GEOARCHIVING USE CASE AND DATA COLLECTION TEMPLATE GUIDANCE

GENERAL USE CASE INTRODUCTION

Developing clear and compelling use cases for geoarchiving are an important component in making the business case for the importance of archiving geospatial data. Use cases provide an organized method to assess the value of a project or dataset in a way that will help a geoarchiving team identify financial or organizational benefits for preserving individual datasets or groups of data, and will help tell the compelling story of why these data need to be preserved. Use cases identify a goal-oriented set of interactions between actors and the geospatial components under consideration. Coupled with the cost analysis of the geoarchiving process presented elsewhere in the business planning toolkit, use cases can be a powerful tool to develop cost/benefit analysis.

Geospatial data produced by governmental entities may be subject to records retention and disposition schedules and/or other public laws, which may require that datasets be permanently saved. Users of geospatial data may also recognize the long term value of preserving data for future use and analysis. While legal retention requirements are an essential consideration for why a state or other institution needs a geoarchiving program, the use case guidance and data collection examples in this document concentrate to a greater degree on identifying the long-term *value* of historical geospatial datasets. The spatial-temporal dimensions associated with historical collections of geospatial data are important for geographically-oriented analyses, interpretation, and understanding of phenomena across time.

Historic data availability is a key dependency for this type of analysis. Analysis incorporating historic data enables a GIS to provide modeling and visualization tools for detecting trends, cycles, or other patterns and changes in the areas and features being analyzed. The results of such analysis can be used to understand current conditions or forecast future trends, which in turn may be used to inform policy setting, legislation authoring or business planning.

The development of use cases for geoarchiving will articulate important business needs including assessing organizational and monetary value for current and future geospatial data that are shared by stakeholder groups The compelling case for geoarchiving is best achieved when the need for it is presented in the context of actual organizational mission requirements, business processes, and scientific studies, strategically communicating the tangible value of historical geospatial data in real-life

user scenarios, thereby reinforcing the need for geoarchiving to ensure that historic data exists and is managed over a long period of time.

By articulating a stakeholder's business needs, use cases should, in large part, document the value of geoarchiving which would then be applied to a cost-benefit analysis. Examples of monetary benefits can be demonstrated through examples of cost savings, cost avoidance, or other benefits provided by the ongoing access to historical data, such as the following examples:

- **Cost Savings:** "We currently spend $100,000 per year searching for data" to defend against lawsuits on which we have inadequate recordkeeping to prove the state's contentions; with geoarchiving, "we could save $50K on what we spend" by going to one source for reliable records
- **Cost Avoidance:** Because snapshots were not retained of property parcel boundaries at different times for an area of the state, "$100,000 was spent to recreate missing records" from multiple disparate sources; we could have avoided this spending if the missing records were retained in a geoarchival system
- **Penalty Fees:** Because histories of hydrography data were retained for a fast developing area of the state, changes to stream channels and the construction of impoundments without permits were discovered that "allowed the state to assess $100,000 in fines" to help pay for restoration

Additional value characteristics of datasets that may warrant preservation include:

- Affect on and importance to multiple constituencies
- Significance over a long time span
- Document data points as they change over time
- Importance to governance
- Transparency in government and public rights
- Evidentiary assurance

As an example, if a Public Health Department was studying a particularly high rate of breast cancer in a specific geographic area, they would be interested in knowing the historic land use in the area.

Continuing with this example, perhaps the current use is residential, but historically, the epidemiologists discover that there were cranberry bogs in the area, and that pesticides were once used – including pesticides that contained DDT, which is especially persistent in the environment. Even though it has been banned for many years, the epidemiologists may be able to find trace amounts in the residential

area where the cancer rates are unusually high. This is not just a hypothetical scenario – it is the subject of multi-year studies in several parts of the country.

HOW TO USE THIS USE CASE GUIDANCE

The approach presented in this guidance document can be applied to help build thorough and persuasive business cases for geoarchiving and to assist in developing compelling stories for the overall business planning documentation.

It is important to note that the case for geoarchiving may be raised by any stakeholder at any time so there isn't an ordered workflow to develop a geoarchiving use case. For example, as in the case above, the user (Public Health) may approach GIS data producers and/or archivists about a need they have to maintain historic GIS data and thus the discussion is born. Conversely, the data producers (those who produce data regarding pesticide use) may inherently know that there is value in the data they produce and they make seek out users (Public Health and others) to help build the case for geoarchiving. Finally, archivists, charged with a statutory mandate to preserve geospatial data, may partner with both data producers and users to develop use cases to inform a geoarchiving program. In each scenario it is important to take a wide view of possible data users and all possible data that will enhance their business processes in order to make the most compelling cases for geoarchiving.

The examples provided in this document address a specific business need and then provide examples of the types of information that might be gathered to support geoarchiving under that particular scenario. Each area of the template provides question prompts in addition to actual input examples.

The following overall use case development recommendations are based on materials written by Alistair Cockburn in his book "**Writing Effective Use Cases**" (Addison-Wesley, 2001), modified based on experience:

- Name the scope of the use case
- Brainstorm and list the stakeholders and actors
- Use terse, active prose
- Include the main success criteria
- Address contingencies and preconditions

- Receive input from all the important stakeholders and actors, including government or regulatory entities and the general public; seek out the anticipated users of geospatial data that spans long periods of time to ensure that the anticipated users applications are properly capture
- Write a narrative story to learn the material before filling in the template (such as the Public Health example at the beginning of this guidance document)

This high-level guidance is reflected in the examples below.

Each organization may have its own approach to capturing use case information and for gathering benefits. The following examples provide a starter template and sample results that can be used and customized by organizations to identify compelling use cases for preserving geospatial data and to assess the monetary value of datasets to be preserved.

As a general workflow, organizations can begin their assessment by either:

A. Formulating a diverse list of relevant and compelling geospatial project or workflow use cases that require or could benefit from historic or superseded geospatial data

    i. Each use case should be analyzed and assessed to ascertain stakeholders and actors, required datasets, and access value

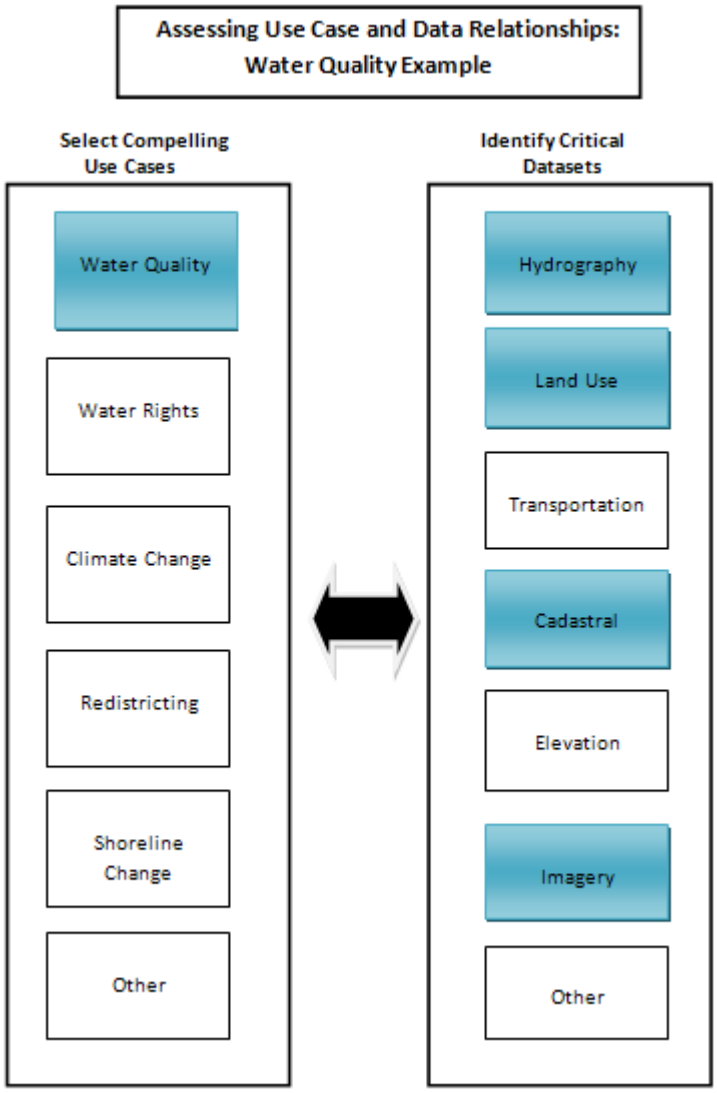
    ii. Each required dataset should be assessed for usage, attribution, and monetary value

OR

B. Creating a list of critical datasets that need to be preserved to determine a list of potential list of use cases that could take advantage of each dataset and then perform the use case assessment as listed above

Whether an organization starts its assessment with assessing potential use cases or by investigating individual datasets, actual ongoing data collection will likely be a hybrid approach as defining use cases may identify datasets that may not have been previously considered for preservation and while assessing datasets, other unexpected use cases may arise that may not have been initially identified.

The following graphic shows the relationship between the example use case of 'Water Quality' and its related datasets. The model scenario below pulls out 'Hydrography' to demonstrate an approach to documenting datasets as part of the use case development, even while other datasets (Land Use, Cadastral, etc.) would have an impact on the 'Water Quality' use case example, and any of them could be important to the 'Water Rights,' 'Climate Change' and/or 'Shoreline Change' use cases.



SAMPLE USE CASE SCENARIO FOR GEOARCHIVING

In this example, a data user identifies a business need that articulates the benefit to preserve hydrologic data. It is important to note that, as is demonstrated below, the geoarchiving team should also engage other possible stakeholders who might also benefit through access to preserved hydrologic data and, additionally, the team should identify additional datasets that would support the same use case to be included in a holistic approach to geoarchiving. Doing so will strengthen the overall business case for geoarchiving.

Each section provides a question or prompt for information and example responses are included as *blue italicized text.*

USE CASE NAME. What is the name of this use case? ***Supporting Water Quality***

USE CASE NUMBER. What is a unique number for tracking purposes?

Geoarchive_UseCase_WaterQual001

TECHNICAL USE CASE TRIGGER. What is the driver for the use case? *Water quality issues pursuant to the Clean Water Act (CWA) and specific manifestations of CWA, such as the EPA Phase II regulations and National Pollutant Discharge Elimination System (NPDES) permit requirements, which apply to Municipal Separate Storm Sewer Systems (MS4s) and anyone disturbing between 1 acre and less than 5 acres of soil.*

BRIEF DESCRIPTION. Provide a short explanation of the use case's background/purpose. *The regulatory requirement to map stormwater systems includes both above and below ground infrastructure. The goal of the mapping is to know where all storm water flows and where it discharges. Stormwater management also entails the ability to map and model land cover and land use characteristics to show the progression of cumulative effects of land use on hydrologic characteristics, including impervious surface, runoff, and contamination. The ability to address this regulatory requirement is dependent on data which provides knowledge of historic storm water flows and their possible shifting locations over long timeframes.*

STAKEHOLDERS AND INTERESTS OR GOALS (TWO SETS):

APPLICATION STAKEHOLDERS	
Application Stakeholders	**Interests or Goals**
Who are the use case's stakeholders?	Explain the Group's interest or relationship with the use case
Inter-local or Metropolitan Stormwater Group	*Regionally responsible group for guiding implementation of water quality regulations at the municipal level and ensuring best practices*
Municipality	*Locally responsible for implementing the regulations*
Authorized State Authority	*The state's Department of Environmental Protection or equivalent, with delegated responsibility for implementing the federal regulations*
Subject Matter Experts	*For example, county soil and conservation districts; departments of public health; others as needed*
EPA (Regional Office)	*Federal regulator*
Municipal Stormwater (MS4) Authority	*Manages fee collection to help build, operate, and maintain storm sewers via equitable fee assessment*
Local Water Districts	*Concerned about water quality protection issues*

APPLICATION STAKEHOLDERS	
Engineering and Environmental Consultants	*Calculate impervious surface areas, write stormwater management plans, recommend water quality solutions*
State Department of Transportation	*Responsible for stormwater management of their storm sewers statewide*

GEOARCHIVING STAKEHOLDERS	
Geoarchiving Stakeholders	**Interests or Goals**
Who are the geoarchiving process' stakeholders	What is their role in the geoarchiving process for this use case
State GIS Coordinator	*Geospatial information assurance and future access and applications*
State Archivist	*Geospatial records retention and preservation and future access*
Departmental GIS Manager(s)	*Continuity of operations and redundancy*
IT representative	*Provisioning hardware, software, network devices*
GIS Data Users	*Access to both current and historic GIS data*

PROBLEM SCENARIO(S). Describe the needs/usages of temporal data within the use case. *Temporal and spatial land use patterns affect water quality. For example, urbanization increases the area of impervious surfaces (e.g. pavement, roofs), thereby decreasing absorption of rainfall, and increasing runoff during storms. Increased runoff during big rain events increases peak discharges and flooding. It also increases contamination of water resources due to increase runoff of pesticides, herbicides, heavy metals, and volatile organic compounds; these hazardous substances can also come from agricultural, rural, and suburban areas. Some contaminants are especially persistent in the environment, such as DDT. Even though it has been banned for many years, trace amounts can be found in areas where it was used in the past. Finding these areas requires knowledge of historic land use and activities.*

SPONSOR. Who will pay for the necessary geoarchiving action to support this activity? *Preservation of these datasets will be covered under the state geoarchiving program sponsored by State Archives and the State GIS Clearinghouse. The State Transportation Department has requested the preservation of these datasets and will participate in the appraisal discussion.*

CANDIDACY FOR PRESERVATION: What are the characteristics of this business process that call out for geoarchiving?

- ★ *Affect on and importance to multiple people*

- *Significance over a long time span*
- *Importance to governance*
- *Transparency in government and public rights*

MAIN ACTORS. The following *human* actors perform mapping and assessment of impervious areas and water flow for managing storm water with a variety of data sets, including historical data:

- *Municipal or Water District GIS technician*
- *Municipal or Water District CAD technician*
- *Tax Assessor's Office*
- *Engineering consultant*
- *Users of historical data sets*

The following system actors are used to perform mapping and assessment of impervious areas and water flow for managing storm water with a variety of datasets:

- *Local GIS systems and current data*
- *State GIS systems and current data*
- *State Archives systems and historical data*

DATA DEPENDENCIES. The following data is needed to perform mapping and assessment of impervious areas and water flow for managing storm water with a variety of data sets.

For the "Historical Datasets there are many key questions that need to be asked:

Is there appropriate data available to support this activity?
How regularly has historic data been captured?
How frequently has older data been updated or overwritten?
Is there enough usable historic data to support this activity?
What are the results of a detailed inventory of historic data to support this action?

The results of these questions will provide primary inputs into supporting a geoarchiving use case.

Example:

Dataset	**Purpose in Use Case**	**Authoritative Data Source**	**Frequency of Update**	**Available Snapshots**
List datasets utilized by the use case	Describe the dataset's role in the use case	Which agency is responsible for this dataset	How often is the dataset updated or recreated?	Names/dates of available current/historic snapshots of the dataset
Orthoimagery	*Needed to observe changes in land use due to run-off*	*State GIS Clearinghouse*	*~Every 4 years*	*Statewide 6in pixel orthos : 2006, 2010 Statewide 1ft pixel orthos: 1995, 2000*
Land Use	*Needed as basis for analyzing*	*State*	*~Every 5 years*	*Statewide Land Use:*

	changes in land-use over time	*Environmental Department*		*2000, 2005, 2010*
Hydrography	*Used for locating stream flow networks and determining change of networks over time*	*State GIS Clearinghouse*	*~10 years*	*Statewide 1:24k NHD: 2001, 2011*

Some other relevant data sets that could be included in water quality and hydrologic applications are steam gauges, rainfall information, political boundaries, elevation, watersheds, transportation networks, and demographic data.

DATA QUALITY

Each snapshot of data needs to be assessed for:

- Accuracy
- Integrity
- Timeliness
- Consistency
- Completeness
- Validity

USE CASE PRE-CONDITIONS/PREREQUISITES

- ***Historic imagery with sufficient resolution*** *- Municipalities support sewer-engineering activities with available State imagery when it has sufficient resolution or currency. If the imagery is deemed to be too coarse or unavailable individual towns might need to re-purchase imagery which they once had possession of, resulting in duplicated costs.*
- ***Access to imagery and other geospatial data*** *- Actors will typically download local copies of data depicting areas they use and access on a regular, frequent basis. Is historic data available for download (or other modes of accessibility), and can it be authenticated?*
- ***Historic imagery and land use data*** *– Needed to assess impacts of historical land uses on pollution runoff – may need to be retrieved from State Archives*

CONTINGENCY SCENARIO(S). What are the consequences if required data were not archived and are no longer available for temporal and spatial analysis? *Historic snapshots of hydrological data are required for conducting this temporal analysis. If no data existed, then stream networks could be extracted from archived DEMs or by manually digitizing paper stream maps. Both options would be very costly.*

MAIN SUCCESS FACTORS AND BENEFITS.

Attempts should be made to articulate a benefit (as defined above) for the success factors described. These benefits would be used as inputs to the cost/benefit template provided elsewhere in the business planning toolkit.

- *Improved stormwater management and the attainment of regulators' goals*
- *Cost effective maintenance, operating and construction costs for engineered solutions*
- *Cost avoidance of need to recreate or repurchase historic geospatial data*
- *Cost avoidance of potential legal challenges*
- *Availability of accurate and complete geospatial data record over time*
- *Cost savings in reduced number of trips to perform field assessment*
- *Successful assessment of impervious areas within a jurisdiction*
- *Perception of fairness of equitable assessments*
- *Reduced appeals from reliable and defensible assessments*
- *Current and historic data to support analysis*
- *Improved water quality*
- *Wise development (smart growth)*

ESTIMATED VALUE OF ARCHIVED DATA, What is an approximate dollar value of the archived datasets for this use case (as a portion of revenue derived from the use case from individual dataset valuations)

SAMPLE DATA DEPENDENCIES EXAMPLE

In this example, a data producer inherently understands the value of the hydrography data produced. By answering a series of questions about the dataset, the geoarchiving team should be able to identify numerous data users who have business needs for the data and would, therefore, be candidates for the creation of use cases. Again, the geoarchiving team should identify as many stakeholders as possible and should consider complimentary datasets to include in holistic use cases.

In addition to providing the basis for use cases, this data dependency analysis may be useful to document for posterity why importance was given to particular datasets in the geoarchiving process.

Each section provides a question or prompt for information and example responses are included as *blue italicized text.*

DATA SET NAME: What is the name of this dataset? ***Hydrography Data***

(Note: The specific information in this example is based on MassGIS data set and associated metadata documentation, with the inclusion of additional content pertinent to geoarchiving that is not specific to Massachusetts. In addition to the value for preparing use cases, detailed information about the datasets will need to be gathered as part of the long-term archiving process. This example demonstrates some of that detail, which may not be applicable in all instances.)

DATA SET CLASS. Which ISO 19115 Topic Category does the dataset belong to? *Inland Waters*

DATA SET NUMBER. What is a unique number for tracking purposes,? *Geoarchive_Data_InlandWaters_Hydro001*

GEOARCHIVING SOURCE. Who is the data distributor? *State GIS Clearinghouse*

ORIGINATOR. Who is the data originator? *State Environmental Department*

STEWARD. Who is the data steward, responsible for maintenance and updates? Who has long-term responsibility for the data? *State Environmental Department*

DATES. What are the dates of data collection in the field, the data set creation and the last update? *For 2010 copy: field collection 2008-9, Dataset crated in 2009, Attribute updates in 2010*

FREQUENCY OF UPDATES. How frequent are updates performed and published? *~Every 10 years*

DATA SET DEFINITION. Provide some background about the dataset. *The MassDEP Hydrography layer is a hybrid of data based on US Geological Survey (USGS) 1:25,000 Hydrography layer enhanced with the MassDEP Wetlands layer and field verification for many areas. It represents hydrographic (water-related) features, including surface water (lakes, ponds, and reservoirs), wetlands, bogs, flats, rivers, streams, and inland water features.*

GENERAL FUNCTION. Describe the purpose of the dataset. *This layer is intended as an interim product that will be incorporated into the USGS's National Hydrography Dataset (NHD). It is useful for statewide applications and for where local data is not available. It is the authoritative small-scale (1:25,000) hydrography data for the Commonwealth of Massachusetts, representing the geographic location of the state's inland water resources.*

SPECIFIC USES. List of applications or use cases for this dataset.

Water Rights; Planning; Outdoor Recreation; Flood Risk Assessment; Topographic Mapping.

COST. What was the cost of creating this data set, and how much has been spent on maintaining it?

$500k initial investment in dataset creation. ~$20k per year for dataset maintenance/ updates.

PUBLIC ACCESSIBILITY. Is this data accessible to the public? If so, where is it accessible? *Yes, dataset is publically available on the state GIS clearinghouse*

SECURITY REQUIREMENTS. Is this data set or any of its subsets restricted in any way? If so, what are the restrictions? *No restrictions*

CANDIDACY FOR PRESERVATION: What are the characteristics of this data theme that call-out for geoarchiving?

- *Affect on and importance to multiple people*
- *Significance over a long time span*
- *Importance to governance*
- *Transparency in government and public rights*

RELATED DATA SETS. List other related datasets. *From ISO Inland Waters: 100-Year Floodplains; 500-Year Floodplains; Base Flood Elevations; Boat Pump-out Locations; Channel Cross Sections; Coastal Barrier Resource Areas; Dam Inventory; Ditches & Drainage Structures; Docks; Dredged Material Placement; Field Drainage Tiles; Flood Control Structures; Flood Insurance Rate Map; Flood Zones; Flow Rates; Hurricane/Tsunami Inundation Areas; Navigable Waters; Navigation Aids; Navigation Channels; Piers; Port Facilities; Protected Areas; Public Access Points; Public Beaches; Recreation Facilities/Location; Riparian Zone; Sediment Types; Shorelines; Springs; Submerged Grasses; Suspended Sediments; Swamps; Water Management Plans; Water Recreation Areas; Wetlands & Deepwater Habitat*

DATA SET FORMAT/CONTENTS (ARC AND POLYGON FEATURE CLASSES). List the file format of the dataset and available products. *Available statewide, the hydrography data is stored in ArcSDE as HYDRO25K_POLY and HYDRO25K_ARC. The USGS Digital Line Graph (DLG) quadrangles were converted into Arc/INFO coverages and projected into the Massachusetts State Plane Coordinate System.*

METADATA COMPLIANCE AND AVAILABILITY. How good is the metadata, and is it readily available? What is needed to bring the metadata up to geoarchiving standards? *Metadata is FGDC compliant, but needs additional details in the attribute section.*

ATTRIBUTE INFORMATION. Describe the attributes for this dataset. *This layer contains both a polygon and arc feature class. The modified DLG coding scheme is extensive and includes a wide variety of features, including ponds, cranberry bogs, impoundments, wetlands, tidal flats, dams, streams, and aqueducts.*

- *Only data from the DLGs have been coded this completely. Data from other sources have been coded to include ponds and streams and in the case of data from the scanned quads, wetlands.*
- *Pond and Lake Identification System (PALIS) IDs are unique codes which were added to ponds and lakes by DEP GIS in conjunction with the DEP Division of Watershed Management using identification codes developed by the Pond and Lakes Information System.*
- *For historical reasons, some wetland polygons have PALIS IDs.*
- *PALIS IDs were also given to impoundment areas along rivers and when necessary closure lines were added.*
- *Since the data now comes from different sources, the attribute SOURCE was added to differentiate which program/entity provided the feature.*
- *An attribute for the approximate NHD resolution was added to aid in future input into NHD.*

The items in the polygon attribute table are:

MINOR_TOT	*Text (15)*	*Concatenated feature code*
POLY_CODE	*Number (10)*	*Generalized code based on MINOR_TOT simplified to these 9 codes:* ***0*** *- LAND/ISLAND/DAM/AQUEDUCT* ***1*** *- RESERVOIR (with PWSID)* ***2*** *- WETLAND, MARSH, SWAMP, BOG* ***3*** *- SUBMERGED WETLANDS* ***4*** *- CRANBERRY BOG* ***5*** *- SALT WETLANDS* ***6*** *- LAKE, POND, WIDE RIVER, IMPOUNDMENT* ***7*** *- TIDAL FLATS, SHOALS* ***8*** *- BAY, OCEAN* ***9*** *- INUNDATED AREA*
PWSID	*Text (11)*	*DEP public water supply identification number*
PALIS_ID	*Number (6)*	*A unique ID from the Ponds and Lakes Information System*
SOURCE	*Text (12)*	*Program Source for the feature:* ***USGS/MGIS*** *– Original 1:25,000 Hydrography* ***DEPGIS*** *– Added from DEP Wetlands by DEP GIS staff* ***DCRGIS*** *– From DEP Wetlands and edits/verified by DCR staff*
MINOR_NUM	*Number (15)*	*Same as MINOR_TOT, as integer*

RESOLUTION	*Text (12)*	*NHD corresponding resolution:* ***HIGH*** *- Nominally > 1:25,000* ***LOCAL*** *- < 1:25,000, nominally 1:12,000, some at 1:5,000*

The items in the arc attribute table are:

MINOR_TOT	*Text (12)*	*Concatenated feature code*
ARC_CODE	*Number (10)*	*Generalized code based on MINOR_TOT simplified to these codes:* ***1*** *- SHORELINE* ***2*** *- CLOSURE LINE* ***3*** *- APPARENT WETLAND LIMIT* ***4*** *- STREAM* ***5*** *- INTERMITTENT STREAM* ***6*** *- DITCH, CANAL* ***7*** *- AQUEDUCT* ***8*** *- DAM* ***9*** *- INTERMITTENT/INDEFINITE SHORELINE* ***10*** *- MAN-MADE SHORELINE* ***11*** *- CHANNEL IN WATER* ***99*** *- TRANSPORT ARC* ***0*** *– UNKNOWN*
PWSID	*Text (11)*	*DEP public water supply identification number*
SOURCE	*Text (12)*	*Program Source for the feature:* ***USGS/MGIS*** *– Original 1:25,000 Hydrography* ***DEPGIS*** *– Added from DEP Wetlands by DEP GIS staff* ***DCRGIS*** *– From DEP Wetlands and edits/verified by DCR staff*
MINOR_NUM	*Number (12)*	*Same as MINOR_TOT, as integer*
RESOLUTION	*Text (12)*	*NHD corresponding resolution:* ***HIGH*** *- Nominally > 1:25,000* ***LOCAL*** *- < 1:25,000, nominally 1:12,000, some at 1:5,000*

MAIN SUCCESS FACTORS AND BENEFITS.

Through a thorough analysis of the data, the geoarchiving team would then be able to identify data users who could assign financial value to any number of success factors associated with the preservation of this dataset

- ★ Improved storm water management and the attainment of regulators' goals
- ★ Cost effective maintenance, operating and construction costs for engineered solutions
- ★ Cost avoidance of need to recreate or repurchase historic geospatial data
- ★ Cost avoidance of potential legal challenges
- ★ Availability of accurate and complete geospatial data record over time
- ★ Cost savings in reduced number of trips to perform field assessment
- ★ Successful assessment of impervious areas within a jurisdiction

- Perception of fairness of equitable assessments
- Reduced appeals from reliable and defensible assessments
- Current and historic data to support analysis

ESTIMATED VALUE OF ARCHIVED DATA, What is an approximate dollar value for this dataset if archived? (cost avoidance in recreating, value for various use cases, etc)