Chesapeake Bay Program Guidance for Data Management

August 2018

Chesapeake Bay Program
A Watershed Partnership

Revision History

NAME	REVISION DATE	REASON FOR CHANGES	VERSION
Brian Burch	8/20/18	Document Review and Update for C4	2.0

Review History

REVIEWER	TITLE	VERSION REVIEWED	DATE REVIEWED

1. Introduction

The Chesapeake Bay Program (CBP) is a unique regional partnership leading and directing restoration of the Chesapeake Bay since 1983. The CBP partners include the states of Delaware, Maryland, New York, Pennsylvania, Virginia and West Virginia; the District of Columbia; the Chesapeake Bay Commission, a tri-state legislative body; the U. S. Environmental Protection Agency, which represents the federal government; and participating citizen advisory groups.

In 1996, the Chesapeake Executive Council adopted the Chesapeake Bay Program's "Strategy for Increasing Basin-wide Public Access to Chesapeake Bay Information." The approach calls for the development of a shared resource of information that is available through the Internet and based on standards and protocols that facilitate access across agency and jurisdictional boundaries.

This document describes the guidelines and policies for submitting data to the Chesapeake Bay Program (CBP) Data Center in Annapolis, Maryland. Organizations funded by CBP, including grantees and contractors, are required to submit deliverables including reports, graphics, spreadsheets, imagery, data files, audio, and digital video products in electronic format. These deliverables must be submitted per the schedule specified in the grant or contract.

All data and information funded by CBP agencies, through direct CBP funding or indirect (matching) funds, is the property of the program and public information unless there is a grant or contract condition that specifies otherwise. In addition, source data collected and processed in the creation of a deliverable shall also be submitted, if practical. Deviations from these guidelines and policies must be negotiated with the CBP Grant/Contract Officer and documented within the grant or contract agreement.

Specific policies and guidelines described in this document include:

- Data, Information and Document Delivery Policy
- Deliverable Serving vs. Submission Policy
- Locational Data Policy
- Metadata Policy
- Common Station Names Guideline
- Common Data Dictionary Guideline
- Common Database Design Guideline
- Calendar Date Policy
- Common Method Codes Guideline
- Data Reporting Guideline
- ITIS Biological Nomenclature Policy

These policies and guidelines are intended to assist contractors and grantees in effectively collecting, processing, and submitting data and information to the Chesapeake Bay Program (CBP). Recipients of Chesapeake Bay Program funding shall comply with these policies and guidelines unless otherwise negotiated and documented with the relevant Chesapeake Bay Program representatives.

2. Chesapeake Bay Program Data Center

The CBP maintains a Data Center at its office in Annapolis, Maryland. The purpose of the Data Center is to provide data management and technical support to program participants in accomplishing the goals set forth by the Chesapeake Executive Council. The Data Center coordinates the management of the Chesapeake Center for Collaborative Computing (C4) which provides the computing resources to the broader partnership. Recipients of Data Center services are the CBP committees, goal implementation teams, Bay Program managers, the watershed's scientific community, and the general public.

3. Chesapeake Center for Collaborative Computing

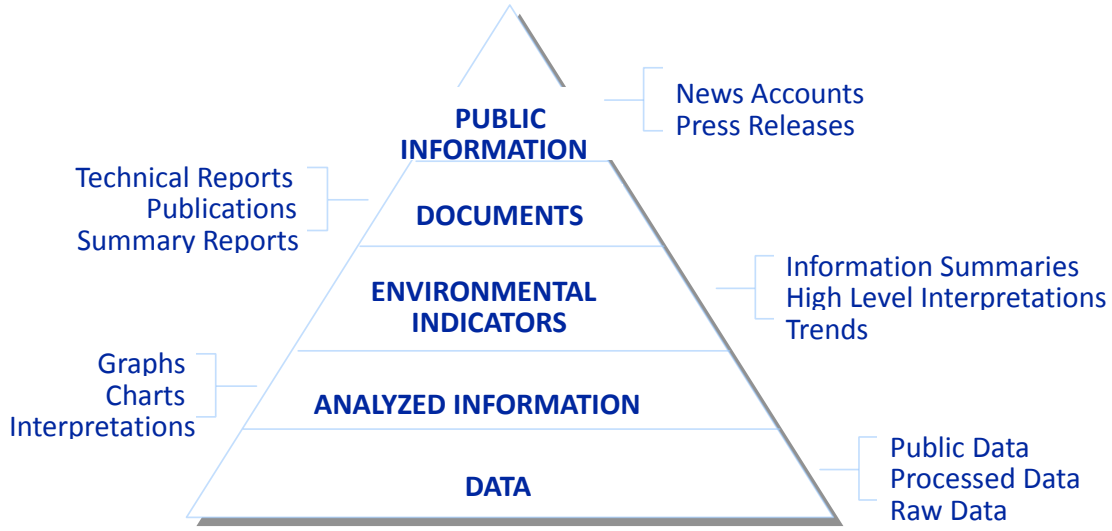
In 2013, EPA's Chesapeake Bay Program Office awarded a cooperative agreement to establish the Chesapeake Center for Collaborative Computing (C4). The University of Maryland's Center for Environmental Science (UMCES) was the recipient of the award and is fully responsible for the management and maintenance of C4 on behalf of the Chesapeake Bay Program partnership.

C4 provides cloud-based infrastructure to CBP partners for collecting, aggregating, storing, analyzing, and disseminating Chesapeake Bay data and information. C4 also provides on-premises infrastructure supporting the non-federal partners located at the Chesapeake Bay Program in Annapolis, Maryland.

C4 receives and manages the authoritative data used by Chesapeake Bay scientists and managers in their science-based decision making. This same authoritative data is disseminated to key Chesapeake Bay stakeholders, including CBP partners, academia, non-governmental organizations working on Bay restoration, and the general public.

4. Types of Information

Various types of data and information are collected, used or generated by the Chesapeake Bay Program and its participants. Each of the major categories of information represented in the information pyramid diagram below serves specific audience needs.



5. Chesapeake Bay Program's Web Resources

The CBP has operated the Chesapeake Bay Program website (http://www.chesapeakebay.net) since April 1995. This site is the authoritative source of Chesapeake Bay data and information. It has been designed to provide comprehensive information at all levels of the Information Pyramid. Authoritative data is available for download from the CBP website through the DataHub application (http://data.chesapeakebay.net)

6. CBP Information Management Guidelines and Policies

The following guidelines and policies must be followed by all agencies, institutions, and organizations participating in data and information collection, processing, document generation and submittal to the Chesapeake Bay Program under grant or cooperative agreement funding. The Chesapeake Bay Program has adopted these guidelines and policies to improve coordination, compatibility, standardization, and information access across all the Bay Program partners. In addition to these guidelines and policies, any activities funded with Federal Government funds, must also adhere to applicable federal guidelines, policies and executive orders, such as the Federal Information Processing Standards (FIPS) (https:/www.nist.gov/itl/itl-publications/federal-information-processing-standards-fips/) and Executive Order 12906.

CBP-specific policies and guidelines include:

- Data, Information and Document Delivery Policy
- Deliverable Serving vs. Submission Policy
- Locational Data Policy
- Metadata Policy

- Common Station Names Guideline
- Common Data Dictionary Guideline
- Common Database Design Guideline
- Calendar Date Policy
- Common Method Codes Guideline
- Data Reporting Guideline
- ITIS Biological Nomenclature Policy

6.1 Data, Information, and Document Delivery Policy

Recipients are required to submit data, information, and document deliverables in electronic format unless exceptions are specified in the grant, cooperative agreement, or contract work plan. Electronic deliverables include, but are not limited to, reports, graphics, spreadsheets, imagery, data files, audio, and digital video products.

All data, information, and documents funded by the Chesapeake Bay Program, whether through direct Chesapeake Bay Program funding or indirect matching funds, are public information and shall be made available to the public unless there is a grant/cooperative agreement award condition that specifies otherwise. In addition, all deliverables must have associated metadata and the source data used in the creation of a deliverable shall also be submitted when practical. If source data is submitted, it should also be delivered in electronic format.

Electronic deliverables shall be submitted to the Chesapeake Bay Program utilizing the acceptable formats documented in the table below:

Document Type	Acceptable Formats
Text	**Preferred:** Microsoft Word (DOC; DOCX) Portable Document Format (PDF) * ASCII Text Extensible Markup Language (XML)
Spreadsheet	**Preferred:** Microsoft Excel 2003 or higher Comma Separated Values (CSV) **With Prior Approval:** Tab delimited text files

Database	**Preferred:** Microsoft Access 2003 or higher Comma Separated Values (CSV) Microsoft SQL Server 2008 or higher Extensible Markup Language (XML) ASCII delimited text files
Graphics	**Preferred:** PNG TIFF GIF JPEG SVG PDF
Geographic Information System	**Preferred:** Personal or file geodatabase ESRI Grids Shape files **With Prior Approval:** KML, KMZ

** Data tables delivered within PDF documents must be delivered in one of the spreadsheet formats.*

6.2 Deliverable Serving vs. Submission Policy

Recipients can submit deliverables directly to the Chesapeake Bay Program or provide deliverables from their organizational data/web servers. State grant recipients are required to submit non-point source BMP data using EPA's National Environmental Information Exchange Network (NEIEN). Recipients who plan to directly provide their grant/cooperative agreement deliverables through their organizational data server/website must include relevant documentation, including access instructions, within their work plan and/or progress reports.

Recipients intending to deliver applications to be hosted by the Chesapeake Bay Program must coordinate directly with the CBP Data Center to ensure those applications are within the Data Center's capacity to operate and maintain. The technology used to develop those applications must be included in the CBP baseline technology portfolio. Recipients delivering computer applications must participate in the CBP Data Center release and deployment planning boards. This requirement only applies when client/server

or web applications are submitted as part of the deliverables with the expectation that CBP will host those applications within the CBP Data Center.

6.3 Locational Data Policy

The Chesapeake Bay Program adheres to the EPA's locational data policy that requires consistent use of latitude/longitude coordinates to identify the location of entities. All data, containing spatial and/or specific geographic locations, collected or assembled under a CBP grant or cooperative agreement must have latitude and longitude information for each entity. Projects not creating or reporting spatial data, but confined to one or more given project locations, shall include the latitude/longitude of the locations within the study/final report.

The recipients agree to ensure that latitude and longitude coordinates (given in decimal degrees) are provided for all data collection sites and accurate to the level required for the application of the data. Field measured locations shall be accurate to the best practical geographic positioning method. Applications such as station monitoring locations should provide locational data with accuracy to that level. Other applications, such as digitizing points or watershed boundaries from Mylar media maps, cannot provide accuracy better than that of the original map, and cannot match the accuracy of GPS or surveyed locations.

Remote sensing platforms can now collect sub-meter resolution data (6 decimal places in decimal degrees). Therefore, it is required that metadata be provided for all data and must include a measurement of the accuracy of the coordinates submitted and the original source material and methods for obtaining the coordinates. It is the responsibility of data generators/providers to include coordinates accurate to the level that is practical for the intended application, and to document the accuracy of those coordinates.

The recipient further agrees to document, in writing, that locational data was derived using an approved method and recorded in accordance with federal regulations and other EPA requirements, noted in the "Authorities" section of the EPA's policy. Recipients shall include in their work plan an assurance to comply with this requirement.

6.4 Metadata Policy

The Chesapeake Bay Program has adopted the policy, consistent with Executive Order 12906, that all data generated or collected using federal funds or submitted to the Chesapeake Bay Program shall be accompanied by metadata that conforms to the Federal Geographic Data Committee's (FGDC) requirements. Metadata created for CBP shall also be delivered to the EPA or other federal clearinghouses as a requirement for fulfilling this policy and related grant or contract conditions. The FGDC guide for creating metadata is the *Content Standard for Digital Geospatial Metadata Workbook* (http://www.fgdc.gov/metadata).

The Chesapeake Bay Program has adopted the policy that all data generated or collected using federal funds or submitted to CBP shall adhere to the National Biological Information Infrastructure's (NBII) metadata standard, where applicable. The NBII metadata standard, popular for environmental programs, provides extensions to the FGDC metadata for documenting biological data and information. The NBII Biological Data Profile can be found at: https://www.fgdc.gov/grants/2006CAP/projects/IA660110253.

6.5 Common Station Names Guideline

The Chesapeake Bay Program has adopted the guideline that all data generated, collected for, or submitted to CBP shall utilize a consistent set of common station names for identifying and reporting monitoring station locations. It is the data provider's responsibility to comply with this policy.

The purpose of this guideline is to create a single master table of station names, to the extent possible, to reduce confusion among cooperating agencies. The Station Names table, maintained on the Chesapeake Bay Program website, should serve as the master list. Data submitter requests to update the table shall be coordinated with the CBP Water Quality Data Manager.

6.6 Common Data Dictionary Guideline

The Chesapeake Bay Program has adopted the guideline that all data generated, collected for, or submitted to the CBP shall utilize the CBP common data dictionary for defining all data elements and units of measure. It is data provider's responsibility to comply with this policy.

The purpose of this guideline is to create a single data dictionary, to the extent possible, to reduce confusion among cooperating agencies. Data submitter requests to update the data dictionary shall be coordinated with the CBP Water Quality Data Manager.

6.7 Common Database Design Guideline

The Chesapeake Bay Program has adopted the guideline that all data generated, collected for, or submitted to the Chesapeake Bay Program shall utilize the CBP common database design for managing data. It is the data provider's responsibility to comply with this policy.

The purpose of this guideline is to create similar database designs, to the extent possible, to simplify data formatting and sharing. Modifications to the common database design shall be coordinated with the CBP Data Center to maintain consistency in the database structure.

If CBP partner organizations do not have existing database designs that are acceptable for the work being conducted, the grantee/contractor should work with the funding agency to develop a suitable database design. The database design should maintain maximum compatibility with other Chesapeake Bay Program database designs.

6.8 Calendar Date Policy

The Chesapeake Bay Program has adopted the standard that all data generated, collected for, or submitted to the Chesapeake Bay Program shall adhere to the Federal Information Processing Standard, Representation for Calendar Date and Ordinal Date for Information Interchange (FIPS PUB 4-1).

This standard states, "For purposes of electronic data interchange in any recorded form among U.S. Government agencies, National Institute of Standards and Technology (NIST) highly recommends that four-digit year elements be used." The year should encompass a two-digit century that precedes, and is contiguous with, a two-digit year-of-century (e.g., 1999, 2000, etc.). In addition, optional two-digit year time elements specified in ANSI X3.30-1985(RI991) should not be used for the purposes of any data interchange among U.S. Government agencies.

Therefore, it is required to report and store all dates using four digits for the year.

6.9 Common Method Codes Guideline

The Chesapeake Bay Program has adopted the guideline that all data generated, collected for, or submitted to the CBP shall utilize the CBP Method Codes tables. The method codes are defined in the *Guide to Using CBP Water Quality Monitoring Data.* It is the data provider's responsibility to comply with this guideline that requires the use of standardized method codes, to the extent possible, to simplify data coding and sharing.

The methods used by monitoring agencies and analytical laboratories are critical in providing accurate measurements. Knowing the field and laboratory methods used is critical, therefore capturing the methods is a high priority during database development. Modifications to the CBP Method Codes shall be coordinated with the CBP Water Quality Data Manager and QA Officer to maintain consistency in the table contents. If CBP agencies do not have existing method codes that are acceptable for the work being conducted, the grantee/contractor should work with the funding agency to develop suitable method codes, while maintaining maximum compatibility with other CBP codes.

6.10 Numeric Data Reporting Guideline

The Chesapeake Bay Program has adopted the guideline that all data generated, collected for, or submitted to the Chesapeake Bay Program shall report numeric data elements at the same level of

precision as that of the original measurement. This guideline has a significant impact on data analysis and the decisions made based on these analyses.

The exact precision of recorded values must be maintained. Values should not be zero-filled to greater precision than actually recorded. For instance, if the measured value is 0.03, then the reported value should be 0.03 and not 0.030, which would imply precision to the third decimal place. For values that are recorded as below or above detection, a detection flag (in a separate data field) shall be used to identify the value as below or above the detection limit of the method, and the value shall be reported as the detectable limit. Values should be reported as zero, only if the measured or recorded value is zero. Values that are missing shall be reported as missing or null or nil, to identify values that were sampled but no value was obtained. Missing, null, or nil values are different than those that were never sampled, which should be recorded as a blank field, if they are recorded at all. It is the responsibility of the data submitter to record in the metadata how measurements are coded, as well as the accuracy of the measurements.

It is important to note that some software tools used in data processing may represent the data internally with more precision than the original measurement, and/or may round the value. For instance even though a value of 0.3 was entered, the value may be stored and reported as 0.299999.

6.11 ITIS Biological Nomenclature Policy

The Chesapeake Bay Program has adopted the policy that all data generated, collected for, or submitted to the Chesapeake Bay Program shall utilize the Integrated Taxonomic Information System (https://itis.gov/) biological names for identifying and reporting species. It is the data provider's responsibility to comply with this requirement.

The purpose of the policy is to create a single master table of species names, to the extent possible, to reduce confusion among cooperating agencies. The ITIS taxonomy table, maintained on the ITIS website, should serve as the master list. Data submitter requests to update the table shall be coordinated with the CBP Water Quality Data Manager and QA Officer to maintain one consistent species name list.

Appendix A - Reference Material

Chesapeake Bay Program. *Chesapeake Bay Program Home Page* (URL: http://www.chesapeakebay.net/). Chesapeake Bay Program, Annapolis, MD.

Chesapeake Bay Program. *Chesapeake Bay Program DataHub* (URL: http://data.chesapeakebay.net/). Chesapeake Bay Program, Annapolis, MD.

U.S. Environmental Protection Agency. September 1996. *Designing an Integrated, Accessible Information Management System for the Chesapeake Bay Region*. Chesapeake Bay Program, Annapolis, MD. SAIC Contract 68-C4-0072, Work Assignment EC-1-8.

National Biological Service. December 1995. NBII Biological Metadata Standard (URL: https://www.fgdc.gov/grants/2005CAP/projects/IA560110409).

Federal Geographic Data Committee. June 1994. *Content Standards for Digital Geospatial Metadata.* (URL: https://www.fgdc.gov/metadata/csdgm/). Federal Geographic Data Committee. Washington, D.C.

Chesapeake Bay Program. March 1993. *Guide to Using Chesapeake Bay Program Water Quality Monitoring Data*. Chesapeake Bay Program, Annapolis, MD.