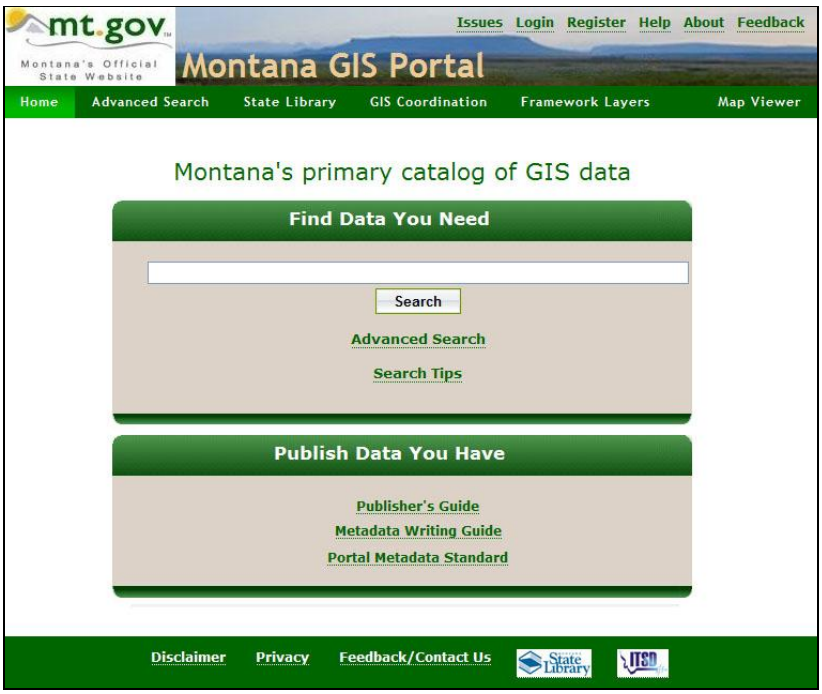
Finally, MSL and MHS share a joint license to the digital content management system, CONTENTdm, which is the software that supports the Montana Memory Project[8]. The Montana Memory Project is a

[7] Montana strategic plan: http://sos.mt.gov/Records/committees/erim_resources/Strategic_Plan_Version_6.pdf .
[8] Montana Memory Project: http://mtmemory.org

resource for all Montana libraries, archives, museums and historical societies to use to make accessible the digital Montana content that they manage. Examples of electronic records currently accessible through the MMP Montana Memory Project include Governors' records, Constitutional Convention records, livestock brand records, and state prison records.

Existing GIS architecture

NRIS has served as the state GIS clearinghouse for almost two decades. In this capacity, NRIS manages a large GIS data collection and makes GIS data available via web applications, web mapping services, and as downloadable data. For years MSL has maintained a GIS data list and in 2008 MSL launched the Montana GIS Portal based on the Esri GeoPortal Toolkit. This portal provides discovery and access to more than 400 Montana-related GIS datasets served both by MSL as well as other state agencies and local governments. MSL is also the theme steward for twelve of the fourteen framework layers that comprise the Montana Spatial Data Infrastructure (MSDI).



The Montana GIS Portal

NRIS has long been a statewide leader and advocate for the creation of GIS metadata. The State of Montana created a technical standard[9] that requires agencies that wish to publish metadata to the Montana GIS Portal, to do so following Technical Specifications drafted by MSL. These technical specifications[10] support the functionality of the Esri GIS Portal Toolkit and are FGDC CSDGM-compliant.

The technical architecture that supports NRIS and the MSDI is a hybrid model. MSL maintains a SQL database and ArcGIS Server environment in a local data center. Additionally, the State Information Technology Services Division hosts an enterprise ArcGIS Server environment to host web services. More details about MSL's infrastructure are found elsewhere in this report.

[9] MT Metadata Standard: http://itsd.mt.gov/content/policy/policies/Infrastructure/1200-XS3
[10] MT Geoportal Tech Specs: http://itsd.mt.gov/content/policy/policies/Infrastructure/1200-XS4

Existing geoarchives activities

Montana does not currently include GIS data in agency records retention schedules. MSL has long been recognized as the “archives” for GIS data, a role that is supported by state Records Management, the State Archives, the State Geographic Information Officer and the Montana Land Information Advisory Council (MLIAC).[11] Prior to joining GeoMAPP, MSL’s process to archive GIS data was to simply not throw any data away. In the past, NRIS had made attempts to inventory the GIS data collection. The last inventory took place in approximately 2002. Unfortunately, after completing the initial inventory no attempt was made to develop a system to manage this information in a way that was useful long-term. The majority of the NRIS data collection has good descriptive metadata but no thought was given to archival metadata. Backups of the data exist but are not integrity checked and no thought was given to format migration. Access to data in certain formats is in question. Finally, no effort was made to distinguish archived or superseded data from current data.

As MSL approached the GeoMAPP project it was understood that MSL would take a “library collection development policy” approach to managing a GIS data archive rather than a “records management” approach. The authority to undertake this role is granted under the statute that governs state publications as well as MSL’s collection development policy. This policy states that MSL will manage state government information and natural resource information about Montana including GIS data. Further work needs to be completed to more formally define MSL’s GIS data collection development policy. Emphasis is placed on Montana GIS Clearinghouse data which has a statewide focus and MSDI data which incorporates both local and state data.

Outreach

MSL staff spent the early part of 2011 reviewing the research and documentation prepared by GeoMAPP members during previous years of the grant. Prior to participating in a two-day mentoring session conducted by other participating states, MSL completed the *GeoMAPP Geoarchiving Self-Assessment*. This self-assessment will be shared with the State Records Manager, the State Archivist, the State Geographic Information Officer and MLIAC. MLIAC meets quarterly and is now chaired by MSL. NRIS staff regularly updates the Council on GeoMAPP activities at their meetings.

MSL offered an introduction to GIS data archiving session at the State Information Technology Conference hosted by the State Information Technology Services Division in Helena the first week of December 2011.

Benefits/drawbacks of being the GeoMAPP guinea pig

Montana was in the unique position of joining GeoMAPP over three years after the project’s inception. This gave MSL staff the benefit of taking advantage of the considerable research that had already been completed by other partners while at the same time being able to lend a new perspective to research that was underway. Because the other partners know each other so well, MSL staff found the June face-to-face meeting to be particularly valuable to learn about the group dynamics that were not always conveyed in conference calls. This meeting gave staff the comfort level to participate on an equal level with other partners.

[11] The MLIAC is the council that makes recommendations on statewide GIS policy for Montana.

For new states that wish to take on the challenge of archiving GIS data, the value of mentoring sessions with subject matter experts should not be underestimated. MSL found these sessions to be most valuable when they were short and focused on a specific topic, such as appraisal; sessions that addressed a critical information need at the point in time when the need was most apparent.

Equally valuable is the wide array of documentation prepared by the GeoMAPP partners that share a variety of perspectives and options to explore on a number of different topics. MSL appreciated knowing that for many issues, different partner states chose different approaches to similar challenges. States new to these concepts should find it encouraging to know that they can follow a pre-defined formula or develop their own customized approach to developing systems to support a successful GIS data archives program. Two documents that MSL found particularly useful as staff prepared for the data transfer demonstration were the *Geospatial Metadata Elements for Preservation* and GeoMAPP *Data Transfer* whitepapers which were both in draft form when reviewed by Montana.

The only real drawback from MSL's participation in GeoMAPP was that the time for participation was condensed to less than a year. This timeline created pressure on staff to get up to speed on the research already conducted while at the same time applying draft practices to local situations in order to verify and/or draw new conclusions about the research at hand. MSL was fortunate that the staff working on the project already knew one another so that the timeline was not impacted by the need for extensive relationship building and cross-training. Other states should not underestimate the time it takes to build sustainable relationships between partner organizations within the state.

Positive changes from project

As a result of MSL's participation in GeoMAPP, MSL staff has a better understanding of archival concepts and how to apply them to GIS data management. MSL now has a sustainable plan in place to develop an inventory system and a workflow that can be used to manage both current and archived data. Decisions such as what metadata is required, how frequently should data be captured and how should data collections be managed no longer seem overwhelming. MSL has procured and configured storage and a file system that will serve as a dark archive. Beyond MSL, Montana officials including those who sit on MLIAC, have a raised awareness of the importance of GIS data archiving. Most importantly, MSL now belongs to a network of professionals who understand and value GIS data archiving and who can be relied on to continue this dialogue in the future.

North Carolina[12]

North Carolina was the principal investigator (PI) and lead state for the GeoMAPP effort. The North Carolina team paired staff from the North Carolina State Archives Electronic Records Branch and the North Carolina Center for Geographic Information and Analysis (CGIA). The North Carolina State Archives is part of the North Carolina Department of Cultural Resources (NC DCR) which has responsibility for archival records created by state and local government agencies in North Carolina. CGIA manages NC OneMap [13] North Carolina's geospatial data portal and program for data sharing, and is responsible for the project management, coordination and contracts administration for GeoMAPP. CGIA began the project organizationally aligned with the state Department of Environment and Natural

[12] For more background information on North Carolina see the GeoMAPP Interim Report:pp9-12

[13] NC OneMap http://www.nconemap.net/.

Resources, but in late 2009 was transitioned to the Office of the State Chief Information Officer. North Carolina State University (NCSU) Libraries assisted GeoMAPP and the North Carolina team in a technical advisory role, sharing lessons learned from their experiences with the North Carolina Geoaptial Data Archiving Project (NCGDAP)[14] and involvement with national geospatial organizations such as the Open Geospatial Consortium (OGC). [15]

Electronic Records Program Background

At the beginning of the GeoMAPP project, the archives had 2.5 staff dedicated to collecting and managing electronic records including:

- E-mail from the Superintendant of Public Schools and the Governor of North Carolina;
- Records from the Governor's office released at the end of each administration;
- State Agency website archives (since 2005);
- Audio files from the State Senate;
- Files from the State Office of Information Technology Services.

These data were typically stored on CDs, DVDs, or on agency servers. Despite losing a staff analyst in late 2008 and the Government Records section head in early 2009, the electronic records program has continued to grow. In January 2009, the archives received over 200,000 files (90 GB) from the outgoing Governor's administration and collected 50,000 e-mails, while also continuing to capture websites, accessioning senate audio files and actively participating in the exploration of ingesting and preserving geospatial content. Additionally, items such as archiving state government-wide e-mail and the capture of state government maintained Web 2.0 tools such as Twitter and Facebook have arisen as archival challenges for the state.

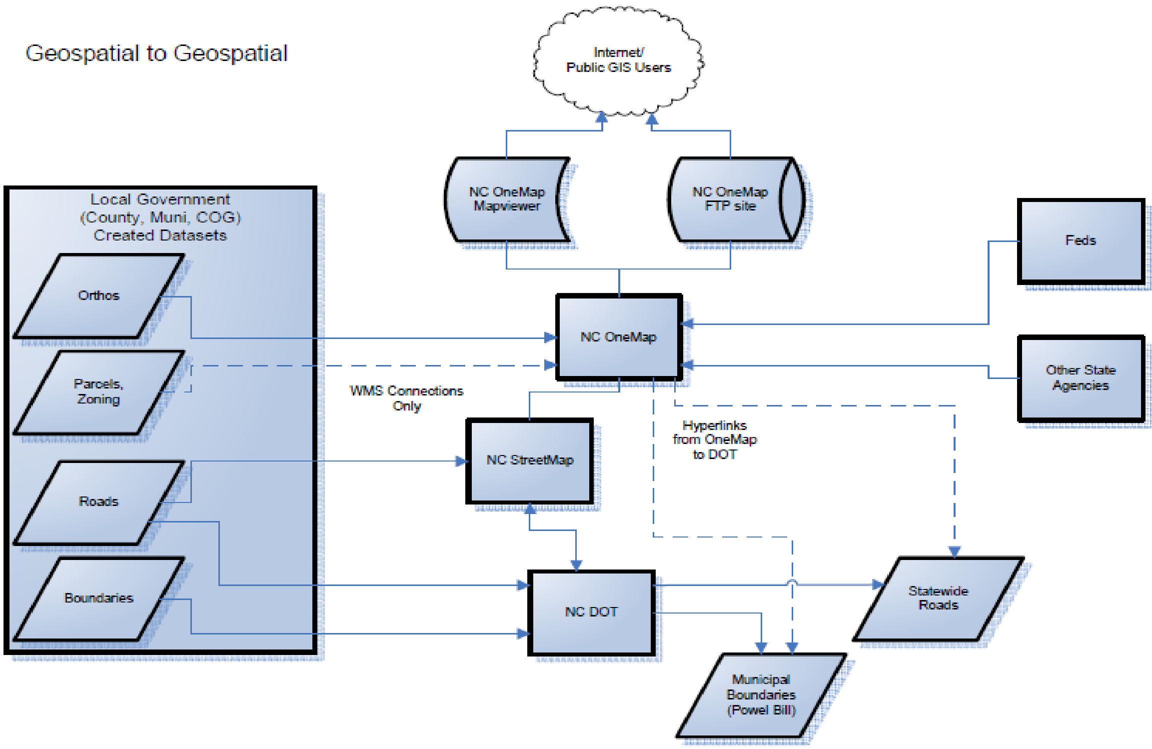
North Carolina's Geospatial Architecture

North Carolina's geospatial portal, known as NC OneMap, enables data discovery and access to web map services, image services, and downloadable vector datasets in shapefile format and raster datasets in TIFF with JPEG compression, MrSID, and IMG formats. NC OneMap is a clearinghouse, a repository for a portion of the data, and a portal for distributed map and image services. Datasets accessible through NC OneMap are developed and managed by state, local, and federal agencies, coordinated and/or hosted by the OneMap program, and made available online free of charge. In 2011, the site provided access to dozens of statewide map services, a statewide image service, and File Transfer Protocol (FTP) download access to over 110 vector and 125 raster geospatial datasets. NC OneMap's datasets feature FGDC CSDGM-compliant metadata records. If a metadata record is not included when data is submitted for posting, staff will create a new metadata record with input from the data creator. The OneMap team will also enhance or refine existing metadata records transferred with datasets when they are missing critical information, with input from the data creator. Before data is posted it is also opened and checked to assess file validity, dataset projection and geographic extent.

North Carolina has both a robust centralized repository and access to decentralized map services. This hybrid approach takes advantage of Web Map Services (WMS) to provide access to remotely created and managed datasets via the Internet. The NC OneMap program is in a planning process to determine the most effective ways to provide access to geospatial content from a diverse mix of federal, state and local

[14] NCGDAP: http://www.lib.ncsu.edu/ncgdap/
[15] For more info about the OGC, see: http://www.opengeospatial.org/.

government agencies and academic institutions that produce and manage data within the state. This hybrid approach to hosting services and data has made archiving geospatial content and determining a location of capture a significant challenge for the North Carolina team.



Mapping the movement of geospatial content in North Carolina

Project Accomplishments

GeoMAPP participants in North Carolina focused their efforts on a number of fronts. The NC team continued to focus on outreach within the state by attending regional conferences and legislative outreach efforts. Team members engaged with GIS data creators at the North Carolina Arc Users Conference and presented at the well-attended biannual NC GIS conference. In November 2011, the team also gave two presentations at a statewide electronic records conference and had in excess of 150 state and local government records creators attend both sessions. A session on business planning was also attended by the Archivist of the United States.

The team also continued to attend working group meetings of the NC Geographic Information Coordinating Council (GICC), Council meetings, and conducted a data appraisal and transfer demonstration with the City of Charlotte. The project team also engaged with the NC DCR Historic Preservation Office and its fairly mature GIS program. While a transfer of GIS records from this group may not result from this engagement due to confidentiality and business use cincerns, the hope is to influence their record keeping practices to include best practices for file format and file naming. The team also began discussions with the North Carolina Department of Transportation (NCDOT) regarding assessing their GIS data holdings. Due to workloads, however, the team continues to try to establish talks and meetings to discuss how to approach DOT's large collections of geospatial content

In 2011, the Statewide Mapping Advisory Committee (SMAC) of the GICC formed a Working Group for Standards. The chair of the working group invited staff from the State Archives to be a part of the working group and to contribute to the group's advisory role to the GICC regarding geospatial data content standards and related practices.

While both the Archives and CGIA were heavily involved in the administration of GeoMAPP due to their co-Principal Investigator designation and had a significant role to provide technical leadership for several workplan research tasks, both agencies were able to focus resources to enhance the geoarchiving workflow within the state. The team continued to transfer content from CGIA to the Archives and by the end of the grant, NC OneMap's entire collection of superseded vector files had been archived. These files include 75 unique datasets that had been saved by the NC OneMap database administrator and were added to the complete collection of over 500 vector datasets preserved in the NC geoarchive. The Archives also continued to add datasets to its CONTENTdm access solution and tweaked some of the data discovery functionality. The team also modified the collection to ensure that it took advantage of product enhancements made to CONTENTdm. The Archives installed and began testing the Audit Control Environment (ACE) toolkit on its geospatial holdings. ACE manages the bit level verification of files in a digital repository and periodically calculates and compares hashes for files in a collection to make sure that files have not been accidently altered or have suffered from degradation or 'bit rot'.

Responding to the Strategic Plan adopted by the NC Archives, the State Archivist created the Electronic Records Branch in June 2010. Employees of this branch work in collaboration with the other branches within the section to promulgate best practices guidelines[16], establish a digital repository, issue practice guidance and consult staff and other state employees.

CGIA was also witness to a number of organizational and staff changes as GeoMAPP transitioned into its final phase. In addition to facing the retirement of one of the project's founders and co-PIs in the fall of 2009, the agency was also organizationally and physically relocated from the state's Department of Environment and Natural Resources to the Office of the State Chief Information Officer (CIO) in September 2009. While the move aligned CGIA with the state's technology decision makers, it also introduced new administrative policies and procedures which impacted the administration of this complex multi-state, federally funded grant project.

Technologically, CGIA benefitted from a unique 2010 project that it led to capture detailed aerial imagery for the entire state. In addition to creating 17 terabytes of new imagery and immediately superseding the agency's vast existing imagery collection, the project also featured an implementation of Esri's Geoportal Server for data discovery and access to NC OneMap's data resources. The resulting NC Geospatial Portal[17] was a significant improvement over the existing NC OneMap data discovery infrastructure and could provide future beneficial linkages to superseded data being preserved at the Archives.

Sustainability

During the first quarter of the second phase of the project, the NC General Assembly passed legislation requiring a five dollar fee be collected on all deeds recorded with the exception of mortgages. This fee, the Archives and Records Management Fee (ARM), is to go to directly support the work of the Archives and Records organization. This funding helped the Archives make additional investments in its IT

[16] NCDCR records best practices: http://www.records.ncdcr.gov/erecords
[17] NC GeoPortal: http://data.nconemap.com/

infrastructure. The Archives purchased additional storage for the Raleigh site as well as storage located in the Western Office of Archives and History. The storage in the Western office serves as a disaster recovery copy of records and items in the digital repository. Additionally, the Archives began to build tools as well as incorporate existing tools to perform the digital preservation steps necessary for the preservation of and access to geospatial data files.

The NC team is also looking forward to testing the results of the business planning contractors' efforts to potentially generate future business justification documentation. As GeoMAPP winds down the agencies will continue to investigate other grant opportunities to potentially fund future enhancements to the NC geoarchiving process.

Utah[18]

The Utah GeoMAPP team was comprised of staff from the Division of Archives and Records Service and the Automated Geographic Reference Center (AGRC). AGRC manages the State Geographic Information Database (SGID)[19], Utah's geospatial data clearinghouse. The Archives is a division within the Department of Administrative Services, while AGRC is part of the Department of Technology Services.

Electronic Records Program Background

Prior to kicking off the GeoMAPP effort, Utah was in the early stages of building an electronic records program. Selected records were submitted to the archives from a variety of sources, usually on compact discs placed in boxes with paper records. Utah Archives also received governors' records in electronic form and stored them on a hard drive. The files were typically desktop files, such as Microsoft Word documents or spreadsheets. Additionally, the archives contracted with the Internet Archive to harvest state websites, but the archives have had only limited interactions with this data which is typically managed and harvested by the Utah State Library. Catalyzed by GeoMAPP project efforts, the Archives made a concerted effort to identify individual electronic datasets and record them in a catalog database.[20] The catalog functionality has expanded so it can be used for multiple formats including geospatial data. The archives staff has had ongoing discussions with its IT department with regard to preserving e-mail. The Archives have also begun a pilot project with the state's Purchasing Division to classify agency e-mail messages and export them out of the existing proprietary e-mail system.

Utah's Geospatial Architecture

Utah began the project with a fairly federated approach to managing their state's geospatial holdings. Relationships between AGRC and state agencies and local governments were traditionally formed on a project-by-project basis. AGRC has managed large road and parcel data collection efforts, which has allowed for unprecedented opportunities to interact and build relationships with county governments. Many of the state agency relationships are built between people in each office. Because of these outreach efforts, the reputation and purpose of AGRC as a data clearinghouse has encouraged participation without prompting.

---

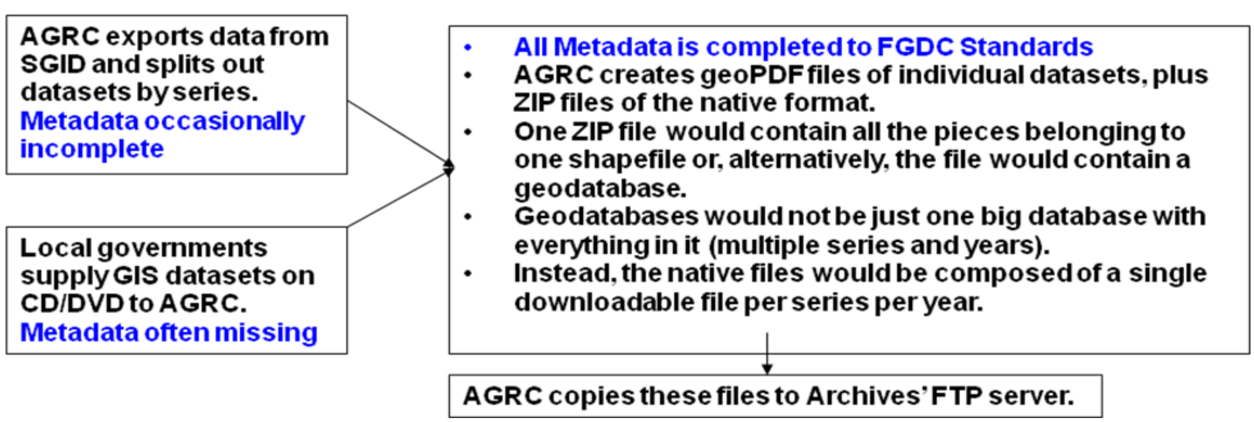
[18] For more background information on Utah, see the GeoMAPP Interim report: pp14-17.
[19] Utah SGID, http://agrc.its.state.ut.us/
[20] Utah Archives e-records catalog: http://images.archives.utah.gov/cdm4/search.php

AGRC hosts any public or private data that data producers are willing to share, whether this data is from the local, federal or state level. The data focus has also shifted[21] for the SGID from being project driven to being more varied in type and focus.

AGRC receives and ingests raster and vector datasets ensuring that metadata is both complete and FGDC CSDGM-compliant. AGRC staff will enhance or refine existing metadata records transferred with datasets when they are missing critical information with input from data creator. If metadata is absent, AGRC will contact the owner or steward of the data so that the metadata is completed to meet FGDC CSDGM standards. Additionally, the AGRC staff opens and checks the dataset to assess file validity, dataset projection and geographic extent. Once the dataset and metadata record have been validated, the data is made available for public access via FTP. The data listed can be downloaded for free and can be used by anyone without restriction.

The SGID is required to provide an accurate representation of all civil subdivision boundaries of the state. Each state agency that acquires, purchases, or produces digital geographic information data is required to inform AGRC about the existence of the data layers and their geographic extent and allow AGRC access to all data classified public. Additionally, the State Tax Commission annually delivers data relating to the creation or modification of the boundaries of political subdivisions. AGRC has also created a data sharing Memorandum of Understanding (MOU) with the federal government that was accepted by 13 federal agencies.

GeoMAPP has enabled the Archives and AGRC to extend their relationship with local data creators by supporting travel to localities and regional agencies statewide. During these visits, data were inventoried and added to the GIS Inventory, and targeted data were copied and transferred to the SGID and the Archives.



Utah's process for capturing data for the archives

Project Accomplishments

The Utah State Archives made significant progress in its ability to accession and preserve electronic records in general and geospatial records in particular. Prior to GeoMAPP, the procedures for ingesting

[21] The SGID Legislative mandate can be found here: http://www.le.utah.gov/UtahCode/getCodeSection?code=63F-1-507

electronic records into the Archives were to simply accept whatever media the record creators sent in, and then eventually write a finding aid describing those records. Now, the ingest process has become highly sophisticated and records are run through Utah's AXAEM application, which captures a checksum, extracts metadata, and associates these records with their retention schedules and finding aids automatically, all while providing an access interface and preservation features.

In the race to complete as much application programming as possible, the actual ingest of geospatial records had barely begun by the project's closure. In the project's first phase, geospatial records were inventoried and organized on an FTP server, so the Archives has access to them. Now the Archives needs to take all of those files and run them through the formal ingest process, largely following the processing procedures outlined by GeoMAPP's Preservation and Data Transfer working group.

More adjustments need to be made to the design of the finding aid when it comes to geospatial records. Previously, the finding aid was designed to display records item-by-item with one column in the container list dedicated to technical metadata. This design could be made much more readable. Once the design work is done, the finding aids for the ingested records need to be published.

In addition, much more development is required to develop a map-based interface to allow searching of archival records, including geospatial records. AGRC assisted the Archives to plan and develop one interface. While this end product looks promising, more work needs to be done before it is ready for public use. Another map-based interface uses location-based metadata found in the electronic records, in association with the search engine Solr, which has been integrated into AXAEM. That software appears to offer features that could be incorporated into a map-based search, although specifics have not yet been explored.

Storage space at the Archives is a large looming issue. The Archives has no independent storage capacity, which is why Archives and AGRC share the FTP server where the geospatial records currently reside. Most other data is stored on portable hard drives, which have a known high failure rate. A few digital collections are online. Storage services offered by Utah's Department of Technology Services (DTS), including SAN and NAS technology, have been deemed too expensive for the Archives' budget. This service is the only online storage offered that connects to the Archives' web server and ingest process.

The Archives is currently seeking extra funding to pay for online storage, as well as exploring other options for a dark archive of offline storage. One technology that seems promising for offline use is Millenniata disks. Although storage space on these DVDs are limited to the standard 4.7 GB, Millenniata disks appear to be less fragile than other media and less prone to bit-rot. This media stability and reliability could save the Archives money in the long-term due to the low cost to obtain the media and being able to avoid the monthly maintenance fees associated with leasing SAN from DTS. Additionally, media migration would likely need to happen less frequently. If used, the ingest process would be adapted to only store records on the server until the ingest was complete, then be moved to offline storage, with no direct download ability provided to the public.

In contrast, AGRC's approach to the storage problem is to participate in exploring cloud-based storage along with Montana, Oregon, and Colorado. If this proves to be a financially viable solution, the Archives may choose it, too. Other goals that AGRC completed during the GeoMAPP project include:

- Consistency in applying ISO-naming convention standard;
- Ease of export of individual datasets from the GIS Clearinghouse database, the SGID.

GeoMAPP Working Groups: The Engines for Project Research and Discovery

Working Group Background

In addition to the engagement, exploration and implementation efforts occurring within individual states, each of the GeoMAPP partners also participated in the project's technical working groups. These teams took the lead for conducting GeoMAPP's collaborative detailed research investigations and generated many of the project's deliverables and white papers during the project's final phase.

As tasks and project scope began to be finalized for 2010-11 efforts, the project team reassessed the exiting working group structure and membership established during the first phase of work. Based on these considerations, the project established new working groups that better aligned with the project's new focus areas of enhanced preservation techniques, providing diverse access options to superseded content, developing a suite of tools and templates for justifying the investment in preservation, and enhanced national and state outreach and new partner mentoring. The original six working groups (Business Case, Inventory and Metadata, Appraisal and Access, Content Lifecycle and Data Transfer, Communications, Industry Outreach)[22] were condensed and transitioned to form the new GeoMAPP 2010-11 Working Group slate including:

- Preservation and Data Transfer
- Storage and Access
- Business Planning
- Outreach and Mentoring

A number of early-stage GeoMAPP efforts, such as the GIS/archive knowledge exchange and developing demonstration geoarchiving workflows, were largely experiential and executed within the individual states. The 2010-11 project workplan included many more tasks that called for collaborative research and led to the creation of periodic white papers. This focus on providing an ongoing publication of tools and white papers allowed for GeoMAPP to share its findings incrementally. This incremental release benefitted the community of interest by making tools and documents available that could be reviewed and used immediately rather than providing an avalanche of information and documentation at the project's closure.

This section of the report details the background, focus areas, key findings, lessons learned and recommended practices from each of the GeoMAPP Working Groups and proposes items for a future research agenda:

Exploring the Mechanics of Geoarchiving- Preservation and Data Transfer

The long-term preservation of digital data, whether geospatial or not, is more than just copying files to some type of portable storage media and placing it on a shelf. It involves a set of processes to ensure that data is transferred in a manner that accurately encapsulates and moves all of the data's elements, and maintains the data in manner that allows it to be accessed and used over time. Preservation systems must

[22] For more information about the initial Working Group efforts see the GeoMAPP Interim Report: http://www.geomapp.net/docs/GeoMAPP_InterimReport_Final.pdf

be designed in ways to facilitate this capture by ensuring the following characteristics of the data are maintained:

- **Authenticity**: the data is what it purports to be, and may be verified by assessing the identity and the integrity of the record. "It must be possible to ascertain at all times what a record is, when it was created, by whom, what action or matter it participated in, and what its juridical/administrative, cultural, and documentary contexts were;"[23]
- **Reliability**: the data is trustworthy and represents the record as a statement of fact. "It exists when a record can stand for the fact it is about, and is established by examining the completeness of the record's form and the amount of control exercised on the process of its creation."[24] It will be difficult for archival organizations to ascertain the reliability of received datasets (i.e., the factual accuracy of the data/attributes stored in the datasets), and will largely be dependent upon the data contributors to deliver reliable datasets. However, once the datasets are received, the archives can ensure the ongoing reliability through maintaining the datasets' integrity;
- **Bit-level Integrity**: the data is complete and protected from data loss or damage. Preservation systems must have a way of monitoring the bit-level integrity of data when datasets are received, and over time as the data resides in the archival storage system (e.g., through a mechanism such as checksums), to ensure the datasets and their associated files remain digitally intact;
- **Security**: the data is protected from unauthorized access;
- **Usability:** the data is discoverable and accessible in a meaningful way now and into the future, such that it can be used to its fullest potential.[25] Migration, transformation, and/or emulation strategies and plans are in place to protect against obsolescence in data formats and/or software applications that would restrict or exclude use.

Preservation and Data Transfer Tools and Whitepapers

- Geospatial Data File Formats Reference Guide
- Utilizing Geospatial Metadata to Support Data Preservation Practices
- Best Practices for Geospatial Data Transfer for Digital Preservation
- Metadata Capture and Geospatial Records (IS&T Archiving, 2011)
- Montana Data Transfer Design
- BagIt User Guide, Quick Reference Summary
- Best Practices for Archival Processing for Geospatial Datasets
- Archival Metadata Elements for the Preservation of Geospatial Datasets
- Archival Challenges Associated with the Esri Personal Geodatabase and File Geodatabase Formats
- Emerging Trends in Content Packaging for Geospatial Data

"The first unofficial Content Lifecycle and Data Transfer working group was formed at the GeoMAPP kickoff meeting in March 2008."[26] Thus began the section of the *GeoMAPP Interim Report* covering the

[23] Luciana Duranti (2001), "International Research on Permanent Authentic Records in Electronic Systems (InterPARES): Experiential, Interactive and Dynamic Records," SSHRC MCRI InterPARES 2 Project Proposal, 412-2001, 1.1-11 (emphasis in original). Available at http://www.interpares.org/display_file.cfm?doc=ip2_detailed_proposal.pdf.
[24] Brent Lee (2005), Authenticity, Accuracy and Reliability: Reconciling Arts-related and Archival Literature (InterPARES 2 Project)." Available at: http://www.interpares.org/display_file.cfm?doc=ip2_aar_arts_lee.pdf
[25] ISO 15489-1: Information and documentation — Records management — Part 1: General
[26] GeoMAPP Interim Report: http://www.geomapp.net/docs/GeoMAPP_InterimReport_Final.pdf

actions of the Content Lifecycle and Data Transfer working group. The report goes on to say that the Content working group's "official" actions evolved to a "focus on investigating and documenting the lifecycle of geospatial content and data transfer methodologies."

With that focus in mind, the Preservation and Data Transfer working group became almost a direct continuation of the work of the earlier working group from the previous phase of GeoMAPP, with the final phase's efforts digging deeper into the technical factors influencing the transfer and long-term preservation of geospatial content. The working group was tasked with two broad objectives:

- To explore and document the processes and requirements for preserving geospatial data over time;
- To assist the new partner (ultimately the Montana State Library) in conducting a similar data transfer demonstration that the original partners accomplished in the first phase of GeoMAPP (2007-2009), armed with best practice guidance from the existing partners' previous experiences.

To meet these challenges the working group addressed several tasks detailed in the project workplan, and produced several technical white papers, including a suite of whitepapers that together document the end-to-end process of describing, transferring, and processing geospatial datasets for long-term preservation :

- *Utilizing Geospatial Metadata to Support Data Preservation Practices*: identifies important geospatial metadata fields that facilitate the long-term preservation of geospatial datasets.
- *Best Practices for Geospatial Data Transfer for Digital Preservation:* offers strategies, processes and approaches for successfully and reliably transferring geospatial datasets to the archival organization.
- *BagIt User Guide*: provides detailed instructions on how to use BagIt to package, transfer, and validate transferred files.
- *Best Practices for Archival Processing for Geospatial Datasets:* offers strategies for establishing archival metadata, and presents workflow processes and technologies to reliably move geospatial datasets into preservation and access repositories.
- *Archival Metadata Elements for the Preservation of Geospatial Datasets:* offers an extensible metadata model and dictionary to use as the basis for describing and managing archived geospatial datasets.

Preservation systems must deal with the ongoing issues of hardware and software obsolescence inherent in the constantly shifting state of technology. File formats come and go, features are added and removed, and conversion from one format to another is problematic as features of the old format may not be available in the new. The complex, often multi-file nature of geospatial formats and the unfamiliarity of archivists not versed with the myriad of geospatial data file formats in GIS systems potentially lead to further challenges in maintaining all of the characteristics and accessibility of preserved geospatial datasets. The working group also produced whitepapers that address these issues:

- *Archival Challenges Associated with the Esri Personal Geodatabase and File Geodatabase Formats:* describes specific challenges associated with these rich, complex data file formats;
- *Emerging Trends in Content Packaging for Geospatial Data:* explores the latest advancements and research areas for geospatial content packaging.

The following sections describe some of the specific research items the working group investigated during the project's final phase:

Moving Content and Documenting Findings

A significant portion of GeoMAPP's initial efforts during the first phase of work (2007-09) were dedicated to designing, developing and implementing geoarchiving workflows and systems to preserve each state's archival GIS datasets. The original Content Lifecycle and Data Transfer working group and the state partners allocated significant resources to:

- Discover and inventory superseded geospatial datasets;
- Appraise these datasets for their archival worthiness;
- Design and then implement workflows and systems to prepare geospatial datasets for preservation;
- Transfer this content to the archival organization;
- Review and ingest submitted geospatial content;
- Store and provide access to archived data.

Each state developed its own unique approach to the data transfer process, with a focus on leveraging existing workflows and technologies within the state.

After processes were established within each state and demonstration data transfers were validated, each of the states transferred select datasets among the other two state partners. This interstate transfer of content helped to validate each individual state approach to data packaging and file naming, and also helped to assess the possibilities for developing geospatial archive content exchanges across state borders. Detailed findings from these initial data transfer efforts can be found in the project's Interim Report and project website.[27]

The new phase of GeoMAPP work, beginning in January 2010, added two new focus areas to the project's data transfer research efforts: developing a concise and inclusive data transfer best practices document; and conducting a data transfer demonstration with new state partner Montana.

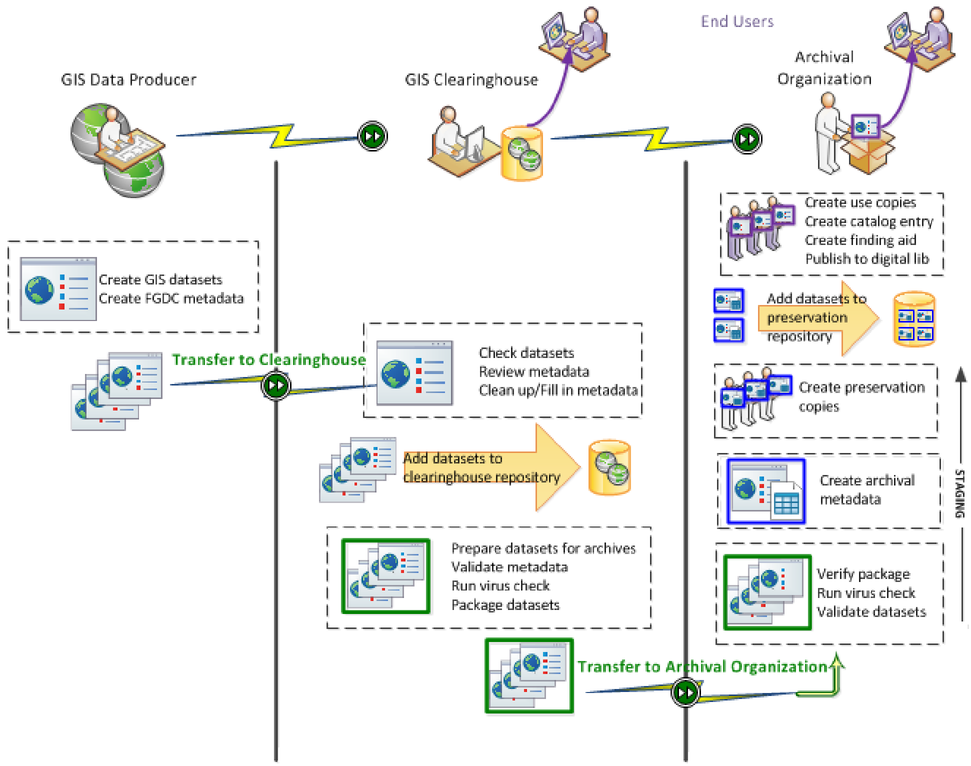
Documenting the Data Transfer Lifecycle and Recommending Best Practices

As archival organizations are dependent upon outside contributors to build their collections, a critical aspect of the acquisition of archival materials is the physical transfer or conveyance of those materials to the archives. Transferring files can be as simple as the GIS professional packaging them up in a zip archive and FTP-ing them across the network to the archival organization, or copying files to a portable disk for off-network transfers. Since geospatial datasets are typically complex, multi-file digital entities, successfully transferring "archives-ready" geospatial datasets may require a bit more planning and preparation.

The GeoMAPP team developed the *"Best Practices for Geospatial Data Transfer for Digital Preservation*"[28] to offer guidance and suggestions for GIS and archives professionals who will collaborate to transfer geospatial datasets to the archival organization for long-term preservation. The *Data Transfer*

[27] GeoMAPP Data Transfer Resources: http://www.geomapp.net/publications_categories.htm#xfr
[28] GeoMAPP Best Practices for Geospatial Data Transfer for Digital Preservation Whitepaper: http://www.geomapp.net/docs/Geo_Data_Transfer_BestPractices_v1.0_final_20111201.pdf

document offers a general model for geospatial data flow, running from the data creator through a GIS clearinghouse and ultimately to the long-term archival organization. Most states, including each of GeoMAPP's four partners, have established a GIS clearinghouse, which often serves as a central distribution point for current geospatial datasets. While the *Data Transfer* document focuses on a clearinghouse to archives workflow, this does not preclude states from conducting a transfer of geospatial assets to an archival organization directly from the originator.



General Model for Geospatial Data Flow

The *Data Transfer* document lists several planning activities that facilitate and systematize the data transfer process:

- Identify the technical data transfer method (e.g. over the network or with a removable storage device);
- Understand technical infrastructure constraints (e.g. time of day to transfer files, or file size limits for what can be transferred across the network);
- Define the naming conventions and file organization scheme that facilitate the arrangement of and, ultimately, access to the geospatial datasets;
- Define how files will be packaged for transfer, and define how those transferred files will be validated.

The *Data Transfer* document lays out a roadmap for GIS and archival organizations to follow when planning and executing the transfer of their geospatial assets. It offers:

- A series of planning activities, including anecdotal experiences from the GeoMAPP partners as they conducted their own planning and preparations for transferring geospatial datasets;
- Role-based task lists to facilitate and successfully perform the data transfer including geospatial data creators, geospatial data contributors (often in a GIS clearinghouse), and the archivists;
- Appendices that include checklists and resources to assist other to plan and execute their data transfer activities.

Applying GeoMAPP's Data Transfer Best Practices in the Real World: Data Transfer in Montana

In early 2011, after a formal selection process, the state of Montana was selected to join the GeoMAPP project as a new fully-funded partner to help test the project's initial findings and best practices, and to offer a new unique approach and outlook on project tasks. The Montana team had participated in the project as an Informational Partner and was aware of some of the project's initial findings and research. Full partnership allowed the state to take a more active role in the project and to become a guinea pig for the project to help GeoMAPP's data transfer recommendations, which had been created to help states like Montana get started with geoarchiving. As with the initial three states, Montana was encouraged to develop their own unique approach to geoarchiving that took advantage of existing workflows and processes. The Montana team relied on a draft of the GeoMAPP *Data Transfer* document as an initial guide to learn from the successes and shortfalls from initial GeoMAPP data transfer efforts to develop their own internal data transfer design. After assessing the document and completing their infrastructure design work, Montana provided feedback about the document and integrated unique findings from their data transfer demonstration.

Background:

Montana State Library (MSL) serves as both the state's GIS Clearinghouse and the state's GIS archive. MSL manages its own internal data center that serves these functions. Storage is comprised of an SQL database for managing active datasets and a file structure created on a Storage Area Network (SAN) for archives storage. Data discovery and access is made possible through an ArcGIS Server environment as well as web-based tools such as the Montana GIS Portal[29]. The Portal is based on the Esri GeoPortal Toolkit and enables discovery of GIS metadata (and data access instructions) describing the clearinghouse data as well as data made available by state agencies and local governments across the state. MSL also provides discovery and access to its GIS clearinghouse data via web mapping services and applications, and web pages offering data download options.

Data Transfer Demonstration:

MSL staff members spent a significant portion of 2011 reviewing GeoMAPP documentation and envisioning how to fit archiving into the state's geospatial Clearinghouse workflows. MSL made active use of live GeoMAPP discussions during bi-weekly conference calls, the June 2011 face-to-face meeting in Montana, and online mentoring/demonstration sessions related to archival tools and appraisal considerations to work through the details of the existing documentation. The value of those dynamic discussions should not be underestimated—the discussions set a context for the team's GeoMAPP work and highlighted the need for GIS archiving leadership. Specific to the task of Montana's data transfer, the

[29] Montana GIS Portal: http://gisportal.msl.mt.gov/GPT9/catalog/main/home.page

team found the GeoMAPP *Geoarchiving Self-assessment Tool*[30], the *Best Practice for Geospatial Data Transfer for Digital Preservation*, and the *North Carolina Intrastate Data Transfer Design*[31] documents to be very useful as they initiated their geoarchiving efforts.

The Montana data transfer demonstration (and ultimately accessioning and archiving processes) was likely more streamlined than the other states, which have separate archives and geospatial organizations, since both responsibilities are housed in the MSL. The team extracted needed steps from the aforementioned GeoMAPP documents and assembled and modified them to fit MSL's environment and workflows. The GeoMAPP documents proved to be very useful to MSL even though their organizational and technical environments differed from the other three states. The MSL team reviewed and provided feedback to the final draft of the *Best Practice for Geospatial Data Transfer for Digital Preservation*.

The best practice and design documentation review taught the Montana team about practices and challenges applicable to archiving spatial data, and also spawned an objective critique of their existing data storage, data access tools, and accessioning processes in light of what's needed to consistently and professionally archive spatial data. A significant portion of MSL's work during 2011 involved envisioning a new accessioning process as well as developing storage and access mechanisms needed to bring the process to reality. This new accessioning process guided the team as they assisted with the project's technical workplan tasks and will continue to guide their geoarchiving efforts after the grant's completion. MSL documented their new accessioning process graphically and in a written procedure as the *Montana State Library Spatial Data Accessioning Design*[32] document.

Given the abbreviated nature of Montana's engagement as a full GeoMAPP partner (11 months), the team was only able to scratch the surface in developing new accessioning tools and preservation architecture. To complete the data transfer task and leverage what could be learned from such a task, the team:

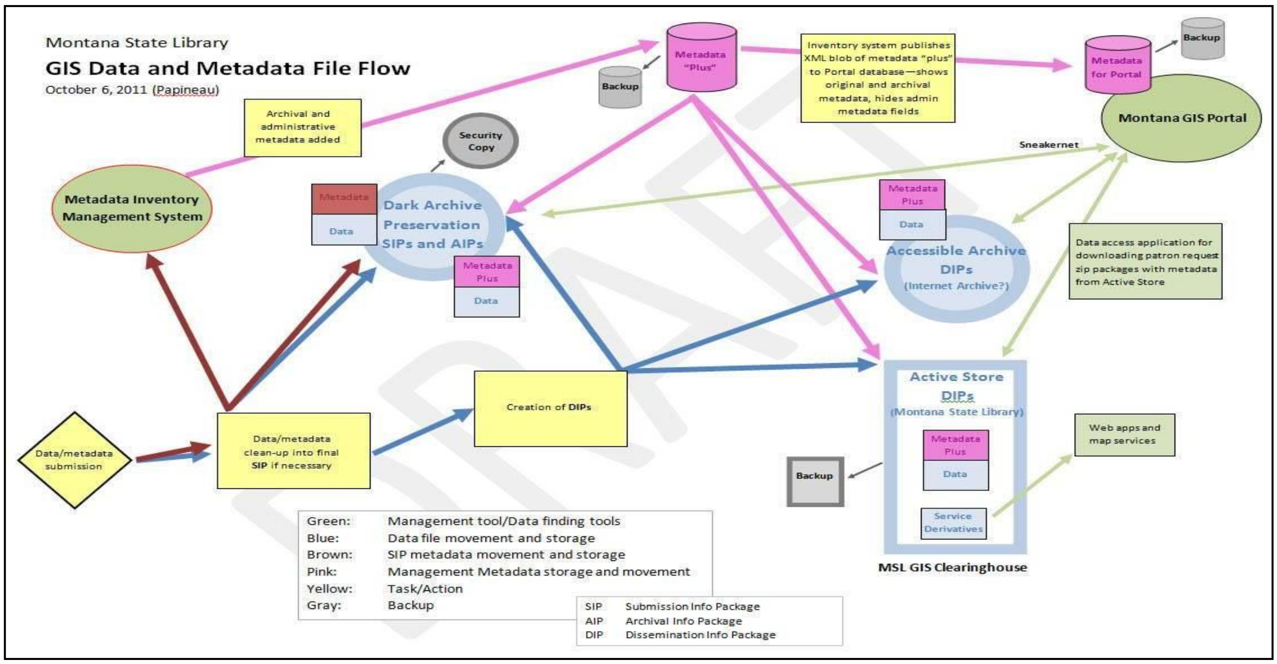
1. Extracted a stripped down set of steps that mimicked the proposed accessioning process to complete the demonstration.
2. Created a spreadsheet representing a new metadata management system (for original, archival, and administrative metadata).
3. Wrote specifications and requirements for the new data management system.
4. Selected three datasets to transfer: Montana Land Cover File Geodatabase (100MB), Montana Census Blocks shapefile (57MB), and Montana Towns shapefile (.08MB).
5. Installed BagIt and determined how the team would use it (packaging dataset bags, creating checksums, and validating).
6. Completed a Local Area Network (LAN) data transfer inside MSL for all three datasets.
7. Completed a Wide Area Network (WAN) data transfer to the Kentucky State Archives for all three datasets.

---

[30] GeoMAPP Self Assessment Tool: http://www.geomapp.net/docs/GeoMAPP_GeoArchiving_SelfAssessment_20100914.xls
[31] NC Intrastate Data Transfer Design: http://www.geomapp.net/docs/NC_Intrastate_Geoarchives_Final_20090914.pdf
[32] MSL Spatial Accessioning Design document: http://www.geomapp.net/docs/MSL_Data_Transfer_Design_final_20111231.pdf

8. Documented difficulties encountered and tied them to the related step in the procedure and documented any adjustments made during the process of completing the transfers.

The data transfer demonstration task served two important purposes within Montana's GeoMAPP effort. First, it inspired MSL to think through, as a team, how they would factor archiving into their existing geospatial workflows. Second, it provided the opportunity to test this envisioned approach to accessioning with archiving and record any difficulties and successes. This strengthened the proposed process; the lessons learned were incorporated into the *Montana State Library Spatial Data Accessioning Design* document. The most challenging part of this demonstration was planning how archiving could fit into, and even change, current workflows and infrastructure. The team mostly found the actual act of transferring the data to be fairly easy.



Proposed Montana Accessioning Process

Content Packaging for Geospatial Data

An individual geospatial data resource may be composed of a complex, inter-related set of data files as well as metadata and other supporting file objects, all of which need to be arranged in a certain fashion in order to be understood by the software and the humans that are involved in the exchange, management, and use of the data. In order to facilitate automated exchanges of complex data and avoid costly and error-prone human intervention, two organizing components are needed: a physical and/or logical package to encapsulate and structure data objects, and well-structured metadata or manifest information that is associated with that package. In the *"Emerging Trends in Content Packaging for Geospatial Data"* [33] the working group engaged in an investigation of emerging content packaging approaches in order to:

[33] Emerging Trends in Content Packaging for Geospatial Data – http://www.geomapp.net/docs/ContentPackaging_v1.0_final_20111202.pdf

1. Characterize the role that content packaging is coming to play with regard to geospatial data management and access;

2. Document emerging content package types that have appeared in the geospatial community;

3. Explore preservation challenges that may arise when these packages are expected to persist over time or when the packaging process itself results in changes to packaged data.

While complex XML wrapper formats have emerged to support content packaging in some other industry sectors, within the geospatial community archive formats such as ZIP or TAR commonly function as rudimentary content packages for multi-file datasets or groups of related datasets. Formalized approaches to the use of ZIP files with geospatial data have emerged in connection with specific types of content or software. The working group's investigation addressed four such formalizations of ZIP that have been developed to address specific needs:

- KMZ – for packaging a specific file format (Keyhole Markup Language- KML);
- Metadata Exchange Format (MEF)– for packaging metadata and data in connection with specific geoportal software (GeoNetwork);
- Layer Package (LPK) – for packaging data with display information within a suite of a specific vendor's software tools (various Esri software packages or online tools);
- Map Package (MPK) – for packaging data and finished maps within a specific desktop GIS tool (Esri's ArcGIS).

Although each of these formats addresses specific packaging problems within the geospatial domain, the examples provide some insight into preservation opportunities and challenges related to content packaging. The working group identified the following issues:

- Content packaging is increasingly being used to capture and make persistent cartographic representations and other data representations whether at the individual dataset level or the level of a data project encompassing many datasets.
- While the packages themselves may be transparent, long-term viability of these packages depends upon ongoing software support for individual components that might provide core functionality within the package.
- There may be dependencies on external resources that are ephemeral in nature. In cases where content packages are expected to be viable over a long period of time it will be necessary to make a concerted effort to limit the number of external dependencies.

GeoMAPP Geospatial Data File Formats

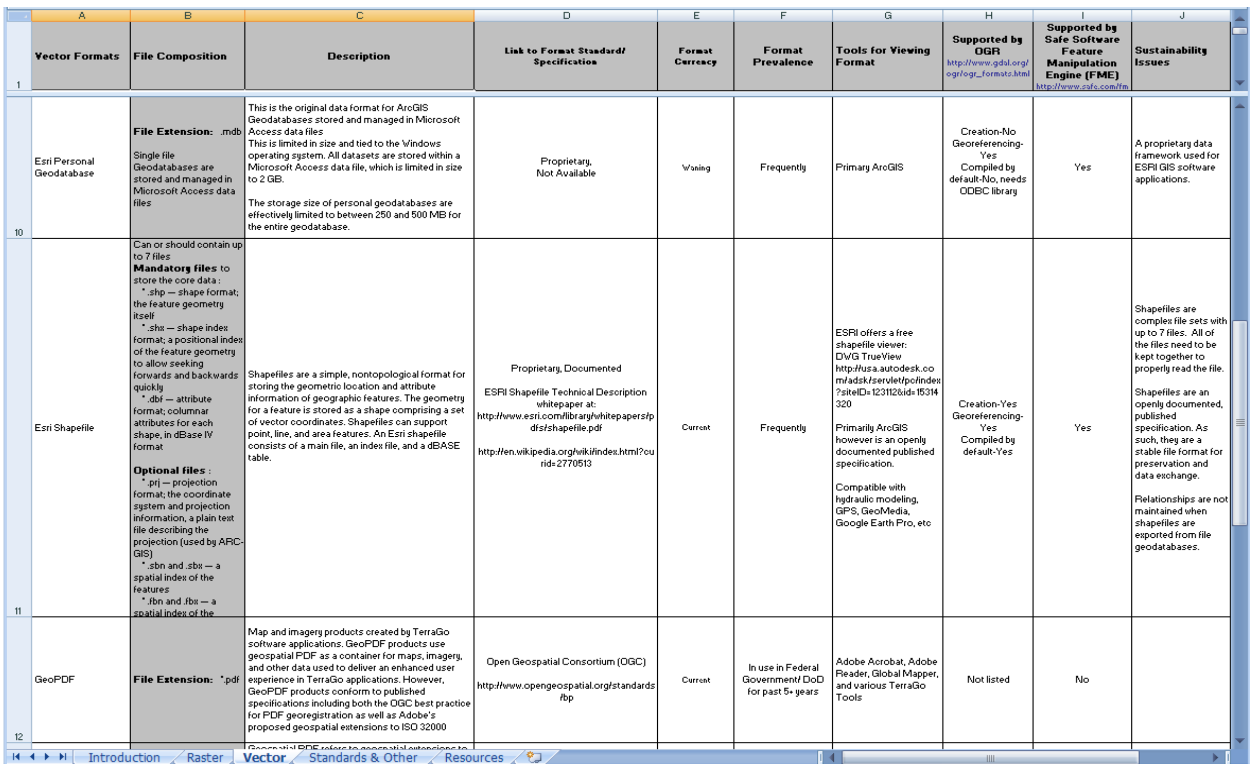
One of the major objectives of the GeoMAPP project was to address "at risk" digital spatial content, and electronic file format support is a fundamental challenge in the long-term preservation of digital materials. This issue is especially relevant for geospatial datasets as they are created, shared, and stored in many different file formats, many of which are proprietary to a specific vendor and/or software application.

[34] Geospatial Data File Formats Reference Guide- http://www.geomapp.net/docs/GeoMAPP_Geospatial_data_file_formats_FINAL_20110701.xls

In an attempt address some of these concerns the working group developed the *Geospatial Data File Formats Reference Guide*[34]*;* a detailed description of various legacy, current and emerging geospatial data formats that are encountered in state governments. This guide is intended to familiarize archivists with the wide variety of geospatial data formats that they might encounter, as well as to provide some assistance in identifying unknown geospatial assets that may be in their holdings. The *Reference Guide* provides:

- A listing of file extensions associated with each format;
- An assessment of the format's currency and adoption;
- Other descriptive information that are important considerations for preservation of the dataset.

The *Reference Guide* also provides information regarding the composition of geospatial data formats and offers suggestions on tools that may be used to view the geospatial dataset. Archivists may choose to transform or migrate a certain data file format to another format either to conform to an archive's collections requirements (e.g. using only shapefiles for vector archives) or to convert access copies of legacy data to a current usable format. Additionally, the *Reference Guide* provides a summary of the Geospatial Data Abstraction Library (GDAL)[35] support for the translation of any particular raster format; the OGR[36] Simple Features Library support for any particular vector format; and conversion support provided by Safe Software's Feature Manipulation Engine (FME)[37] for each of the data file formats.



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/fm	Sustainability Issues
Esri Personal Geodatabase	**File Extension:** .mdb Single file Geodatabases are stored and managed in Microsoft Access data files	This is the original data format for ArcGIS Geodatabases stored and managed in Microsoft Access data files This is limited in size and tied to the Windows operating system. All datasets are stored within a Microsoft Access data file, which is limited in size to 2 GB. The storage size of personal geodatabases are effectively limited to between 250 and 500 MB for the entire geodatabase.	Proprietary, Not Available	Waning	Frequently	Primary ArcGIS	Creation-No Georeferencing-Yes Compiled by default-No, needs ODBC library	Yes	A proprietary data framework used for ESRI GIS software applications.
Esri Shapefile	Can or should contain up to 7 files **Mandatory files** to store the core data : • .shp — shape format; the feature geometry itself • .shx — shape index format; a positional index of the feature geometry to allow seeking forwards and backwards quickly • .dbf — attribute format; columnar attributes for each shape, in dBase IV format **Optional files** : • .prj — projection format; the coordinate system and projection information, a plain text file describing the projection (used by ARC-GIS) • .sbn and .sbx — a spatial index of the features • .fbn and .fbx — a spatial index of the	Shapefiles are a simple, nontopological format for storing the geometric location and attribute information of geographic features. The geometry for a feature is stored as a shape comprising a set of vector coordinates. Shapefiles can support point, line, and area features. An Esri shapefile consists of a main file, an index file, and a dBASE table.	Proprietary, Documented ESRI Shapefile Technical Description whitepaper at: http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf http://en.wikipedia.org/wiki/index.html?curid=2770513	Current	Frequently	ESRI offers a free shapefile viewer: DWG TrueView http://usa.autodesk.com/adsk/servlet/pc/index?siteID=123112&id=15314320 Primarily ArcGIS however is an openly documented published specification. Compatible with hydraulic modeling, GPS, GeoMedia, Google Earth Pro, etc	Creation-Yes Georeferencing-Yes Compiled by default-Yes	Yes	Shapefiles are complex file sets with up to 7 files. All of the files need to be kept together to properly read the file. Shapefiles are an openly documented, published specification. As such, they are a stable file format for preservation and data exchange. Relationships are not maintained when shapefiles are exported from file geodatabases.
GeoPDF	**File Extension:** *.pdf	Map and imagery products created by TerraGo software applications. GeoPDF products use geospatial PDF as a container for maps, imagery, and other data used to deliver an enhanced user experience in TerraGo applications. However, GeoPDF products conform to published specifications including both the OGC best practice for PDF georegistration as well as Adobe's proposed geospatial extensions to ISO 32000	Open Geospatial Consortium (OGC) http://www.opengeospatial.org/standards/bp	Current	In use in Federal Government/ DoD for past 5+ years	Adobe Acrobat, Adobe Reader, Global Mapper, and various TerraGo Tools	Not listed	No	

Geospatial Data File Formats Reference Guide

[34] Geospatial Data File Formats Reference Guide-
http://www.geomapp.net/docs/GeoMAPP_Geospatial_data_file_formats_FINAL_20110701.xls
[35] GDAL: www.**gdal**.org/
[36] OGR: http://www.gdal.org/ogr/
[37] Safe Software FME: http://www.safe.com/fme/fme-technology/

In developing the *Reference Guide,* the working group engaged a large community of reviewers, including the GeoMAPP Informational Partners, U.S. National Archives and Records Administration (NARA), and the FGDC Users/Historical Data Working Group (U/HDWG).

The GeoMAPP team also contacted the National Archives of the United Kingdom, which manages PRONOM[38], a technical registry of file formats and their supporting software products, to see if they would be interested in using the *Reference Guide* to expand their geospatial coverage, and the PRONOM organization responded positively. Prior to GeoMAPP's submission, PRONOM had limited coverage of geospatial data file formats. A potential geo-centric expansion of the PRONOM registry could be useful to extend the capabilities of file identification tools such as JHOVE[39] which uses PRONOM as the basis for reporting determined file type identities.

Due to their wide adoption within each of the partner states, the working group dedicated extra attention to Esri-originated data file formats (e.g. shapefiles, Geodatabases), while attempting to document all current and historic geospatial formats widely encountered in state government.

This effort just scratches the surface of the documentation possibilities for geospatial file formats. Future geospatial preservation projects might go into greater detail in exploring data file formats from other vendors, or still emergent open standards-based geospatial data file formats. There are also many opportunities to explore the capabilities of transformation tools, and to assess their potential utility for archival organizations in transforming submitted dataset formats into an alternate format the archival organization has identified as the basis for is geospatial preservation.

Archival Challenges Associated with the Esri Personal Geodatabase and File Geodatabase Formats

Spatial databases play a prominent role in geospatial data production and management, storing a range of data types including geographic features, attribute information, satellite and aerial imagery, surface modeling data and survey measurements. In addition to storing data, some types of spatial databases can model the relationships between data, handle data validation, and support complex data models, versioning, and multi-user editing, all of which greatly improve data integrity and analysis capabilities. Spatial database formats are playing an increasingly prominent role in the distribution of data to end users. These formats also provide an option for transferring geospatial content to archives.

Due to the pervasiveness of Esri Geodatabases within the geospatial community and within each of the partner states, GeoMAPP focused its research efforts on Esri's Geodatabase formats. Within the Esri community, two proprietary geodatabase file formats have emerged: the Personal Geodatabase and the File Geodatabase. Since the nature of long-term software support for any particular database format will always be an unknown, archivists will need to plan to make format conversions in the future to whatever new data or database formats and architectures arise, so that the content produced today will not be inaccessible and lost. One area of investigation for the GeoMAPP project, and the working group specifically, concerned the long-term sustainability of Esri Geodatabase formats in an archival context, and whether it would be better to convert these databases to open formats or retain them in a particular

[38] PRONOM website: http://www.nationalarchives.gov.uk/PRONOM/Default.aspx
[39] JHOVE - JSTORE/Harvard Object Validation Environment. http://hul.harvard.edu/jhove/

Esri Geodatabase format. More detailed analysis is found in the *Archival Challenges Associated with the Esri Personal Geodatabase and File Geodatabase Formats.*[40]

Near-term software support from Esri and other data translation software vendors such as Safe Software appears strong for all of the Esri Geodatabase formats, including support for managing File Geodatabases through Esri software version upgrades as a hedge against potential loss of support for older versions. Testing done as part of earlier GeoMAPP efforts indicated that dataset conversions through incremental software upgrades of ArcGIS may prove to be less subject to data loss and errors than conversions that involve skipping versions.

One alternative approach is to convert selected datasets or feature classes stored within the geodatabase into to the shapefile format, which is openly specified and much more widely supported than the File Geodatabase. However, the inability of the shapefile to support the advanced features found in geodatabases poses limitations as an archival format for more complex content. Emerging spatial database formats such SpatiaLite warrant following for potential future value in an archival context, but these open formats cannot currently support the more complex aspects of Esri Geodatabases and may not be directly supported by Esri software.

In order to make File Geodatabase content more accessible outside of Esri software, Esri released the File Geodatabase API in June 2011[41]. The API arrives with a number of well-documented limitations, including lack of support for various dataset types as well as most raster-related database components, yet the API may provide some utility in an archival capacity, especially with regard to providing access to metadata. As part of the OGC Web Services, Phase 8 (OWS-8) initiative[42], the File Geodatabase API was used to facilitate bulk transfer of data to and from an open source PostGIS database. A resulting engineering report highlighted some of the possibilities and limitations with regard to conversion of File Geodatabases into open source databases using the API.

Metadata's Importance in Geoarchiving

Creating and managing rich metadata records that document important details about datasets is a key component of both geospatial and archival preservation workflows. Researching these workflows and the underlying structures of geospatial and archival metadata, and investigating metadata content has been a large area of focus for the working group. The second phase project metadata research follows previous GeoMAPP Metadata efforts from the project's first phase including the development of a comparison and crosswalk between the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSGDM)[43] standard and Dublin Core.

The following sections provide details about geospatial metadata elements that are critical for the preservation process and describe a framework for integrating certain GIS metadata elements with key preservation and administrative items to form a geo-centric archival metadata record.

[40] GeoMAPP Geodatabase Whitepaper: http://www.geomapp.net/docs/Geodatabase_Report_v1.0_final_20111206.pdf
[41] ESRI API: http://resources.arcgis.com/content/geodatabases/10.0/file-gdb-api
[42] OGC Web services:http://www.opengeospatial.org/projects/initiatives/ows-8
[43] FGDC CSDGM: http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/index_html

Utilizing Geospatial Metadata to Support Data Preservation Practices

The geospatial community has long embraced and encouraged the creation of rich, descriptive metadata to document the background information about how, when and why a dataset was created, technical details about how the dataset was assembled, its projection and coordinate system information and database attributes, as well as information about provenance and ownership. The four GeoMAPP partners require that each dataset transferred to their state clearinghouse include a fully compliant FGDC CSDGM record.

The FGDC CSDGM is a rich metadata standard made up of around 300 possible data fields and several dozen required elements. Due to the expansiveness of the metadata standard, the working group investigated the each element of the standard in detail and highlighted fields that would be of greatest importance for the long-term preservation of a dataset to provide focus areas for the metadata creation and review process. Though not widely used as a metadata standard in the state government context, the team also investigated the International Organization for Standardization (ISO) - 19115:2003 Standard for Geographic Information Metadata.[44] While the team did not implement ISO as a metadata standard, the team utilized the ISO standard's 19 Topical Categories to help categorize their datasets. Topic categories range from Biota and Boundaries to Structures and Utilities and each state has their geoarchival holdings organized within ISO Topic Categories to help group, organize and manage their data.

After conferring among the project partners and seeking outside guidance from the Informational Partners, the project published a white paper *Utilizing Geospatial Metadata to Support Data Preservation Practices*[45] that:

- Provides some background on the team's metadata research;
- Describes the seven sections of an FGDC CSDGM metadata record (Identification, Data Quality, Spatial Data Organization, Spatial Reference, Entity and Attribute, Distribution and Metadata Reference Information sections);
- Includes a table of the GeoMAPP recommended FGDC CSDGM fields that should be richly populated to benefit preservation and data utility.

	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*

Snapshot from Table of Recommended CSDGM fields

[44] ISO 19115:2003 http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020
[45] CSDGM Items for Preservation http://www.geomapp.net/docs/GeoMetadata_Items_for_Preservation_2011_0110.pdf

While a thoughtfully completed, fully compliant FGDC CSDGM record should always be the goal of geospatial data creators, the white paper identifies and describes around 50 fields that should be an area of focus for geospatial data creators. While the primary purpose of the document was to identify fields that are important for preservation and for possible inclusion into descriptive archival metadata, these fields are also extremely useful for everyday production data use and sharing. Given that it can often be a challenge to get data creators to create rich and detailed metadata records, this list can serve as a guide for key elements that they should focus on during metadata creation, hopefully making the process less intimidating.

Administrative archival metadata elements for long-term preservation, management and access

A critical enabler for the successful archiving of any digital materials is defining a metadata element set that supports the long-term management and access of the materials. Dublin Core provides a core set of primarily descriptive metadata. Frameworks such as the Reference Model for an Open Archival Information System (OAIS)[46] offer a conceptual model for the universe of metadata needed to ensure long-term preservation, management and access to digital materials. However, archivists are generally on their own to define the particular metadata elements that will comprise their archival data models. In addition, archivists of digital materials are likely preserving materials in a variety of digital formats. As such the archival metadata model should be extensible enough to:

- Accommodate different formats archivists are managing today;
- Accommodate different formats that archivists will be managing in the future;
- Be flexible enough to meet the special needs of each archival organization.

The working group adopted a two-tier strategy to identify archival metadata elements:

1. Identify archival metadata elements applicable to all materials, regardless of digital format.
2. Identify digital format-specific metadata elements, including the particular archival metadata needs for managing archived geospatial datasets based on the FGDC CSDGM metadata elements that facilitate the preservation objective.

The working group published a paper, "*Archival Metadata Elements for the Preservation of Geospatial Datasets*,"[47] that provides digital preservationists a preliminary archival metadata element set for a multi-format digital archival repository, as well as for one specific to digital geospatial data. Based on the OAIS model, the metadata organization scheme is organized around the following categories:

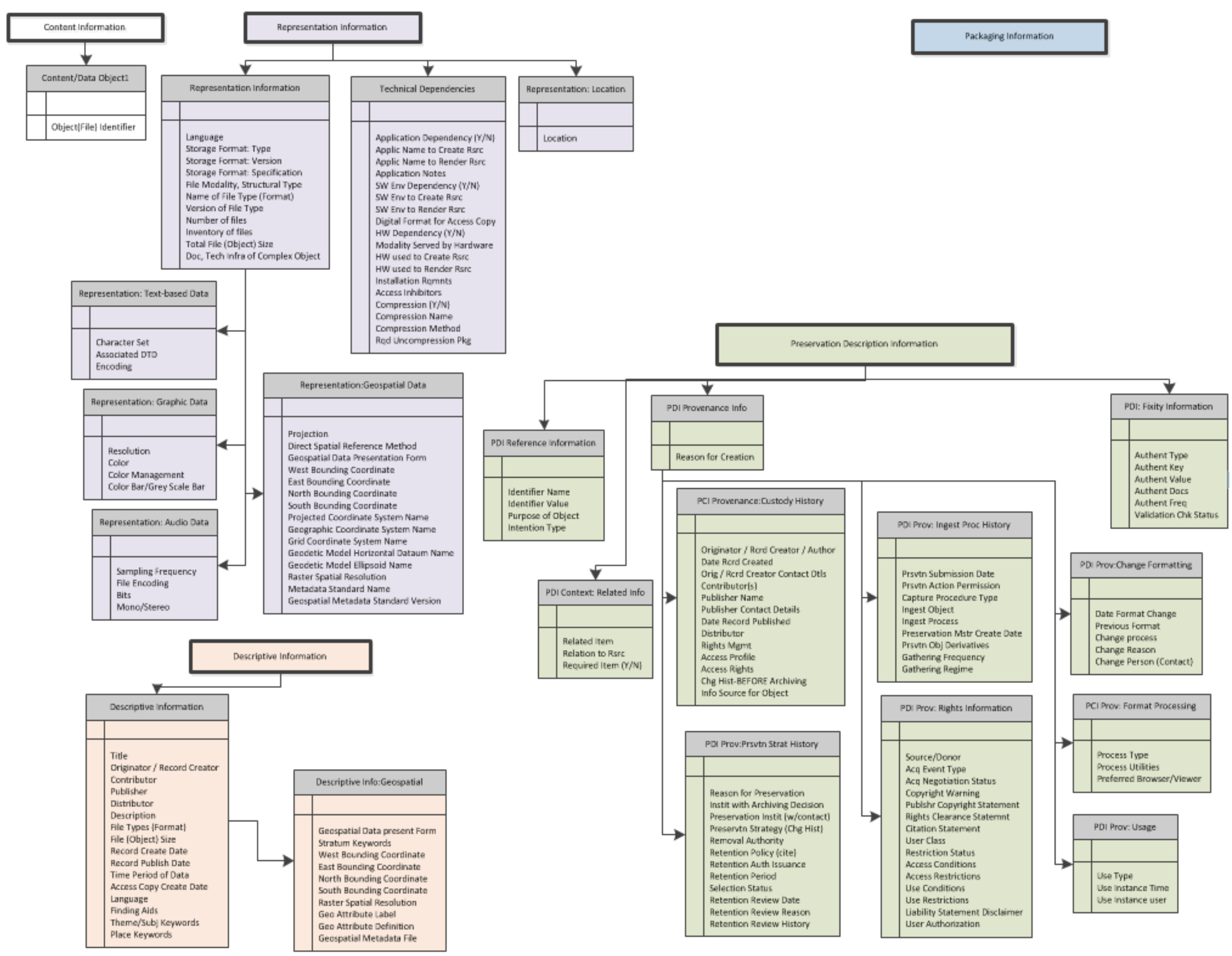
- **Content Information -** identifies the object being preserved;
- **Representation Information -** describes the structural and semantic information of the object;
- **Preservation Description Information -** facilitates the ongoing management of the digital object, such as provenance and fixity information;
- **Packaging Information -** describes how the Content Information is packaged in a particular environment;

[46] OAIS: http://public.ccsds.org/publications/archive/650x0b1.PDF
[47] GeoMAPP Archival Metadata Elements for the Preservation of Geospatial Datasets: http://www.geomapp.net/docs/GIS_OAIS_Archival_Metadata_v1.0_FINAL_20110921.pdf

- **Descriptive Information -** describes the digital object, and is often used as the basis for access points.

The team consulted several sources in preparing the metadata element set, including general purpose digital archives as well as other geoarchiving projects. These references are available in an appendix of the *Archival Metadata Elements* paper for readers whom may want to explore a wider set of metadata elements.

The working group constructed the *Archival Metadata Elements* paper to assist GIS practitioners and archival staff with metadata planning and design. It can serve as an archival metadata documentation tool to construct the data models and metadata dictionaries for digital repositories in general and geospatial archives in particular. It includes an archival metadata table that has been designed so that archivists can augment it with their own metadata as well. The paper also identifies potential issues and questions archivists may encounter as they ingest and process geospatial datasets and their metadata. To assist archivists in identifying the geospatial metadata applicable to include in the archival metadata record, the archival metadata table also includes a column that notes cross references to the FGDC CSDGM metadata element, where applicable. A column providing a mapping to Dublin Core is also included.



Archival Metadata Model for Preserved Digital Assets and Geospatial Datasets

Many FGDC CSDGM metadata elements are vital to the long term preservation of a geospatial dataset, including traditional descriptive metadata elements such as title, creator, date published, date(s) of the data, abstract, and purpose. The FGDC CSDGM metadata model also includes:

- Valuable technical metadata characterizing technical properties of the geospatial dataset;
- Descriptions of the data elements in the dataset;
- Data sources for the data (important for archival provenance);
- Processing steps describing the development of the dataset.

The GIS community strongly advocates for, and in some cases requires, GIS professionals to enter the companion geospatial metadata for their datasets, and is often distributed with the geospatial dataset as a supplemental .xml file. Some GIS software packages, such as Esri's ArcCatalog, include options to automate the population of some of the metadata elements such as the geospatial technical properties of the dataset further assisting GIS professionals in this task. Many of these metadata elements are valuable for the long term preservation of the dataset and should be harvested, either manually or through automated programs, and assigned to the archival metadata record.

The geospatial metadata also includes fields to record the geospatial dataset distributor, as well as access policies and distribution methods associated with that distributor. When an archival organization accessions geospatial datasets, it effectively becomes the dataset distributor. The *Archival Metadata Elements* paper also provides some thought provoking questions for archivists as they take ownership of these geospatial datasets, and how they might consider creating an archives-specific version of the geospatial metadata file, especially if the archival organization will publicize the availability of their archived geospatial collection through a national geospatial inventory. However, as an archival organization is charged with preserving the original records it receives, it is also imperative that the archival organization retain and preserve the originally received geospatial metadata as an historical record related to the dataset.

Originally, the working group wanted to identify metadata fields they felt were required for an archive. However, upon review, the group decided to simply leave the elements in the table format and defer the identification of required metadata fields to the archival organizations, as they are best informed regarding their own local practices and policies. However, the metadata table in the *Archival Metadata Elements* paper includes a column where archival organizations can record their own required fields.

Also, through the metadata review process, the team noted that this collection of metadata elements may be managed or collected by one or more computing systems in an archival organization. While there are benefits to a centralized data store, the metadata element set does not necessitate it. Instead, the set offers a collection of archival metadata elements. Each archival organization can determine, based on its own technical infrastructure and technical and archival processing resources, what system(s) will store and manage the metadata. For example, some archives have applications to manage their accessioning activities. Therefore, some of the acquisition and provenance-related metadata may be stored in the accessioning system rather than the preservation system. Likewise, the descriptive metadata may be largely managed in the system that provides access to the archived datasets.

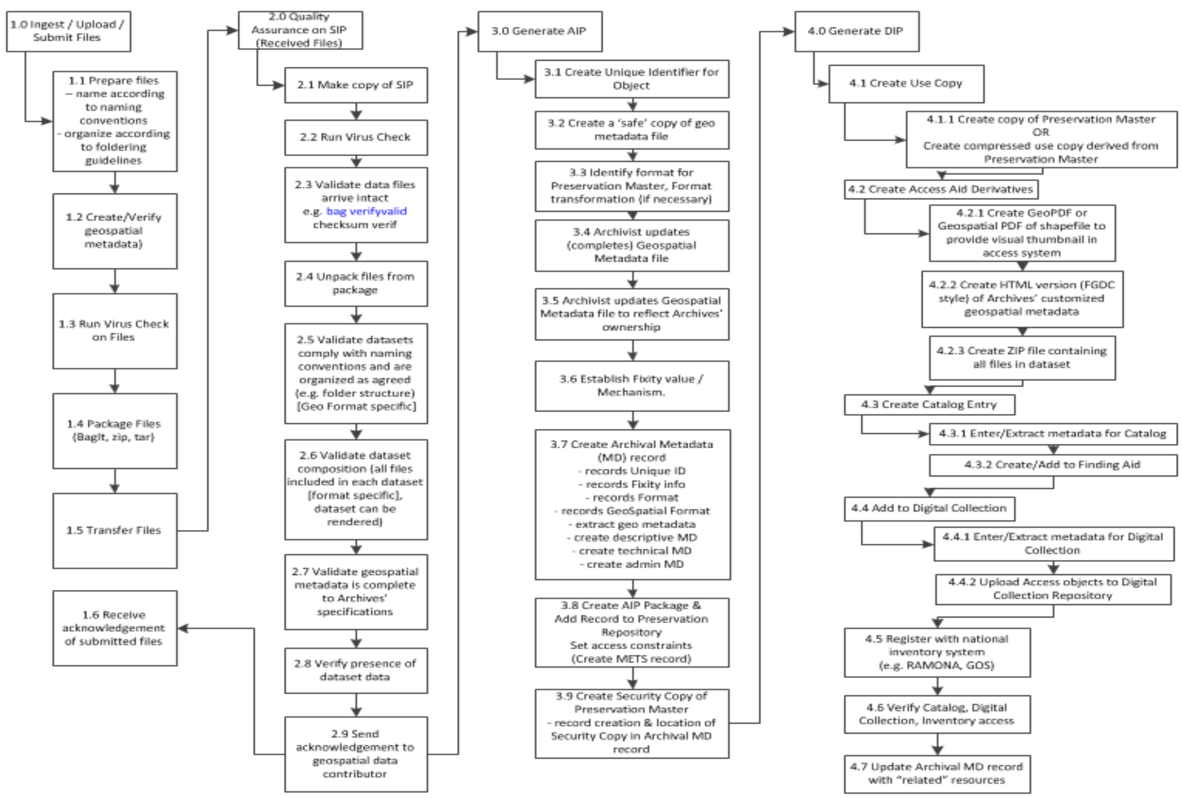
The working group intends for this paper to be an actively used tool by the geospatial and archival communities as they proceed with their geospatial archiving initiatives. This paper also serves as a companion document to the *GeoMAPP "Best Practices for Archival Processing for Geospatial Datasets"*[48] white paper.

Archival Processing of Geospatial Datasets

The storage and management of preservation and/or administrative metadata is conducted as part of an overarching archival processing process, and the complexities of geospatial datasets introduce special archival processing considerations.

The *Best Practices for Archival Processing for Geospatial Datasets* provides digital preservationists with a suggested process flow, based on the OAIS model to process archival geospatial datasets. It illustrates the phases of processing dataset files from deposit to ingest into preservation and access storage repositories, as illustrated in the figure below. The document is organized into the general processing phases, suggested by OAIS[49], of:

1. Ingesting the Submission Information Package (SIP)
2. Performing quality assurance of the SIP
3. Generating the Archival Information Package (AIP)
4. Generating the Dissemination Information Package (DIP)



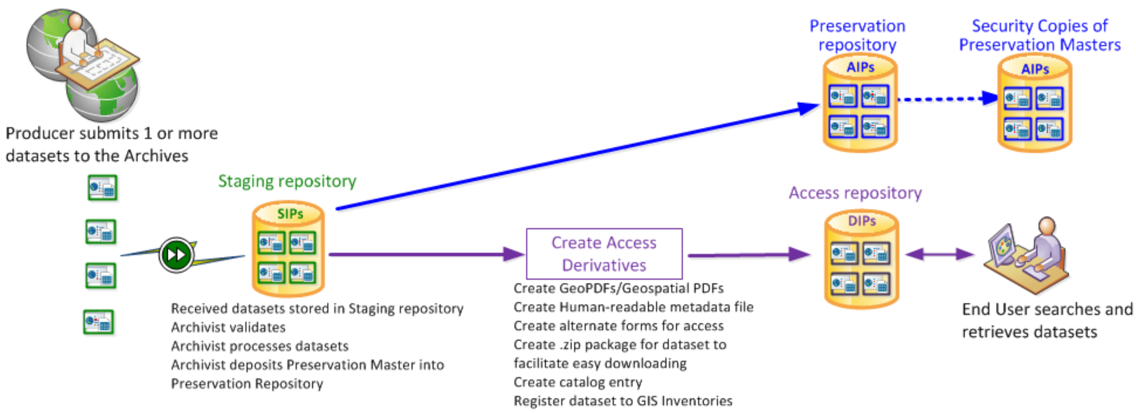
Geospatial Archival Processing Workflow

---

[48] GeoMAPP. Best Practices for Archival Processing for Geospatial Datasets: http://www.geomapp.net/docs/GIS_Archival_Processing_Process_v1.0_final_20111102.pdf

[49] Consultative Committee for Space Data Systems. "Reference Model for an Open Archival Information System (OAIS). Blue Book. CCSDS 650.0-B-1." 2002. http://public.ccsds.org/publications/archive/650x0b1.PDF

The document also recommends a generalized storage architecture for managing the processing, storage and access of archived geospatial datasets as seen in the figure below.



Generalized Storage Architecture for Preservation Environment

The *Archival Processing* document is intended to be a tool GIS and archives practitioners can use to facilitate their geospatial archival processing, planning and execution and includes processing task checklists and key planning questions for GIS staff and digital archivists to consider. The paper also offers a suggested object-oriented approach to managing the collection of assets and metadata that will eventually comprise the archived object, which frequently consists of multiple files. Each archived asset will also have associated preservation and access copies, and may also have access derivatives that all should be managed as a logical unit. The paper offers mappings of the metadata from the ingested package to the archival metadata needed to manage the preservation package (preservation master, security copies), and the access package (access copy, access derivatives, etc.)

The working group used varying ranges of manual and automated workflows to process geospatial datasets. The group found that existing digital repositories, both in the digital library and geospatial markets provided reasonable functionality and utility to adequately support access (e.g. the GeoMAPP partners explored CONTENTdm, DSpace, and Esri Geoportal). However, the out-of-the-box configuration of these tools often lacked key preservation and archival functionalities such as:

- Support a full range of archival metadata;
- Administer multi-file digital objects and their derivatives;
- Identify and validate digital file types;
- Automatically populate the archival metadata record from the geospatial metadata;
- Manage fixity / integrity checks on the collection;
- Manage the security replica(s) of the preservation master;
- Repair damaged files;
- Migrate files to newer formats;
- Implement policies related to security;

- Facilitate access to and retention of geospatial datasets;
- Auto-generation of finding aids.

Over the course of the project, the Utah State Archives made significant progress with the development and implementation of the APPX-based, Archives Enterprise Manager (AXAEM) archives management software platform, implementing many of these archival functions. Storage solutions functionality continues to develop to address several of the core data integrity needs, so that eventually archivists will not necessarily have to explicitly manage them within their archival solutions (e.g. manually storing checksums in the archival database and periodically checking). Also, more sophisticated archival management software solutions are starting to become available, and while the working group did not get a chance to evaluate them, they might be of interest to future geospatial archivists to explore.

Exploring Tools that Advance Digital Archives from Simple Storage Holdings to Digital Repositories

As mentioned in the introduction, digital preservation systems must have processes in place to validate the authenticity and validity of data transferred to the archives to ensure that a complete data set was transferred and that the data format is what it is claimed to be (the PDF file is indeed a PDF file); and processes to monitor the integrity of the data stored in the repository to protect from damage or loss (by the applications of checksums or hashes). There are several tools to monitor and facilitate the preservation of digital data, many used by archivists to manage other types of digital records. Individual members of the working group explored the use of some of these tools to manage geospatial data sets. These tests were conducted in the individual partner states as part of their overall digital archiving projects. This is not intended to be an exhaustive list, and some of the tools explored are emerging or still in development. Clearly, there is much room for further exploration of these tools.

Format validation-JHOVE2, DROID

The Utah State Archives investigated the use of JHOVE2[50] as a metadata extractor tool, along with other metadata extractors (FITS, New Zealand metadata extractor, DROID, etc.) One of the appealing elements of JHOVE2 is its ability to identify shapefiles. Utah ran the tool against a folder that contained a shapefile, and the software confirmed that it was a shapefile consisting of seven component sources (or files). It had a fair bit of difficulty determining formats for the individual shapefile components, and often responded with errors about unmatched PUIDs (PRONOM IDs). The version of DROID (Digital Record Object Identification)[51] that the tool was using was not up-to-date, so perhaps some signature files were missed, or were not registered in the PRONOM database. The data that JHOVE2 returned that seemed most useful were: the overall determination that the group of records consisted of a shapefile; the number of files in the group; file names, the format(s) of the files; and file sizes. Other metadata specific to shapefiles included attribute information from the dBASE file, the number of records, as well as other information from the shapefile header. Much of that was duplicative from what can be obtained from the FGDC CSDGM metadata that is part of the shapefile dataset; however the record count from the .dbf was captured by JHOVE2 but was not in the geospatial metadata record.

[50] JHOVE2 wiki found at: www.jhove2.org
[51] The DROID tool is a product of the National Archives of the UK/ part of the PRONOM project.: http://droid.sourceforge.net/

When JHOVE2 runs against individual files rather than a folder, the relationship between files is lost and JHOVE2 no longer recognizes the collection of component shapefile files as a shapefile. It identifies files with a .shp extension as either a 3D Studio Shape or a shapefile, but then never runs the shapefile module since the other required related files are not present. When using this tool on individual files during an automated ingest process, the output from JHOVE2 is much less satisfying than the output received from FITS. When running JHOVE2 against the contents of an entire folder, then the output includes nice details for shapefiles as noted above. The number of modules that JHOVE2 supports for various file types is currently quite limited.

During the testing process, an earlier version of JHOVE2 was first installed, and then upgraded to version 2.0. The output from the two versions was very different. While the newer version did fix a problem that the older one had with recognizing files that included spaces in their names, the older version was probably a little more readable. The majority of the new details that the tool produced appeared more focused on the internal functioning of the tool itself (such as modules used or time elapsed during processing), rather than displaying data that an archivist would need to capture to assist in documenting the technical attributes of the record at hand. However, when a module was available for a particular format type, the details, although buried among all the processing details, were helpful. When metadata extraction is automated, then the processing details only need to be deciphered and dealt with the first time they are encountered. If an archivist is manually running and reviewing the JHOVE2 output, they should be prepared to wade through a mountain of extra "stuff."

In Utah, the ingest process takes the output from the metadata extractors used, plus FGDC CSDGM metadata, if present, and maps the output to fields in the Utah State Archives' database. Programming the base functionality to accomplish this required a few months' effort. But with that functionality in place, new tools can be added and new data mappings completed with only a couple of days' effort per map, depending on the complexity of the XML source and how easy the tool itself is to install.

From the Utah team's testing, it appears that the use of JHOVE2 could be worthwhile for identification of shapefiles if the tool is run against a folder rather than individual files. If FGDC CSDGM metadata is not present, then the tool can also capture details that might otherwise be difficult for an archivist to discern. When comparing JHOVE2 against other metadata extractors, the results are less satisfying, although that may improve as development of modules continues. The data structure of the output appears to be changeable from one version to the next, so automated extraction may be impacted until the tool matures.

Integrity Checking- ACE

The North Carolina State Archives downloaded and installed the Audited Control Environment (ACE)[52] tool developed by the University of Maryland's "An Approach to Digital Archiving and Preservation Technology" (ADAPT) group. The ACE tool is "a system that incorporates a new methodology to address the integrity of long term archives using rigorous cryptographic techniques."[53] The tool has two key features:

- The Audit Manager is the client component that runs on Java and Tomcat, and uses a MySQL database. This piece runs locally and can be installed on nearly every modern computer platform, including Windows Desktop, Windows Server, Mac OS, UNIX, and Linux.

[52] ACE wiki: https://wiki.umiacs.umd.edu/adapt/index.php/Ace
[53] From https://wiki.umiacs.umd.edu/adapt/index.php/Ace:Main

- The Integrity Management Service (IMS) is the second piece that is server based and runs on the Java Enterprise Edition, using a Java Application Server. This service provides a token, stored remotely, that the Audit Manager can use to verify that the files are true.

Natively, the tool runs on a Linux platform. The NC DCR IT department loaded the software onto a MySQL server after consulting with the ADAPT group. The tools worked and hashed the files; however the process was extremely slow (7 hours to process 1 TB). After investigating, the IT staff determined that this occurred because of the type of hardware the department chose to purchase (best cost rather than high performance).

In looking at implementation, the first way in which ACE could be leveraged is to help detect files that have changed between what should be two identical sets of files. IT staff can write software to access the MySQL database that the Audit Manager uses, and from this create a simple text-based manifest of files at each location. Additional software will then compare those manifests to determine which files have been added, deleted or changed on the primary set so that these changes can be reproduced on the secondary set. If the tool detects changes on the first set, the staff will be alerted to investigate further and determine the cause. At the project's end, State Archives, in collaboration with the IT staff, was in the process of trying to determine the path forward using this tool. Since the Bag-It tool provides an initial hash to files, the NC Archives want to leverage the BagIt generated hash and use it as the baseline for thier digital preservation efforts. In that vein, NC State Archives considered using BagIt bags as the basis for the information package. Those bags will allow verification of the hash and certify that the content has not changed in the transfer as well as contain copies of the "original" files. Once records are placed in the repository, they will be re-hashed using the ACE tool and continue to monitor and audit them in the repository.

Future plans include possibly creating multiple ACE collections, purchasing a better SAN controller that has more data access channels and purchasing additional front-end servers. In this way, multiple ACE jobs can be run simultaneously. At GeoMAPP's closure, the ACE service was not in production as a true service. As such, one would have to rely on the servers in Maryland to provide that service and thus provide proof that the files have not changed. Depending on resource availability, NC Archives may pursue building a similar service or may contact the ADAPT group to formalize the relationship through a Service Level Agreement or a Memorandum of Understanding.

Rules-based preservation tools

At the beginning of the GeoMAPP grant Kentucky planned to explore the use of the integrated Rule Oriented Data System (iRODS)[54] as a preservation environment that might integrate with DSpace, the acquisition and access software that KDLA is using for its digital repository. Kentucky has continued to examine this avenue through participation in the Distributed Custodial Archival Preservation Environments (DCAPE)[55] project that looks to create rule based open source software (using iRODS) that could provide the basis for a preservation environment. IRODS "rules" are invoked on the servers internally to enforce/execute management policies for the system. Examples of policies include data management policies such as enforcement of authenticity, integrity, access restrictions, data placement,

[54] iRODS project site: https://www.irods.org/index.php/IRODS:Data_Grids,_Digital_Libraries,_Persistent_Archives,_and_Real-time_Data_Systems
[55] DCAPE project site: http://salt.unc.edu/dcape/

data presentation, replication, distribution, pre- and post-processing, and metadata extraction and assignment. For each user a set of rules can be adopted to establish a workflow.

KDLA did preliminary testing of loading the Kentucky GeoMAPP demo records into an iRODS test bed created by DCAPE. The project is developing a “virtual loading dock” containing a configurable workflow using a web interface. The loading dock consists of micro-services or rules to automatically generate checksums, replicate data, and scan for virus. It also contains metadata templates for repeatable information, but it has not been developed to the point where it can be fully tested. Since the DCAPE project is focused on the creation a hosted web based preservation service, it will not support integration with locally based repository software. KDLA is also looking at other software to develop its own a preservation environment for its holdings in DSpace.

Several plug-in tools have been used by the DSpace community to broaden the DSpace capabilities. Additionally, the DuraSpace community, which DSpace shares with Fedora, has embarked on DuraCloud[56], a cloud based open source preservation service that manages repository holdings stored both locally and in multiple commercial cloud locations. This service, which is designed to integrate with DSpace version 1.8 and some versions of Fedora, currently has replication, integrity checking, data validation, and audio/video conversion capabilities. It was moved from a test environment to full production (Version 1.1) in October 2011. KDLA is considering testing a 30 day trial of DuraCloud. In an effort independent to GeoMAPP, NCSU Libraries participated in an initial DuraCloud pilot, and has subsequently entered into an agreement to store 1 TB of non-geospatial content in DuraCloud.

Working Group Conclusions

The Preservation and Data Transfer Working Group explored numerous technical aspects of assessing and transferring geospatial content, as well as various techniques to preserve geospatial data over the life of the GeoMAPP project. Contributing factors to the group’s success include engaging people from a variety of backgrounds who were truly interested in the exploration of trying to preserve geospatial data. Many of the papers produced by the working group document insights that the GeoMAPP team gained over the life the project, and can be applied to the larger world of digital preservation. The products produced by the group were intended to benefit to the community at large, and the inclusion of geospatial data file formats described by the team into the PRONOM registry was a major achievement.

Key Findings

- A thoughtfully populated FGDC CSDGM metadata record contains rich material that can benefit both a detailed understanding of the dataset and for preserving that dataset for the long-term.
- Outreach to data creators on the importance of creating complete metadata and providing guidance on important fields to populate makes their data more useful and can benefit the preservation process.
- Dialog between Archives and GIS professionals is essential. Developing a good working relationship is critical. Few archivists are versed in the many files and formats used by the

[56] Duracloud wiki - https://wiki.duraspace.org/display/DURACLOUD/DuraCloud

geospatial community and need the GIS staff to help maintain this valuable data. Few GIS professionals fully understand how their actions can impact the survivability of their data for future use.

- Given the fact that technology constantly changes, new GIS professionals may not be familiar with older legacy formats that are still "floating around" out there and may be archived.
- File size, complexity, and the dominance of one major vendor in the GIS marketplace are critical issues impacting the preservation of geospatial content.
- It is critically important in the archiving process that institutions define archival metadata elements, how those elements will be stored, and how they will be populated, whether manually or through tool-assisted automation.
- Organizations should define and document archival processes for ingesting, processing, and making files available, and identifying processes and tools for how preserved geospatial data will be made accessible
- It is critical to define processes and tools for validating the integrity of the preserved geospatial datasets.
- It is critical to define the access derivatives that will be created, and to formalize the processes for creating them.

What's Next and What Still Needs to be Done

The field of digital preservation is dynamic due to the constant changes in technology. This opens the door to more research and development of long-term storage and preservation systems. The complexity of geospatial data (multipart files, Geodatabases, etc.) combined with the large files sizes especially with raster data will make preservation an ongoing issue. Some quick notes are:

- The *Geospatial Data File Formats Reference Guide* will need updates as more formats will be encountered. The working group obviously did not produce an exhaustive list, and purposefully left out some things like representation formats (such as Layer files) as out of scope. The group discussed how to keep the document living beyond GeoMAPP but reached no conclusions within the working group as to how this document would live beyond GeoMAPP.
- Continued discussion with Esri is a must. Digital preservation archivists generally support open-source or broad, standards-based formats as essential for long-term preservation. Given the GIS industry's reliance on Esri and their position in the marketplace, and the considerable limitations in functionality with many current open source solutions, Esri formats (like the File Geodatabase) will likely continue to be the primary formats for storing geospatial data. As noted in the *Archival Challenges Associated with the Esri Personal Geodatabase and File Geodatabase Formats* whitepaper, there are still ongoing concerns about the long-term sustainability of Esri Geodatabase formats.
- Continued development of a comprehensive archival software solution platform to assist the archivist to configure and manage the archival processing workflow and increase automation of its associated tasks, such as virus checking, identifying and validating, ongoing integrity checking and data recovery, archival metadata extraction and assignment, and managing related assets that enhance access (e.g. derivatives, catalog entries, finding aids, etc.)

Finding a Home for your Data and Sharing it with the World-Storage and Access

In the wake of the project's early explorations into moving content within and between the states, there was a realization that to strengthen the geoarchiving process the project needed to:

- Better understand storage infrastructures that support digital preservation environments;
- Research tools that could benefit discoverability of archived geospatial datasets;
- Investigate mechanisms to provide easy public access to these data.

To address these aims, the project established the Storage and Access working group to study a variety of tools and services that support access to geospatial records, as well as investigate storage technologies that can house archived geospatial holdings for the purpose of providing public access and for long-term preservation. The project's focus on storage was driven from the need to understand the differing storage media requirements for online storage for public access versus secure and restricted preservation storage for long-term preservation, as well as to strategize for the hefty storage requirements for preserving orthoimagery. While investigating data access solutions, effort was made to research tools from both the GIS and archival communities.

Before delving into technical analysis of storage and access solutions, early team efforts were dedicated to developing an evaluation template[57] that could be used by the project as a whole to capture critical information such as the features, cost, usability, and available support for storage and access solutions that the team was investigating. Divided into three unique tabs for archives, geospatial, and storage technology assessments, the tool provided a common template from which different reviewers' assessments of various tools could be reviewed and compared.

The following sections provide some background and information about some of the tasks that this working group investigated:

Storage and Access Tools and Whitepapers

- National Inventory of Archived and Superseded Geospatial Records
- GeoPortal Toolkit Evaluation
- Storage Primer and GeoMAPP Partners' Storage Architectures
- APPX-based Archives Enterprise Manager (AXAEM) Review
- GeoMAPP AXAEM Testing Instructions

Discovering Geospatial Data Holdings with Collaborative Inventory Tools

With the volume of geospatial data created on a daily basis, knowing where to find that data can be a challenge. GeoMAPP's interest in data discoverability drove an effort to look at this problem in more detail. The project focused on two approaches: 1) what national systems have been set up to provide an access portal to geospatial data from a variety of contributors, and 2) what resources are available from regional providers who have large geospatial collections. The team gave greater emphasis to those services that specifically included non-current data.

The results from this task are outlined in the white paper, *National Inventory of Geospatial Datasets*[58]. One national geospatial inventory system, called RAMONA[59], allows contributors to share metadata

[57] GeoMAPP Technology Assessment Template: http://www.geomapp.net/docs/Technology_Assesment_Template.xls
[58] GeoMAPP National Inventory of Geospatial Datasets: http://www.geomapp.net/docs/National_Inventory_of_Geospatial_data_final_20111231.pdf

about their geospatial records. Geospatial One-Stop (GOS)[60], another national geospatial inventory system that announced in September 2011 that it was shutting down its website and moving its data to the more generic www.data.gov website. The impact of this move on geospatial accessibility is unknown. The paper also provides linkages to large geospatial collections that may otherwise be unknown to the archival community and general public.

Researching Free Access Tools for Hosting Consolidated Archived Geospatial Content

During GeoMAPP's initial efforts from 2007-2009 each of the original partners transferred content from their state GIS clearinghouse to their archives. Each of the archives initially developed dark archives that were inaccessible to the general public to store this transferred content. After transfer was complete, the archival organizations began to develop cataloging solutions to identify which data was included in their geoarchives holdings and also began to provide some limited access to selected holdings.

One of observations that followed these transfer demonstrations was that none of the archives initially had the capability or resources to develop and host access copies of this superseded geospatial content in interfaces that could take advantage of the spatial nature of this archived data. There was also no easy mechanism to provide public access to the combined partners' data from a single consolidated location. As a result, the team decided to look into free web-based resources where superseded GIS content from multiple states could be uploaded to a common location and could be discovered or displayed on a map interface. The intent of this research was to discover existing tools and resources that archivists could freely use to easily publish archived geospatial content to the web without having to implement costly in-house spatial data infrastructure technologies.

The team chose to investigate two tools: GeoCommons[61] and Esri's ArcGIS Online[62]. Both tools allow individuals to create free user accounts to upload certain types of geospatial content for the intended purpose of online map creation using uploaded and other free content provided on the site. While not intended to act as remote storage locations, both applications support limited storage of a user's datasets and the creation of both public and private groups which either allow or restrict access to a user's uploaded content. Below are some unique details about both tools:

GeoCommons

GeoCommons supports the upload of KML, shapefiles (.shp, .dbf, .shx files only), and .csv data. Raster datasets can only be uploaded as KML. GeoCommons also allows for the creation of a minimal metadata record for dataset description and discovery that is automatically populated from information extracted from an associated uploaded geospatial metadata record. The tool allows a user to upload files up to 20 megabytes (MB) in size and has no size or count restrictions on the storage of publicly accessible datasets, but the user is limited to 20 MB of private data. Datasets can be downloaded either as KML, shapefile or .csv data.

[59] RAMONA http://gisinventory.net
[60] GOS: www.geodata.gov
[61] GeoCommons: http://geocommons.com/
[62] ArcGIS Online: http://www.arcgis.com/home/

GeoCommons User Library View

Positives include:

- Unlimited storage of publicly accessible datasets;
- Easy to use map building tools with some GIS analysis functionality;
- Ability to create a basic metadata record for data that may not have existing metadata;
- Easy to use search functionality for dataset discovery;
- Easy to use data management tools.

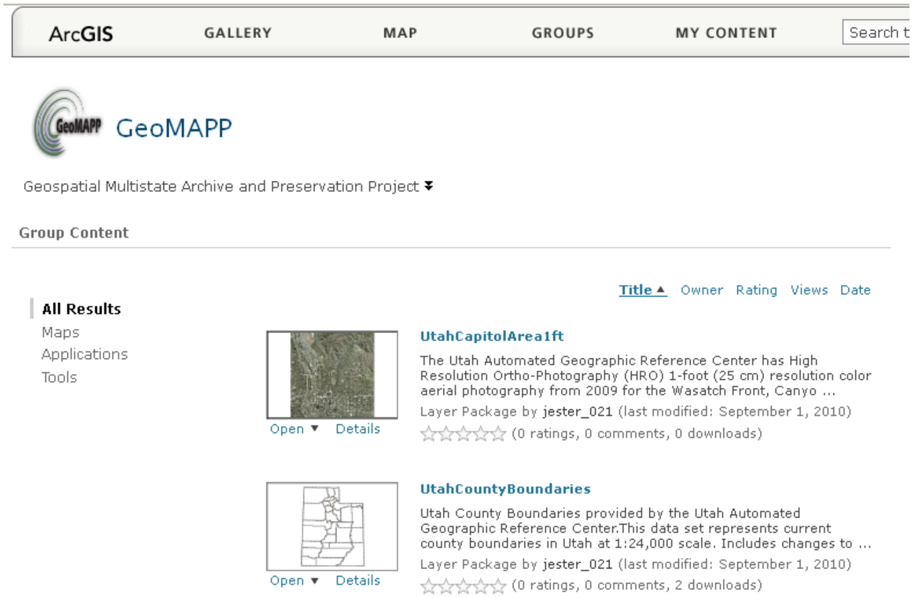
Negatives include:

- Shapefile upload only supports the requisite .shp, .dbf, and .shx files, and does not include a spatial reference or coordinate system which means that data will have to be projected when loaded into a GIS system;
- The minimal GeoCommons metadata record is included with a data download, but any unique metadata provided with the original dataset is not uploaded to GeoCommons and is subsequently not included in the download package
- Limitations on private storage size;
- The team experienced some application difficulties (timeouts, etc) with uploading datasets.

GeoCommons offers a nice tool for uploading datasets for the purpose of building maps. However, with its limitation regarding uploading only a subset of the shapefile dataset files, GeoCommons would have limited value as an access and/or distribution solution for archives. The GeoMAPP team did meet with GeoCommons technical staff during the project and shared some of these concerns. GeoCommons hopes to address many of these items in future product releases.

ArcGIS Online

ArcGIS Online supports the upload of KML, .csv/text files, GPS files (.gpx), zipped shapefiles and LPK/MPK files as well as Esri map documents (.mxd), layer files (.lyr) and OGC and Esri web services. Raster data can be exposed via web service or as part of a KML file. ArcGIS Online allows the user to create some minimal documentation for each file (background info, access/use constraints, and tags) but does not appear to extract information from the geospatial metadata file. Each ArcGIS Online account comes with 2 GB of total storage space, with an upload limit for individual items of up to 1 GB in size and no limitations on private versus public data.



ArcGIS Online Group Content Section

Positives include:

- The upload of a wide range of data types including the ability to expose web services;
- 2 GB of storage for private data holdings;
- The ability to download data exactly as it was uploaded;
- Easy to use data management tools;
- Lots of free data available for map creation.

Negatives include:

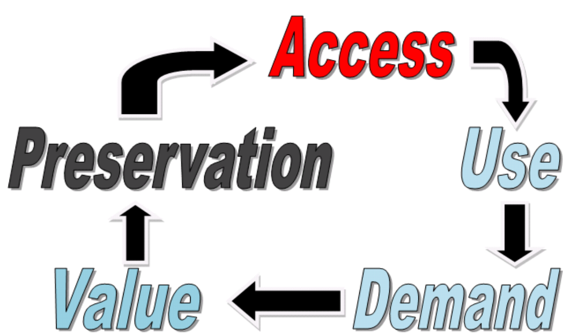
- Geospatial metadata associated with uploaded files is not integrated into data discovery or management tools;
- It is difficult to have uploaded shapefiles display in map making interfaces without associated layer files;
- Limited overall storage which would preclude the posting of uncompressed image files;
- Difficulties making maps with uploaded data.

Conclusion

Both tools were able to meet the basic requirements to support the upload of geospatial content from multiple agencies to a common location, making the data publically discoverable and viewable using both text and web search functionality, and supported some sort of download functionality for hosted datasets. While neither tool seems to be a viable option to provide access to and discovery of large archival collections, they both have promise for providing limited access to smaller "demonstration" collections and offer unique, user-friendly map making tools that could take advantage of uploaded superseded content.

<u>Improving Access to Archived GIS Content at each Partner's Archives</u>

In addition to the collaborative efforts to research technologies that support data access within the geospatial and archives domains, each state partner also worked to improve access to their archived geospatial content. A significant portion of GeoMAPP's project outreach centered on 'liberating' some preserved content from dark archives, initially implemented within each state to support their preservation activities, and making this data accessible to the public.



Data Access Benefit Loop

The benefit loop for providing free and easy access to archived datasets is based on the notion that as people begin to use these data and realize the value of this information, demand will increase for other archived datasets. Increased usage may also promote an increase in the value-add to existing datasets, that in turn may be submitted to archival organizations for preservation.

The following sections detail advancements each state has made in providing access to their archived GIS content during the final two years of the project:

The **Kentucky** Department for Libraries and Archives uses DSpace as an ingest and data discovery tool for most of the electronic records that have been accessioned into the electronic records archive (e-archives). Due to access issues and file size limitations, only some shapefiles, geospatial PDFs, image files and very small databases can be compressed for download from this web-enabled, open-source software. During the ingest stage the following archival features are possible via DSpace or DSpace plug-ins:

- Automatically captures a checksum and records it in DSpace. Verifying and monitoring the checksum is possible using a plug-in script in DSpace, but not yet implemented by Kentucky.

- After pre-ingest processing, files are moved to archival storage within DSpace. Source files are tracked via a separate accession record. The accession number recorded in DSpace enables the archivist to locate source files.
- Metadata extraction and format validation may be integrated with DSpace ingest, but Kentucky has not found time to implement appropriate scripts. During ingest Kentucky uses a template within DSpace to auto-populate fields that are constant within a series grouping.

After ingest the following archival features or descriptive fields are used for geospatial records:

1. A basic parent-child structure permits simple handling of complex objects such as a shapefile. Within DSpace records, whether single or multi-file, are placed within a records series under an agency hierarchy.
2. All electronic records inherit basic metadata from the series or collection-level in DSpace.
3. Item-level metadata stores are based on the Dublin Core (DC) metadata standard, including specialized DC fields for content relevant only to geospatial records (geographic bounding box, projection, etc.). Bounding box coordinates are added only for shapefile records, not for image or PDF maps where the coordinates are not available.

Kentucky provides the public access using the following features and limitations:

4. DSpace allows keyword searching on a set number of fields and browsing by issue date, creator, and title. Customization of the fields used in various types of search is possible; however Kentucky has not yet done so. Since geospatial subjects and Dublin Core field usage is specialized, the user will be guided by an introductory page designed to help the user to identify and use the geospatial holdings. DSpace allows for creation of customized interfaces to specific records at the agency or series level. Kentucky has begun customizing the interface for some records, but has not implemented this for geospatial records.
5. Files can be viewed within DSpace if the desktop is configured to handle the format (images and PDF files open automatically. If shapefiles are downloaded from DSpace to a desktop, that desktop needs software (Esri or other) to read it.
6. EAD finding aids are **not** produced from DSpace and Kentucky has no pressing need for that product currently.
7. Using patterned searches at the county level, Kentucky has created a map-based search in DSpace for Confederate pension records (see screenshot below in Web Mapping section).

Customization of DSpace by archives staff to improve the user interface, enhance processing, or ensure preservation is dependent on limited technical resources to adapt scripts written by other institutions using DSpace.

Montana used their work with GeoMAPP to define and begin initial development of an archives and inventory system. In the past, NRIS conducted initial inventorying efforts but those efforts have never been sustained and did not include an archival element. This new system will inventory data upon arrival at the Montana GIS Clearinghouse and a master copy of the data will be archived at that time. Over the summer of 2011 the MSL researched several content management systems including APPX to determine if a vendor system might serve as an inventorying system but, in the end, felt that these systems were overly complex given Montana's needs. Instead, MSL developed system requirements for a system that is being developed in-house. The intent is to develop a near-fully automated system that will place data in its dark archive location and will include auto-generated checksums, archival metadata and automated data integrity verification. However, the initial release of this system, slated for winter 2012 will require that these steps be done manually.

The process to develop system requirements required MSL to:

- Develop a file structure for the archive;
- Test audit software;
- Create metadata requirements;
- Test initial data transfers following the planned requirements documentation.

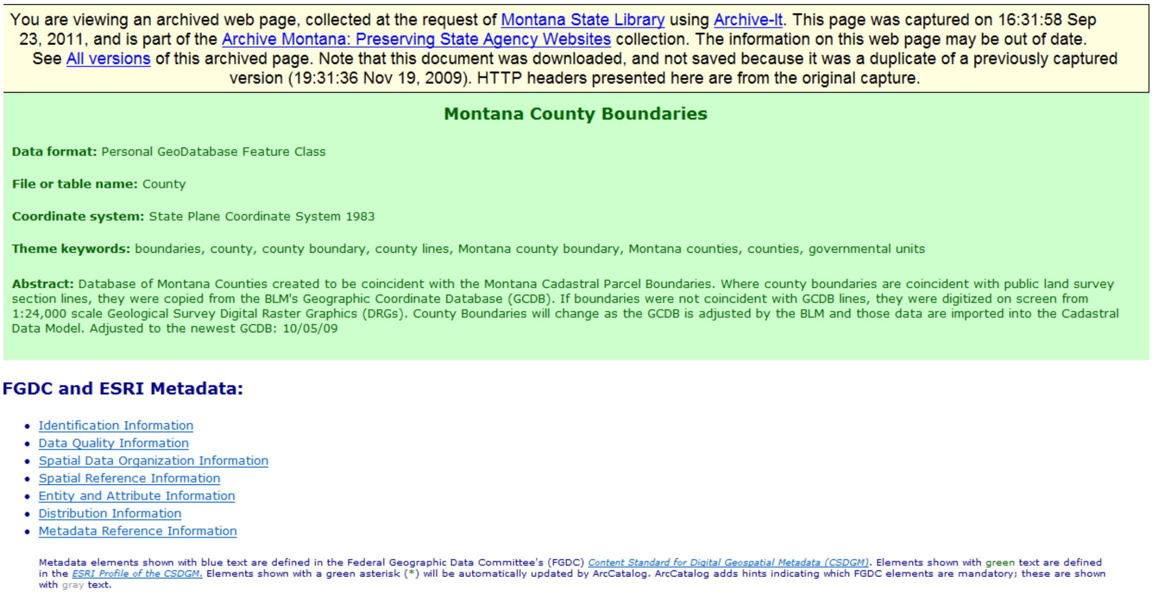
MSL tested the use of both standard Zip files and BagIt to generate and verify manifests and checksums. This testing was completed successfully and MSL determined that the BagIt tool will be used to validate data upon ingest into the preservation system.

MSL developed a folder system that will be used to organize archived master GIS data files. This folder system organizes general GIS data by date and/or by series data (i.e. Cadastral) or collection data (i.e. Butte Superfund Collection) as appropriate.

MSL also developed a set of administrative and archival metadata requirements that will be used to augment the metadata received by the publisher or created by MSL staff. At this time this metadata is primarily being recorded manually, however much of this is planned to be automated in the future. MSL intends to work with a statewide metadata working group to update Montana technical metadata standards to include metadata useful for archiving purposes.

Although this process to archive GIS data will likely require further refinement, MSL believes it is ready to be implemented for any new data that the Clearinghouse acquires. There will be an ongoing process to inventory, appraise and archive MSL's existing GIS data collection.

During the summer of 2011, MSL tested the Internet Archive's web archiving service to archive downloadable GIS datasets currently maintained on the MSL website. Test results were mixed and the viability of ongoing web harvesting is currently being evaluated. MSL also has plans to test the ability to upload GIS data to the Internet Archive for availability through http://archive.org. Initial research revealed that a handful of shapefiles are available for download from the Internet Archive. However, a process to upload GIS data and metadata to the Internet Archive will require research beyond the scope of the GeoMAPP period.



You are viewing an archived web page, collected at the request of Montana State Library using Archive-It. This page was captured on 16:31:58 Sep 23, 2011, and is part of the Archive Montana: Preserving State Agency Websites collection. The information on this web page may be out of date. See All versions of this archived page. Note that this document was downloaded, and not saved because it was a duplicate of a previously captured version (19:31:36 Nov 19, 2009). HTTP headers presented here are from the original capture.

Montana County Boundaries

Data format: Personal GeoDatabase Feature Class

File or table name: County

Coordinate system: State Plane Coordinate System 1983

Theme keywords: boundaries, county, county boundary, county lines, Montana county boundary, Montana counties, counties, governmental units

Abstract: Database of Montana Counties created to be coincident with the Montana Cadastral Parcel Boundaries. Where county boundaries are coincident with public land survey section lines, they were copied from the BLM's Geographic Coordinate Database (GCDB). If boundaries were not coincident with GCDB lines, they were digitized on screen from 1:24,000 scale Geological Survey Digital Raster Graphics (DRGs). County Boundaries will change as the GCDB is adjusted by the BLM and those data are imported into the Cadastral Data Model. Adjusted to the newest GCDB: 10/05/09

FGDC and ESRI Metadata:

- Identification Information
- Data Quality Information
- Spatial Data Organization Information
- Spatial Reference Information
- Entity and Attribute Information
- Distribution Information
- Metadata Reference Information

Metadata elements shown with blue text are defined in the Federal Geographic Data Committee's (FGDC) *Content Standard for Digital Geospatial Metadata (CSDGM)*. Elements shown with green text are defined in the *ESRI Profile of the CSDGM.* Elements shown with a green asterisk (*) will be automatically updated by ArcCatalog. ArcCatalog adds hints indicating which FGDC elements are mandatory; these are shown with gray text.

Additionally, MSL licensed the EBSCO Discovery Service (EDS) integrated search tool to harvest standardized metadata from a variety of tools and centrally index the metadata to provide integrated discovery to a wide variety of library information resources. The initial launch of the EDS system is slated for fall 2011. After this initial launch, MSL plans to test harvesting metadata from the Montana GIS Portal for inclusion in this centralized discovery system.

North Carolina continued to ingest, process, and provide access to newly received datasets. In addition, they made previously ingested datasets accessible online through the North Carolina State Archives Manuscript and Archives Reference System (MARS) [63] and the North Carolina Digital Collections, which utilizes CONTENTdm for dataset discovery and download. Significant consideration was made regarding managing metadata, and creating crosswalks between MARS and the CONTENTdm system, while considering other collections' metadata models, to promote consistent use of the metadata across all of the collections.

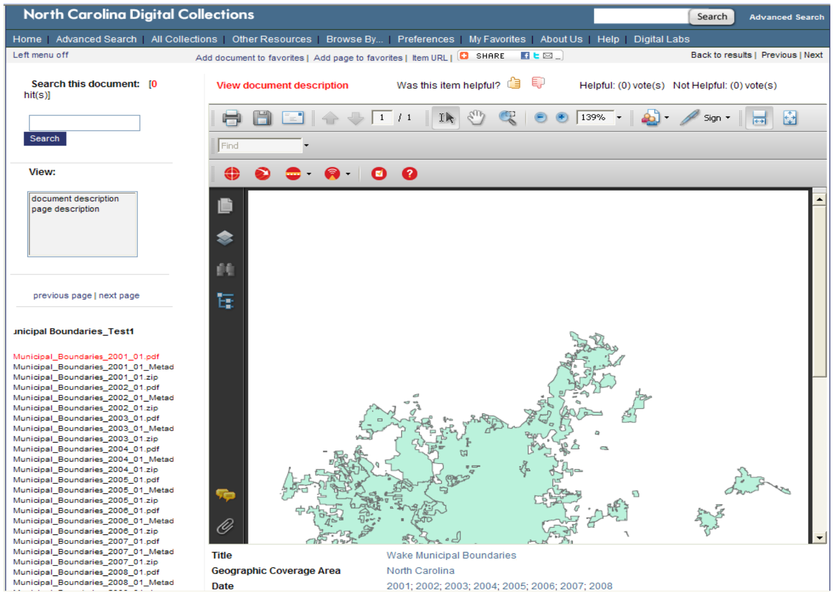
North Carolina uses a file-system-based approach to managing its geospatial collection. It has designated staging, preservation, and access spaces to manage the archival processing lifecycle. File transfer and dataset validation occurs in a staging area. A clean file transfer is validated using the BagIt verifyvalid option. Esri ArcCatalog is used to verify that each dataset is functional, its attribute table is reasonably populated, and the geospatial metadata is complete. As part of the processing, the originally received geospatial metadata file is copied, and a processing step is added to document the transfer of the dataset to the archives. Access derivatives are produced:

- ArcMap is used to create a geospatial PDF;
- ArcCatalog is used to create a styled HTML formatted version of the geospatial xml metadata file;

[63] MARS Catalog: http://mars.archives.ncdcr.gov/AdvancedSearch.aspx

- A zip archive file is created that contains all of the dataset files.

A preservation master version of the dataset is copied to the preservation repository and the access materials are copied to the access repository. Then the access files are uploaded to CONTENTdm as a compound object and CONTENTdm metadata is populated. The North Carolina team prototyped an archival series-like CONTENTdm entry by loading several years' variations of a particular dataset into a single CONTENTdm object. This has the benefit of providing the user easy access to the temporal progression of a single dataset over time through a single CONTENTdm entry.



North Carolina Digital Collection - built on CONTENTdm

In addition, North Carolina explored publishing its archived datasets to the national geospatial inventory system, Geospatial OneStop (GOS). North Carolina registered the Archives as a "Clearinghouse" entity, registered datasets available through the North Carolina Digital Collection as "downloadable" datasets, and registered datasets held in the access repository but not yet available through the North Carolina Digital Collection as "offline". In registering the datasets with GOS, they were explicitly labeled as "archived," and this seemed to offer a viable approach for publishing archived geospatial datasets through a national geospatial inventory system. It is unfortunate that GOS has an uncertain future, as this seemed to offer a viable distribution point for archived geospatial datasets through a well-known geospatial distribution channel. The RAMONA GIS Inventory system offers another alternative, and would likely support a similar sort of strategy, through item naming. However, when registering the archived geospatial datasets with a national geospatial inventory system, the archival organization will likely need to update the geospatial metadata file to reflect the archival organization as the distributor for the archived geospatial dataset, so that when a dataset is downloaded, and the metadata file consulted, it is evident that the dataset was retrieved from the archival organization, as opposed to the dataset contributor to the archives.

Utah made significant strides in their ability to manage electronic records in general as well as geospatial records in particular. The software the Utah State Archives uses to manage all of their work, APPX-based

Archives Enterprise Manager (AXAEM), has been benefitted from a number of updates during the final years of the project.

When electronic records are ingested into AXAEM, a checksum is captured at the desktop level. Files are copied to the server, and then the checksum is rerun to make sure all files were safely transferred. The process writes new records in the database to reflect entries for each file, called an Electronic Record, as well as objects at the folder level, called an Object Group, and sub-objects within the folder as needed (in the event that more than one multi-file object is stored in the same folder). Details in these records include the location of the source file as well as ingested files, whether on the server or in hosted locations. Functions include the ability to send copies of the files back to the desktop for viewing. Items in Object Groups are zipped so that all files arrive at the desktop as a unit and the desktop auto-launches a viewer that the user has associated with that file type. From the Object Group record, users can drill down to see details from individual Electronic Record files, sub-groups, or parent objects. Object Groups record structural relationships between folders, and allow archivists to write descriptive metadata at any folder or object level of their choosing.

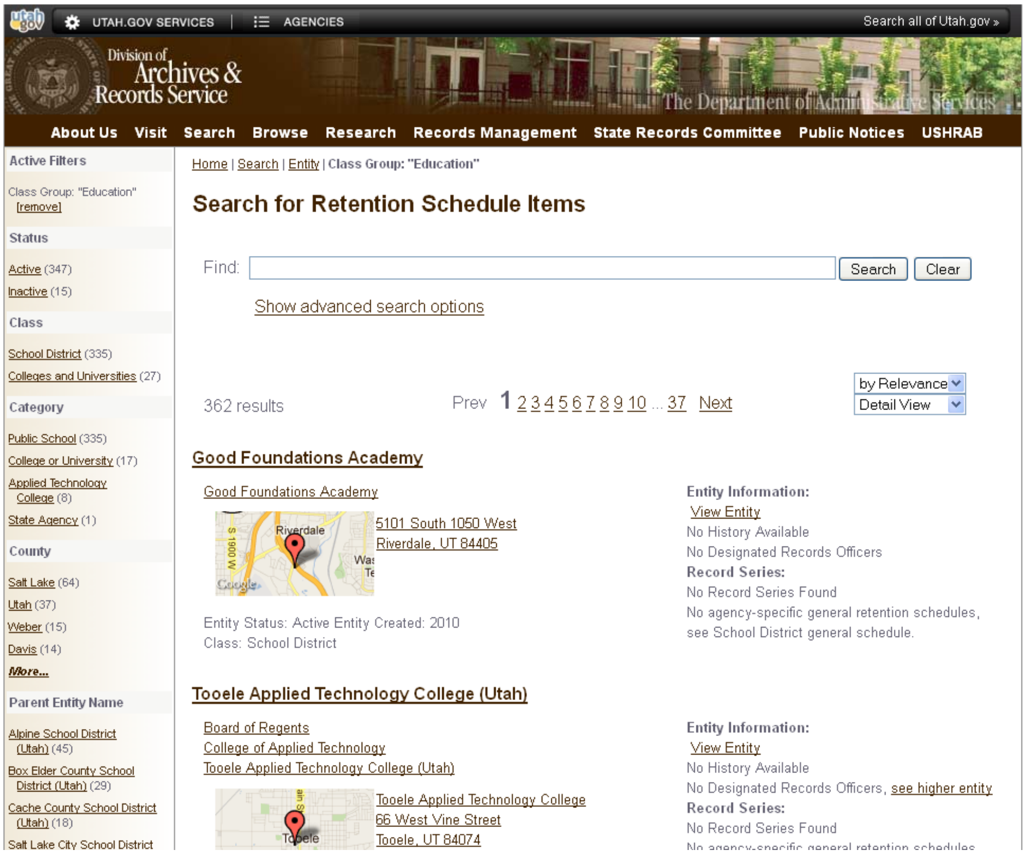
During ingest, users may extract metadata from each file using tools FITS and JHOVE2. Typical metadata captured includes file extent in bytes, format(s) of files, creation and modification dates, and folder names from the original directory tree for re-use as possible subject headings.

Some records come with their own descriptive metadata, such as FGDC CSDGM-compliant GIS metadata files. Information from the metadata file can be mapped to the description, subjects, contributors, dates, and geo-specific metadata fields within the Object Group. All original geospatial metadata source files are stored with the ingested electronic objects. Metadata extractors may be run against these files more than once as tools are updated, and each iteration of metadata is stored as a separate file with the record. When the ingest process is completed, AXAEM reports back to the user what was successfully ingested and any errors encountered.

AXAEM database also supports the ability to record archival metadata. All Electronic Records inherit metadata from their series or collection level, including provenance, custody history, scope and content, and technical access notes. Item-level metadata stores details consistent with Dublin Core, as well as more specialized fields for subject-specific content such as geospatial records (geographic bounding box, datum, projection, and attributes).

Public access to ingested records is a critical component. AXAEM has been integrated with the Solr search engine, which allows broad keyword searching capabilities as well as facets and filtering by field types. While still in development, the intent is that the contents of ingested records will be as searchable as possible, not just the metadata contained within the Electronic Record or Object Group. Search results will display metadata and allow users to download datasets. The software does support map integration, which could utilize coordinate details from the metadata. Another access tool includes finding aids.

AXAEM automatically produces EAD-formatted finding aids for all collections, and they also include links to geospatial datasets stored on Utah’s FTP server.



: *Prototype Search Interface based on Solr*

Future development for this system, which should be available after December 2011, includes the ability to ingest container files from tools such as from BagIt, zip, and tar. In the case of BagIt, the files in a bag will be bit-level validated, and files extracted as needed. Also, virus checking will be added as an optional step in the ingest process. Tools will be added that can transform from a specific data format to one or more formats either upon ingest or some other timetable. One likely result from that feature will be the ability to create a .jpg file for viewing records directly in the desktop client as well as providing thumbnail images for web server access. Ongoing automatic checksum audits will also be deployed.

Using Web Mapping Techniques to Provide Access to Archives other Digital Holdings

GeoMAPP's collaboration between geospatial and preservation staff created a natural interest in how the work product of one group could benefit the work product of the other. To that end, the workgroup evaluated a handful of technologies that use map interfaces to facilitate discovery of archival records. These technologies included Brigham Young University's Mappify[64], North Carolina State University's

[64] Mappify: http://lib.byu.edu/DigitalMaps/

WolfWalk[65], and Utah State Archives' AXAEM system using a Geodatabase web service [66] back-end. Kentucky also explored the use of a map interface [67]for a specific collection.

Mappify

The Mappify project at BYU uses a mash-up between Google Maps and CONTENTdm that allows the users to run a search on specific metadata fields and see responses on the Google Map with clickable links that take the user to the record in question, such as a photograph. This tool is no longer in development and the code has not been made public. The tool as designed does not offer support for discovery or presentation of superseded geospatial datasets, but does offer an easy to use map interface to discover and view digital materials that have a defined X,Y location.

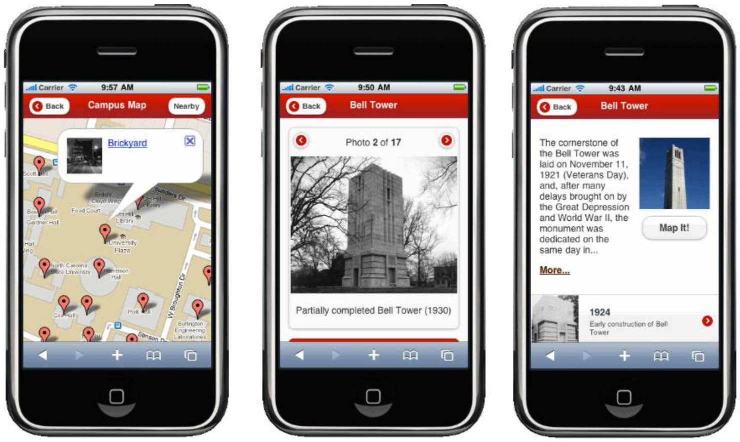
WolfWalk

WolfWalk is a mobile library project that enables users to explore North Carolina State University campus history using a location-aware interface optimized for mobile devices. The application supports a map view with geo-tagged place marks for over 90 major sites of interest on the NCSU campus, and a browse view for quickly locating a known site by name. Each site has several historical images associated with it that are sourced from NCSU Special Collections Research Center digital archives. WolfWalk is a pilot project to explore new user interaction models with digital collections on mobile devices.

WolfWalk capitalizes on the location awareness of mobile devices to allow users to give themselves a self-guided historical walk through NC State's campus. As users stroll around the grounds, their mobile devices detect their current locations and deliver a tour of nearby buildings and other historically interesting locations. Users who use a device other than the iPhone or iPad can access material through the web version of the tool, and owners of devices that don't support GPS or other location-detection methods can still manually navigate through the web site to enjoy a tour of campus.

The materials that form the basis of WolfWalk's database are heavily drawn from the resources of the University Archives in the NCSU Libraries' Special Collections Research Center, a vast array of documents, photos, audio files, and other historical materials from the founding of the school up through the present. While, not directly associated with GeoMAPP, the WolfWalk application offers a unique example of providing access to historical materials via a mobile device-enabled map interface.

[65] WolfWalk: http://www.lib.ncsu.edu/dli/projects/wolfwalk/
[66] Utah Webmap: http://archives.utah.gov/mapsearch.html
[67] Kentuck Webmap: http://www.e-archives.ky.gov/Confederate/kdla/index.html



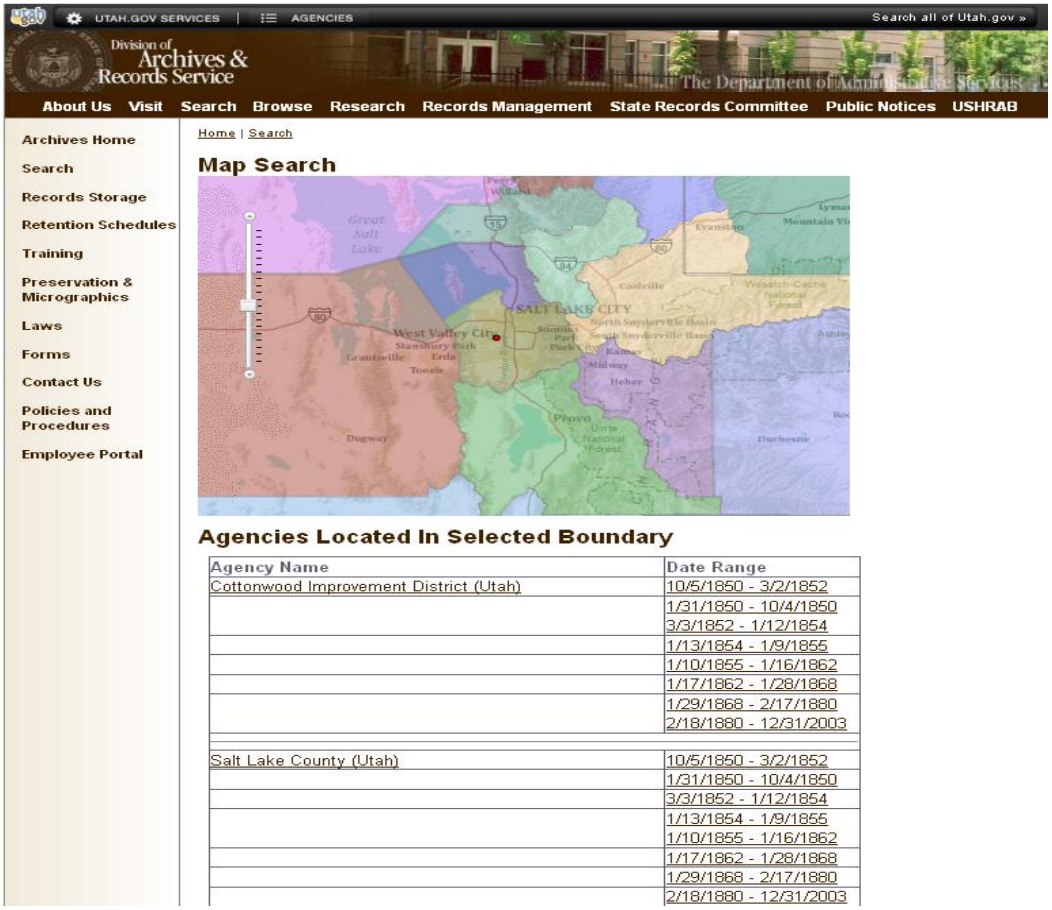
WolfWalk sample screens

AXAEM Map Interface

The AXAEM map interface approaches the task from a different angle. It associates locational information derived from historic boundaries stored in a geodatabase with various record creators in the Archives' database. Since the record creators include institutions with understood geographic boundaries, such as counties, municipalities, school districts, court districts, mining districts, etc., it's possible to click on a point on a map and produce a list of records tied to the agencies whose boundaries surround that point.

One interesting aspect to this approach is that over time, boundaries change, yet records are still tied to the name of the organization that created them. Clicking in Utah's current Garfield County will yield search results tied to Iron County, since records that were created prior to 1860 for that area belonged to Iron County. This provides the user with a methodology to find records based on location that they might otherwise think aren't available if they searched solely based on the current entity name. Since the map also displays a broader area than Utah's current boundaries, territorial records may also be searched. For example, Utah State Archives records include holdings from Carson County, located in current-day Nevada.

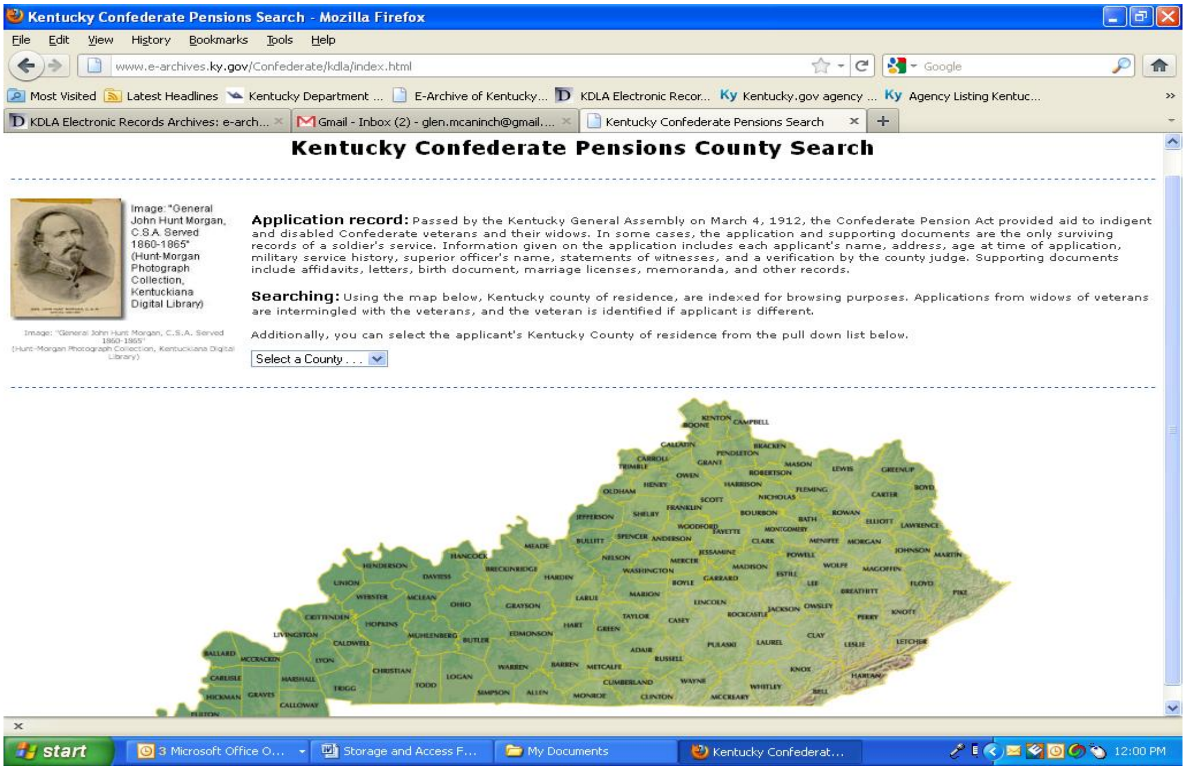
As a proof-of-concept, the AXAEM map interface shows what's possible, especially as more location-based metadata are included in the database for all types of records. Eventually, the goal is to be able to zoom into a street level, choose a date span, and find all records related to that time and place, even drawing in other archival resources from other institutions through OAI-PMH harvesting. Records could include geospatial datasets. For this to happen, much development needs to take place. Currently, the agency names in the Archives' database should have more beginning and ending dates added so that they can be tied to the correct boundary, and not be tied to boundaries that existed before the entity itself existed (e.g. the GIS office of Tooele County did not exist prior to 1990, and so should not be a valid hit for boundaries associated with 1910). Other development would need to happen on the geodatabase side, so that changes in the AXAEM database could be reflected in the GIS attributes. For the initial test, a spreadsheet linking agency names and boundary shapes was provided to Utah AGRC, and then incorporated manually into their geodatabase. A web service then hooked the two systems together at runtime so that the map search could draw upon live AXAEM data. For a production environment, the two systems will need to update each other a little more directly.



Proof-of-concept Map-based Search Interface for Utah AGRC

The other map interface being explored via AXAEM relates to the use of electronic record metadata and the Solr search engine which has been integrated into the application. Any record which includes coordinate data or place names can be connected to Google Maps. Specific features for this type of integration have not been explored yet. However, since the technology appears to readily support location searching, the possibilities look promising.

The Kentucky team has also implemented a web mapping interface as an additional search tool to reference the county of residence for Confederate pensioners whose records have been scanned and stored in the e-archive. The tool was created by DGI and can be reused for other appropriate search applications.

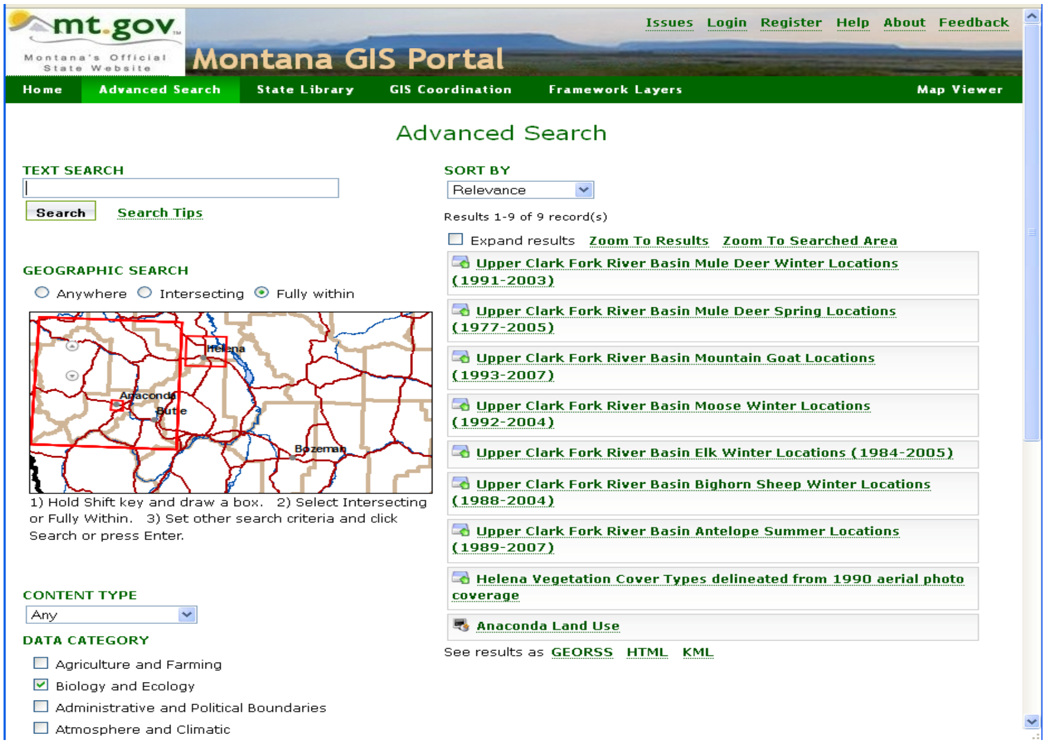


Kentucky Confederate Pension Records Web Mapping Application

<u>Leveraging Esri's Geoportal Server as a Tool for Accessing Archived Geospatial Content</u>

Esri's Geoportal Server (formally named GeoPortal Toolkit) is a software suite that allows agencies to store information about spatial dataset and web service holdings and provides both map and text based searching mechanisms to both discover these holdings and provide linkages for data access. The tool itself includes a web front end and a database that acts as a metadata catalog, which stores and manages metadata records describing an agency's GIS data holdings without actually storing physical GIS datasets.

Geoportal has a user-friendly search tool that features map-based searching to specify geographic regions of interest and provides visualization of the geographic extents of datasets discovered in the search results. If the metadata stored within Geoportal contains an access path for locating the physical dataset, then the Geoportal search results page provides a link for downloading it. Out of the box, Geoportal Toolkit is essentially a development kit whose default web interface requires customization in order to satisfy the needs of its users and any required custom functionality.



Montana GIS Portal search results - built on Esri GeoPortal

Three states (Kentucky, Montana, and Utah) installed Geoportal and described the installation process. The findings of this investigation may be found in the white paper, *GeoPortal Toolkit Case Studies.*[68] Some common findings from these explorations included trouble installing the software among each of the states, although each of the implementations were eventually successfully installed. The software requires specialized expertise for proper installation and for customization of the web interface. The main advantages this tool provides over typical archival inventory and access systems are its ability to extract information from existing geospatial metadata, and the use of a map-based search interface accessible via a web-browser.

An organization which does not provide access to GIS data as one of its primary missions may find the Geoportal Server to be an overly sophisticated and complex solution to implement for basic access functionality. Archival organizations, however may wish to partner with a state GIS clearinghouse that uses Geoportal Server. The archival organization could enter metadata for its holdings in the clearinghouse's Geoportal, or the clearinghouse could implement another instance of the Geoportal dedicated to the discovery of archived data collections.

Exploring Storage Options for Geoarchiving Needs

The costs of storage for geospatial records are a common concern for both archivists and GIS professionals. Imagery files require significant storage capacity, and geoarchiving collections that include

[68] GeoPortal Toolkit Case Studies: http://www.geomapp.net/docs/Geoportal_Toolkit_Evaluation_final_20111231.pdf

imagery can run into the multi-terabytes. A 2010 statewide collection of aerial imagery from North Carolina totaled over 17TB in size. There are a variety of storage technologies and services available on the market today. Their capabilities range from simple disk-based storage devices to complex storage networks, and they continue to evolve and extend their storage management functionality.

Individual adoption of a particular technology should take into account a variety of factors including: budget; internal IT resources and skills; available physical space; policies that may restrict options; a willingness to use third-party services and store records off-site; and the ability to maintain and/or migrate any chosen media. Due to these variables, the working group chose to develop a storage primer that describes several possible storage solutions, along with their applicability for archival use, rather than champion any particular solution as a "best practice." This list may be found in the white paper, *GeoMAPP Storage Primer*.[69] Each state's archival storage architecture is also presented, offering examples of storage solutions the partner states implemented to manage their geospatial collections.

Exploring Partnership Opportunities for Distributed Storage

At GeoMAPP's closure, several states were exploring the opportunity to collaborate on a cloud-hosted solution for geospatial records. In 2010, a request for information (RFI) was issued by Montana, Utah, Oregon, and Colorado to determine the feasibility of this plan, and a follow-up request for proposal (RFP) was issued in December, 2011. The intent of the RFP is to lead to a contract that will allow states to receive volume discounts for storage if they are all using the same cloud service. This would help preserve at least one copy of the data, although security copies may yet use more traditional local technology.

As of the end of the grant, the findings for this collaboration are not yet known since the cloud service is not yet in use. However, the willingness of the various states to enter into such a partnership does show that such endeavors are possible and could result in cost savings. The GeoMAPP team shared archival considerations with the cloud services team to be used as the RFP was drafted.

Other GeoMAPP partners have explored the use of Duracloud. Investigation of this technology has so far been limited to seeing presentations about it. The DuraSpace community, which DSpace shares with Fedora, has embarked on Duracloud, a cloud based open source preservation service that manages repository holdings stored in both local and multiple commercial cloud locations. This service which is designed to integrate with DSpace version 1.8 and some versions of Fedora currently has replication, integrity checking, data validation, and audio/video conversion capabilities. It will be moved from a test environment to full production (Version 1.1) in October 2011.

While not directly driven by GeoMAPP, as part of NC CGIA's efforts with the 2010 North Carolina Statewide Orthoimagery project, the state engaged with the US Geological Survey (USGS) to share a copy of statewide color 6-inch resolution orthoimagery in native GeoTIFF format with the USGS' EROS Data Center. EROS will copy and process the images for use in The National Map as well as retain a copy of the original imagery files at the Data Center.

[69] GeoMAPP Storage Primer- http://www.geomapp.net/docs/GeoMAPP_Storage_Primer_final_20111231.pdf

Investigating Systems for Finding Geospatial Datasets Ingested into an Archive

The archival management systems offering access to GIS datasets being used by North Carolina, Kentucky, and Utah include: CONTENTdm; DSpace; and AXAEM. All of these tools have shown the ability to provide access to many types of electronic records, including geospatial records. Different levels of data management are also available.

In the case of AXAEM, some preservation management features are available to specifically help the archivist identify and manage archived geospatial records. One important feature is its ability to manage multi-item objects. Records may be described as individual files, or as part of a group. Using PRONOM-enabled JHOVE2, the system automatically recognizes a shapefile when it sees certain naming patterns upon ingest. It also provides the user with the ability to declare a set of files as a geodatabase, orthoimagery set, or project. FGDC CSDGM metadata is captured and maintained with these files, along with traditional archival metadata. Other features include the ability to record format types and associate those types with format history and tools which can be used to interact with the format and migration rules. Future development will include the ability to auto-migrate or transform one format type to another and/or update the version of the format to a later release, provided appropriate tools are available and robust enough for integration with AXAEM.

DSpace, the data management software used to disseminate archival records in Kentucky, can ingest shapefiles either as a single .zip file or as separate parts of a single record. Kentucky found it impractical to ingest snapshots of File Geodatabases or statewide raster data into DSpace because the file size makes it difficult for users to download, and many users are unlikely to have the necessary GIS tools to use the records. Individual layers in the File Geodatabase that did not change more than once or twice during a three year period were extracted as shapefiles and made available as single records associated with the agency that created them. These extracted shapefiles are also cross referenced to the listings under the Department for Geographic Information in the agency browse section of DSpace, where there is a description with each archived vector snapshot. While the File Geodatabase including the vector snapshots cannot be downloaded from DSpace, a report of the contents of the database can be downloaded and the user is encouraged to contact the archive if specific feature classes are needed. Plans are being made for users to have the ability to come to the archive and be provided direct access to each of the File Geodatabase snapshots using Esri software available in the research room. The DSpace e-archive also contains numerous maps in geospatial PDF format, raster images, and shapefiles taken from state and local agencies' web sites that were never included in the state's centralized database system.

Investigating APPX as an Archives Management Platform

At the beginning of the GeoMAPP grant, the Utah State Archives had a web-based archives management application written with software called APPX. This was used to manage all archival functions and Utah's unique workflows in particular. At the time it did not really include electronic records, other than in passing as part of a retention schedule or a line item on a finding aid. With the need to support geospatial records, the idea was posed to adapt the software to handle both GIS and other e-records holdings.

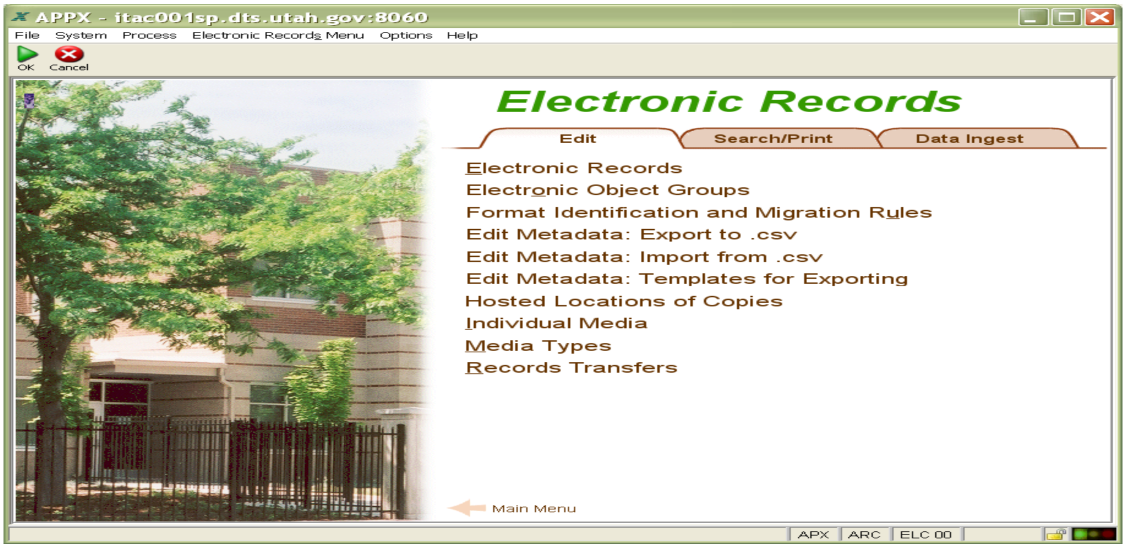
In September of 2009, the Archives made their code open source and named it APPX-based Archives Enterprise Manager (AXAEM). With that, a project was launched with their vendor, APPX Software, to evolve the system to a more general purpose archival management platform. At GeoMAPP's end, that adaptation was still in process. In the resulting two years, hard-coded Utah-specific features were removed, parameter files were added to record institution-by-institution values for certain fields, and a plethora of functionality was added to accommodate many types of electronic records and geospatial records in particular. Menus were changed and data entry wizards were added to assist new users to navigate the application interface.

In the midst of these rapid changes, the Utah State Archives invited various institutions to test the system. Early testers ran into some difficulties when parameter files were not in place yet, which meant that features not problematic in Utah's production system suddenly didn't cooperate in the beta test environment. Later testers, including institutions outside of GeoMAPP, have had more success. System documentation was still being finalized in late 2011, but some of it was made available to testers. In North Carolina and Kentucky's final review of AXAEM, the conclusions were positive.

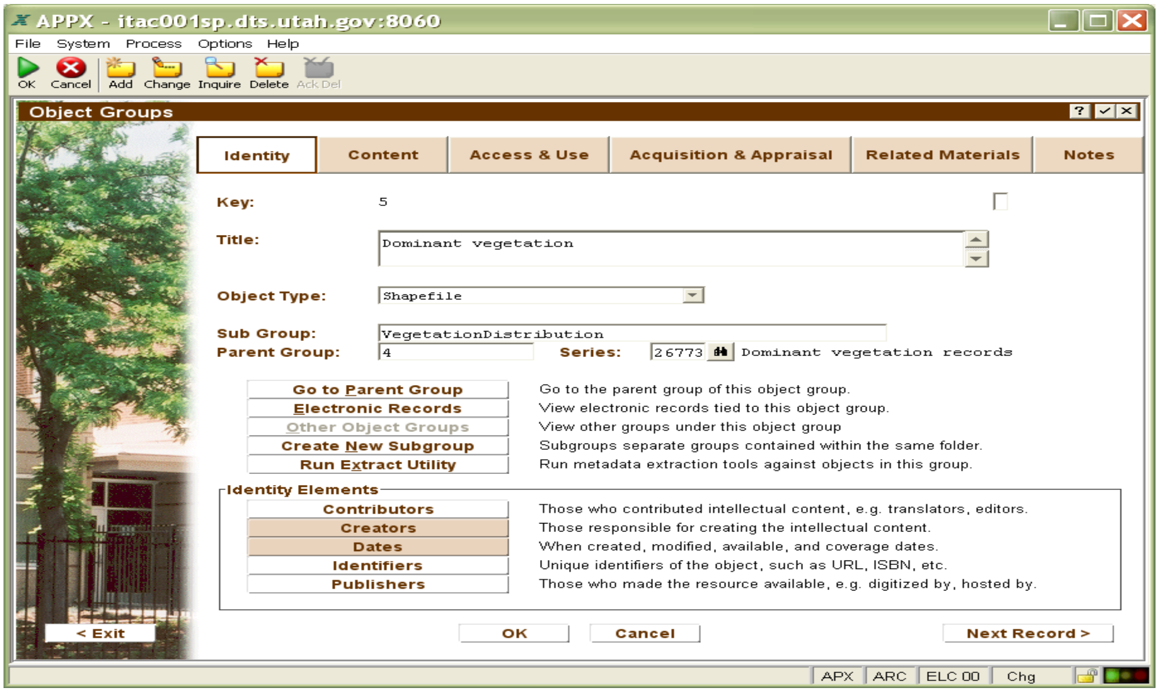
North Carolina said, "In general, I can see the potential for AXAEM as a preservation platform ... with the caveat that to pursue a real implementation would require more formally defining our organization's particular end-to-end workflow, and investment in the associated custom development/customizations/ configurations and refinement based on the particular of how NC State Archives would want to store and tag its geospatial data collection." Kentucky said, "While Kentucky's testing of the system was less involved than North Carolina, the value of the generic application was apparent enough to Kentucky to recommend it to archives and records management units, particularly ones with data repositories, to investigate and examine its adaptability to their environment."

While the open source version continues development, public release is expected in 1st Quarter 2012. The more feedback that the team received from testers, the more features have been added to improve usability. Many features will yet be added, even after initial release. Due to time constraints within the grant project, actual integration of this system with other states' existing systems was very limited. For a full report of the AXAEM project, see the white paper, "APPX-based Archives Enterprise Manager (AXAEM)."

Below are some sample screenshots displaying AXAEM functionality:



AXAEM: Menu for electronic records



Data entry screen for an object group representing a shapefile for dominant vegetation.

<u>Working Group Key Findings:</u>

- Traditional cataloging and access tools used by archival organizations can accommodate geospatial records, with various levels of success and ease of use.
- Traditional GIS tools, such as national inventories, can also be adopted by archival organizations to assist with access.
- Specific GIS tools such as the Geoportal Server may pose some challenges when it comes to installation, but can be useful for dataset discoverability.
- Storage of records is still problematic for many institutions. The perfect, affordable answer has not yet presented itself, but the options which are available have been discussed and documented, so that institutions may assess based on their own internal archival and access storage needs.
- So much could yet be done to integrate tools into the archival environment and make data management and public data discovery easier for those managing the repositories and for the public seeking access:
    - More geospatial metadata could be made available to archival systems and utilized in creative ways;
    - The development of spatial-based searching would be beneficial to encourage GIS users to take archival interfaces and historical datasets seriously, which may in turn encourage them to produce better metadata.

What's Next and What Still Needs to be Done

- Inventory and data discovery services with better support for superseded or historic materials are needed.
- It would be beneficial to have more access systems that utilize or can interface with true digital preservation systems (including OAIS compatibility, and support of hashing, multiple data copy, format upgrade functionality) that also can also take advantage of the spatial nature of the data being preserved with map interfaces and viewing tools, etc.
- Follow ongoing developments in storage technology that will address some of the critical data preservation needs (e.g. bit-level-integrity tracking and auto-repair/auto-recovery when corrupted files are discovered, replication for security copies, WORM or CAS functionality that enforces the immutability of the preserved file, audit and change management tracking functions that log and tracks changes to files).

Developing the Business Case to Support Geoarchiving

A significant outcome from the initial phase of the GeoMAPP project was establishing the value in developing a business case for geoarchiving. Through internal GeoMAPP team discussion and feedback from outreach activities, the importance of creating a business case document to support geoarchiving, along with the foundational activities required in developing the business case document was established as an important focus for the second portion of the project.

The Business Planning working group was tasked to engage an external contractor with the goal of jointly developing a suite of tools to provide comprehensive guidance in the development of a business plan, including:

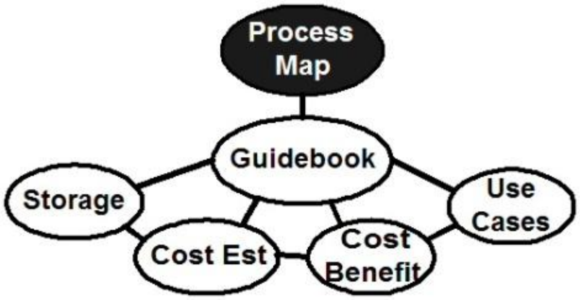
- Establishing a strong-multidisciplinary team of managers, practitioners and champions to develop the strongest business case possible;
- Articulating use cases for compelling and relevant business applications;
- Linking costs estimates to geoarchiving workflow processes and sources of geospatial data required for geoarchiving;
- Developing a documented, metrics-based estimate for project costs and benefits over a rational performance period;
- Calculating credible cost-benefit analysis reporting and return on investment metrics for decision makers based on the compiled documentation;
- Generating a business plan summary suitable for consideration by agency leadership and budget analysts.

The Business Planning working group engaged the contractor team of Applied Geographics and AECOM to support these efforts over a three and one-half month period. The GeoMAPP Business Planning workgroup and contractor team developed a suite of documents representing the *business planning toolkit*. The materials provided in the toolkit strive to find a balance in providing templates that can be adapted by users of the toolkit, while acknowledging toolkit users likely have existing templates and business planning templates and requirements in their respective states. The focus of individual documents is to provide a comprehensive understanding of the business planning process, and the points of emphasis that will develop the strongest final product.

Using the Toolkit

The *business planning toolkit*[70] is presented as a series of six interrelated documents. The Business Planning team recommends users thoroughly review and understand the contents of **all** documents prior to embarking on developing a business plan. In lieu of a singular linear document, the elements presented in the documents do share some overlap in principles and content due to the interrelated nature of the material. The filename for each of the toolkit documents is preceded by a numeric prefix to provide additional guidance in the recommended order of reviewing each of the documents.

Throughout the documentation, a reference graphic is used on the cover page or first section of each of the documents in the toolkit to orient each document in the overall process. The bubbles on the map reference the contents covered by specific documents in the toolkit.



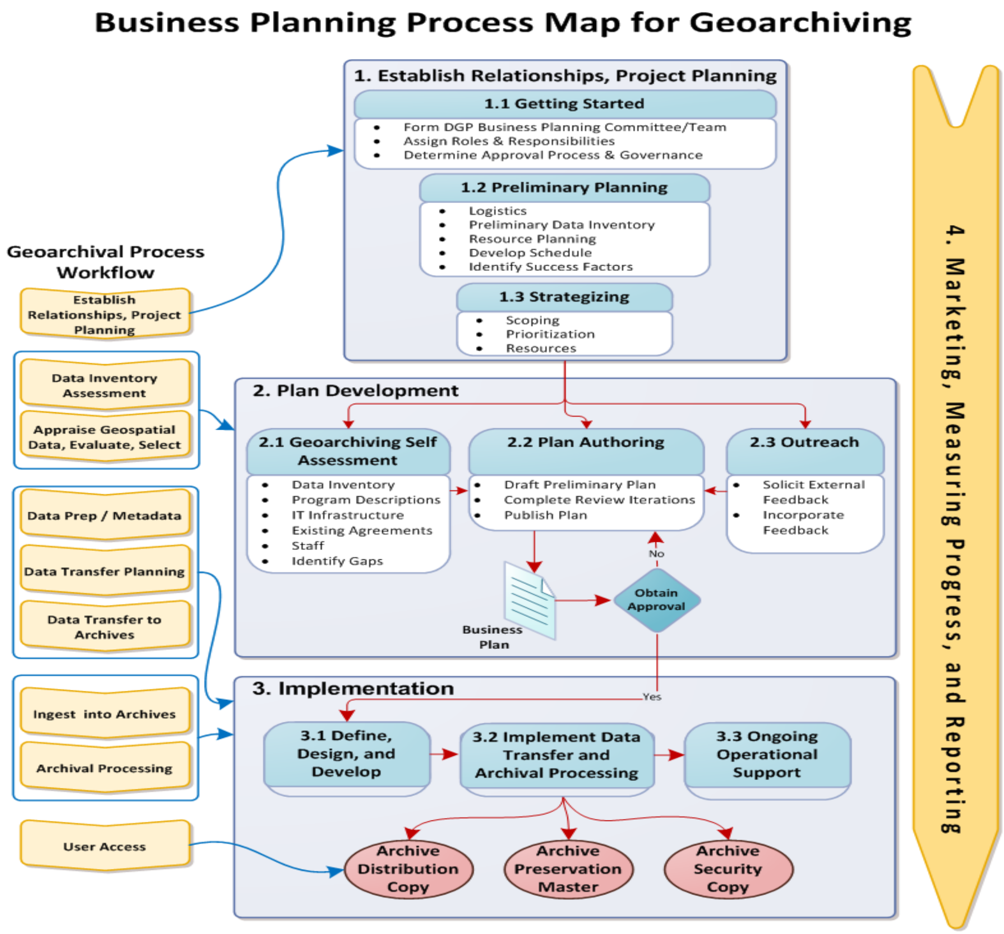
The documents in the toolkit include:

1. Geoarchiving Business Planning Process Map and Checklist[71]
2. Geoarchiving Business Planning Guidebook[72]
3. Geoarchiving Cost-Benefit Analysis Guidance[73]
4. Geoarchiving Cost-Benefit Analysis Tool[74]
5. Geoarchiving Use Case Guidance and Rationale Documentation[75]
6. Geoarchiving Business Planning Bibliography[76]

[70] Complete Business Planning Toolkit: http://www.geomapp.net/docs/00_Geoarchiving_Business_Toolkit_20111231.zip
[71] Business Planning Process Map: http://www.geomapp.net/docs/01_Geoarchiving_Business_Planning_Process_Map_and_Checklist_20111231.pdf
[72] Business Planning Guidebook: http://www.geomapp.net/docs/02_Geoarchiving_Business_Planning_Guidebook_20111231.pdf
[73] Cost Benefit Analysis Guidebook: http://www.geomapp.net/docs/03_Geoarchiving_Cost-Benefit_Analysis_Guidance_20111231.pdf
[74] Cost Benefit Analysis Tool: http://www.geomapp.net/docs/04_Geoarchiving_Cost-Benefit_Analysis_Tool_20111231.xlsx
[75] Use Case Guidance: http://www.geomapp.net/docs/05_Geoarchiving_Use_Case_Guidance _20111231.pdf
[76] Business Planning Bibliography: http://www.geomapp.net/docs/06_Geoarchiving_Business_Planning_Bibliography_20111231.pdf

Geoarchiving Business Planning Process Map and Checklist

The *Process Map and Checklist* provides a high-level overview of the entire toolkit and recommended process. The process map graphic (below) integrates the geoarchival process workflow steps in the left-most column along with the primary working sections to move from project planning, to business plan development and preparation for implementation. Each of the subcomponent sections within the process map and checklist includes recommendations for action and guidance for information to be captured as part of preparing to do the foundational work in documenting decisions and developing metrics in support of the geoarchiving business plan.



Geoarchiving Business Planning Guidebook

The *Guidebook* provides the execution-level details for developing a business plan for geoarchiving. The guidebook is broken into sections that provide the general outline for chapters to be developed in a successful geoarchiving business plan document. Each section includes an initial discussion on the relevant contents to be developed for the business plan chapter, the interrelations between sections, and a series of recommended discussion points that should be addressed in that chapter of the business plan. Most sections include an embedded bibliography for referencing more detailed information on the topics in that section for additional information.

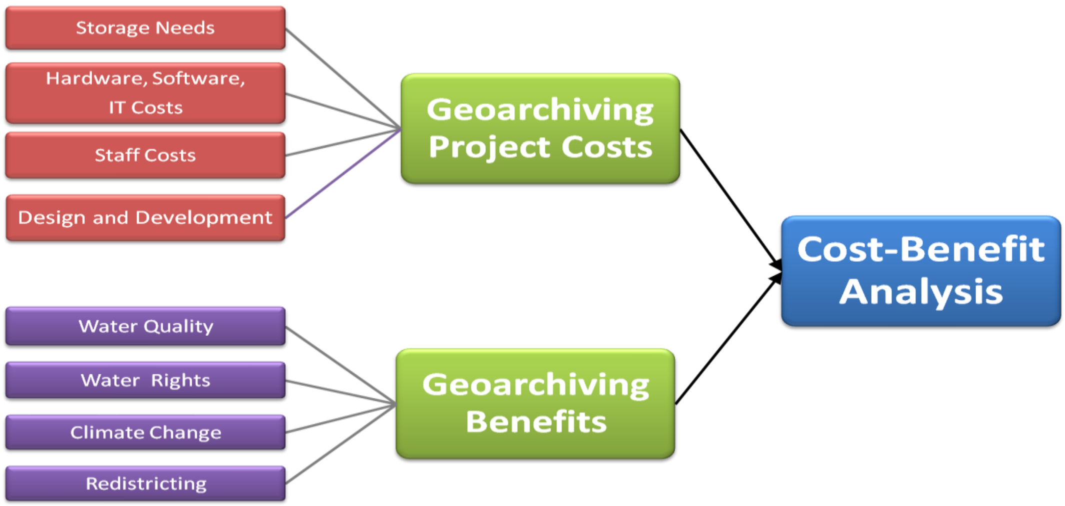
The table below encapsulates the section-by-section breakdown of the guidebook and the recommended context to be covered in each corresponding chapter of the geoarchiving business plan:

MAIN SECTION	ESSENCE
1. Executive Summary	What outcome(s) are you proposing to accomplish?
	Why do you need to do it?
2. Geoarchives Self-Assessment	What are the current conditions and assets
3. Customers and Stakeholders	Who is this for and who is making the case.
4. Program Goals	What are the specific 'Programmatic Goal(s)' for this Business Plan?
	For each goal, what are the 'Success Factors' (or supporting objectives)?
5. Benefits and Justification	What is the primary reason 'why' you need to do what you are proposing?
	What benefits and return on investment will accrue if it is done?
6. Requirements and Costs	What is your organizational approach?
	What are the estimated total costs of your proposal?
7. Implementation Overview	Phasing and milestones
	Budget Plan
8. Measuring Success and Feedback for Recalibration	Establish cost and benefit metrics and process for regular update/review

Geoarchiving Cost-Benefit Analysis Guidance and Geoarchiving Cost-Benefit Analysis Tool

The *Cost Benefit Analysis (CBA) Tool* is a collection of spreadsheets in workbook format that will support the toolkit user in developing comprehensive costs analysis, benefits summaries, and other fiscal metrics suitable for reference in the geoarchiving business plan. The worksheets provide a concise summary and are linked so that updates made within the tool flow to subsequent worksheets. For instance, storage estimates are linked to the page compiling project costs, as updates are made to storage estimates, these changes are linked and applied to overall project costs. Similarly, as changes that drive project costs are made, these changes are represented in the fiscal analysis sections for cost-benefit analysis and return on investment metrics.

The *CBA Guidance Document* is a companion to the *CBA AnalyisisTool*, and provides detailed guidance in using the *CBA Analysis Tool*. Each category of all of the included tabs in the *CBA Tool* spreadsheet are discussed in detail to provide toolkit users a comprehensive guidance of what is to be calculated and documented. As referenced earlier, many toolkit users may have internal tools that will be required for their respective organizations for documenting estimates related to costs and/or benefits.

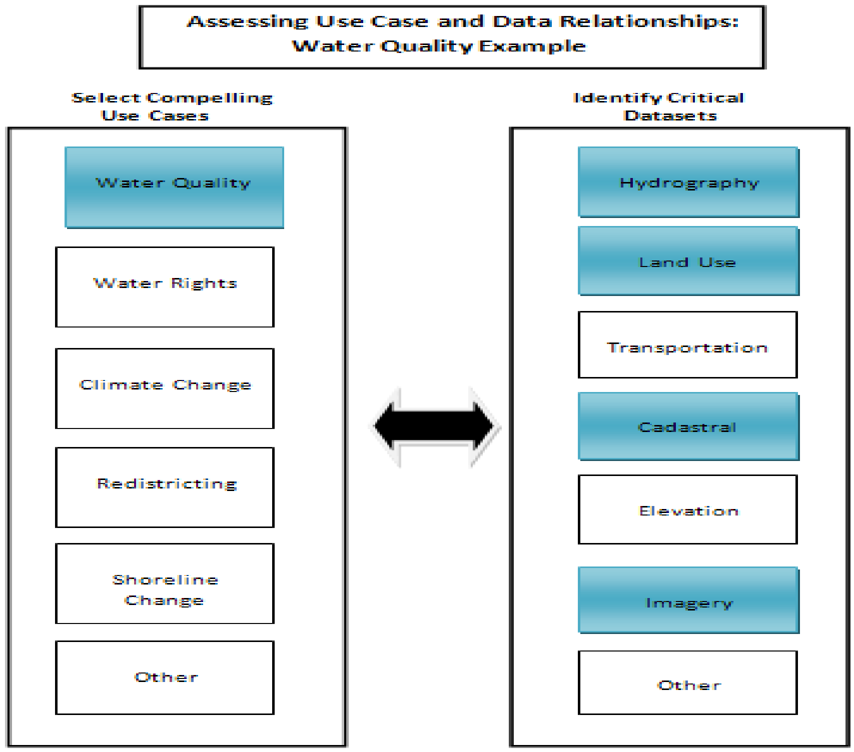


Within the toolkit reference graphic, the *Geoarchiving Cost-Benefit Analysis Guidance* and *Geoarchiving Cost-Benefit Analysis Tool* are referenced in multiple bubbles in the bottom row of the graphic.

<u>Geoarchiving Use Case Guidance and Rationale Documentation</u>

The *Use Case Guidance* document in the toolkit provides an overview and sample template for documenting the foundational use cases that will articulate the prioritized business needs that underpin and justify the need for establishing a geoarchiving business plan. Use cases can be used to tell the compelling stories of why it is important to preserve certain geospatial datasets and to help capture the value of this information. The document provides a strong, template based approach that can be used "as-is", while the guidance behind the use case can provide support and be adapted for organizations that already have a use case template within their respective organization.

Additional templates are provided for the development of documentation that captures rationale and details behind geoarchiving workflow processes and geospatial dataset details. The documentation of these details will be important for providing a framework for developing a comprehensive geoarchiving program over time. As use cases and datasets are added beyond the initial implementation, this collection of knowledge will provide a touchstone for providing performance measures against initially planned metrics for projects costs and benefits over time. Managing these performance measures into implementation is the most direct path to demonstrating project success, and expectations for growing the geoarchiving process through time.



Identifying Datasets Associated with a 'Water Quality' Use Case

<u>Geoarchiving Business Planning Bibliography</u>

The *Bibliography* includes references to a number of existing documents used by the GeoMAPP Business Planning working group and the contractor team in the development of the toolkit. At the time of developing the GeoMAPP final report, this represents a solid reference compendium of relevant documents in the area of business planning for geoarchiving.

<u>Key Findings:</u>

- **Understanding where you are and where you'd like to go--** Before conducting any formal business planning activities, organizations should conduct a rigorous self assessment to determine the current maturity of their geoarchiving workflow. This assessment needs to address the strengths of the implemented portions in a business oriented context.
- **Determining the value of the data**—Geospatial data contains valuable and critical information that can influence policy decisions of today and inform policy decisions of tomorrow. If information is overwritten or no longer exists, staff will have to take valuable time and resources to try to document and recreate data that once existed. GIS and archival staff should work together to identify the most critical and at risk data in order to determine begin calculating the value of it.
- **Metrics matter –** It is crucial for business planning users to develop quantitative metrics to estimate and demonstrate the benefits of geoarchiving. Decision makers evaluating the business plan are rarely subject matter experts in geoarchiving. Part of telling the compelling story of why

this data should be preserved is establishing relevant and approachable metrics for comprehensive costs and benefits. Tracking and reporting these metrics after implementing a geoarchiving solution is a tremendous opportunity for demonstrating success.

- **If you can't count it, it doesn't count** – There are numerous qualitative benefits underpinning digital preservation and by extension geoarchiving. Business planning toolkit users need to develop surrogate quantitative metrics to demonstrate these social goods. Capturing these benefits into the analysis is essential for establishing the return on investment figures. Cost avoidance in recreating historic datasets, and risk abatement in potential future penalties for not curating geospatial resources are two examples of these surrogate metrics
- **Assesses the costs**—A key ingredient in building a business plan is documenting the comprehensive costs for implementing a geoarchiving program. These costs include the start up costs for initial planning of a geoarchiving project as well as the long-term costs of maintaining a program including support labor, storage sizing and cost estimates and other technology costs associated with data prep, transfer, preservation and access.
- **Understand the audience of the business plan**—The audience of the final business plan document are agency heads, budget analysts, and legislators. The individuals of this audience, except in rare circumstances, are not subject area experts in geoarchiving. The business plan document needs to establish the business justification for geoarchiving in a strong metric-based context while understanding the funding support for geoarchiving justification will also be competing with other requests for government services that are also being presented in this business-oriented decision making environment.
- **Adapt business planning tools for other preservation scenarios–** The overall focus of the GeoMAPP project and partnership is geoarchiving, thus the business planning deliverable products is focused on preserving geospatial data. Geoarchiving can be a strong centerpiece for developing a comprehensive digital preservation initiative, but only in extreme cases will it be the singular focus. Toolkit users can adapt the principles within the toolkit to other relevant themes to achieve larger goals.

Next Steps

Business planning and establishing quantitative metrics to demonstrate the value of a geoarchiving program are still relatively novel ideas in both government geospatial and archives organizations whose missions are often focused on creating and providing access to data to benefit the common good or to meet statutory requirements, rather than focusing on business needs or requirements. However as state government budgets continue to be constrained and business justifications become a larger part of GIS and archives practitioners existences, easy to use tools to determine cost of ownership and business value will be needed to help support the viability of these agencies.

Since many of these tools were delivered late in the project timeline, GeoMAPP partners were not able to fully test the tools to develop unique business plans for each state. Additional review and testing of the tools, including state specific customizations and refinements, and the creation of additional example output materials from the tools could benefit the overall utility of the tool suite and could help other organizations interested in using the tools in getting started with the business planning process.

Broadcasting the GeoMAPP Message and Educating- Outreach and Mentoring

The Outreach and Mentoring team was very productive during the second phase of the GeoMAPP project. During the final phase of the project, the Outreach team increased their efforts at national and regional conferences, conducted industry outreach, established an active web and social media presence, and promoted an Informational Partner program. The Outreach group also conducted a series of onboarding activities for the new state partner, the Montana State Library. Team members presented at over 18 national or regional conferences, engaging both geospatial and archival professionals to raise awareness about the opportunities and challenges related to the preservation of and access to geospatial data.

Geoarchiving Self-Assessment

The GeoMAPP team developed a geoarchiving self assessment tool[77] to assist GIS practitioners and archival organizations in assessing their readiness for establishing a digital geoarchiving program. The assessment criteria are organized into a progressive series of readiness levels ranging from Basic Needs to Advanced GIS Archiving. Each level is further broken down into several evaluative categories:

- Plan Sponsorship and Project Governance;
- Current Programs;
- Human Resource Requirements;
- Data Requirements;
- Technology Requirements.

	A	B
1	***Directions**: Check the box next to the statement(s) that apply to your current situation. If the statement(s) do not apply skip to the next header.*	
2		
3	**1. PLAN SPONSORSHIP AND PROJECT GOVERNANCE**	
4		**1.1 Do you have an impetus for a geoarchiving program? Check ALL that apply**
5		a) By statute.
6		b) Directive (departmental management decision)
7		c) Staff-driven
8		d) Customer-driven
9		e) Is included in published agency mission statement
10		f) Process is informal -- no formal program exists
11		**1.2 Is there a commitment from both GIS and Archives management to a geoarchiving program?**
12		**1.3 Is there a project leader/champion for a geoarchiving program?**
13		**1.3.1 Which agency is going to lead the effort to implement/manage program?**
14		a) Archives
15		b) Geospatial agency
16		**1.4 Do you have the financial commitment to establish a geoarchiving program?**
17	**2. CURRENT PROGRAMS**	
18		**2.1 Do you have a formal process for ingesting digital information that includes verification and authentication?**
19		**2.2 Do you have an active and established workflow for intake and ingestion of data sets into STATE-LEVEL archives? (Check ALL below that apply)**
20		a) From local data creators (municipal governments/agencies)?
21		b) From state agencies?
22		c) From state-wide clearinghouse?
23		d) Other
24		**2.3 Are you currently archiving geospatial data within state?**
25		**2.3.1 Description of efforts**
26		a) Actively receiving and archiving geospatial data records from multiple sources (state agency, county, and/or municipal government)
27		b) Actively receiving and archiving geospatial data records from one or a few (less than 5) sources (state agency, county, and/or municipal government)
28		c) Geospatial data is included in agency records retention schedules
29		**2.3.2 Geo-centric records retention schedules with (Check ALL that apply):**
30		a) State agencies
31		b) Counties
32		c) Municipalities

Introduction | Basic needs | Intermediate GIS Archiving | Advanced GIS Archiving | Glossary

Snapshot of the GeoMAPP Self-Assessment Tool

[77] GeoMAPP. "Geoarchiving Self Assessment."
http://www.geomapp.net/docs/GeoMAPP_GeoArchiving_SelfAssessment_20100914.xls

Web Presence and Communications:

During the final phase of the project, the team revised the GeoMAPP website (http://www.geomapp.net) giving it a new look and organization. The web pages are structured to better reflect the content and to make information easier to find. Visitors to the site can find project publications based on the phases of the geospatial archival project life cycle. Presentations are listed by event but can also be accessed chronologically. The project team also made a commitment to keep content up to date on the site, ensuring that site visitors would have access to the most recent GeoMAPP related content. This more regular update cycle was supported by the working group approach of publishing periodic whitepapers as research elements had been completed.

In addition, the team established a social media presence on Facebook and Twitter. Team members took turns posting to these sites to announce events such as Informational Partners meetings and the addition of Montana as a new state partner, to announce presentations and conferences such as the Esri International Users Conference, and to announce the release of publications and tools such as *the Geospatial Data File Formats Reference Guide*.

Courtesy of Kentucky Department for Libraries and Archives (KDLA), the team began using Constant Contact, an e-mail marketing tool, to reach the GeoMAPP audience via monthly e-newsletters. In addition to sending the e-newsletters to established informational and interested partners, the e-newsletters were posted to the website so that visitors can access both the current and older issues.

For the second phase of the project, the team refreshed the brochure produced in the first part of the project. The updated brochure reflected the updated color scheme, information regarding the active Informational Partner group and meetings as well as the addition of Montana as a new full state partner. The team traveled to every event with copies of the brochure to be included on literature tables and to hand out to people to promote the GeoMAPP project.

GeoMAPP team members also participated in several national initiatives including the Federal Geographic Data Committee (FGDC) Users/Historical Data Working Group and the Library of Congress' National Digital Stewardship Alliance (NDSA) [78]. The team also engaged with staff from the National Archives and Records Administration regarding their efforts and share information with them.

Industry Outreach

The Outreach and Mentoring group helped to support the GeoMAPP efforts by engaging several companies in the corporate sector to raise awareness of the unique issues surrounding preserving geospatial data. The team hosted demonstrations, including:

- GeoCommons, an online map building tool and repository;
- PTFS' ArchivalWare[79], a data management and access tool;
- Tessella's Safety Deposit Box,[80] a digital preservation platform.

[78] NDSA: http://www.digitalpreservation.gov/ndsa/
[79] Archivalware: http://www.archivalware.net/
[80] Tessella: http://www.digital-preservation.com/

All of these industry outreach efforts led to a greater project awareness of the preservation tools that exist in the marketplace and also served to inform the vendors about the unique considerations for preserving geospatial datasets. Some of the tools demoed by the team could also benefit access to archived content.

As a result of the conversations held, several of the companies that GeoMAPP engaged with worked to structure their tools to better process geospatial data to aid discoverability. Tessella took the feedback provided to them by the GeoMAPP group and incorporated much of it into the production tool and established an ingest workflow for geospatial data. New functionality includes the automated extraction of metadata from the geospatial metadata file to populate the archival metadata file, management of establishing the fixity value for ingested files, handling multi-file geospatial datasets as a single archival record, and can generating and managing the various "manifestations" associated with a single archival record(e.g. SIP and AIP.)

The team also continued to informally engage with Esri. GeoMAPP team members met with both Esri senior management and technical staff at various events to discuss matters relating to metadata support, geodatabase file formats, Geoportal functionality and general concerns related to preservation. While there appears to be recognition of the preservation challenges facing data created by their tools and products, it is unclear how future product releases from Esri may benefit data preservation.

Several GeoMAPP team members also engaged in a multi-state effort to investigate the “cloud” for GIS data management. This effort, which is being led by Montana, Oregon, Utah and Colorado, began with a RFI that generated responses from IBM, Google, Amazon, Microsoft, Esri and Skygone. The RFI process culminated in a meeting of interested data IT representatives and vendors in Helena, Montana in June 2011. This meeting was attended by GeoMAPP representatives from Montana. Utah AGRC drafted a Request for Proposal (RFP) that was just recently issued. The GeoMAPP team drafted a document of “archival considerations” for the advisory committee to review as the RFP was drafted.

<u>New Partner Enablement</u>

During this phase of the project, the Montana State Library joined the team as a full partner. The project launched a national search via an RFP in December 2010 and the Montana State Library was selected and joined the partnership in February, 2011. Due to the shortened time frame, the Outreach group made a concerted effort to expedite “onboarding” activities and information sessions. Topics covered included an introduction to the project wiki, working groups, appraisal of geospatial content, data transfer best practices, and archival systems for ingesting and processing archived geospatial data.

<u>Informational Partners</u>

The GeoMAPP partnership established an Informational Partners group at the end of the first phase of the project. The Informational Partnership was a voluntary collection of state federal government archives and GIS professionals who met every other month to discuss topics related to the work the GeoMAPP team was doing. GeoMAPP sought this group’s input and perspectives on developing best practices guidance for the long-term stewardship of geospatial data. At the project’s conclusion, the Informational Partnership consisted of 25 members of the archives and GIS professions from thirteen states and the District of Columbia.

The Outreach team engaged other working groups in order to address topics of interest and concern to the Informational Partners. During the second phase of the project, the team held quarterly meetings with the Informational Partners and presented on Storage, File Formats, Appraisal of Geospatial Data, Processing of Archival Processing of Geospatial Datasets, and Project Lessons Learned. For a more details on the Informational Partner sessions, please consult the Informational Partners webpage on the GeoMAPP website[81]

Key Findings and Next Steps

Outreach and education are core components of a successful geoarchiving program. The GeoMAPP partners focused their outreach efforts on engaging members of the GIS and archives industries as well as conducting outreach to data creators and consumers. In doing so, they raised the level of awareness of how older data can inform future decisions as well as educate data produces on how small changes to naming conventions and data stewardship can help make data more "preservation ready" at the source. Data producers truly a care about the content they create. They are willing to make small, incremental changes to their data creation workflows and can be open to new ideas. Data producers are also a key advocacy group to ensure the success of a geoarchiving program going forward as they will likely be users of this system in addition to being contributors.

Many of the findings from GeoMAPP can also be applied on a national scale. The Outreach and Mentoring team spoke in different national venues with great success. The team continued to raise the awareness of the need to preserve geospatial data in both the national geospatial and archival communities and highlighted the consequences of not preserving this content.

The GeoMAPP partnership intends to continue to collaborate in the future. All four full partners as well as over half of the Informational partnership intend to continue to meet and hold discussions after the project's closure at the end of 2011, although no definite structure has been established for that collaboration as of yet. All of the members gained quite a lot from this experience and have expressed an interest to have a venue to continue these discussions into the future.

GeoMAPP Outreach Events 2010-2011

The Outreach and Mentoring group actively pursued outreach and education opportunities in 2010 and 2011. Team members pursued speaking engagements with both local and national organizations. Below is a list of presentations and speaking engagements. For a more detailed description, please consult the website at http://www.geomapp.com/presentations_chrono.htm.

2011

Montana State IT Conference, December 5, Helena MT

GIS Day 2011, November 16, Helena, MT, Lexington, KY, Raleigh, NC

Best Practices Exchange, October 20-22, Lexington KY

NSGIC Annual Conference, September 28, 2011, Boise, ID

FGDC Users/Historical Data Working Group, July 28, 2011

NDIIPP/NDSA Partners Meeting, July 19-21, 2011, Washington, DC

NAGARA/CoSA Joint Annual Meeting, July 13-16, 2011, Nashville, TN

Esri International User Conference, July 11-15, 2011, San Diego, CA

[81] GeoMAPP Informational Partners sessions: http://www.geomapp.net/infopartners.htm#present

New Partner Mentoring: Introduction to Archival Appraisal, June 2011

IS&T Archiving 2011, May 16-19, 2011, Salt Lake City, UT

North Carolina Legislature's GIS Day, May 18, 2011, Raleigh, NC

Nevada Geographic Information Society, May 9, 2011, Reno, NV

NAGARA E-records Forum, April 7, 2011, Austin, TX

NSGIC Midyear, February 27-Mar 2, 2011, Annapolis, MD

North Carolina GIS Conference, February 16-18, 2011, Raleigh, NC

NC State Government GIS Users Committee, February 3 2011, Raleigh, NC

Esri Federal User Conference, January 19-21 2011, Washington, D.C.

--

2010

Library of Congress Geosummit: GIS Data Appraisal, December 2010, Washington, DC

Best Practices Exchange 2010: Libraries and Archives in the Digital Era, October 2010, Phoenix, AZ

North Carolina Arc Users Group, September 2010. Carolina Beach, NC

Society of American Archivists, August 2010, Washington DC

NDIIPP Partners Meeting, July 2010, Washington DC

Esri International User Conference, July 2010, San Diego, CA

Open Geospatial Consortium, June 2010

NC State Government GIS Users Committee, May 2010, Raleigh, NC

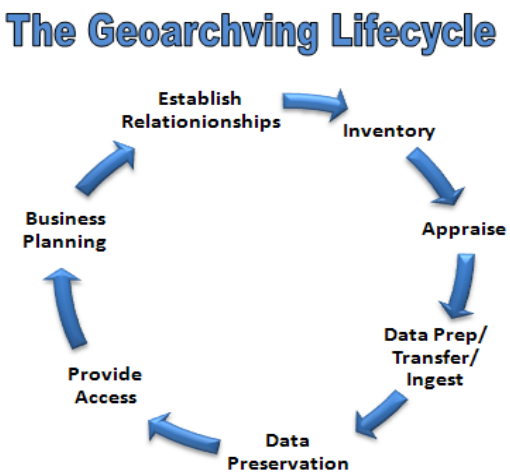
Charlotte, NC Metro GIS Users Group, May 2010, Charlotte, NC

American Society of Photogrammetry and Remote Sensing, April 2010, San Diego, CA

Esri Federal User Conference, February 2010, Washington, DC

GeoMAPP Key Findings and Best Practices

A key GeoMAPP effort has been to test and implement preservation workflows, document key findings and compile useful recommendations and observations that can be shared with state and local governments to assist with addressing geospatial preservation concerns. As the project evolved, the GeoMAPP team determined that these workflows could be subdivided into several unique and critical interrelated steps that would combine to encompass the geoarchiving lifecycle process.



This section distills the most important of the project's key findings and recommendations derived from the project's four year effort. These key takeaways are grouped by steps from the geoarchiving project lifecycle.

Establishing Key Relationships

- **Build a cross-functional team to address the geoarchiving challenge-** Due to the complexities of geospatial data and the unique processes required to preserve it, agencies seeking to preserve geospatial data need to establish a cross-functional team that includes representative practitioners and project champions from the archival, GIS and IT organizations in partnership to establish practices and policies for preservation.
- **Get to know your partners –***GeoMAPP's success is a result of the partners embracing new concepts and being actively engaged and committed to the mission*. GeoMAPP's partners forged relationships between their state's archival and GIS organizations and engaged in cross-training among partners on archival processes and GIS tools and technologies. All key staff participated jointly in outreach activities, held regular formal and informal meetings, and built familiarity with each discipline's standard terms, jargon, workflows and responsibilities. These interactions highlighted similarities and began the process to build a common understanding among the partners in each state. By understanding each other's language, responsibilities and goals, the state teams were able to develop and share best practices to educate constituent communities and tackle the challenge of preserving geospatial content together.

- **Leverage existing networks --** Leveraging existing geospatial/archival relationships is critical for developing a unified approach to preserving geospatial content. Each state partner reevaluated the relationships between the statewide GIS coordination bodies and the staff within state and with government staff responsible for managing records (i.e., chief records officers, clerks, etc.) for those state and local government agencies that produce GIS data. Tapping into these relationships can catalyze a strong interest in preserving geospatial content.
- **Engage data creators --** Conducting extensive face-to-face outreach efforts with state and local government GIS data creators and regional professional organizations can yield positive results and is a crucial element to develop a successful geoarchiving strategy. Sharing best practices about metadata creation, data formats and file naming, and data management techniques will help make geospatial data more preservation-ready at the source and reduce data preparation and archival processing efforts on the back end.
- **Influence change--** Active communication with GIS and archival policy-making bodies, as well as the vendors that create the tools and technologies used in the archives and GIS communities, can influence future product changes or policy decisions. Engage policy and coordination bodies to develop standards around metadata, data formats, and file naming that can benefit both current data utility and the geoarchiving process.
- **Use a phased approach--** It is best to take a modular approach to archiving geospatial data, starting with small steps and building the program over time. Trying to address all of the challenges of inventory, appraisal, outreach, system design, system implementation, data transfer, long-term management, and data access at one time may be overwhelming.
- **Conduct a pilot program first—** Test, validate, and sharpen your geoarchiving procedures using a small subset of data before starting full-scale production.

Justifying the Investment-- Developing a Geoarchiving Business Plan

- **Understanding where you are and where you'd like to go.**--Before conducting any formal business planning activities, organizations should conduct a rigorous self assessment to determine the current maturity of their geoarchiving workflow. This assessment needs to address the strengths of the implemented portions in a business oriented context.
- **Determining the value of the data—**Geospatial data contains valuable and critical information that can influence policy decisions of today and inform policy decisions of tomorrow. If information is overwritten or no longer exists, staff will have to take valuable time and resources to try to document and recreate data that once existed. GIS and archival staff should work together to identify the most critical and at risk data in order to determine begin calculating the value of it.
- **Metrics matter –** It is crucial for business planning users to develop quantitative metrics to estimate and demonstrate the benefits of geoarchiving. Decision makers evaluating the business plan are rarely subject matter experts in geoarchiving. Part of telling the compelling story of why this data should be preserved is establishing relevant and approachable metrics for comprehensive

costs and benefits. Tracking and reporting these metrics after implementing a geoarchiving solution is a tremendous opportunity for demonstrating success.

- **If you can't count it, it doesn't count** – There are numerous qualitative benefits underpinning digital preservation and by extension geoarchiving. Business planning toolkit users need to develop surrogate quantitative metrics to demonstrate these social goods. Capturing these benefits into the analysis is essential for establishing the return on investment figures. Cost avoidance in recreating historic datasets, and risk abatement in potential future penalties for not curating geospatial resources are two examples of these surrogate metrics
- **Assesses the costs**—A key ingredient in building a business plan is documenting the comprehensive costs for implementing a geoarchiving program. These costs include the start up costs for initial planning of a geoarchiving project as well as the long-term costs of maintaining a program including support labor, storage sizing and cost estimates and other technology costs associated with data prep, transfer, preservation and access.
- **Understand the audience of the Business Plan**—The audience of the final business plan document are agency heads, budget analysts, and legislators. The individuals of this audience, except in rare circumstances, are not subject area experts in geoarchiving. The business plan document needs to establish the business justification for geoarchiving in a strong metric-based context while understanding the funding support for geoarchiving justification will also be competing with other requests for government services that are also being presented in this business-oriented decision making environment.
- **Adapt business planning tools for other preservation scenarios–** The overall focus of the GeoMAPP project and partnership is geoarchiving, thus the business planning deliverable products is focused on preserving geospatial data. Geoarchiving can be a strong centerpiece for developing a comprehensive digital preservation initiative, but only in extreme cases will it be the singular focus. Toolkit users can adapt the principles within the toolkit to other relevant themes to achieve larger goals.

Documenting Processes and Data through Inventory and Appraisal

- **Survey the landscape --** While time-consuming to develop, manage and analyze, conducting surveys targeting GIS data producers as well as GIS and archival leadership and documenting the results can identify the current state of geospatial archiving and the preservation landscape within communities of interest. Surveys also perform critical outreach as they inform groups that the preservation of geospatial data is important and worthy of being investigated.
- **Establish GIS data as an archival record --** Geospatial data has enduring historical and evidential value and needs to be preserved and made available to future generations. Review records retention policies and public records laws to establish legal requirements for preservation.
- **Keep track of things with an inventory --** Having a method to identify and track preservation-eligible geospatial datasets is an essential first step in deciphering the preservation puzzle. An ideal inventory should contain basic information such as: the title of the dataset; the creation date and time period the dataset represents; who created it/owns it; where the dataset physically

resides; how often the dataset is updated; the dataset's size and format; and should ideally provide a method to categorize/organize the data based on a widely recognized standard and/or significant keywords. An inventory can be as simple as spreadsheet or a complex national inventory database such as the Ramona GIS Inventory

- **Establish a cross-functional GIS data appraisal team --** Engage internal state and local government archives staff to work with data producer organizations to investigate organizations' geospatial data holdings. Use these discussions as an opportunity to begin the conversation about which datasets need be preserved and integration of selected data into records retention schedules. These discussions not only open lines of communication and identify quantities and types of content, but also engage the primary participants who can most inform the preservation appraisal process. Establishing these relationships also helps to shape approaches to data ingest and management. Important things to capture during these conversations include: an identification of the business value of targeted datasets; the frequency of capture of targeted datasets; and the mechanisms by which datasets can be transferred to the archives.
- **Develop a repeatable process to appraise GIS data—** Due to limited storage and processing resources, most organizations are not able to preserve every item in their geospatial holdings. To aid in the selection of data, a geoarchiving program needs to develop a formalized, documented appraisal process to assess GIS data for their archival worthiness. Factors to consider include: legal/evidentiary, historical, or analytical value; the dataset's utility (including an identification of the primary data layers and geographic extents); frequency of update; as well as processing constraints (data format and sizing footprint relative to available capacity).
- **Create formal preservation policies for GIS data --** Developing geo-centric records retention schedules or other formalized collection or preservation policies is an effective way to ensure that geospatial data worthy of long-term preservation are retained and transferred to the archival organization. Retention schedules or other policy documents can help give data creators' guidance on identifying, organizing and transferring geospatial records once they've reached the end of their active lifecycle.

Data Preparation, Transfer and Validation

- **Get content 'in-motion'-** Designing and testing workflows to move geospatial content between agencies provides opportunities for both GIS and archival agencies to address data-related challenges and develop lasting data transfer and geoarchiving workflows.
- **Leverage existing workflows—**Developing a geoarchiving process doesn't require building new unique systems from the ground up. Take advantage of existing data management and access workflows and systems in place within organizations to assist in getting a geoarchives off the ground.
- **Require standards-compliant geospatial metadata--**A thoughtfully populated Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSGDM)-compliant metadata record contains rich material that is critical to understand the dataset for current use as well as for long-term preservation. A compliant metadata record must be included with any dataset that is to be shared or archived. Prior to any movement of data, a

decision must be made to determine the level of completion that is acceptable to both the GIS data consolidator and the archives.

- **Use logical and consistent file naming conventions**-- Assigning a logical file name to a geospatial dataset is important for the dataset's identification and management. Use a system that works best for your workflow, but include information about geographic extent/location, data theme, and creation date that are useful attributes worth capturing in the file name (e.g. SaltLakeParcels2006).
- **Be aware of GIS software and data format versions--**The type and version of the GIS software being used to view geospatial content does matter. Interoperability between different vendors is always a concern in the digital world. The differences between product releases can impact data preservation and access.
- **Establish data format standards--** Selection of a type of geospatial data format for preservation depends on the goals established for long-term preservation. A series of sustainability criteria should be identified and applied. These might include: format currency, openness and prevalence, community uptake, data portability, and the ease of data migration and data access.
- **Understand data packaging challenges**--Geospatial datasets are often comprised of a number of related files that must be present in order for the set to be complete and functioning. One fundamental challenge of archiving geospatial data is the lack of existing standards for packaging these disparate files in a consistent way. While each state had a unique approach for determining what information should be included in the archival package, GeoMAPP universally agreed that the most complete metadata available for the dataset must accompany the data.
- **Assess data storage requirements--**Knowing the overall size of your geospatial content, understanding your state's network infrastructure and capacity, determining how the data is to be transferred between entities (i.e., over the network versus external hard drives), and examining the pros and cons of different storage media and their costs are critical pieces of information when selecting data storage. It is essential to work with information technology staff from the beginning of the process to ensure the most effective outcome for data storage.
- **Establish a geoarchiving storage environment—**Preserving multiple copies of data helps to defend against data loss, and a tiered archival storage that supports multiple copies of data on discrete storage systems benefits unique preservation processes. Establish: (1) a staging area where ingested files can be reviewed and processed; (2) a preservation storage 'vault' where copies of data can be stored and managed untouched; and (3) access storage where a copy of the data and access derivatives can be stored for public access. Ideally all three environments should be on isolated storage systems and preservation and access data should be backed up.
- **Define data transfer validation mechanisms and process**: Before any content is transferred both the GIS and archival organizations need to establish procedures to review and validate data both before the data is transferred and after it is received by the archival organization. Key factors to consider include virus checking, bit-level verification (hashing), functional review of the dataset using GIS software, and metadata validation.

Archival Processing

- **Define an archival metadata model** and determine the sources that will populate the archival metadata record**.** Populate geo-centric archival metadata by developing crosswalks from appropriate FGDC CSDGM metadata fields that can be harvested to populate the archival metadata record, identify file properties that can be harvested to populate the metadata record, and identify how other fields will be populated (manually or external harvest).
- **Define the archival processing process -** Determine the actual tasks and tools involved in processing geospatial datasets to prepare the datasets for preservation storage. Steps should include: capture and record fixity value, create archival metadata, create security copy(ies) for preservation, data access preparation, create access derivatives, and deploy access copies. Automated tools will benefit the various activities, but human interaction is unavoidable. It is advisable to establish a database to manage the repository's process information and archival metadata, and develop a dataset file organization scheme.

Long-term Data Preservation and Access

- **Perform preservation actions against your data—**Datasets in your preservation environment should be tracked, managed and audited. Schedule and run periodic integrity checks on the data to ensure at least bit-level integrity. Consider strategies for dataset migrations and for refreshing media.
- **Provide access to your archived holdings--** A primary goal of long-term preservation is to provide appropriate long-term data access. Make data appropriately accessible to end users via the Internet, and search–enable geospatial collections so that users can locate the datasets that meet their needs. Register archived geospatial metadata with geospatial inventory systems to advertise the availability of archived, superseded geospatial datasets.
- **Develop finding aides and archival metadata--** Archives should develop full-context finding aids and/or archival metadata records to track accessioning and support data preservation. Both finding aids and archival metadata can provide a direct link and context to the data regardless of where and how it is stored, or what it is named.

Next Steps: the Future of Geoarchiving in the Wake of GeoMAPP

After four years of productive collaboration and exploration, the GeoMAPP partnership concluded at the end of December 2011. This section touches on external environmental issues that informed the GeoMAPP project, and considers activities that will influence for future investigators, including:

- What the technical geoarchiving landscape outside of the project looks like at GeoMAPP's closure;
- Other bodies that are currently concerned with the preservation of geospatial data that could be good sources for information and forums for discussion for those interested in geoarchiving in GeoMAPP's wake;
- What the post GeoMAPP world may look like for the project's four partners.

The State of the Geoarchiving Landscape Outside of GeoMAPP

During GeoMAPP's four year performance period (2007-11) many technological changes impacting geoarchiving have been observed in both the preservation and geospatial communities. This section looks at the 'state of the union' of the geoarchiving landscape outside of GeoMAPP's domain.

Exploring the Long-Term Preservation of Proprietary Geospatial Formats

GeoMAPP dedicated significant effort into the exploration of both proprietary and open geospatial formats in the *Geospatial Data File Formats Reference Guide* [82] assessing factors such as the format's openness (specification, standard), currency, prevalence, file composition, convertibility and other sustainability issues. Highly prevalent proprietary formats such as GeoPDF, MrSID, and several Esri formats may pose challenges to the preservation community due to their closed proprietary nature.

The File Geodatabase has come into wide use as a format for management and transfer of geospatial data and is now recommended by Esri as the native ArcGIS format for data that is stored in file form. While the format has some advantages, including support for the full information model of the Esri Geodatabase including maintaining relationships between stored feature classes and business data, there are also some disadvantages, including lack of an open specification and the use of binary files of unknown construction. The release of an API in June 2011 creates an opportunity to make data available in a more open manner, and OGC experiments involving of conversion of File Geodatabase information to a PostGIS open source database, via an abstraction layer built on the OGR Simple Feature Library open source data translation toolset, points the way towards possible new curatorial approaches.[83]

Improving Bulk Transfer Methods

A major challenge of geoarchiving lies in finding efficient means to reliably transfer large amounts of complex data between agencies. GeoMAPP encountered a variety of technical challenges in the course of intrastate and interstate data transfer tests carried out in the course of the project. The issue of bulk data

[82] GeoMAPP Data File Formats Guide: http://www.geomapp.net/docs/GeoMAPP_Geospatial_data_file_formats_FINAL_20110701.xls

[83] http://www.opengeospatial.org/standards/per

transfer was a key thread in the OGC Web Services Initiative, Phase 8 (OWS-8) test bed activity in 2011, a component of which was focused on advancing the state of geospatial data sharing and synchronization.[84] OWS-8 included an experiment involving data exchange using an approach called Geodata Bulk Transfer, which defines a container format that is ZIP compressed and contains feature data, schema, and topology as well as metadata.[85] A second experiment involved transfer of data via the File Geodatabase API. Further experiments in this area, and related work on an OGC Geosynchronization standard[86], may benefit the efforts of data custodians and archives to improve data exchange practice.

Capturing and Preserving Complex Data Representations

While the main target of geoarchiving efforts will continue to be the geospatial data itself, efforts will also be made to capture representations that are built on top of the data. These representations, which are created through a collection of choices and application of current methods with regard to data selection, symbolization, classification, data modeling and annotation, are in many ways analogous to the old paper map. The recent emergence of packaging schemes, such as LPK (Layer Package) files for packaging of data layers, and MPK (Map Package) files for packaging of finished GIS maps, provide a means to capture and exchange representations, but may also involve data transformations that make the packaged data less authentic than the original data. Geospatial PDF documents, including Terrago GeoPDF[87] documents, are another way to capture representation, although significant data conversion is involved. The potential utility as well as preservation impacts of these approaches will need to be evaluated and assessed over time.

Capturing Data Context in a Web Services Environment

As data users increasingly interact with geospatial data through web services or APIs, and as decision-makers increasingly make use of data in this way, it becomes increasingly important to find ways to capture these interactions in order to document decisions or capture data representations. The OGC has established an OWS Context Standards Working Group to work on an approach for defining the application state of a client that is utilizing services (i.e., service state), but it is not clear if or how the issue of data state will be addressed.[88] One interesting development is the creation of the MBTiles open source specification for storing tiled map data in SQLite databases for immediate use and transfer.[89] Although MBTiles was not developed to support persistent access to data, the specification is interesting in that it delves into the issue of capturing dynamic data such as map tiles in a storable and exchangeable form. The issue of capturing and documenting data state in a services-oriented environment will continue to require exploration.

84 http://portal.opengeospatial.org/files/?artifact_id=41689
85 http://www.opengeospatial.org/standards/per
86 http://www.opengeospatial.org/projects/groups/geosyncswg
87 Terrago: http://www.terragotech.com/
88 http://www.opengeospatial.org/projects/groups/owscontextswg
89 http://mapbox.com/mbtiles-spec/

The Challenges of Preserving Three-Dimensional Information

The convergence of geospatial data with three-dimensional information can be expected to create new opportunities and challenges. While effective integration of geospatial and 3D information has eluded the geospatial and CAD industry for decades, concerted efforts to create integrated infrastructure information are under way, and the rise in the use of mobile technologies and the emergence of augmented reality applications is creating new uses for integrated data. The OGC 3D Information Domain Working Group is working on facilitating the definition and development of interface and encoding standards, and cooperation has been established with organizations such as the National Institute for Building Sciences.[90] The emergence of new types of data products can be expected to raise new data archiving challenges.

The Future of Geospatial Metadata

With minimal support for the FGDC CSDGM metadata standard, Esri's release of ArcGIS version 10 signaled a warning shot about the long-term viability and support of this ubiquitous metadata standard. Driven by Esri's waning support for CSDGM metadata tools and the growing adoption and acceptance of the North American profile (NAP) of the ISO 19115, in December of 2011 FGDC began advising federal agencies to begin transition planning to migrate from the CSDGM to the ISO 19115 NAP. It is unclear what the timeline for metadata format transition will look like and how this change will be received in the federal and state GIS communities where the FGDC CSDGM has long been embraced and implemented.

Support Efforts to Address Geoarchiving Issues

The following bodies, organizations and groups of interest have an established interest in preservation and geoarchiving and may be good sources for information and discussion post GeoMAPP:

The Federal Geographic Data Committee(FGDC), Users/Historical Data Working Group[91]

This working group has been established under the auspices of the FGDC to coordinate activities among Federal agencies who are primarily users of geospatial data. In recognition of the fact that access to historical data is a prominent aspect of user needs, a major thrust of this working group is focused on facilitating the long-term retention, storage, preservation and accessibility of selected geospatial data. Some specific, relevant areas of work include: inventory of federal government data stewards and record officers and the materials they preserve; inventory of geospatial appraisal documents; and encouraging agencies to schedule disposition of data. The working group liaises with external groups such as the International Council for Science Committee on Data for Science and Technology Task Group on Data at Risk.

The Open Geospatial Consortium (OGC) Data Preservation Domain Working Group[92]

The purpose of this working group is to address technical and institutional challenges posed by data preservation, to interface with other OGC working groups that address technical areas that are affected by data preservation challenges, and to invite dialog with geospatial community and archival community

[90] http://www.opengeospatial.org/projects/groups/3dimwg
[91] FGDC UHDWG- http://www.fgdc.gov/participation/working-groups-subcommittees/hdwg
[92] OGC Preservation Domain WG- http://www.opengeospatial.org/projects/groups/preservdwg

constituents. Possible future areas of work include conceiving, designing, coordinating, and implementing demonstration and pilot projects that demonstrate technical approaches to data preservation within the context of OGC standards and relevant technologies emerging within the preservation and archiving communities. Such efforts might inform the development of OGC specification profiles and application schemas for archival purposes.

The Earth Science Information Partners (ESIP) Federation Preservation and Stewardship Cluster[93]

The ESIP Federation facilitates interactions among scientists, engineers, information technologists, and user communities with the goal of establishing and continuously improving the quality and value of Earth science products and services throughout their lifecycle. The Preservation and Stewardship Cluster provides a forum for ESIP members to collaborate on data preservation issues and works to support the long-term preservation of Earth system science data and information. Areas of work include: defining data stewardship principles and recommended practices, creation of data citation guidelines, test bed activities focused on implementation of persistent identifier schemes, development of a provenance and context standard, and definition of a preservation ontology.

European Spatial Data Research Network (EuroSDR) Data Archiving Working Group[94]

The European Spatial Data Research Network is a not-for-profit organization that links national mapping and cadastral agencies with research institutes and universities for the purpose of applied research in spatial data provision, management, and delivery. The Archiving Working Group brings together librarians, researchers and national mapping agencies to discuss digital archiving issues, and plans to publish a best practices guide by the end of 2011. This group works in close collaboration with the OGC Data Preservation Domain Working Group and will focus on the needs of European national mapping agencies.

National Digital Stewardship Alliance

The National Digital Stewardship Alliance is an initiative of the National Digital Information Infrastructure and Preservation Program (NDIIPP). It is a collaborative effort among government agencies, educational institutions, non-profit organizations and business entities to preserve a distributed national digital collection for the benefit of citizens now and in the future.

The objectives of the NDSA are to identify and recruit institutions with shared commitment to digital preservation and access, along with complementary content, technical, business, and legal expertise. The Library of Congress and Members of the NDSA are committed to serving as digital stewards of America's national digital collection and employing standards, systems, and cooperative relationships that advance digital stewardship. Members commit to pursuing and participating in collaborative stewardship endeavors for at-risk digital content. To continue the efforts of the NDIIPP-supported geospatial projects, a geospatial "action team" will be formed under the auspices of the NDSA "Content Working Group."[95]

---

[93] ESIP Federation- http://wiki.esipfed.org/index.php/Preservation_and_Stewardship

[94] Euro SDR- http://bono.hostireland.com/~eurosdr/start/index.php?option=com_frontpage&Itemid=1

[95] National Digital Stewardship Alliance Content Working Group - http://www.digitalpreservation.gov/ndsa/working_groups/content.html

Other Venues

Special interest group sessions are being held in connection with a range of geospatial industry events. For example, a Data Preservation Special Interest Group session has been held at each of the last two Esri International Users Conferences, and data archiving sessions have been held at the last two Cambridge Conference events, the quadrennial meeting of chief executives from national mapping organizations around the world. Such venues will continue to provide opportunities to engage larger industry audiences in discussion of data archiving issues.

Additionally, Columbia University's Center for International Earth Science Information Network (CIESIN), in partnership with NDIIPP has developed the Geospatial Data Preservation Resource Center web portal[96] to provide a consolidated access point to best practice information and technical documentation for geospatial data archiving.

What Comes Next for the GeoMAPP Partner States

December 31st 2011 marked the official end for the GeoMAPP project, its four years of technical research and collaboration and generous funding from the Library of Congress. In the wake of this groundbreaking effort, the project's four partner states will transition from funded project participation to an independent continuation of development and support of their geoarchiving systems. Each state will highlight geoarchiving as a critical component of their digital preservation efforts.

Despite the end of grant funding and a migration of project staff to other activities, the future of geoarchiving with GeoMAPP's partners looks promising. The following sections describe each state's anticipated next steps. In addition to pursuing these activities, the project partners committed to continuing to support the Informational Partnership and also holding regular "reunion" meetings for project team members after the project concludes at the end of 2011.

Kentucky

With the continued cooperation of DGI and records creators, KDLA will continue to archive geospatial records, work to expand the number of participants, particularly on the local level, and to refine the acquisition process. The greatest challenge will be implementing processes that address preservation issues, such as integrity checking and data migration. This will involve developing a comprehensive strategic preservation plan for all electronic record holdings with non-grant sources of funding. As the collection is used more, KDLA will refine collection practice and access methods to meet the needs of its customers through focus groups, face-to-face interactions and presentations at user group meetings. KDLA is a member of the NDSA, and will continue to support its preservation initiatives.

Montana

MSL has drafted system requirements for an initial release of an inventory system that will be developed in-house and should be released in winter 2012. It will be used to continue the ongoing process to inventory and archive the extensive MSL GIS data collection. Additionally, MSL staff will work with all MSDI theme stewards to develop archival plans for each theme.

---

[96] Geopreservation Portal: http://geopreservation.org

More research will also be conducted on ways to improve discovery and delivery of archived GIS data. An initial review of the Internet Archive's site[97] shows that a handful of shapefiles are available for download. MSL plans to test distribution of GIS data via this site in conjunction with the MSL State Publications program.

MSL also plans to modify the GIS technical specifications document to include required archival fields. This change will improve discoverability of archived GIS data through the Montana GIS Portal. Because the process to update this document will require input from the Montana Association of Geographic Information Professionals and the State of Montana GIS Managers Forum, it will provide a good opportunity for MSL staff to educate GIS professionals about the criticality of metadata to the archiving process.

Montana is currently exploring different funding models to fund GIS activities including ongoing development of the MSDI. An initial funding proposal includes an annual budget for a GIS archives program. This funding need will continue to be pursued as part of the larger funding discussion to be taken to the 2013 Montana Legislature.

North Carolina

The NC GeoMAPP team met with CGIA's Coordinator Program Manager to discuss the project and next steps after the grant ends. During the first phase of GeoMAPP, the NC team drafted retention schedules for both NC OneMap and for local governments. The team met to discuss the feasibility of going forward with the implementation of these schedules.

During phase two of GeoMAPP, the team transferred records (mainly vector) over three periods. To date, the State Archives holds slightly less than one terabyte (TB) worth of data. The State Archives is willing to move from a demonstration environment and into an active collection and preservation process for geospatial data. The NC team discussed the best path to proceed. At this writing, the NC Team began discussions about implementing the records retention and disposition schedule for NC OneMap as well as the records retention and disposition schedule written for local governments. CGIA assisted with final drafts of the schedules. The GICC's NC OneMap Governance Committee will consider and define next steps for archiving geospatial data in the context of a business plan for NC OneMap.

The NC team will also continue to participate in the NDSA. Team members serve on each of the committees from this group and will seek to engage practitioners and continue to examine the preservation of and access to geospatial data. NC project leads will also help lead future sessions with the GeoMAPP partners and determine a path forward with the Informational Partnership.

Utah

The MOU between AGRC and the Archives has not yet really been put into practice, although both groups still intend to adhere to the MOU. Next steps will likely involve staff training at AGRC so that everyone knows their role and how to produce the output needed for sending data to the Archives. As the SGID becomes more decentralized, other data creators will need similar training, such as at the Department of Natural Resources.

[97] Internet Archive: http://archive.org

Glossary of Archival and GIS Terms

Archival Information Package (AIP) - based on the OAIS Reference Model, an AIP is the digital equivalent of an archival item such as a book, a record album, or a motion picture. AIPS are used to transmit and/or store archival objects within a digital repository system. It consists of one or more data files that contain the digitized content of the archival item. In more sophisticated archival management systems, in addition to the data files, the AIP also contains the metadata that describes the structure, content, and meaning of the data files. The data files and metadata are packaged (encapsulated) either logically or physically as a single entity. (Source: Library of Congress. "Archival Information Package (AIP) Design Study - LC-DAVRS-07." 2001. http://www.loc.gov/rr/mopic/avprot/AIP-Study_v19.pdf

An AIP represents the digital content that is being stored and preserved, and may consist of a single logical or physical package representing the data and metadata that describes the structure, content, and meaning of the data files.

Archival Record - materials created or received by a person, family, or organization, public or private, in the conduct of their affairs that are preserved because of the enduring (or permanent) value contained in the information they contain or as evidence of the functions and responsibilities of their creator. Archival records may be in any format, including text on paper or in electronic formats, photographs, motion pictures, videos, sound recordings. (Society of American Archivists [SAA] Glossary)

Archival Value - the ongoing usefulness or significance of records, based on the administrative, legal, fiscal, evidential, or historical information they contain, justifying their continued preservation. In general, records with archival value are estimated to make up only three to five percent of an organization's records. (SAA Glossary)

Attribute Data (GIS) - generally defined as additional information about each spatial feature, usually arranged in a tabular format (e.g., the address and emergency services offered by each hospital in a hospital dataset, the area and perimeter of a county in a counties dataset, endangered species population counts in a wildlife dataset) .

Checksum Value - a computed value that is used to verify the integrity of a file or data, often in the case of ensuring that files are stored or transmitted without error. There are a variety of algorithms available to compute checksums. It is derived by calculating the binary values in a block of data, and storing that result with the data. When the data are retrieved or received at the other end of a transfer, a new checksum may be computed and checked to verify it matches the existing checksum. Also known as the ***Hash Value***.

Conversion - the archival management process of changing a digital object from one format to another, especially data files from a format becoming obsolete to a currently supported format. (SAA Glossary)

Content Standard for Digital Geospatial Metadata (CSDGM) - a content standard defined by the FGDC to provide a common set of terminology and definitions for the documentation of geospatial metadata. (FGDC. CSDGM Introduction. http://www.fgdc.gov/metadata/csdgm/introduction.html)

Digital Preservation - sustaining the validity, reliability, accessibility, and readability of digital content for the long term, also stated as the preservation of digital content, and not to be confused with reformatting physical content into digital form (which is more commonly referred to as "digitization"). (Source: Library of Congress. "Archival Information Package (AIP) Design Study - LC-DAVRS-07." 2001. http://www.loc.gov/rr/mopic/avprot/AIP-Study_v19.pdf)

The process of active management by which we ensure that a digital object will be accessible in the future. (Beagrie. "Digital Preservation Policies Study." 2008. http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf

Dissemination Information Package (DIP) - based on the OAIS Reference Model, the information package delivered to the end user, consisting of elements drawn and/or derived from the Archival Information Package, and reshaped to fit the end user's requirements and needs.

EAD (Encoded Archival Description) –standard used to mark up (encode) finding aids that reflects the hierarchical nature of archival collections and that provides a structure for describing the whole of a collection, as well as its components. (SAA Glossary)

Esri (Environmental Systems Research Institute) - developers and sellers of GIS software (such as Esri ArcGIS, EsriArcMap, EsriArcCatalog).

Feature - natural and man-made geographic features represented by points/symbols, lines, and areas on a map. Object in a geographic or spatial database with a distinct set of characteristics. For example, a road segment, manhole, building, or area designated having the same soil type. (http://www.kansasmappers.org/kam/services/gisdictionary.cfm#P)

FGDC (Federal Geographic Data Committee) – promotes sharing of the nation's geospatial resources. The FGDC metadata standard was adopted in 1994, and is a standard developed to determine the robustness, the method of accessing, and the successful transfer of geospatial data. (http://www.fgdc.gov/metadata)

File Integrity (or File Fixity) - a crucial long-term digital preservation consideration where the bit-level representation of a file or data block can be validated and confirmed over time to not have changed, thereby, confirming the file has not been altered through either human intervention, or technical action or decay. There are several methods available to record a fixity value for a digital object that can be periodically recomputed and checked to verify the file continues to be unaltered.

Finding Aid - a description of records that gives the repository physical and intellectual control over the materials and that assists users to gain access to and understand the materials. (SAA Glossary)

Framework Datasets – geospatial datasets deemed to be the most critical or commonly used for a wide variety of mapping and analytical purposes. The RAMONA GIS Inventory database delineated the following 23 datasets as being "framework": **Boundaries**- American Indian Reservation, Cities/Towns/Villages, Civil Township, Counties/Parishes, and State; **Elevation-** Bathymetric Contours, Contours, Digital Elevation Model (DEM), Digital Surface/ Terrain Models (DSM/DTM), and Spot Elevations; **Imagery/Base Maps/Earth Cover-** Digital Orthophotography/Orthoimagery, and Land Cover; **Inland Waters-** Hydrography, and Watershed Boundaries; **Location-** Address Points, Geodetic

Control Points, Geodetic Networks, and Geographic Place Names; **Planning/Cadastral-** Parcel/Cadastral/Land Ownership, and PLSS Townships & Sections; **Transportation-** Airports & Airfields, Railroad Lines, and Roads/Streets/Street Centerlines.

FTP (File Transfer Protocol) - is used to transfer files between computers on a network, such as the Internet. You can use FTP to exchange files between computer accounts, to transfer files between an account and a desktop computer, or to access software archives on the Internet. Keep in mind, however, that many FTP sites are heavily used and require several attempts before connecting.

Geoarchive – refers to a digital records repository designed to ingest and manage archived geospatial content.

Geodatabase - a common data storage and management framework containing a collection of geographic datasets for use by Esri's ArcGIS. There are various types of geographic datasets, including feature classes, attribute tables, raster datasets, network datasets, and topologies. (Tasha Wade & Shelly Sommer. "A-Z GIS." 2006.)

Geospatial Metadata – provides a way to describe geospatial data and other related records. It contains such information as the coordinate system, when the data was created, when it was last updated, who created it and how to contact them and definitions for any of the code attribute data.

GIS (Geographical Information Systems) – incorporates graphical features with tabular data in order to assess real-world problems (e.g., prioritizing sensitive species' habitats to determine optimal real estate locations for new businesses).

At the simplest level, the combination and visualization of GIS datasets and analysis can be thought of as a high-tech equivalent of a map. The key word to this technology is Geography – this usually means that the data (or at least some proportion of the data) is spatial, in other words, data that is in some way referenced to locations on the earth. Coupled with this data is usually tabular data known as attribute data. Attribute data is generally defined as additional information about each of the features, which then can be tied to spatial data *(GIS Lounge: http://gislounge.com/what-is-gis /).* Examples of attribute data might include the address and emergency services for a hospital; street address, property ID, appraised value, and property taxes for a real estate parcel; the land area and district number for a congressional district; the locations of fire hydrants; or the dimensions for the pipes for the public water system.

Ingest – The term associated with processes related to receiving information from an external source and preparing it for storage (SAA Glossary). After the SIP has been accepted, the AIP is prepared, the transfer of the AIP into the archival storage system, along with the associated archival metadata to ensure the long-term preservation and management of the data held in the AIP.

Information Package - based on the OAIS Reference Model, a digital content package transmitted between subsystems. There are three variants of the Information Package: 1) Submission Information Package (SIP), representing the data package submitted to the archival system for storage and preservation, 2) Archival Information Package (AIP), representing the data package holding the data and associated metadata to be preserved, and 3) Dissemination Information Package (DIP), the data package that will be distributed to an end user for access and use. (Source: Library of Congress. "Archival

Information Package (AIP) Design Study - LC-DAVRS-07." 2001. http://www.loc.gov/rr/mopic/avprot/AIP-Study_v19.pdf)

ISO (International Organization for Standardization) 19115 - defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. (*http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020*)

Layer - the visual representation of a geographic dataset in any digital map environment. Conceptually, a layer is a slice or stratum of the geographic reality in a particular area, and is more or less equivalent to a legend item on a paper map. On a road map, for example, roads, national parks, political boundaries, and rivers might be considered different layers. (ESRI Glossary: *http://support.Esri.com/*)

MARC (Machine Readable Cataloging) - data communications format that specifies a data structure for bibliographic description, authority, classification, community information, and holdings data. (SAA Glossary)

Metadata - data that describes other data or data objects. It may be descriptive in nature (e.g., title, subject, date created), it may be technical (e.g. file format, software used to create file), it may be structural (e.g. paper comprised of 500 pages, 10 chapters), or it may be administrative (e.g., file ingested into the archives on mm/dd/yyyy). The collection of metadata associated with an archived digital object will be used to manage and maintain it over its life in the archives, as well as facilitate its accessibility by end users.

Metadata Schema - A metadata schema defines a framework or specification for representing metadata. In general it includes definition of terms used in the schema, structural constraints and data structure definitions, and bindings to physical description syntax. Metadata schema descriptions are generally given in RDF (Resource Description Framework) Schema language. For additional information, see: http://www.ieee-tcdl.org/Bulletin/v3n1/nagamori/nagamori.html.

Migration - the archival management process of moving data from one information system or storage medium to another to ensure continued access to the information as the system or medium becomes obsolete or degrades over time. (SAA Glossary)

Network - a number of computers connected together to share information and hardware. A Local Area Network (LAN) is small, usually confined to a single building or group of buildings. A Wide Area Network (WAN) is a system of LAN's. It is large, with many computers linked.

Network Attached Storage (NAS) - a network attached storage (NAS) device is a specialized file server that connects to a local area network (LAN) and is dedicated solely to file sharing. A NAS device contains a slimmed-down operating system and a file system, and processes I/O requests for popular file sharing protocols, primarily CIFS for Windows and NFS for Unix. NAS does not provide any of the functionality that a typical application server typically provides, such as e-mail, authentication, or file management. NAS allows more hard disk storage space to be added to a network that includes application servers without shutting them down for maintenance or upgrades. Multiple NAS devices can be attached to a single LAN. (Webopedia: http://www.webopedia.com/TERM/N/network-attached_storage.html; PC Magazine Encyclopedia: http://www.pcmag.com/encyclopedia_term/0,2542,t%3DNAS&i%3D47631,00.asp)

Non-Framework Datasets – see Framework Data. Any geospatial datasets that are not included in the list of 23 Framework datasets listed above.

OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) – published by OAI, the protocol defines an application-independent interoperability framework based on metadata harvesting. The framework is used by data providers, who expose metadata about information held in a repository, and by service providers, who use that metadata to build value-added services. See http://www.openarchives.org/.

OAIS (Open Archival Information Systems) Reference Model - an International Organization for Standardization (ISO) reference model for an Open Archival Information System (OAIS). A high-level model that describes the components and processes necessary for a digital archives, including six distinct functional areas: ingest, archival storage, access, data management (for all of the (meta)data held in the archival system), preservation planning, and administration. (SAA Glossary, CCSDS. "Recommendation for an OAIS Reference Model." 2002. http://public.ccsds.org/publications/archive/650x0b1.PDF)

Orthoimagery – digital imagery in which distortion from the camera angle and topography have been removed, thus equalizing the distances represented on the image. A rectified copy of a photograph (typically an aerial photograph), showing image features corrected for variations in scale and height displacements. (*From http://www.websters-online-dictionary.org/or/orthophotography.html*) Aerial photographs that more precisely show the features of the landscape, including those that might be important for agriculture such as slope or size of gullies, because they are corrected for distortion caused by tilt, curvature, and ground relief.

Permanent Record – see Archival Record.

Projection - a system to portray all or part of the earth, which is an irregular sphere, on a planar or flat surface.

Raster Data – cell-based data such as aerial imagery and digital elevation models. Raster data is characterized by pixel values. Basically, a raster file is a giant table, where each pixel is assigned a specific value from 0 to 255. The meaning behind these values is specified by the user- they can represent elevations, temperatures, hydrography, etc. Satellite imagery uses raster data to record different wavelengths of light. Raster data is advantageous to vector data in constructing 3D images, as the values for every pixel are calculated through a process called *interpolation (http://www.umich.edu/~ipcaa/GIS/General%20GIS%20Concepts.htm).*

Record - data or information in a fixed form that is created or received in the course of individual or institutional activity and set aside (preserved) as evidence of that activity for future reference. A record has fixed content, structure, and context. (SAA Glossary)

Retention and Disposition Schedule - a document that identifies and describes an organization's records, usually at the series level, provides instructions for the disposition of records throughout their life cycle. (SAA Glossary)

Shapefile - an openly documented vector data storage format, originated by Esri, for storing the location, shape, and attributes of geographic features. A shapefile is stored in a set of related files and contains one feature class. (Tasha Wade & Shelly Sommer. "A-Z GIS." 2006.)

Spatial Data - also known as *geospatial data* or *geographic information* it is the data or information that identifies the geographic location of features and boundaries on Earth, such as natural or constructed features, oceans, and more. Spatial data is usually stored as coordinates and topology, and is data that can be mapped. Spatial data is often accessed, manipulated or analyzed through Geographic Information Systems (GIS). *http://www.webopedia.com/TERM/S/spatial_data.html.*

Spatial data = Spatial (Where) + Data (What)

Spatial Data Clearinghouse – repository structure, physical or virtual, that collects, stores, and disseminates information, metadata, and data. A clearinghouse provides widespread access to information and is generally thought of as reaching or existing outside organizational boundaries. (Wade, T. and Sommer, S. eds. *A to Z GIS)*

Storage Area Network (SAN) - is a high speed network for connecting and managing storage devices through a dedicated networking infrastructure. SAN is usually deployed over a Fibre Channel network to create a shared pool of storage that may include disk arrays, tape libraries, and optical jukeboxes.

SDE (Spatial Database Engine) - refers to ESRI's spatial database engine. It is a relational database management system that provides a formal structure for storing and managing information in tables. For additional information, see: http://www.Esri.com/software/arcgis/geodatabase/storage-in-an-rdbms.html.

Submission Information Package (SIP) - based on the OAIS Reference Model, an Information Package that is delivered by the Producer (Contributor) to the OAIS for use in the construction of one or more AIPs. (source: Consultative Committee for Space Data Systems. "Reference Model for an Open Archival Information System (OAIS)." 2002.)

Temporal – of or pertaining to time. As an example, the Dublin Core includes a "temporal" element that can be defined by a data, date range, or named period to provide the time period associated with the described object. A temporal GIS analysis may be an analysis of the series of a particular geospatial dataset over time, or the assessment of the impact of a collection of geospatial datasets' impact over time.

Vector Data –spatial data represented as points, lines or polygons. This system of recording features is based on the interaction between arcs and nodes, represented by *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. (*http://www.umich.edu/~ipcaa/GIS/General%20GIS%20Concepts.htm*).

Web Map Service (WMS) – is an Open Geospatial Consortium (OCG) Web service standard for exchanging map information as map images. WMS allows a user to request map images over the web using open standards. WMS supports the use of datasets without the need to keep a local copy. *(http://www.lib.virginia.edu/scholarslab/resources/class/mlbs)*