Chesapeake Bay Program

Geospatial Data Quality Assurance Project Plan

Standard Operating Procedures for Managing Geospatial Data

Chesapeake Bay Program
A Watershed Partnership

Richard Batiuk, EPA
Associate Director CBPO
QA Officer

Date

Mike Fritz, EPA
Associate Director CBPO

Date

John Wolf, NPS
GIS Team Leader

Date

CONTENTS

A. Project Management

1 Project distribution List

R. Batiuk, EPA (Associate Director, CBPO, and QA Officer)*
B. Burch, EPA (Data Center Manager)
M. Ley USGS (QA Coordinator)
J. Wolf, NPS (GIS Team Leader)
P. Claggett, USGS (Research Geographer)
H. Weinberg, UMCES (GIS Analyst)
K. Hopkins, UMCES (GIS Analyst)
Vacant, UMCES (GIS Analyst)
*indicates approving authority

2 Project/Task Organization

The Chesapeake Bay Program GIS Team is comprised of a GIS Team Leader and four full-time GIS Analysts. The GIS Team Leader is employed by the National Park Service (NPS), the full-time GIS Analysts work under a cooperative agreement with the University of Maryland Center for Environmental Science (UMCES) and through an interagency agreement with the U. S. Geological Survey. The Team is primarily responsible for analyzing and mapping geo-spatial data. The source of the geo-spatial data is almost exclusively from other organizations as the Team creates very little data itself.

Other individuals responsible for implementing the Geospatial QA Project Plan include the Data Center Manager, who provides guidance and oversight for Data Center projects (including GIS), the QA Officer, and the Associate Director for Science within the Chesapeake Bay Program Office.

In addition, intended users of geospatial products are not included in the listing or organizational chart because of their high numbers. Users of geospatial analyses include environmental management agencies in the Bay watershed jurisdictions (PA, MD, VA, DC, WV, NY, DE), members and participants of the Chesapeake Bay Program Subcommittees and Workgroups, federal agencies, state and federal contractors, academic researchers, non-profit environmental organizations, and the press.

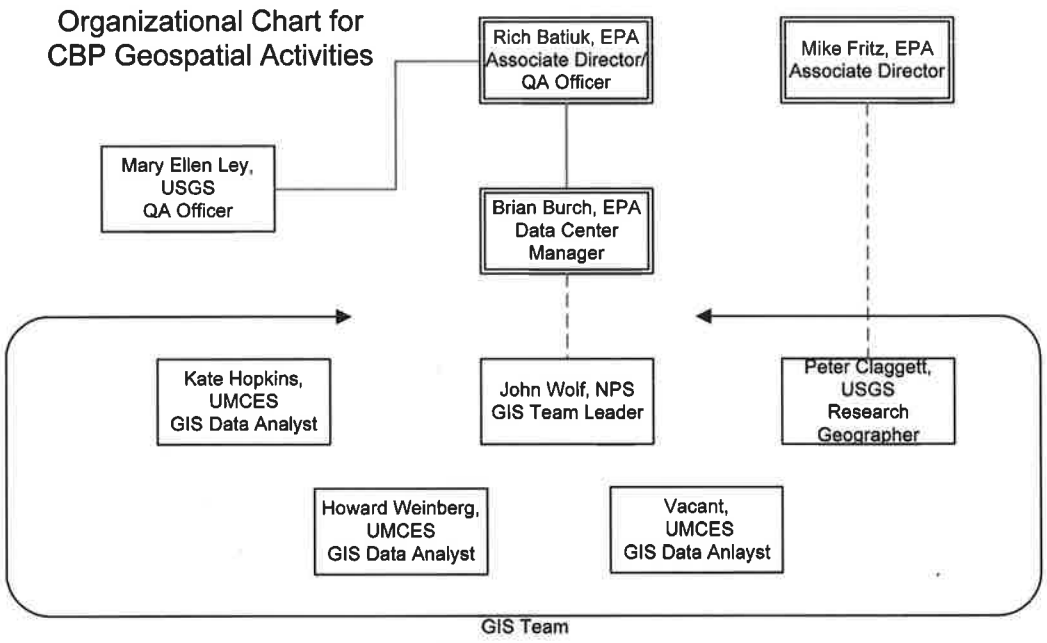


Figure 1. Chesapeake Bay Program Geospatial Staff

GIS projects originate from a variety of sources, including CBP Subcommittees, Workgroups, and EPA Staff. Projects are typically conceptualized to the Data Center Manager or GIS Team Leader and then assigned to appropriate GIS Team staff depending on technical expertise required, subject or application area, and existing workload of Team members.

3 Problem Definition/Background

The GIS Team provides geospatial analysis services for the Chesapeake Bay Program subcommittees and partners. These services are primarily conducted to address the commitments outlined in the Chesapeake 2000 Goals. Specifically, the GIS Team is charged with the following activities:

Geographic Analysis

- Conducts integrated, multimedia analyses of geographical information used in regional land and watershed planning; and targeting nutrient, sediment, and chemical contaminant reduction and prevention programs in an optimal manner.

Spatial Data Development

- Coordinates with CBPO Data Managers and Subcommittee Coordinators on development of spatial data bases needed to support Chesapeake 2000 implementation and spatially enabling existing CBP data bases.

Spatial Data Access

- Develops the technical design and implementation of distributed, networked, spatial data bases necessary for ensuring that the Chesapeake Bay Program partners and stakeholders have efficient, direct internet access to the environmental spatial data and information required to achieve the goals of the Chesapeake Bay Agreement.

Geospatial Application Development

- In cooperation with the CBP Applications Team and Data Team, develops programming solutions for integrating and automating spatial data analyses and associated results.

Web-Based GIS Applications

- In cooperation with the CBP Web Team, develops and implements Internet based GIS applications (e.g. Bay Atlas, Watershed Profiles, Modeling Data Viewer)

Indicator Development and Geographic Profiles

- In cooperation with the CBP Web Team and Data Team, develops multi-scale geographic profiles that encompass environmental quality, implementation actions, and other related restoration and protection data and information.

Technical GIS Support

- Provides expert GIS technical support to CBP partner regional and State agencies staff on the use of GIS to address Chesapeake Bay restoration efforts.

Grant and Project Review

- Reviews grants and project deliverables for adherence to Chesapeake Bay Program Guidance for GIS Data Management and related procedures.

GIS activities undertaken at the Chesapeake Bay Program Office follow the directives outlined in the National Geospatial Data Policy (NGDP). The NGDP establishes principles, responsibilities, and requirements for collecting and managing geospatial data used by Federal environmental programs and projects within the jurisdiction of the U.S. Environmental Protection Agency (EPA). This Policy also establishes the requirement of collecting and managing geospatial metadata describing the Agency's geospatial assets to underscore EPA's commitment to data sharing, promoting secondary data use, and supporting the National Spatial Data Infrastructure (NSDI).

The NGDP addresses geospatial data represented by points, lines, polygons, and complex geographic features. Additionally, it encompasses original and interpreted geospatial data derived through remote sensing including but not limited to images and raster data sets, aerial photographs, and other forms of geospatial data or data sets in both digitized and non-digitized forms.

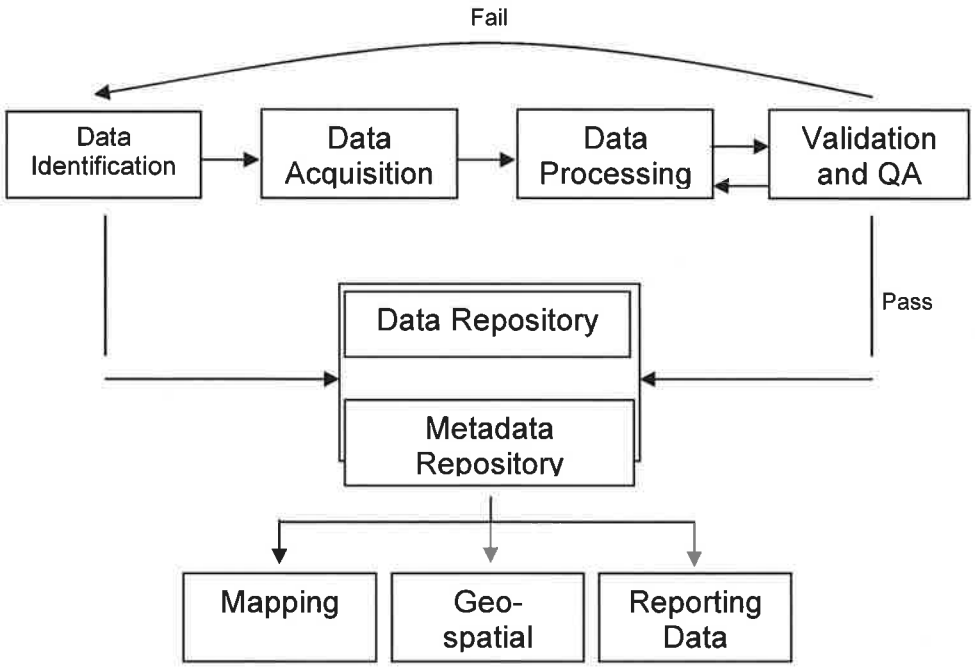
NGDP addresses geospatial data operations in the context of data life cycle phases, consisting of (1) planning, (2) collection and acquisition, (3) processing and documentation, (4) storage and access, and (5) maintenance and retirement.

NGDP applies to all EPA organizations, grantees, agents working on behalf of EPA, tribes, localities and partner states of EPA that design, develop directly or indirectly, compile, operate, or maintain EPA information collections developed for environmental program support.

4 Project/Task Description

CBP GIS project work generally begins with the identification of data to be used for a specific project. Often the data required to support the project is contained within the CBP data repository (ArcSDE geodatabase data residing on CBP database server). Data contained in the CBP data repository has passed GIS Team quality assurance checks and is considered published data. Occasionally, the data to support a project needs to be collected and processed into the CBP repository from an outside entity. This data is then processed into GIS data using the above referenced CBP standards. If the data passes quality checks and has FGDC-compliant metadata, it can be published along with metadata in the data repository.

Maps, geo-spatial analysis, and data reporting are the three primary outputs from a GIS project. The process flow for a GIS project is represented by the following diagram:



For the majority of CBP geospatial operations, only EPA's internal geospatial standards are necessary. These standards are upwardly compatible with standards that apply at the federal or international level. They apply whenever EPA or its partners collect or acquire

data for EPA program or research purposes. Applicable EPA standards include the following:

- EPA Latitude/Longitude Data Standard: This existing standard (http://www.epa.gov/edr/LatLongStandard_08112006.pdf) defines what is required for recording latitude/longitude information for point data.
- Institutional Control Vector Profile Technical Specification: This specification (http://www.envdatastandards.net/files/698_file_IC_Vector_Profile_Version_01_06_2006__Final_.pdf) establishes the requirements for documenting geographic lines and boundary coordinates and related method, accuracy, and description data for places of interest to the EPA.
- National Geospatial Data Policy - Procedure for Geospatial Metadata Management: Establishes procedures, requirements and responsibilities to implement a data life cycle, as defined in the National Geospatial Data Policy (NGDP), for all geospatial metadata used by federal environmental programs and projects within the jurisdiction of the U.S. Environmental Protection Agency (EPA).

In addition, EPA's National GIS Workgroup has created an EPA Metadata Template as an aid in metadata collection. This tool facilitates the metadata documentation process within ArcCatalog.

As jurisdictions are, in part, funded by the Bay Program through grants and cooperative agreements, they are subject to QA requirements as established in the "Chesapeake Bay Program Quality Assurance Guidelines and Requirements" section of the *CBP Grant and Cooperative Agreement Guidance*. This section is relevant to projects involving the collection of environmental data.

The CBP Quality Assurance Program ensures that projects include sufficient up-front planning for the development of well defined project goals and data quality objectives. These objectives need to be supported by implementation of sampling design, collection, and analysis protocols such that the resultant data completely and accurately addresses the project's goals. In order for the data to be useful to efforts to restore and protect the Chesapeake Bay, they must be of known and documented quality, having sufficient supporting documentation such that subsequent data users can evaluate if the data meets their needs.

Additionally, data submission requirements for all grant or cooperative agreement funding are documented in the section "Chesapeake Bay program Guidance and Policies for Data, Information and document Deliverables Submission" of the *CBP Grant and Cooperative Agreement Guidance*.

5 Quality Objectives and Survey Measurement Performance Criteria

The work completed by the GIS Team follows the guidelines set forth by the *Quality Management Plan for the Chesapeake Bay Program Office,* the *Chesapeake Bay Program Guidance for Data Management*, the *National Geospatial Data Policy*, the draft *National Geospatial Data Policy Implementation Guidance,* and the Federal Geographic Data Committee *Content Standard for Digital Geospatial Metadata.* Large projects require development of a comprehensive CBP Data Center Project Plan.

Locational data shall adhere to spatial data models and spatial data standards, and promote the design of data sets that will support data sharing and promote secondary data use, quality information, and otherwise support the NSDI and be publishable for cross-government use on the GeoData Gateway and Geospatial One Stop portal. All geospatial data must adhere to all EPA data standards as well as the ANSI and International Organization for Standardization (ISO) standards that apply.

CBP GIS Analysts must ensure that the geospatial data they collect is structurally consistent with fully documented metadata that is compliant with applicable EPA, FGDC, and GOS standards. In developing procedures under the NGDP, project managers must define their programmatic geospatial data needs in detail and adopt complete data specifications to support their project needs.

In recognition of the varying levels of accuracy and precision appropriate for different program business needs, wherever possible Project Managers should explicitly state their requisite accuracy and precision goals for specific programmatic purposes based on the adopted Geospatial Accuracy Tiers (see Table 1). In the absence of program-specific procedures addressing required minimum accuracy for geospatial data, the NGDP requires a minimum accuracy of Tier 5.

Tier Level	Accuracy and Precision	Examples of Horizontal Collection Method	Example CBP Application Area
Tier 1	<1 m	Classical Surveying Techniques; plus GPS Carrier Phase Static Relative Position	Generally not applicable to CBP projects
Tier 2	1 – 5 m	GPS Carrier Phase Kinematic Relative Position	Generally not applicable to CBP projects
Tier 3	6 – 25 m	GPS Code (Pseudo Range) Standard Position	CBP Shallow Water Monitoring Data
Tier 4	26 – 100 m	GPS unspecified; Photo/GIS Interpolation	SAV mapping, land cover ,mapping
Tier 5	101 – 200 m	Urban style address matching	Toxics, Point Source
Tier 6	201 – 999 m	Public Land Survey – Sixteenth Section	Protected Lands Boundaries

Tier Level	Accuracy and Precision	Examples of Horizontal Collection Method	Example CBP Application Area
Tier 7	1000 – 2000 m	Address Matching – Block Face	Coarse Scale geographic targeting tools
Tier 8	2001 – 5000 m	Census Block Centroid	Watershed (HUC11) statistical summaries and indicators
Tier 9	> 5000 m	Zip Code Centroid	Airshed impacts, Priority Living Resource Areas (derived from Habitat Requirements)
Tier 10	Unknown	N/A	Relative contextual data

Table 1. Geospatial Accuracy Tiers

6 Training/Certification

The GIS Team members are required to have extensive experience and knowledge of geo-spatial mapping and analysis techniques. This collective experience is the first step in assuring quality for products generated by the Team. Team members regularly participate in professional development and continuing education opportunities to keep abreast of the latest technology and software capabilities.

7 Documentation and Records

Project Documentation
Each project requiring a geospatial analysis or reporting requires project documentation. The documentation details the methods, assumptions and rationale that were required to complete the project. This documentation will be stored within the project directory. The project directory will reside on a common, shared network drive that allows full access to the GIS Team. An index file named README.TXT will reside in the project directory that provides an overview of the data and documentation files. The GIS Analyst is responsible for the creation and maintenance of the project documentation. If requested, the GIS Analyst will provide project documentation to the project lead.

Mapping
Often the output from a GIS project is in map form. Because map production is part science and part art, liberal latitudes are given to the GIS Analyst to develop their own style within the boundaries of sound cartographic principles. Based on the intended audience for the map and the message the project lead is attempting to convey, sizes, shapes, colors, fonts, line weights, labels, etc. are left to the discretion of the analyst.

Each published map is required to have a title, north arrow, scalebar, legend, CBP logo, and source information. In addition, notes detailing data limitations, appropriate use of the map and caveats are required where appropriate. Each map is visually inspected and reviewed by at least one additional member of the GIS Team. This review focuses on

spelling, map elements, color, and overall readibility. Finally, the project lead is responsible for review and sign-off on the map.

Checklist for mapping:
- ❑ Does the map have a title, north arrow, scalebar, legend, CBP logo, source?
- ❑ Has the map been visually inspected by at least one additional member of the GIS Team?
- ❑ Has the project lead reviewed and accepted the map?
- ❑ Does the map include a production date?

Metadata

Metadata is critically important to fully understanding source data. When collecting data from sources outside of the CBP partnership, metadata is often the only indicator of data quality. The GIS Analyst should thouroughly read and understand the metadata. The Analyst should pay particular attention to the abstract, purpose and references to spatial accuracy. Based on the metadata, the project lead and GIS Analyst will determine if the data is appropriate for use in the project. The Analyst should follow-up with contacts listed in the metadata if questions arise.

CBP GIS Analysts serve as Geospatial Data Stewards. Every geospatial data asset, and all digital geospatial information collected by CBP for inclusion in EPA databases, shall be under the control of the Geospatial Data Steward, whose identity shall be referenced in the metadata for each database or geospatial data asset.

In the planning stages of geospatial data acquisition or collection, Geospatial Data Stewards shall be responsible for researching the existence and availability of suitable data and for posting notices of intent to acquire or collect geospatial data on the EPA GeoData Gateway.

Geospatial Data Stewards shall provide guidance, support, coordination, review, and public release of geospatial metadata for CBP. This includes local determinations of data sets for contribution, overseeing metadata creation, overseeing data updates, and reporting to central EPA Geospatial Data Stewards, if such stewards exist for a particular data asset.

Documentation of geospatial metadata shall be prepared for all spatial data elements entered into EPA's databases by offices and/or agents of EPA that originate or modify spatial data, in accordance with the provisions of FGDC-STD-001-1998, *Content Standard for Digital Geospatial Metadata* and the *EPA Geospatial Metadata Technical Specification.*

B Measurement Data Acquisition

8 Sampling Process Design

Data Identification:
The first step in conducting a GIS project is matching the project specifications with data required to meet the intended goal of the project.

Know the project goals:
Project goals are widely varied. For example, GIS have often been used in publications like the *State of the Bay* report. Often the GIS generated maps are delivered to graphic artists that alter the complex linework to produce a simpler, more graphically pleasing map image. In doing so, the accuracy of the map is often compromised. Using GIS technology the linework can be generalized more accurately prior to delivery to the graphic artist. At the other end of the spectrum, a project may require highly accurate, large scale data. In all cases, it is critical to know the goals of the project and the intended use of the GIS-derived products.

Checklist for project goals:
- ❑ What are the project deliverables (maps, analysis, reports)?

If the deliverable is a map:
- ❑ What will the map be used for? Who is the audience?
- ❑ What size, scale is appropriate?
- ❑ What is the requirement for accuracy?

If the deliverable is an analysis:
- ❑ What will the analysis be used for? Who is the audience?
- ❑ Have the methodologies been defined/provided?

Investigate using CBP data first:
The CBP has extensive data holdings that are well documented within the metadata repository associated with geodatabase. The documentation stored with the data meets the *Content Standards for Digital Geospatial Metadata,* published by the Federal Geographic Data Committee (FGDC). The first place to search for data to match the project goals is the GIS data on the database server. If the project goals cannot be satisfied with CBP data, the next logical place to search is the Federal Geographic Data Clearinghouse (FGDC), which documents published geospatial datasets or EPA's GeoData Gateway

Checklist for CBP data:
- ❑ Have you searched CBP data holdings in the geodatabase repository?
- ❑ Have you read the metadata for candidate datasets?
- ❑ Does CBP data exist that meet the scale and accuracy requirments defined in the project goals?
- ❑ If CBP data does not meet the requirement, datasets must be identified and collected from outside of the CBP.

Geospatial data that is acquired by EPA (including contractors, grantees and vendors), must comply with all procedures and standards applicable to those data as if they were collected by EPA.

Initial documentation of geospatial metadata shall be accomplished during the collection and acquisition phase to provide information on the steps and methods followed in acquiring the spatial information. This information must be posted on the CBP GIS data server along with the data.

9 Sampling and Image Acquisition Methods

Land cover data derived from satellite imagery is a critical element of many CBP GIS projects. The CBP desires updated comparable land cover data every 5-years. However, to date land cover data has been acquired on an opportunistic basis through data sharing or cooperative agreements with federal and state agencies or with academic institutions. Therefore, the CBP has been largely dependent on the acquisition methods and accuracy standards of third party data providers. For data acquired through cooperative agreement, the CBP has specified that acquisition and accuracy assessment methods should conform to national standards as described for the National Land Cover Dataset (http://landcover.usgs.gov).

CBP does not typically initiate major geospatial data and imagery collection activities. In the event that such collection should be deemed appropriate, a systematic planning process will be conducted to ensure that: (1) the measurements, observations, or images accurately represent the portion of the environment to be characterized; (2) the locational coordinates samples are of sufficient accuracy to support the planned data analysis; and (3) the locational coordinates sampled by measurement meet completeness criteria.

10 Sample Handling and Custody

CBP does not typically handle hard-copy aerial photographs or original maps. In the event that such source materials are required of project implementation, careful and deliberate chain-of-custody procedures will be implemented to ensure that original materials are not lost, misplaced, altered, or destroyed.

11 Quality Control

The GIS Analyst needs to assess the spatial quality of the data using the project goals as the applicable requirement. Generally, the spatial accuracy can be assessed using either hard copy maps (usually USGS 7.5' quadrangles), or ancillary digital data. Additional GIS tools are available to assess the topology of the data, these include: checking dangle nodes, verifying tolerances, and checking for label errors.

Any issues related to the spatial accuracy should be fully documented and reported to the generating organization. If the data can be easily corrected (ex. removal of a dangle node), the GIS Analyst may fix the problem and document the process used to fix the problem. If the problem cannot easily be resolved and the generating organization is unable to resolve the problem, the data cannot be published in the CBP data repository. The data will be considered "draft," and the GIS Analyst will seek guidance from the project lead regarding the appropriate use of the data. If the data is used in a GIS project, any derivative product (map, report, etc.) will have a descriptive note identifying the data as "draft".

If the data passes quality assurance checks, the data and associated metadata will be added to the data repository.

It is important to note that the data quality assurance deals with the spatial quality of the data only. It is assumed that "descriptive" attributes have had some level of quality assurance by the generating organization.

Checklist of quality assurance:
- ❑ Review metadata
- ❑ Fully document QA issues
- ❑ Verify spatial locations using ancillary data from CBP data repository.

12 Instrument/Equipment Testing, Inspection, and Maintenance Records

Not applicable. CBP geospatial projects do not involve the use of instruments or measurement equipment.

13 Instrument Calibration and Frequency

Not applicable. CBP geospatial projects do not involve the use of instruments or measurement equipment.

14 Inspection/Acceptance Specifications for Supplies and Consumables

Not applicable. CBP geospatial projects do not involve the use of supplies or consuumables.

15 Data Acquisition

The GIS Team creates very little data. Virtually all data used in GIS projects are obtained from outside sources. These sources include other federal agencies; state, county and local governments; non-profit groups; and academic institutions. Most often, these external data sources are not funded by the CBP. The GIS Team relies on these external data sources to provide adequate metadata. While the GIS Team can assess the spatial quality of the data using several techniques described below, it often cannot assess the descriptive data. For example, if a facility record is provided to the GIS Team with address information, the Team can verify if the latitude/longitude fall within the state and zip code reported in the address information. The Team often has no way of verifying the street address of the facility, (or even if the city, state and zip code are correct). For this reason, the Team relies on the data generating organization to assure the descriptive attributes provided with data.

Geospatial data acquired by CBP (including contractors, grantees, and vendors) must comply with all EPA procedures and standards. Before any data can be collected or acquired, a program must first determine the level of accuracy the project will require. Project Managers may choose to use the *Geospatial Accuracy Tiers Table* (Table 1) to determine the tier the project data collection will require.

Any data being collected for the first time should be of the highest quality possible based on technology and cost of applying that method or technology. This will ensure that secondary users can also use the data and avoid duplication of efforts between projects. Accuracy will need to be determined, which will then in turn determine which technology and or method should be used in the geospatial data collection process.

Metadata shall be collected along with the source data for all data acquisition activities. The metadata will be entered into ArcCatalog, and the data published in the CBP data repository. If the metadata is incomplete, the GIS Analyst needs to recreate the metadata by contacting the generating entity. If the metadata cannot be recreated, the data will be considered draft. The GIS Analyst, in consulation with the project lead, need to use their best professional judgement as to whether the data meets the intended needs for the project. A descriptive note shall appear on the map or report detailing the uncertainty of the data.

Checklist for collecting data:

- ❑ Check Federal Geographic Data Committee (FGDC) web sites first.
- ❑ Collect FGDC compliant metadata
- ❑ If metadata does not exist, contact the generating organization to recreate the metadata
- ❑ If metadata cannot be recreated, use best professional judgement whether the data is still useful.

16 Data Management

Data collected by the CBP will be processed in accordance with the *Quality Assurance Management Plan for the Chesapeake Bay Program Office*, and the *Chesapeake Bay Program Guidance for Data Management.*

Checklist of data processing:
- ❑ Process to data standards set forth in the *Quality Assurance Management Plan for the Chesapeake Bay Program Office,* and the *Chesapeake Bay Guidance for Data Management.*
- ❑ Fully document the data processing steps.
- ❑ Does the data have topology?

The data repository is the location for all published geo-spatial data for the CBPO. Currently, the data repository is on a Windows NT server at the Chesapeake Bay Program Office in Annapolis, Maryland. The server contains over 80 gigabytes of quality checked geo-spatial data collected for previous GIS projects. As data is collected, processed and quality checked to support a project it is added to the data repository. The repository has daily, weekly and monthly backup to ensure no loss of data. Monthly backup tapes are sent to a secure off-site location.

Currently, the geo-spatial data holdings are in Environmental Systems Research Institute (ESRI) ArcInfo coverage, shapefile, grid or Spatial Database Engine (SDE) geodatabase format. SDE uses standard RDBMS technology (Microsoft SQL Server for the CBP) to store, manage and maintain geo-spatial data holdings. Each data layer in the data repository has a corresponding metadata record in the metadata repository.

Checklist of the data repository/metadata repository:
- ❑ Has the geo-spatial data passed GIS Team data quality assurances?
- ❑ Does the geo-spatial data have a corresponding metadata entry in ArcCatalog?

Geospatial data records in the form of coverages, tables, files, working or draft files, in both hard copy and electronic format, are to be treated as official Agency documents for the purpose of records management. Data disposition for archiving must comply with the Records Retention requirements of the program under which the data was collected.

For large GIS projects that involve the interaction of multiple CBP Data Center Teams, the CBP Project Planning – Development Process should be followed. The process describes the activities and documentation prepared within the system development life cycle for geospatial (and other) Data Center projects. The Data Center development process is a user-driven, peer-reviewed process aimed at efficiently solving business problems. Solutions should involve team-based knowledge transfer, conform to Data Center and industry best practices, and utilize standards-based architecture and reusable

logic. The process is not strictly iterative but rather is designed to be a guideline or best practice for application development that demands ongoing communication throughout the project life cycle. The process involves a series of steps: Assess - Plan - Design - Prototype - Develop - Test - Document - Deploy - Maintain.

C Assessment/Oversight

17 Assessment and Response Actions

The purpose of the assessments and response actions is threefold:

- To insure that all elements of this QA Project Plan or project-specific project plans are correctly implemented as prescribed;
- The quality of the data and product generated by implementation of the QA Project Plan is adequate; and corrective action, when needed, are implemented in a timely manner and their effectiveness is confirmed.
- To ensure that data standards set forth in the *Quality Assurance Management Plan for the Chesapeake Bay Program Office,* and the *Chesapeake Bay Guidance for Data Management* are followed.

Checklist of geospatial data analysis oversight:
- Check of locations of features in data sets selected against locations of these features in other data sources.
- Check of attribute data to ensure that it is of acceptable quality.
- Review of processing procedures by others on the GIS Team.
- Tests that compare processed geospatial data to the original or source data sets throughout production.
- Audit of data quality prior to producing a final report.

18 Reports to Management

The GIS Team Leader provides oral project status reports to the Data Center Manager at bi-weekly update meetings. GIS Analysts report project status at monthly GIS Team meetings or via e-mail to GIS Team Leader as necessary. For projects of limited scope, GIS Analysts provide project status/updates to CBP client via e-mail or telephone conversations. More complex projects can include formal written semi-annual, quarterly, or monthly updates to the Data Center Manager or Client as needed.

D Data Validation and Usability

19 Data Review, Validation, and Verification Criteria

For each geospatial project conducted, a review of data is conducted to ensure that it is suited for the intended use. In addition, the final products of geospatial projects should

undergo final checks to ensure that it meets the objectives of the project for usability and quality.

For imagery or other geospatial data products acquired from other sources, a CBP GIS Analyst must verify that the materials are received as originally ordered and that the order is complete. For imagery, the contents of each digital file is checked for coverage and quality upon receipt.

20 Verification and Validation Methods

Upon completion of a geospatial project, final checks are implemented to make sure that the geospatial data or products are usable by the intended audience. For geospatial data products that will be made available to others, the checks include:

- verifying that each output data set falls into the correct geographic location and has the specified coordinate system and precision;
- verifying that the files to be delivered are of the specified format;
- verifying that each data set can be unpackaged, uncompressed, or otherwise configured for use by end-users; and
- verifying that all of the needed database tables and fields are present.

For maps and other cartographic products to be made available to others, the product is reviewed by another member of the GIS Team to ensure the accuracy of the title, geographic extent, legend, map scale, and other elements.

21 Reconciliation with User Requirements

For CBP geospatial projects, products are reviewed by the requesting CBP entity or partner to ensure that they meet the needs of the data user or decision maker. For products derived from general-purpose databases (e.g. - water quality, point source, living resources), CBP GIS Team members work with the appropriate Data Manager to ensure that products meet QA requirements associated with those databases.

For map products, a specific discussion of limitations of the output is included either on the map itself or by reference to a readily accessible narrative (e.g. - map limitation discussion on the CBP web site). Final map products to be included in the CBP map data base must be reviewed and approved by a GIS Team member other than the map author.

E. Graded Approach Examples

The scope and complexity of the geospatial projects drives the scope and complexity of QA procedures for these projects. Geospatial projects that are narrow in scope, that do not result in decisions that have far-reaching impacts, and are not complex suggest a less rigorous or complex approach to QA. For broad-scope projects with significant and widespread ramifications, a more detailed QA is necessary. Major factors in determining the level of detail in a QA assessment include the intended use of the project output and

the project scope and magnitude. The intended use determines the potential consequences or impacts that might occur because of quality problems. Projects that involve potentially large consequences, such as the development of procedures for determining if water quality standards are being achieved, require the application of the most rigorous QA procedures. In contrast, where no litigation or regulatory impacts are expected, more flexibility can be applied to the QA effort.

The following examples illustrate this graded approach to QA based on project complexity.

22 Minimum Complexity Example: Creating a cartographic product from a spreadsheet containing riparian forest buffer (RFB) latitude/longitude coordinates.

In this example, the GIS Team member has been asked to map the locations of riparian forest buffer projects based on the locations provided by the states in an excel spreadsheet. The locations are provided in latitude/longitude format. The following information should be assembled and kept on file with the project files.

a. Project Management

The GIS Team member seeks the following information from the requestor to complete the product.

- Who is the audience for the map?
- What is the purpose of the map?
- What documentation needs to accompany the map?
- What contextual information (state and county boundaries, labels, hydrography) should be included on the map?

The GIS Team member documents the steps necessary to complete the request:

- How will the Excel spreadsheet be converted for use in GIS?
- How will the data be checked for quality?
- What will the map to be generated look like?
- How and when will draft maps be generated, reviewed, and revised?

The quality objectives for the project are described. For the riparian forest buffer map example, they may include the following:

- The latitude/longitude coordinates in the spreadsheet are to reflect the actual locations of projects on the ground. (Such a quality objective is important because the source data may be inaccurate or contain inconsistencies.)
- The original latitude/longitude coordinates are to be converted into a GIS format and displayed on the map without loss of precision or accuracy.

b. Measurement/Data Acquisition

For this project, no new data is being collected. However, data quality indicators and methods for the existing data would be sought for assessment from the data sources.

Information on data sources to be used in the map include:

- The name of the individual who provided the spreadsheet.
- When the spreadsheet was delivered.
- Format of the spreadsheet.
- Origin of the spreadsheet (where did the requestor get the RFB locations).
- Existing information on how the RFB locations were derived.
- The format of the latitude/longitude coordinates.
- The date the locations were derived
- The contents of metadata for the other data layers on the map.

Information on how the data will be managed once acquired include:

- Description of the applications format to be used to store the converted spreadsheet data file (e.g. - Access, SQL Server, shapefile, other)
- Documentation of any changes to field definitions necessary when converting the spreadsheet.
- Documentation of the computer path to the data file(s) along with the names of the original spreadsheet and the names of any files created during the data conversion process.
- Description of the GIS software programs and version used to process the data.

c. Assessment/Oversight

The primary assessments to be described for this minimum complexity project include:

- The method for ensuring that all RFB spreadsheet records were properly translated into GIS. (e.g. - comparing number of records in GIS with number of elements in spreadsheet).
- The method for ensuring that there are no errors (missing elements or typos) in the map itself.
- The method for correcting errors found during the assessment.

For the minimum complexity project, the report to management may simply be an e-mail or informal memorandum, describing the completion of the project, the map deliverables, any problems encountered and their resolution.

d. Data Validation and Usability

For minimum complexity projects such as the RFB example, it may be appropriate to explore the data quality visually and report to the map requestor any omissions, problems, or concerns with the data.

Typical data validation investigations could include:

- Does every RFB project contain a latitude/longitude coordinate?
- Are the latitude/longitude coordinates consistent in their precision?
- Do there appear to be any transpositions of latitude/longitude in the file? Are the use of minus signs ("-") consistent?

After the input data set is reviewed, a summary for the requestor indicating the nature of any omissions, errors, questions, or concerns about the data and their impact on the intended use will be developed. By providing a summary report/e-mail/memo, the requestor will have the options of modifying the map request, seeking clarification from the data originator on questions, and/or withdrawing the request.

23 Medium Complexity Example: Development of a forest conservation targeting application to support the Forest Conservation Directive

In this example, the GIS Team member has been asked to develop a forest resource conservation value model using a variety of watershed-wide data. The model will prioritize forest cover throughout the watershed for its value in protecting water quality in the Bay. The following information should be assembled and kept on file with the project files.

a. Project Management

The GIS Team member seeks the following information from the requestor to complete the product.

- Who is the audience for the analysis?
- What is the purpose of the analysis?
- What documentation needs to accompany the analysis?
- What criteria should be applied to determine high value forests?

The GIS Team member documents the steps necessary to complete the request:

- How should the criteria to determine high value forests be translated into geographic data?
- Should data and criteria be weighted equally or will some factors carry greater weight in the analysis?
- How will the source data be checked for quality?
- At what resolution should the output of the model be presented?
- How and when will draft products be generated, reviewed, and revised?

The quality objectives for the project are described. For the forest conservation prioritization example, they may include the following:

- Each input data set should represent the actual physical location of phenomena in the real world. (Such a quality objective is important because the source data may be inaccurate or contain inconsistencies.)
- The original source data should be converted into a consistent projection and coordinate system without loss of precision or accuracy.

b. Measurement/Data Acquisition

For this project, no new data is being generated. However, data quality indicators and methods for the existing data would be sought for assessment from the data sources.

Information on data sources to be used in the map include:

- The source of the input data.
- When the data set was obtained.
- Format of the data.
- The date the input data were created.
- The contents of metadata for each of the input data layers.

Information on how the data will be managed once acquired include:

- Description of the applications format to be used to store the raw or converted data file (e.g. - shapefile, coverage, grid, geodatabase, other)
- Documentation of any changes to the data made necessary when converting the data to a standard map projection and coordinate system.
- Documentation of the computer path to the data file(s) along with the names of the original file and the names of any files created during the data conversion process.
- Description of the GIS software programs and version used to process the data.

c. Assessment/Oversight

The primary assessments to be described for this moderate complexity project include:

- The method for ensuring that all input data layers were properly translated into a common projection and coordinate system.
- The method for ensuring that there are no errors (missing elements or typos) in the final map or tabular presentation of the analysis.
- The method for correcting errors found during the assessment.

For the moderate complexity project, their will typically be iterative products developed and shared with the originator of the request. These interim products and progress reports may be presented electronically via e-mail or informal memorandum, describing

the completion of the project, the map deliverables, any problems encountered and their resolution.

d. Data Validation and Usability

For moderate complexity projects such as the Forest Conservation Directive example, it may be appropriate to explore the each input data layer visually and report to the map requestor any omissions, problems, or concerns with the data. This step could help explain anomalies in the end product of the modeling process.

Typical data validation investigations could include:

- Does each input data layer cover the entire geographic extent of the proposed model output?
- Did data transformation operations affect the positional accuracy of the source data?

After the input data sets are reviewed, a summary for the requestor indicating the nature of any omissions, errors, questions, or concerns about the data and their impact on the intended use will be developed. By providing a summary report/e-mail/memo, the requestor will have the options of modifying the analysis request, seeking clarification from the data originator on questions, and/or withdrawing the request.

24 Most Complex Example: Developing complex data sets in a GIS for use in assessing water quality attainment in Chesapeake Bay Program segments.

In this example, the GIS Team member has been asked to develop a process for determining if a Chesapeake Bay Segment is in compliance with water clarity attainment guidelines. The results of the analysis will help to determine if a TMDL needs to be prepared for that segment. Water quality experts from appropriate CBP jurisdictional interests would define the specifications for the geospatial products for their assessment procedures working with GIS Team analysts. This project would involve the collection of geospatial data using shallow water cruise (Dataflow) instruments, converting that data to GIS compatible formats, interpolating water quality surfaces (grids) from the field data, calculating a water clarity surface map from a regression equation developed by water clarity experts, and comparing the resulting surface against thresholds determined for segment-specific designated uses. The following information should be assembled and kept on file with the project files.

a. Project Management

The GIS Team member seeks the following information from the requestor to complete the product.

- Names and addresses of participating project staff and technical staff working on the project from each agency participating, including roles and responsibilities.
- The overall QA Project Plan for water quality data or for shallow water monitoring data could serve as a starting point for this element.
- A summary definition of the problem, background of the overall project, as well as specific definitions of the geospatial portion of the project would be documented.
- A task description for the geospatial processing component of the project should be developed and integrated with the overall water quality QA project plan. The task description for the geospatial processing portion can contain general descriptions of the data sources, processing steps, and data outputs to be created. QA/QC procedures to be applied to the project would be described.

Quality objectives and criteria associated with input data and data processing procedures should be specified to the extent possible. For example, reprojections, transformations and other procedures that modify locational information must result in positional data that is accurate to the level of precision of the geospatial software being used. Similarly, with surfaces interpolated from point locational data, a measure of the uncertainty associated with those interpolations should be included wherever possible.

For shallow water interpolations, the GIS Analyst should have training in operating the geostatistical software used to generate continuous surfaces from Dataflow samples. For all shallow water cruise data, FGDC compliant metadata will be captured and stored with GIS layers derived from field data. Specifically, the metadata should reflect configuration, processing, conversion, transformation, and manipulation of data layers. This information will be attached to the resulting GIS data in an XML format using ArcCatalog or similar tools. For models built using ArcGIS Modelbuilder, the model itself should be documented.

b. Measurement/Data Acquisition

For this project, water quality samples are collected and converted to GIS layers. The sampling process should conform to the CBP QA plan for water quality data.

Quality control procedures for the interpolation process should be documented in this element. These include specifications for interpolation techniques and parameter assumptions, including choice of method (inverse distance weighting vs. Kriging), detrending, data transformation, anisotrophic considerations, covariants, and seasonality differences. Each interpolated data set should be compared against the source (sampled) data to ensure that the interpolations produced verifiable and consistent results.

In addition to the water quality interpolations, other data should be documented appropriately as follows:

- The source of the data to be used. Metadata should be available for each layer.

- How each source will be used during the attainment evaluation process.
- Were checks performed on the existing data to ensure that they were generated correctly and have the predicted content, format, and projection? For example, data assembled from state sources may be supplied in another map projection, which could influence calculations.

Data management documentation should include explanations of the following:

- Path names to all data sources used in the project.
- Methods employed to ensure models or subprograms have been developed and tested to ensure they operated as expected.
- A description of the formats of the data sources, interim data sets that are created during the attainment evaluation, and the final data products (e.g. - pass/fail grids).
- A data dictionary that describes the content, type, name and field width of each attribute in the final product.

c. Assessment/Oversight

The primary assessments to be described for this maximum complexity project include methods to ensure that:

- Water quality interpolations reflect actual measured values where they exit.
- Attribute values in output grids and feature classes are complete and accurate
- Each input data set is in the correct coordinate system information.
- Any reprojections/transformations of input data sets were carried out correctly.
- Each processing step or "macro" was performed correctly and was performed on the correct input data.
- Proper coordinate precision (e.g. single or double precision) was maintained throughout each step in the process.
- There was no unacceptable loss of precision or rounding of coordinates through the process

For each of the assessment methods above, the methods to be used to correct the problem and reprocess any resulting data sets would be described.

A final report to management would indicate the overall processing results, identify the products created, and describe the assessment methods to gauge accuracy.

d. Data Validation and Usability

This project results in the generation of geospatial data layers (grids) identifying those areas either in attainment or not in attainment for water clarity standards. If oversight steps identified in section c. above are followed then there should be few problems during the data validation phase.

For this project, the verification and validation methods element should include:

- A description of the method for reviewing each final data set to be delivered (in general terms.
- Specific methods for reviewing each data set.

F. Bibliography

Chesapeake Bay Program. 2006. *Chesapeake Bay Program Guidance for Data Management.* [Online]. http://www.chesapeakebay.net/cims/Guidance%20for%20Data%20Management%20Nov%202006.pdf

EPA. 2003. *EPA Guidance for Geospatial Data QualityAssurance Plans,* EPA QA/G-5g.

EPA. 2004. *Latitude/Longitude Data Standard*, Draft Update, EPA/EDSC Data Standard.

EPA. 2005. *National Geospatial Data Policy*. CIO POLICY TRANSMITTAL 05-002.

EPA. 2005. *National Geospatial Data Policy Implementation Guidance*, Draft.

EPA. 2007. *National Geospatial Data Policy Procedure for Geospatial Metadata Management.*

Environmental Data Standards Council. 2006. *Institutional Control Vector Profile Technical Specification. Addendum To The Institutional Control* [EX0000015.1] Data Standard.

Federal Geographic Data Committee. 1998. FGDC-STD-001-1998, *Content Standard for Digital Geospatial Metadata.*

Land Cover Institute. 2007. National Land Cover Data Set. {online]. http://landcover.usgs.gov/uslandcover.php