

PA
Department of Environmental
Protection

Guidance for
Data Management

September 13, 2005

Prepared for Integration with DEP Grant Guidance

PA Department of Environmental Protection Information Access Strategy

This document provides guidance to contractors and grantees that accept funding from and conduct work for the PA Department of Environmental Protection activities. The guidance in this document is provided to guide the collection, processing, and delivery of data and information products. In addition, it provides guidance to serving information on the Internet.

INTRODUCTION

This Guide is intended to aid contractors and grantees in effectively collecting, processing, and submitting data and information to the Pennsylvania (PA) Department of Environmental Protection (DEP) funded contract and grant activities. In addition, it is intended to aid PA DEP in effectively serving information on the Internet and sharing information with other State Agencies or Federal/Local Governments.

PA DEP Enterprise Data

PA DEP maintains enterprise data at the Commonwealth's PA Data PowerHouse and in a Consolidated Data Center at its central office in Harrisburg, PA. The purpose of the Bureau of Information Technology (BIT) is to provide data management and technical support to environmental programs in order to accomplish the DEP mission:

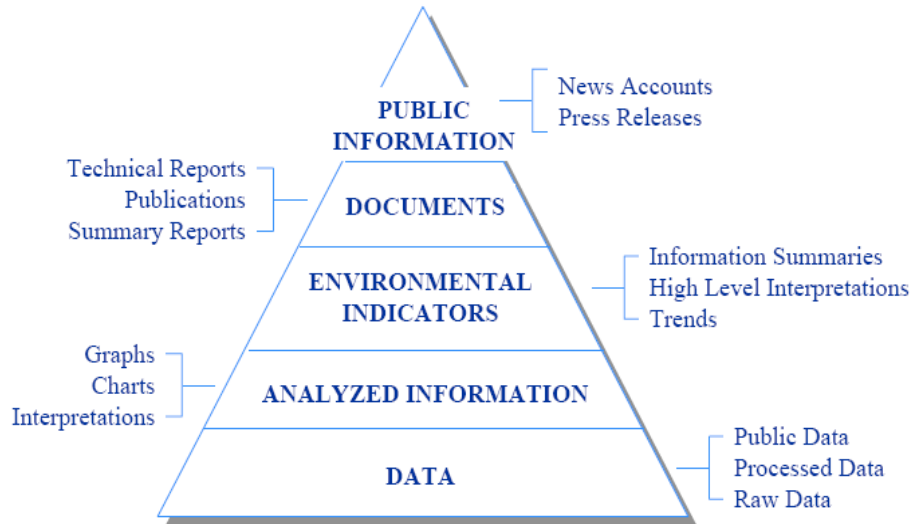
“To protect and improve the quality of the air, water and environment for the health and safety of the citizens of the Commonwealth, to protect the people from dangerous or unnecessary radiation from natural and man-made sources, including occupational and medical exposure, and to manage water and mineral resources in a way which insures against their undue destruction and depletion while allowing economic benefits from their use.”

BIT manages the computer hardware and software, provides user support for computer resources, acquires and stores data sets, and provides data analysis support.

Types of Information

Various kinds of data and information are collected, used or generated by the PA DEP and its environmental programs (see following Information Pyramid). Each of the major types of information -- data, analyzed information, environmental indicators, documents, and public information -- serves one or more purposes aimed at addressing management questions related to the environmental health of the Commonwealth.

Information Pyramid



A goal for PA DEP is to publish all of these types of information for electronic access over the Internet.

PA DEP Information Management Guidelines and Policies

This section discusses the guidelines and policies that must be followed by all agencies, institutions, contractors, and organizations participating in data and information collection, processing, document generation, and data submittal to the PA DEP under grant or cooperative agreement funding. The PA DEP has adopted these data management guidelines and policies in order to improve coordination, compatibility, standardization, and information access for all partners. In addition to these guidelines and policies any activities funded with Federal Government funds, must also adhere to applicable Federal Information Processing Standards (FIPS)

(<http://www.itl.nist.gov/div897/pubs/>).

- Data, Information, and Document Deliverables Requirements
- Deliverable Serving vs. Submission Policy
- Locational Data Policy
- Metadata Policy
- Common Data Dictionary
- Calendar Data Policy
- Numeric Data Reporting Recommendations
- Internet Web Pages

Data, Information, and Document Deliverables Requirements

Recipients are required to submit data, information, and document deliverables in electronic format unless exceptions are specified in the grant or cooperative agreement 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 PA DEP whether through direct 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, source data collected and processed in the creation of a deliverable should also be submitted, when practical. If source data is submitted, it should also be delivered in electronic format utilizing the following standards:

[Environmental Print Shop Document #013-0830-005 lists the Suggested Formats for Required Electronic Deliverable Attachments, version 1.2, dated March 25, 2002.](#)

[Environmental Print Shop Document #013-0830-006 lists the Formats for Required Electronic Deliverable CAD Attachments, version 1.1, dated May 2002.](#)

[Environmental Print Shop Document #013-0830-007 lists the Formats for Required Electronic Deliverable GIS Attachments, version 1.1, dated May 2002.](#)

[Environmental Print Shop Document #013-0830-008 lists the GeoSpatial Digital Data Submission Guideline, version 1.3.5, dated June 14, 2003.](#)

Deliverable Serving vs. Submission Policy

Recipients can submit deliverables directly to the PA DEP or serve deliverables from a data/web server. Recipients who plan to directly serve their grant/cooperative agreement deliverables through their own data server/web site must meet with the Bureau of Information Technology for integration and standards discussions.

Locational Data Policy

The PA DEP adheres to the EPA's locational data policy requiring consistent use of latitude/longitude coordinates to identify the location of physical entities. All data, containing spatial and/or specific geographic locations, collected or assembled under a PA DEP grant or cooperative agreement or to be served on the Internet, must have latitude and longitude information for each entity. Projects not creating or reporting spatial data, but confined to a given project location(s), shall include the latitude/longitude of the location(s) within the study/final report.

In accordance with the PA DEP locational data policy, the recipients agree to ensure that latitude and longitude coordinates (given in degrees and decimal degrees) are provided for all sites for which data are collected and accurate to the level required for the purpose of the application of the data.

[Environmental Print Shop Document #013-0830-003, Pennsylvania DEP Location Data Policy, dated February 10, 2001.](#)

The ultimate goal is for a 10-meter/32.8 feet or better level of locational accuracy. The use of global positioning systems (GPS) is recommended to obtain latitude/longitude of sufficient accuracy to meet program and department goals and to correctly locate facilities.

The Method Accuracy and Description (MAD) Codes are defined in the Location Data Policy. It is the data provider's responsibility to comply with this policy. The purpose is to use 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 and processes used are critical to the department; therefore, capturing the methods is a high priority during database development.

Required information includes: Latitude/Longitude and Accuracy Information, Geometric Type, Horizontal Collection Method, Horizontal Reference Datum, Reference Point, Vertical Location Datum, and Data Collection Datum. The policy lists all codes related to the Location Data Policy.

Metadata Policy

All deliverables must have associated metadata. The PA DEP has adopted the policy, consistent with Presidential Executive Order # 12906, that all data generated or collected using federal funds, submitted to the PA DEP, or served on the Internet, shall be accompanied by metadata (descriptive information about the data, often referred to as documentation), that fully conforms to the Federal Geographic Data Committee's (FGDC) requirements for metadata. The FGDC guide for creating metadata is the Content Standard for Digital Geospatial Metadata Workbook www.fgdc.gov/metadata

Data to be accessed on the Internet must follow the Commonwealth of PA's Governor's Office of Administration/Office of Information Technology's (GOA/OIT) Information Technology Bulletin (ITB): ITB-INF002 Metadata Standards <http://www.oit.state.pa.us/oaoit/cwp/view.asp?A=722&Q=201323>.

PA DEP has an additional document, DEP Metadata Project, further defining the guidelines for the FGDC Content Standard for Digital Geospatial Metadata that can be requested and obtained from the PA DEP.

Common Data Dictionary

PA DEP is in the process of adopting guidelines from its Quality Management Council that all data generated or collected and submitted to the PA DEP, or served on the Internet should utilize the PA DEP common data dictionary for defining all data elements and units of measure. The purpose of this guidance is to create one data dictionary, to the extent possible, to reduce confusion among cooperating agencies. Updates required by data submitters to the dictionary shall be coordinated with BIT to maintain one consistent data dictionary listing responsible data stewards and best management quality practices.

Calendar Date Policy

The PA DEP has adopted the standard that all data generated or collected for and submitted to the PA DEP, or served on the Internet should 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. In addition to facilitating data sharing, this requirement reduces the complications of processing date data after the millennium rollover at year 2000.

Numeric Data Reporting Recommendations

PA DEP recommends that all data generated or collected for and submitted to the PA DEP, or served on the Internet should report numeric data elements at the same level of precision as that of the original measurement. The exact precision of recorded values must be maintained. This guideline has a significant impact on data analysis and the decisions made based on these analyses.

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 limits, 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.

Internet Web Pages

The PA DEP was one of the first state agencies to develop and maintain an Internet Web Page. Currently this web site (<http://www.dep.state.pa.us>) provides the most complete inventory of data and information that is available from the agency. This web site has been designed to provide information at all levels of the Information Pyramid. The PA DEP may also host web sites and pages for organizations that do not have adequate resources to maintain their own Internet servers. The draft policy for web development and maintenance which incorporates all the ITBs issued from the GOA/OIT can be found at:

<http://www.dep.state.pa.us/dep/dsf-policy-091305.doc>

Standard Template for Reviewing Data Management on Submitted Proposals

INFORMATION TECHNOLOGY PROPOSAL REVIEW

1. General Information:

Proposal Title: _____

Organization: _____

Reviewer Name: _____ Review Date: _____

- | Yes | No | |
|--------------------------|--------------------------|---|
| <input type="checkbox"/> | <input type="checkbox"/> | Does the proposal contain a data management plan? |
| <input type="checkbox"/> | <input type="checkbox"/> | Does the proposal contain a project quality assurance plan? |
| <input type="checkbox"/> | <input type="checkbox"/> | Does the proposal identify a data manager? |
| <input type="checkbox"/> | <input type="checkbox"/> | Does the proposal identify project deliverables? |
| <input type="checkbox"/> | <input type="checkbox"/> | Will the deliverables be submitted to the project officer? |
| <input type="checkbox"/> | <input type="checkbox"/> | Will the deliverables be made available on-line? URL _____ |

2. Metadata:

Federal Geographic Data Committee Guidelines <http://www.fgdc.gov/metadata/metadata.html>

Governor's Office of Administration Guidelines

<http://www.oit.state.pa.us/oaoit/cwp/view.asp?A=722&Q=201323>

- | Yes | No | |
|--------------------------|--------------------------|--|
| <input type="checkbox"/> | <input type="checkbox"/> | Will the project submit metadata for each deliverable? |

3. What types of deliverables are identified?

Documents (complete section 4)

- | | | | |
|--------------------------|----------------------|--------------------------|------------------------|
| <input type="checkbox"/> | Booklet | <input type="checkbox"/> | Brochure |
| <input type="checkbox"/> | Fact Sheet | <input type="checkbox"/> | Plan |
| <input type="checkbox"/> | Report | <input type="checkbox"/> | Quality Assurance Plan |
| <input type="checkbox"/> | Curriculum Materials | <input type="checkbox"/> | Other _____ |

Graphics, Images, Multi-Media Presentations (complete section 5)

- | | | | |
|--------------------------|-------------|--------------------------|-------------------------|
| <input type="checkbox"/> | Image | <input type="checkbox"/> | Photography |
| <input type="checkbox"/> | Web Graphic | <input type="checkbox"/> | Map |
| <input type="checkbox"/> | Video | <input type="checkbox"/> | Multimedia Presentation |
| <input type="checkbox"/> | Other _____ | | |

Data (complete section 6)

- | | | | |
|--------------------------|---------------------|--------------------------|-----------------------|
| <input type="checkbox"/> | SAS Transport Files | <input type="checkbox"/> | Delimited ASCII files |
| <input type="checkbox"/> | Relational Database | <input type="checkbox"/> | XML & Schema files |
| <input type="checkbox"/> | Spreadsheet | <input type="checkbox"/> | Other _____ |

Spatial Data (complete section 7)

- | | | | |
|--------------------------|--------------|--------------------------|------------------|
| <input type="checkbox"/> | GIS Coverage | <input type="checkbox"/> | Spatial Database |
|--------------------------|--------------|--------------------------|------------------|

- GIS Shapefile
- Other _____

Software (complete section 8)

- Client Application
- Web Application
- Other _____

4. Document Deliverables:

What is the format of the document deliverable?

- Microsoft Word (preferred)
- Rich Text Format
- Adobe Acrobat * data tables within PDF's must be delivered separately
- Other _____
- WordPerfect
- Page Maker
- Unknown

5. Graphics, Images, Multi-Media Presentation Deliverables:

[Environmental Print Shop Document # 013-0830-005, Suggested Formats for the Required Electronic Deliverable Attachments](#)

[Environmental Print Shop Document #013-0830-006, Formats for Required Electronic Deliverable CAD Attachments](#)

What is the format of the deliverable?

- TIFF
- JPEG
- Microsoft PowerPoint
- PDF (preferred for reports)
- Other _____
- GIF (preferred for web publication)
- SVG
- MPEG (preferred for video)
- Unknown

6. Data Deliverables:

What is the format of the deliverable?

- SAS Transport File
- XML & Schema
- Microsoft SQL Server
- Oracle Database
- Other _____
- Delimited ASCII File
- Microsoft Access
- Microsoft Excel
- Unknown

Yes No Don't Know

- If the deliverable contains geographic locations, will it include latitude and longitude for each entity? [Environmental Print Shop 013-0830-003, Locational Data Policy](#)
- Will the deliverable utilize the DEP's Data Dictionary for defining all data elements and units of measure?
- If the deliverable contains water or land attributes, does it include the Hydrologic Unit Code?
- If the deliverable contains water attributes, does it include a National Hydrography Dataset Reach address?
- If the deliverable holds dates, will it adhere to the Federal Information Processing Standard (FIPS) representation for calendar date and ordinal date form information interchange?
- If the deliverable is numeric data, will it report all numeric data elements at the same level of precision as that of the original measurement?
- If the deliverable is monitoring data, will it be in accordance with the Common Method Codes Guideline that provides standardized Method Accuracy

Description (MAD) codes? [Environmental Print Shop 013-0830-003, Locational Data Policy](#)

- If the deliverable contains species data, will it comply with the National Integrated Taxonomic Information System (ITIS) nomenclature?

7. Spatial Data Deliverables:

*Environmental Print Shop Document #013-0830-007, Formats for Required Electronic Deliverable GIS Attachments.

[Environmental Print Shop Document #013-0830-008, GeoSpatial Digital Data Submission Guidelines](#)

What is the format of the deliverable?

- | | | | |
|--------------------------|-------------------------------|--------------------------|-----------|
| <input type="checkbox"/> | ArcInfo non-compressed (.e00) | <input type="checkbox"/> | Shapefile |
| <input type="checkbox"/> | Exchange file format | <input type="checkbox"/> | GML |
| <input type="checkbox"/> | Other _____ | <input type="checkbox"/> | Unknown |

Yes No Don't Know

- | | | | |
|--------------------------|--------------------------|--------------------------|---|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Will the deliverable conform to the North American Datum 1983 (NAD83) or the World Geodetic System 1984 (WGS84) horizontal reference? |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Will the deliverable conform to the North American Vertical Datum 1988 (NAVD88)? |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Will data be delivered in a Universal Transverse Mercator (UTM), Zone 18, meters projection? |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Will data be delivered in an Albers Equal Area Projection meters? |

8. Software Deliverables:

[Guidelines for Website Development and Maintenance http://www.dep.state.pa.us/dep/dsf-policy-091305.doc](http://www.dep.state.pa.us/dep/dsf-policy-091305.doc)

- Yes No
 Will source code be delivered?

What format is the native source code?

- | | | | |
|--------------------------|------------------|--------------------------|------------------|
| <input type="checkbox"/> | Microsoft C#.NET | <input type="checkbox"/> | Microsoft VB.NET |
| <input type="checkbox"/> | Visual Basic v.6 | <input type="checkbox"/> | Java _____ |
| <input type="checkbox"/> | PL/SQL | <input type="checkbox"/> | Other _____ |
| <input type="checkbox"/> | Unknown | <input type="checkbox"/> | |

- Yes No
 If the deliverable is software, is the IT project being developed under DEP's Project Management Methodology? If not, what methodologies? _____
- Will the software be developed using DEP's System Development Methodology? If not, what methodologies? _____

9. Summary of Findings:

Proposal Rating:

- Acceptable:** The proposal demonstrates in-depth knowledge of the DEP IT requirements and adequately addresses all applicable requirements.
- Acceptable with Minor Modifications:** The proposal meets enough IT requirements to ensure that the deliverables can be accepted. IF SELECTED FOR FUNDING, BIT RECOMMENDS THE PROJECT OFFICER, WORKING IN COLLABORATION WITH BIT, FINE-TUNE THE PROPOSAL TO ENSURE FULL COMPLIANCE WITH REQUIREMENTS.
- Unacceptable:** The proposal does not meet some, or all, of the IT requirements and guidelines. IF SELECTED FOR FUNDING, BIT RECOMMENDS THE PROJECT OFFICER, WORKING IN COLLABORATION WITH BIT, RENEGOTIATE THE PROPOSAL DETAILS TO ADDRESS ALL APPLICABLE REQUIREMENTS.
- Unable to Determine:** The proposal does not contain enough information on data management activities to determine compliance with DEP IT standards. IF SELECTED FOR FUNDING, BIT RECOMMENDS THE PROJECT OFFICER, WORKING IN COLLABORATION WITH BIT, RENEGOTIATE THE PROPOSAL DETAILS TO ADDRESS ALL APPLICABLE IT REQUIREMENTS.

10. Reviewer Comments:
