## GeoMAPP Key Findings and Best Practices

The Geospatial Multistate Archive and Preservation Partnership (GeoMAPP) paired state government GIS and digital preservation archival practitioners from the states of Kentucky, Montana, North Carolina and Utah with the Library of Congress in a four year (1997-2011) exploration into the preservation of "at-risk" geospatial data.

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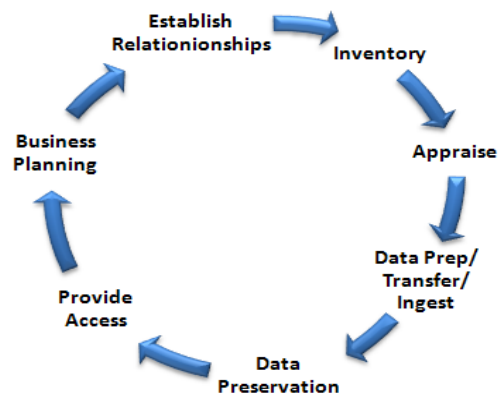
This section distills the most important of the project's key findings and recommendations derived from the project's four year effort. These key takeaways are grouped by steps from the geoarchiving project lifecycle.

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.