



United States Department of the Interior

U. S. GEOLOGICAL SURVEY
Reston, VA 20192

In Reply Refer To:
Mail Stop 412, 440

February 8, 2016

OFFICE OF WATER QUALITY TECHNICAL MEMORANDUM 2016.03 OFFICE OF WATER INFORMATION MEMORANDUM 2016.01

SUBJECT: Policy and Guidelines for Storing Data in the Aquatic Bioassessment Data Management System (BioData)

PURPOSE

This memorandum establishes the Aquatic Bioassessment Data Management System (BioData) as the authoritative repository for U.S. Geological Survey (USGS) Water Mission Area (WMA) aquatic community- and population-level taxon identification and enumeration, and associated habitat data.

BACKGROUND

Data created by or on behalf of the USGS are the property of the Federal Government, and the USGS is required to retain and preserve an authoritative or original copy of all data for which it is responsible ([OSQI Instructional Memorandum 2015-04](#)). Data that are appropriate for public release must be accessible, useable, and available in machine-readable formats ([OSTP, 2013](#); [OMB, 2013](#)). USGS Fundamental Science Practices (FSP) establishes policies and best practices for meeting these responsibilities using the conceptual framework of the USGS Science Data Lifecycle ([OSQI Instructional Memorandum 2015-01](#); [Faundeen and others, 2013](#)). These FSP policies and practices include:

- Carefully documenting methods and techniques used for collecting data, and the methods used to process data to verify, organize, transform, repair, and integrate it with other data sets ([USGS Manual 502.2](#)).
- Incorporating, quality-assurance processes and procedures in all phases of the project ([Faundeen and others, 2013](#)).
- Backing up and securing data from accidental or intentional destruction or alteration ([OSQI Instructional Memorandum 2015-04](#)).

- Preserving data so that it remains accessible and useful to future generations ([OSQI Instructional Memorandum 2015-04](#)).
- Publishing and sharing data in machine-readable standard formats, accompanied by metadata that describe its characteristics ([OSQI Instructional Memorandum 2015-04](#)).

In 2011, [BioData](#) was released for general use within the USGS. BioData is a Director's approved database ([OWQ, 2012](#)) that meets FSP requirements for providing a managed, secure, and curated long-term data repository for USGS taxonomic results and supporting habitat data. BioData also facilitates quality-assurance processes and disseminates data through a variety of channels, including BioData internal and public web sites, the USGS [BISON](#) system for North American taxon occurrence records, and the [Water Quality Portal](#).

Since 2011, BioData has provided support for fish, macroinvertebrate, and algae data collected using standard USGS protocols developed by the National Water-Quality Assessment (NAWQA) Program, and national protocols promulgated by the U.S. Environmental Protection Agency (EPA) National Rivers and Streams Assessment (NARS). Recently, BioData was enhanced to add support for a wide variety of sampling protocols for fish and benthic macroinvertebrate communities. BioData is currently capable of storing the types of community-level bioassessment data that are most commonly collected for USGS water-quality investigations of streams and rivers. Support for other environmental settings and taxonomic groups can be added as need and resources dictate.

BioData has many features that make it a valuable data-management tool and long-term repository for the WMA. Data are protected in a secure, backed-up database housed at the EROS data center and can be easily accessed using a web browser connected to the internet. Projects control access for entering and editing data, and BioData runs a suite of data validation checks to assist with efficient, routine, and repeatable quality-assurance reviews. Data are embargoed from public release until they have passed automated quality checks and the project codes the data as having been reviewed and accepted. Taxonomic nomenclature is maintained and curated by taxonomists to ensure that (1) the original record is preserved and (2) all of the data are also reported using an up-to-date standardized taxonomic system and, when appropriate, Integrated Taxonomic Information System ([ITIS](#)) taxon serial numbers. This makes it very easy for data consumers to harmonize taxonomy across data produced over a long period of time – taxonomic synonymy and ambiguity are resolved for them.

POLICY

All aquatic community- and population-level taxon identification and abundance data produced by or on behalf of WMA Programs, Water Science Centers (WSC), Technical Offices, and National Research Program (NRP) Centers (collectively referred to as "Centers" in this memorandum) must be recorded in BioData unless the data are not compatible with BioData storage capabilities or the Center determines that it is not responsible for preserving the authoritative version of the data. This may be the case, for example, if data were produced by a

collaborator or cooperator who has agreed to assume that responsibility. When biological data are produced in cooperation with other USGS Mission Areas or non-USGS entities, ownership, data management, and preservation responsibilities must be clearly stated in a documented agreement between the parties ([OSQI Instructional Memorandum 2015-04](#)). If responsibility is assigned to a Center the data should be stored in BioData (if compatible) or in an alternative repository that meets FSP requirements for data preservation, documentation, and long-term availability (if incompatible). Duplicating data for storage in BioData and another repository (e.g., a cooperator's database) is discouraged, but if this is necessary, a data sharing agreement must stipulate that BioData contains the authoritative record.

Because BioData is a Director's approved Database, public release of results from BioData fulfills USGS data publication and dissemination requirements.

IMPLEMENTATION GUIDANCE

How does this policy apply to new, ongoing, and completed projects?

Effective immediately, this policy applies to all new projects conducted by Centers with aquatic-biological data-collection activities. Ongoing studies with data collected prior to this memo should consult with the Office of Water Quality (OWQ) to determine the feasibility of importing previously-collected data, and about using BioData for the remainder of the project. This policy does not apply to completed projects, although Centers are encouraged to contact the OWQ to explore the technical feasibility and resources needed to import data that (1) represent a long period of record, (2) describe an important geographic area or environmental setting (e.g., National Park), (3) are at high risk of loss, or (4) are not currently stored in a secure, permanent, accessible repository and the Center wishes to use BioData for that purpose.

How should responsibility for managing and preserving the authoritative version of the aquatic biological data be determined?

If the Center produces the data directly or supervises its production by a third party, the Center is responsible, and if possible, the data should be stored in BioData. This is the simple case in which a Center designs the study, collects the samples, monitors the laboratory, and quality assures the data or directs and monitors another entity performing this work.

If a Center funded by another USGS Mission Area or Region produces the data for use in a collaborative study, responsibility for data management and preservation should be clearly assigned and documented. In this case, responsibility will rest with a USGS entity (another Mission Area, the Region, or the Center) and use of BioData is both appropriate and encouraged.

If a non-USGS partner or cooperator produces the data, long-term data management and preservation is negotiated between the Center and the partner. Responsibilities for data management and preservation should be clearly defined and documented. This case would apply, for example, to projects in which the cooperator collects and analyzes the biological data and the Center collects and analyzes water quality and/or flow data. Each party contributes to the overall

project data set but each party is independently responsible for the data they produced. Data *may* be stored in BioData if it has been thoroughly reviewed and meets USGS standards for data quality. This is consistent with current policy for storing furnished record in NWIS ([WRD Memorandum 2008.01](#)). The decision to accept or not accept furnished data must balance several factors including the benefits and costs associated with accepting the specific furnished data.

In general, it is unwise to add a static copy of data from one database to a second database because later modifications to the original data often fail to reach the copy. A better practice is for the original database to maintain and revise the data as needed, while the second database points to the original and allows users to retrieve combined datasets from both. When evaluating the advisability of storing furnished record consideration of the costs, risks, and alternatives to maintaining a copy of data also stored elsewhere should be considered.

How should compatibility with BioData storage capabilities be determined?

Project planning should include an assessment of the types of biological data that will be collected and how they will be managed, preserved, and published. The [BioData Support website](#) provides up-to-date guidance, including a compatibility checklist that can be used for an initial assessment. Assistance is also available by emailing biodata-support@usgs.gov or by selecting the “contact us” link on the BioData support web site. Inquiries will be fielded by the BioData User Group (BUG) – usually by the next business day. The BUG is a group of WMA and Center biologists that provides user support and training, manages the BioData taxonomic nomenclature, act as liaisons between OWQ and taxonomic laboratories, and guide overall system development.

BioData can accept machine readable files if provided in required formats from laboratories within and outside the USGS. Taxonomic laboratories should be identified as early as possible so the project team, BUG, and laboratory can collaborate to register analysis procedures, reconcile taxonomic nomenclature, and ensure that the laboratory can deliver data that meets the BioData batch loading requirements.

Sample-collection and analytical methods must be available to the general public and should be registered in the National Environmental Methods Index ([NEMI](#)) catalogue. Published, peer-reviewed method documentation is preferred. Data from research, unpublished, and uncommon methods may be used, but the data may be embargoed from public release. Some projects may want to store data before publishing the methods in an interpretive report or other approved information product. In those cases interim documentation must be provided and data must be linked to the approved report when it is published. The BioData Support web site has a list of supported methods and a linked form for requesting additions.

How should data be preserved if they are not compatible with BioData storage capabilities?

As noted in the policy section above, data that are not compatible must be stored in an alternative repository that meets FSP requirements for data preservation, documentation, and long-term availability. An important component of this is to make the data readily available in machine-readable formats. A spreadsheet on a personal computer or a Center server that is not available to the public does not meet that requirement. One way to meet FSP requirements is to (1) describe the data using ISO-compliant metadata, and (2) register and deposit the data in [ScienceBase](#). For more information about best practices see the USGS [data management website](#).

Donna N. Myers /s/
Chief, Office of Water Quality

Nathaniel L. Booth /s/
Chief, Office of Water Information

Distribution: All Water Mission Area Employees

REFERENCES

Faundeen, J.L., Burley, T.E., Carlino, J.A., Govoni, D.L., Henkel, H.S., Holl, S.L., Hutchison, V.B., Martín, Elizabeth, Montgomery, E.T., Ladino, C.C., Tessler, Steven, and Zolly, L.S., 2013, The United States Geological Survey Science Data Lifecycle Model: U.S. Geological Survey Open-File Report 2013–1265, 4 p., available on the World Wide Web at <http://dx.doi.org/10.3133/ofr20131265>.

OMB, 2013, Office of Management and Budget Memorandum M-13-13, Open Data Policy, managing data as an asset: accessed August 7, 2015 at <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>

OSQI Instructional Memorandum 2015-01, Scientific data management foundations: U.S. Geological Survey Office of Science Quality and Integrity, accessed August 7, 2015 at <http://www.usgs.gov/usgs-manual/im/IM-OSQI-2015-01.html>

OSQI Instructional Memorandum 2015-04, Preservation requirements for digital scientific data: U.S. Geological Survey Office of Science Quality and Integrity, accessed August 7, 2015 at <http://www.usgs.gov/usgs-manual/im/IM-OSQI-2015-04.html>

OSTP, 2013, Memorandum for the heads of executive departments and agencies – Increasing access to the results of federally funded scientific research: Executive Office of the President, Office of Science and Technology Policy: accessed August 7, 2015, at http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

OWQ, 2012, Policy for release of computer databases: request director's approval for BioData Database System, Office of Water Quality Memorandum, July 23, 2012, accessed August 7, 2015 at <https://my.usgs.gov/confluence/x/W4MhI>

USGS Manual 502.2, Fundamental science practices - definition of data documentation:
accessed August 7, 2015 at <http://www.usgs.gov/usgs-manual/500/502-2.html>

WMA Memorandum 13-01, Programs and plans – guidelines for preparation, submission, and approval of Water Science Center project proposals: accessed August 14, 2015 at <http://water.usgs.gov/admin/memo/policy/wmapolicy13.01.pdf>

WRD Memorandum 2008.01, Water Resources Discipline policy on accepting furnished records:
accessed October 1, 2015 at <http://water.usgs.gov/admin/memo/policy/wrdpolicy08.01.html>