

BioData: A National Aquatic Bioassessment Database

By Dorene MacCoy

Introduction

BioData is a U.S. Geological Survey (USGS) web-enabled database that for the first time provides for the capture, curation, integration, and delivery of bioassessment data collected by local, regional, and national USGS projects. BioData offers field biologists advanced capabilities for entering, editing, and reviewing the macroinvertebrate, algae, fish, and supporting habitat data from rivers and streams. It offers data archival and curation capabilities that protect and maintain data for the long term. BioData provides the Federal, State, and local governments, as well as the scientific community, resource managers, the private sector, and the public with easy access to tens of thousands of samples collected nationwide from thousands of stream and river sites. BioData also provides the USGS with centralized data storage for delivering data to other systems and applications through automated web services.

BioData allows users to combine data sets of known quality from different projects in various locations over time. It provides a nationally aggregated database for users to leverage data from many independent projects that, until now, was not feasible at this scale. For example, from 1991 to 2011, the USGS Idaho Water Science Center collected more than 800 bioassessment samples from 63 sites for the National Water Quality Assessment (NAWQA) Program and more than 450 samples from 93 sites for a cooperative USGS and State of Idaho Statewide Water Quality Network (fig. 1). Using BioData, 20 years of samples collected for both of these projects could be combined for analysis.

BioData delivers all of the data using current taxonomic nomenclature, thus relieving users of the difficult and time-consuming task of harmonizing taxonomy among samples collected during different time periods. Fish data are reported using the Integrated Taxonomic Information Service (ITIS) Taxonomic Serial Numbers (TSN's). A simple web-data input interface and self-guided, public data-retrieval web site provides access to bioassessment data. BioData currently accepts data collected using two national protocols: (1) NAWQA and (2) U.S. Environmental Protection Agency (USEPA) National Rivers and Streams Assessment (NRSA). Additional collection protocols are planned for future versions.



Highlights

Use the BioData Management System to:

- Capture, review, edit, finalize, approve, and archive bioassessment data collected for local, regional, and national aquatic bioassessment, monitoring, and research purposes.
- Store data collected using two nationally accepted protocols with the capability to add support for additional protocols in the future.

Use the BioData Retrieval System to:

- Retrieve bioassessment data through a single public web site (<https://aquatic.biodata.usgs.gov>).
- Retrieve data using a unified and consistent taxonomic identification system that accounts for changes in taxonomic nomenclature.

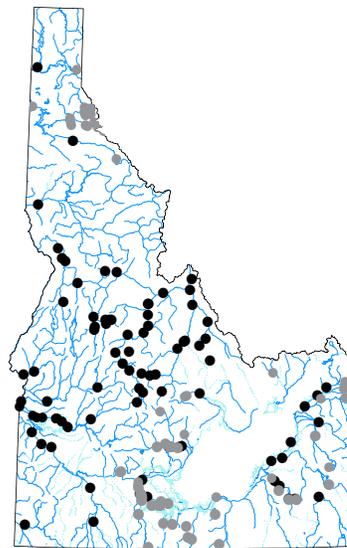


Figure 1. BioData users are able to retrieve data from multiple projects. An example in progress is the capability to retrieve sites in Idaho from the National Water-Quality Assessment Program (gray circles) and the USGS-Idaho Statewide Cooperative Network (black circles). (Idaho rivers and streams are blue lines).

BioData Management System

The BioData management system provides field biologists with data input, review, approval, and management capabilities through a password-protected web application (fig. 2). Data review, revision, and automated data-validation tools make quality-assurance practices more efficient and routine. BioData management system also supports communication and data exchange between field biologists and partner laboratories. Once approved, data are made available from the data-retrieval web site.

Selected Data Input Features

- Manage and organize samples by project. **A**
- Manage data entry and editing rights for each project.
- Enter data using self-guided data-entry forms with built-in help and data validation. **B**
- Auto-save – data are securely and instantly saved for retrieval.
- Review and validate data to find and fix problems quickly and easily.
- Manage laboratory orders and notification of laboratories through email when a project places or modifies an order. **C**
- Review individual taxa records—select and mark identifications for further review and verification. **D**
- Print shipping lists, bottle labels, and USGS National Water-Quality Laboratory (NWQL) Analytical Services Request forms.
- Access to online documentation for laboratories that includes taxonomic reference lists and requirements for data upload.
- Validate data from laboratories automatically – view and download laboratory data-validation reports.

The screenshot displays the USGS BioData web application interface. At the top, the USGS logo and navigation links are visible. The main content area is divided into four sections, each marked with a letter in a blue circle:

- Section A: Select a Project** shows a 'Study Reach List' table with columns for Valid, Actions, Review Status, Project Name, Project Owner, Project Label, Science Center, and Date. It lists projects like 'Silver Creek TNC Project' and 'USGS Idaho Statewide Water Quality'.
- Section B: Community Samples** shows a 'Community Sample List' with filters for Site, Community, Sample Type, Collection Year, Review Status, and Protocol. It includes a table with columns for Valid, Actions, SIDNO, Review Status, Community, Sample Type, Protocol, Collection Date, Site Number, Status, and Lab Order.
- Section C: Manage Lab Orders*** shows a 'Lab Order List' with filters for Community, Status, Review, and Placed Water Year. It includes a table with columns for Valid, Actions, Status, Placed Date, Lab Order, Lab, Contains, Sample, and Last Update.
- Section D: Review Taxa Records** shows 'Results Type' (Fish, Invertebrate, Algae) and a 'Results Selector' (All, Has Results, No Results, Not Reviewed, Partially Reviewed, Reviewed). It includes a table with columns for SIDNO, Review Complet..., Actions, Sample Type, Collection Date, Site Name, and Site ID. Below this is a 'Taxonomic Results' table with columns for Review Status, Method, Common Name, Total Length (mm), Weight (g), Abundance, and Condition Factor.

Figure 2. BioData input screens showing how to (A, B) manage samples, enter, review, revise, and approve data from field and laboratory records; (C) create laboratory order forms; and (D) download reports.

BioData Retrieval System

The BioData retrieval system, <https://aquatic.biodata.usgs.gov>, is a one-stop location for USGS bioassessment data from across the United States (fig. 3). Scientists, resources managers, and the public can use this centralized, web-accessible database to view, browse, access, and download bioassessment data.

Selected Retrieval Features

- Easy step-wise data selection features using a variety of criteria. **E**
- View the number of samples before download. **F**
- Save data-selection criteria for future use, email the file to collaborators. **G**
- Preview datasets before they are retrieved. **H**
- Choose file formats for data retrieval – Excel, tab-delimited, comma delimited, XML. More formats may be added in the future. **I**
- A standardized, unified taxonomic system is used for fish, invertebrates and algae — synonymy is resolved. Fish data sets include ITIS TSN's. ITIS TSN's are planned for invertebrates and algae over time commensurate with advances in ITIS.

USGS
science for a changing world

USGS Home
Contact USGS
Search USGS

BioData - Aquatic Bioassessment Data for the Nation

Select Filter Criteria **E**
Choose filters and set filter criteria from one or more sections, then select "Next".
Filter(s) "State" and "Site Type" and "Taxon Fish Family" will return 38 samples

Biological Community/Data Type and Sample Type Filter

- No filter selected in this section [Filter Definition Help](#)
- Biological Community/Data Type
- Sample Type

Temporal Filter **F**

- No filter selected in this section [Filter Definition Help](#)
- Date Range
- Month/Year

Geographic Location Filter

- No filter selected in this section [Filter Definition Help](#)
- State
 - Hawaii
 - Idaho
 - Illinois
- County
- Latitude/Longitude
- Drainage Basin/Hydrologic Unit Code
- Sites

Organization / Program Filter

- No filter selected in this section [Filter Definition Help](#)
- USGS Program
- USGS Science Center
- Project
- NAWQA Study Unit

Site Characteristic Filter

- Elevation
- Site Type
 - Stream - Stream
 - Stream - Canal
 - Stream - Tidal Stream
- Drainage Area

Taxon-specific Filter (Advanced)

Search by: Phylum Class Order Family

- No filter selected in this section [Filter Definition Help](#)
- Fish
 - Salmonidae
 - Cottidae
- Invertebrate
- Algae

Next >>

Select Data Set(s) **H**
Select the data sets you would like to obtain, then select "Next".

Category	Data Set Name	Matching Records
<input type="checkbox"/> Project	BioData Project	3
<input checked="" type="checkbox"/> Samples	Sample Inventory	38
<input checked="" type="checkbox"/> Physical	Site Information, Study Reaches, Site Conditions	42
<input checked="" type="checkbox"/> Fish	Fish Sampling Information, Fish Methods and Subreaches, Fish Count, Fish Results	1199
<input type="checkbox"/> Invertebrate	Invertebrate Sampling Information, Invert Results, Invertebrate Data Analysis System (IDAS), Invertebrate Collection Point Microhabitat, Invertebrate Snag Microhabitat...	0
<input type="checkbox"/> Algae	Algae Sampling Information, Algae Collection Point Microhabitat, Algae Snag Microhabitat	0
<input type="checkbox"/> Habitat	Reach Level Detail, Longitudinal Detail, Transect Level Detail, Transect Point Detail, Channel Feature...	0

<< Previous Next >>

Preview and Download **I**

Preview: Fish Count
File Type: Microsoft Excel 2007
Download Data

The selected data set and filter criteria **Fish Count** will return record results fitting your criteria

Project Label	SiteNumber	Collection Date	Site Name	Common Name	Abundance
IdahoQWBI	13037500	2009-08-26	SNAKE RIVER NR HEISE ID	rainbow x cutthroat	2
IdahoQWBI	13037500	2009-08-26	SNAKE RIVER NR HEISE ID	cutthroat trout	9
Idaho QWBI	13060000	2009-08-25	SNAKE RIVER NR SHELLEY ID	mountain whitefish	11

1 2 3 4

Figure 3. BioData retrieval web site screens showing how to (E) select retrieval criteria; (F, G) view and save customized retrieval formats; and (H, I) preview and retrieve bioassessment data.

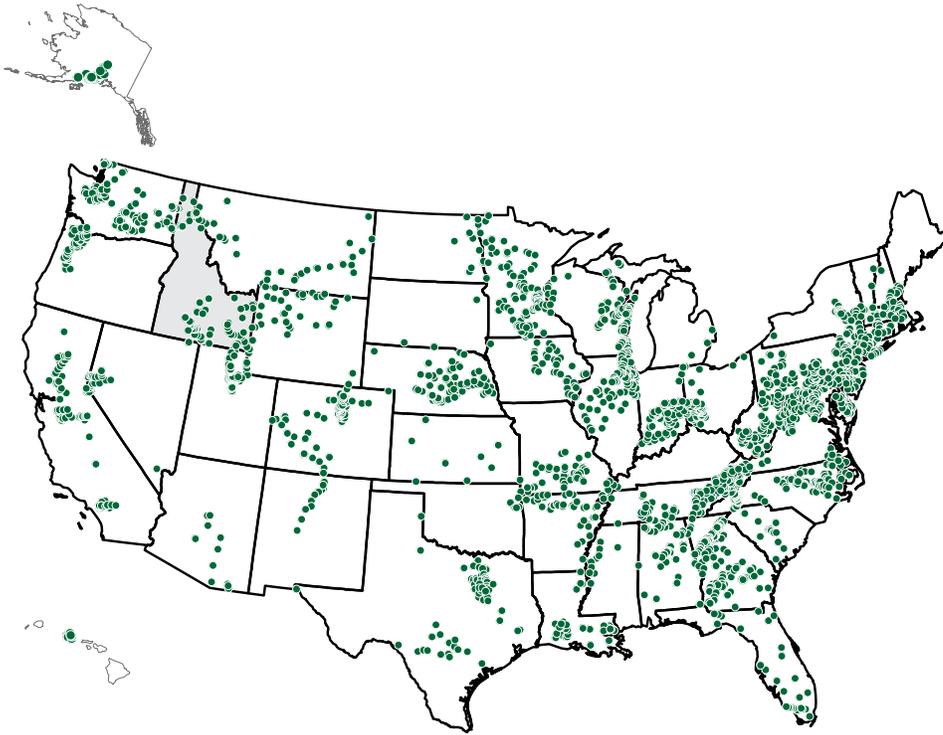


Figure 4. BioData currently stores over 15,000 samples from more than 2,000 sites collected since 1993 by the USGS National Water Quality Assessment (NAWQA) Program.

Get Started Using BioData

Users and uses of BioData:

- USGS can enter, store, and retrieve project bioassessment data
- Government agencies can retrieve the data for national, regional, statewide, and local research and assessment
- Academic professionals and students can retrieve the data for course work and academic research
- Research scientists can retrieve the data to test scientific hypothesis and report findings
- Resource managers can retrieve the data for assessments to evaluate criteria and goals
- Fishermen and outdoor enthusiasts can learn about invertebrate, fish, and algae data for their favorite recreational and fishing locations
- The general public can learn about aquatic organisms for a water body of interest

What BioData delivers to users:

- Tens of thousands of bioassessment sample data that include fish, invertebrates, algae, and supporting habitat data
- Bioassessment data from thousands of sites (fig. 4)
- Taxonomic lists of aquatic organisms found at select sites nationwide
- Step-by-step instruction on data input and retrieval

More information can be found on the BioData web site at <https://aquatic.biodata.usgs.gov> to:

- Obtain a login and become a BioData USGS technical user
- Retrieve data in three easy steps: (1) select filter criteria, (2) choose data sets, (3) preview and select download format
- Become a BioData laboratory

For More Information

Donna Myers, Chief, Office of Water Quality, dnmeyers@usgs.gov, (703) 648-5012

Pete Ruhl, BioData Project Manager, Office of Water Quality, pmruhl@usgs.gov, (703) 648-6841

BioData Users Group (USGS BUG)

Mitchell Harris, Illinois WSC, maharris@usgs.gov, (217) 328-9716

Jana Stewart, Wisconsin WSC, jsstewar@usgs.gov, (608) 821-3855

Scott Grotheer, National Water Quality Laboratory, Denver, sagrothe@usgs.gov, (303)236-3486

Dorene MacCoy, Idaho WSC, demaccoy@usgs.gov, (208) 387-1354

Sarah Spaulding, Fort Collins Science Center, sspaulding@usgs.gov, (303) 492-5158

Tracey Cadkin, Center for Integrated Data Analysis, tcadkin@usgs.gov, (608) 821-3877

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director