



Data Tracking System (DTS) Control Slip

Attach to front of folder

Date: 07/24/2012

DCN: GS12001013		ES No:
Orig Office: AW	Input Date: 07/24/2012	Addressee: Myers, Donna
Due Date: 08/07/2012	Signature Level: AD	
Subject: Policy for Release of Computer Databases		

Comments:

Task Codes:

- | | | |
|-------------------------|--------------------------------|---------------------------|
| 0 - Prepare Draft Reply | 6 - Revise | 12 - Email Draft Reply |
| 1 - Prepare Reply | 7 - Obtain Additional Comments | 13 - Advance Read |
| 2 - Appropriate Action | 8 - Other - See Comments | 14 - File |
| 3 - Surname | 9 - Mail/Distribute | 15 - For Your Information |
| 4 - Signature | 10 - Finalize | 16 - Surname through DTS |
| 5 - Review/Comment | 11 - Simultaneous Surnames | 17 - Required ES Review |

Routing:

Assigned To	Action	Task Assigned Date	Task Completed Date
AW	2 - Appropriate Action	07/24/2012	07/24/2012

ok for Bill should sign - Don C
Mike Henderson - CSIS - MGR 7/26/12
Mark Hoff - CSS - MGR 7/27/12



United States Department of the Interior

U.S. GEOLOGICAL SURVEY
Reston, VA 20192



In Reply Refer To:
Mail Stop 412

MEMORANDUM

July 23, 2012

To: William Werkheiser, Associate Director for Water, Reston, VA
Anne Kinsinger, Associate Director for Ecosystems, Reston, VA
Kevin Gallagher, Associate Director for Core Systems Science, Reston, VA

From: Donna N. Myers, Chief
Office of Water Quality, Reston, VA

Cc: Sandra Cooper, Bureau Approving Official, Norcross, GA

Subject: Policy for Release of Computer Databases: Request Director's Approval for BioData Database System

Reply Requested by August 31, 2012

The Office of Water Quality (OWQ) requests your concurrence for Director's approval of the BioData database system pursuant to USGS Survey Manual (SM) 500.24 - Policy for Release of Computer Databases and Computer Programs. To establish a Director's approved database, the database manager must submit a written request to the Associate Director (formerly Division Chief) demonstrating that the minimum requirements have been met. SM 500.24 states that to receive Director's approval the database manager must demonstrate that the database meets the following minimum requirements

- (a) Contain only data that have undergone rigorous, documented quality assurance. In cases where the data were not collected by the USGS, specific information must accompany the data stating its origin, and the data must have undergone rigorous quality assurance checks by the USGS or the originating organization.*
- (b) Adequately supported by the operating division by designating a database manager to ensure system maintenance, user documentation, and access.*
- (c) Be subjected to periodic review under the direction of the database manager to assure that the data have been properly and accurately entered.*

Section 500.24 also states that the request should document the following:

- Purpose and scope of the database,
- Data collection methods,
- Quality control methods,
- Description of data elements,
- Instructions for operating any software that may accompany the data set, and
- User support.

The attachment to this memorandum documents and details how the BioData System meets the specific requirements of SM 500.24 as described above.

Distribution:

BioData Steering Committee

Peter Ruhl, Database Manager, Office of water Quality/NAWQA Program, Reston, VA

Stephen R. Moulton II, Asst. Program Coordinator, NAWQA Program, Reston, VA

Michael D. Woodside, Asst. Program Coordinator, NAWQA Program, Reston, VA

Katherine Lins, Chief, Office of Water Information, Reston, VA

Robert B. Swanson, Center Director, Nebraska Water Science Center, Lincoln, NE

Gerard Guala, Core Systems Science, Reston, VA

Andrea Ostroff, Core Systems Science, Reston, VA

Barry C. Poulton, Columbia Environmental Science Center, Columbia, MO

Nathan L. Booth, Center for Data Analytics, Middleton, WI

Attachment

Attachment to Memorandum Requesting Director's Approval for BioData Database System

Purpose and Scope of BioData

BioData provides USGS with a secure, web-enabled data management system for capturing, storing, archiving, curating, and distributing aquatic bioassessment data collected or compiled by USGS science projects. Bioassessment data typically describe taxonomic composition and taxa abundances of macroinvertebrate, fish, or algal communities, and often include supporting habitat data. These data are used to characterize and investigate the condition of water dependent and water related ecological resources. BioData use by USGS employees is voluntary.

BioData has two major subsystems. The first is a data management subsystem that provides field biologists with data input, review, approval, and related data management capabilities through a password-protected web application. This subsystem also supports communication and data exchange between field biologists and partner laboratories and serves as a permanent data archive.

The second is a data-retrieval subsystem used to distribute data to scientists, resource managers, and the public. Data are available from the retrieval system after data have been reviewed and approved. Data can be obtained from a public web site (<http://aquatic.biodata.usgs.gov>) or, by authorized users, from an internal web site that provides access to restricted data. In the future, web data services can be built to support automated system-to-system data retrieval. Web data services are currently used to provide data for the USGS Water Mission Area's Annual Data Report (ADR).

Methods for Data Collection and Laboratory Analysis

A requirement of BioData is that all data are produced using methods that meet USGS Fundamental Science Practices. The current version (V. 1.0) of BioData supports scientific investigations of aquatic biology conducted by any USGS scientist and or technicians at any stream site registered in the National Water Information System (NWIS). At this time, BioData supports two national-level sampling protocols: (1) the USGS National Water-Quality Assessment (NAWQA) Program Biological Sampling and Analysis protocols, and (2) the U.S. Environmental Protection Agency (USEPA) National Rivers and Streams Assessment protocols. However, BioData is designed so that it can be enhanced to support additional sample-collection and laboratory-analysis protocols. Additional protocols and methods will be reviewed prior to acceptance to ensure that they meet USGS Fundamental Science Practices as found in Survey Manual chapters 502.3 and 502.4.

The OWQ approves all sample collection and laboratory methods prior to data being included in BioData. Accepted protocols and methods will be reviewed to ensure that they meet USGS Fundamental Science Policy as found in Survey Manual sections 502.3 and 502.4 and Office of

water Quality Technical Memorandum 98.05. Methods are documented within the database using a set of required core metadata elements and citations to external documentation such as published protocol and methods documents.

BioData currently stores data produced using two national protocols: (1) The USGS NAWQA Biological Sampling, Habitat, and Laboratory Protocols (<http://water.usgs.gov/nawqa/protocols/bioprotocols.html>), and (2) the US Environmental Protection Agency (USEPA) National Rivers and Streams Assessment (NRSA) sampling protocols (http://water.epa.gov/type/rsl/monitoring/riverssurvey/riverssurvey_index.cfm).

Field data are input and reviewed using a password-protected web application. Laboratory data are uploaded using the same application. Security controls ensure that only authorized personnel can enter, edit, or view data prior to public release.

Database Review, Quality Assurance Methods, and Quality Control Checks

BioData supports data-review workflows and includes several data quality control features that ensure data comply with USGS Fundamental Science Practices for data quality.

1. Data must be produced using USGS-approved sampling and laboratory methods, and all data stored are traceable to the methods used to produce it.
2. Automated data validation and verification checks flag erroneous and suspect data. This occurs immediately when data are being entered or uploaded so that biologists can immediately address and resolve data quality issues. In addition, particular workflows are blocked until the underlying data pass all of the validation and verification checks. For example, a user cannot transmit an analytical request to a lab if necessary sampling information is incomplete or invalid.
3. Data must be coded as reviewed and accepted by USGS and must pass all validation checks before it is released to the public.
4. Laboratories must complete an authorization process in which they demonstrate compliance with business rules and data standards before data they produce are accepted into the system.
5. Taxonomic nomenclature is controlled and managed by domain experts who maintain name-validation rules for incoming data, map stored data to national and international standards such as the Integrated Taxonomic Information System (ITIS), and curate stored data to account for scientific advancements.
6. At this time, sampling sites must be registered in the USGS National Water Information System (NWIS). This requirement ensures that necessary site information is provided and enforces numerous data-quality rules independently of BioData. This also facilitates integration with other USGS data assets associated with NWIS sites.

Description of Data Elements

BioData stores numerous data elements that document sample collection procedures, locations, conditions, field and lab processing, and results of taxonomic identification and enumeration. The data elements are selected with three goals in mind: (1) to meet the particular data requirements of supported sampling and laboratory methods, (2) to provide information needed to evaluate fitness-of-use for unforeseen analyses, and (3) to preserve the long-term value of the data. Metadata elements ensure that data adhere to USGS Fundamental Science Practice standards for documentation (<http://www.usgs.gov/usgs-manual/500/502-4.html>).

User Support and instructions for operating any software that may accompany the data set

User support and operating instruction are the responsibility BioData project manager and user group. Support is provided through several channels.

1. Context-sensitive help links are embedded in the BioData web application graphical user interface. These provide data entry tips and guidance for specific data fields and data screens.
2. Users can ask questions and make suggestions by submitting an online form while they are using the application.
3. Users can call or contact the USGS Service Desk directly for tier 1 support. If the Service Desk cannot resolve the problem the issue is passed along to the BioData Team for further investigation.
4. A public BioData wiki (<https://my.usgs.gov/confluence/display/biodata/Home>) is used to provide more detailed user documentation and online training materials. Guidance for taxonomic laboratories, for example, is made available in the wiki.

Database Management

BioData is governed by an executive Steering Committee that sets policy, establishes priorities, and identifies and obtains resources needed to develop and maintain the system. The Steering Committee is chaired by the Chief, Office of Water Quality and includes technical representatives from Science Centers and Mission Areas.

Day-to-day management is the responsibility of a project manager who coordinates and directs the activities of a technical development team located at the Center for Integrated Data Analytics (Middleton, WI) and a user group comprised of aquatic ecologists and taxonomists distributed nationwide. The user group is responsible for specifying and prioritizing scientific requirements, conducting user testing, and providing user documentation and training.

Office of Water Quality Technical Memorandum 98.05 – Policy for the Approval of (USGS) Water-Quality Analytical Methods. Accessed July 16, 2012 at <http://water.usgs.gov/admin/memo/QW/qw98.05.html>

U.S. Geological Survey Manual (SM) chapter 500.24 – Policy for Release of Computer Databases and Computer Programs, Dated April 9, 1993. Accessed July 16, 2012 at <http://www.usgs.gov/usgs-manual/500/500-24.html>

U.S. Geological Survey Manual (SM) chapter 205.18 – Authority to Approve Information Products, dated May 24, 2006. Accessed July 16, 2012 at <http://www.usgs.gov/usgs-manual/200/205-18.html>

U.S. Geological Survey Manual (SM) Chapter 502.3 – Fundamental Science Practices: Peer Review, dated May 24, 2006. Accessed July 16, 2012 at <http://www.usgs.gov/usgs-manual/500/502-3.html>

U.S. Geological Survey Manual (SM) Chapter 502.4 – Fundamental Science Practices: Review, Approval, and Release of Information Products, dated May 24, 2006. Accessed July 16, 2012 at <http://www.usgs.gov/usgs-manual/500/502-4.html>