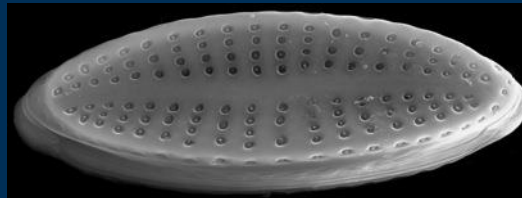


BioData Taxonomy

<http://aquatic.biodata.usgs.gov>
pmruhl@usgs.gov



Agenda

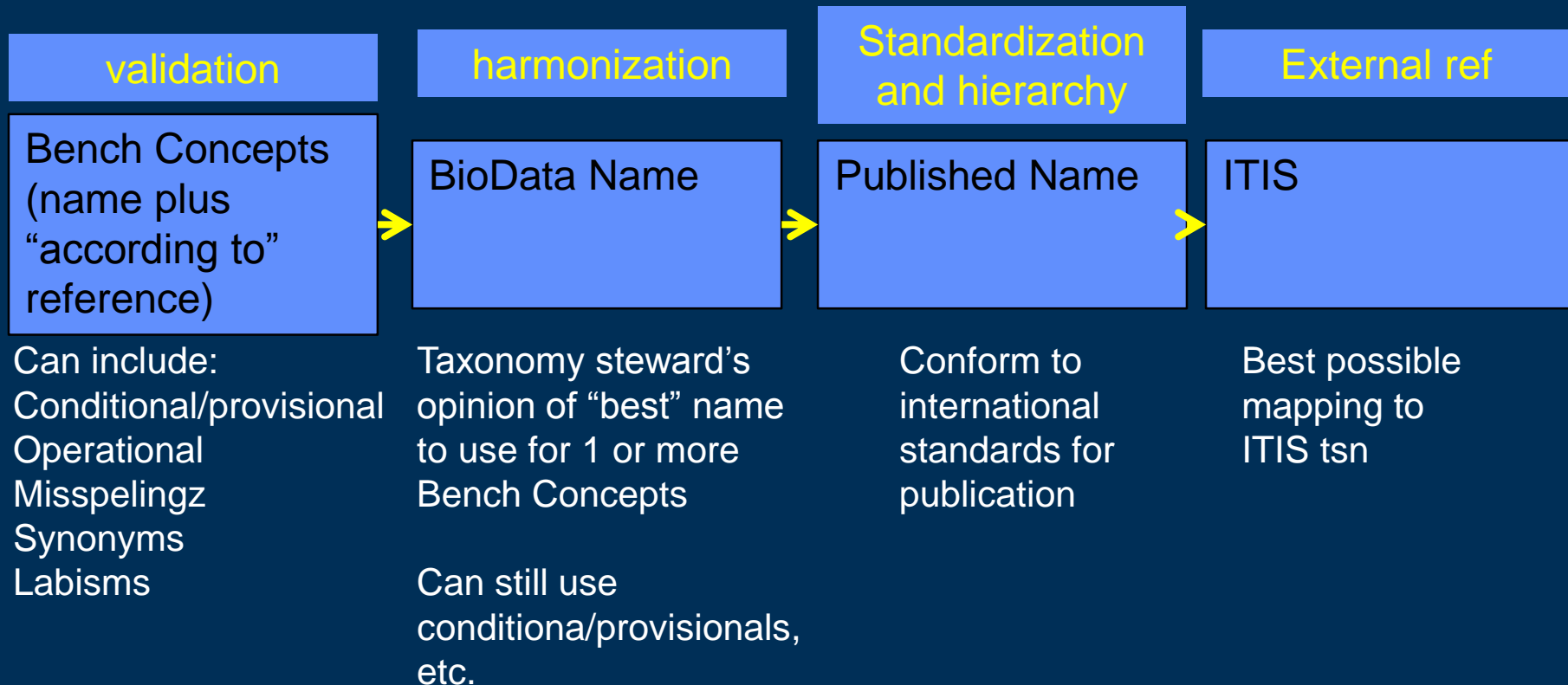
- Our big-picture goals
- What we do
- How we manage it all

Our Goals

- **Validate and document the data stored in BioData**
 - We add taxa “as-needed”
- **Store determinations at the greatest precision possible**
 - Provisional, conditional and operational names
 - Name defined by “according to” reference
- **Keep and report the original record**
- **Resolve synonymies and ambiguities for data users**
(map all of that original record into a “best current” system for reporting)
- **Report using phylogenetic ordering (not alphabetical)**

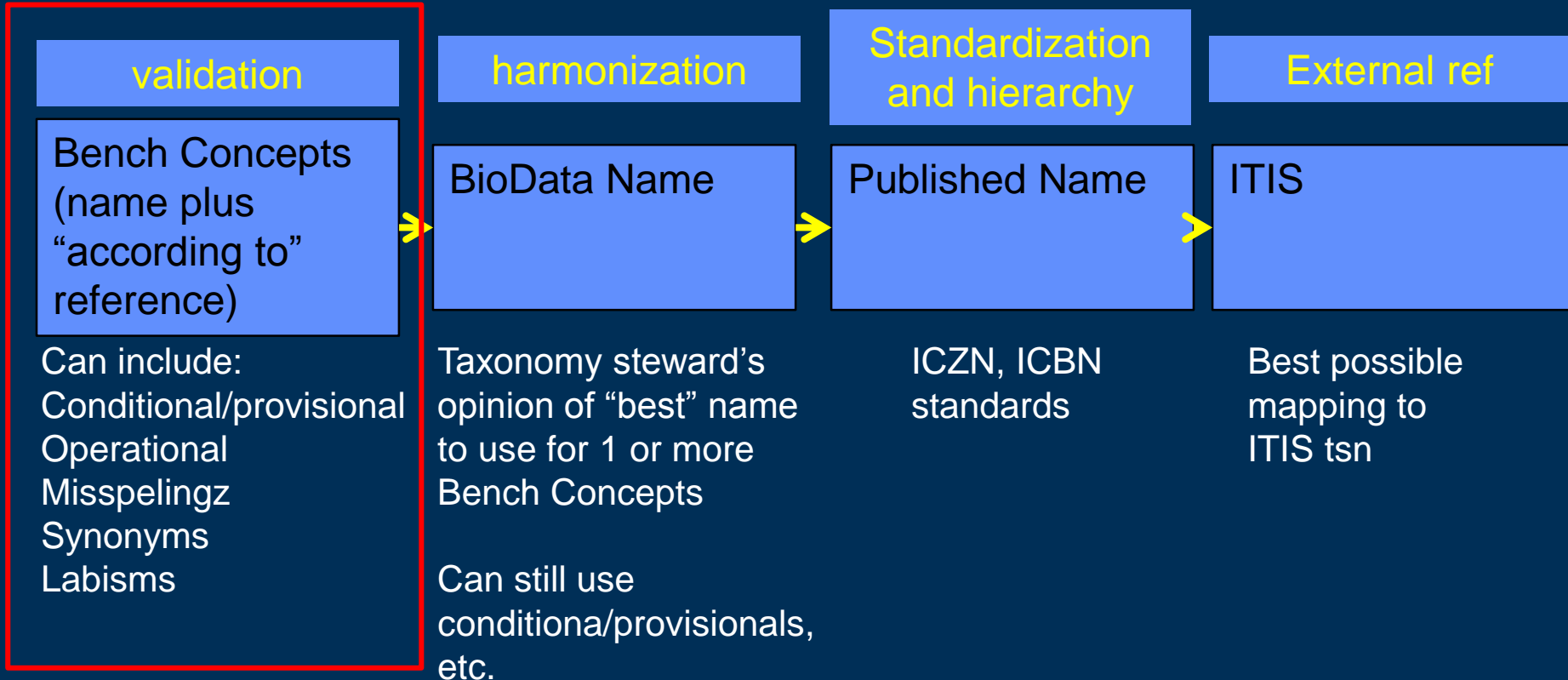
So What Do We Do?

A lot of mapping



So What Do We Do?

A lot of mapping



Bench Concepts

- Used to validate incoming data
- Composite logical key
 - name string + reference code + authorship*
 - only the name string is mandatory, because we still need to be able to accept nominal concepts
- For the macroinvertebrate user community it is customary to use *name string + reference code*
- For the algae user community it is customary to use *name string + reference code OR name string + authorship*
- For the fish user community it is customary to **common name**
Which we want to add to Bench level, but for now common name is taken from BioData Names level

Bench Concepts

- Cannot be deleted. Logical key cannot be edited
- Every Bench Concept is linked to one and only one BioData Name at a time – but which BioData Name it is linked to can change
- A Bench Concept can be linked to identifiers from other taxonomic systems (e.g. NADEDID, BUID)
- We can record additional useful bits of information about a Bench Concept

Bench Concepts

■ Lifecycle States

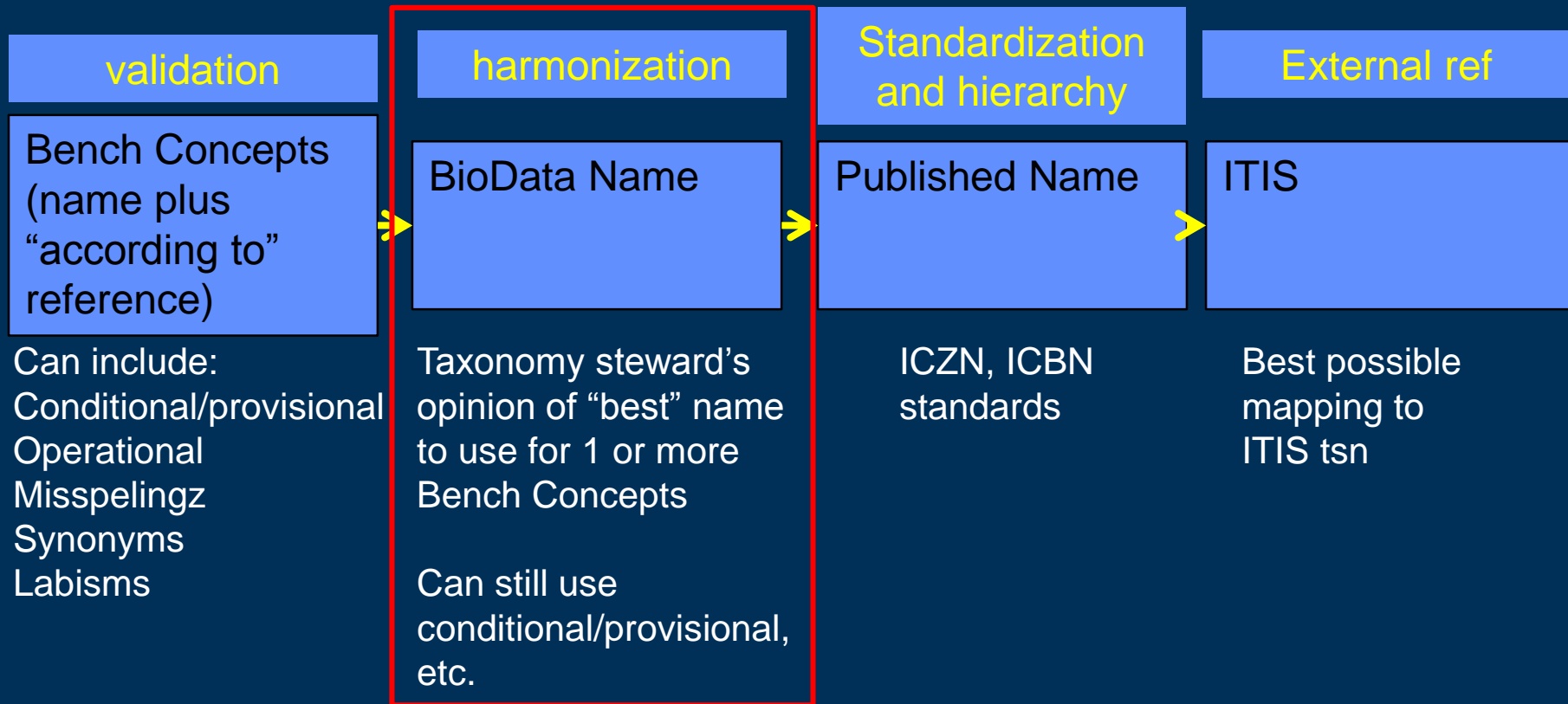
- **Current** – we will accept this concept no questions asked
- **Deprecated** – we will accept this concept but plan to retire it
- **Retired** – we will reject new data that use this concept

Examples of Bench Concepts

Bench Name	Reference	Status
Hydropsyche		RETIRED
Hydropsyche	Geraci and others, 2010	Current
Hydropsyche	Morse and Holzenthal, 1996	Deprecated
Hydropsyche	Morse and Holzenthal, 2008	Current
Hydropsyche	Wiggins, 1996	Current
Hydropsyche	Schuster, 1984	Current

So What Do We Do?

A lot of mapping



BioData Names (our OTU's)

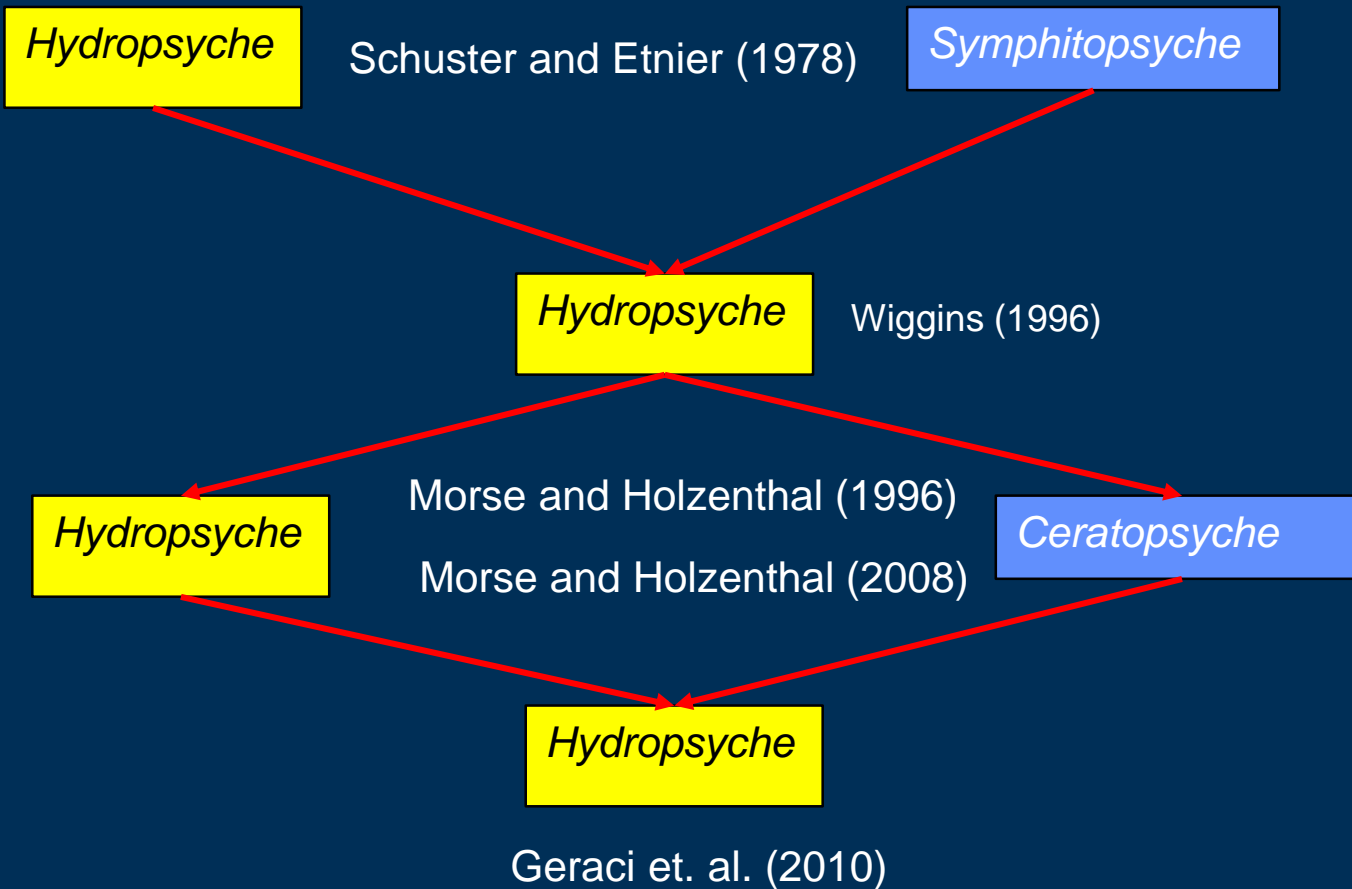
- Resolves synonyms from the Bench Concept level by combining them
- Adjustments to account for ambiguity that may affect older data (e.g. a recent split makes previous lumped nomenclature uncertain)
- Still preserves use of provisional, conditional names, groups, complexes, etc. that are not recognized by codes of nomenclature

BioData Names

- **Must be linked to a Published Name**
- **Can be associated with a common name**
- **Can be associated with a unique short name (code) that may be useful for statistical analysis and charting/graphing labels**
- **Can be designated as hybrids or as provisional/conditional names**

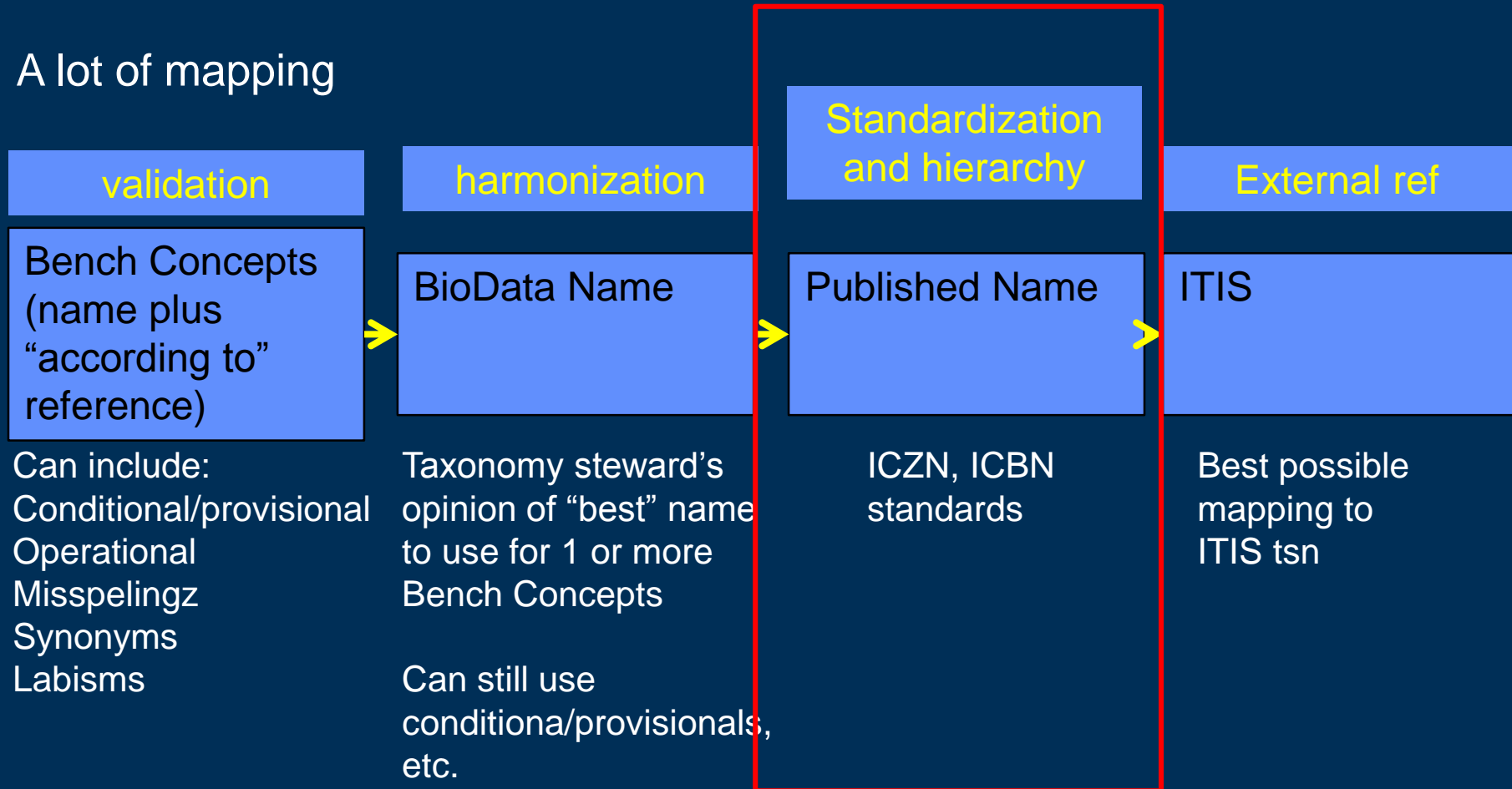
Examples of BioData Names

Bench Name	Reference	BioData Name
Hydropsyche		Hydropsyche
Hydropsyche	Geraci and others, 2010	Ceratopsyche/Hydropsyche
Hydropsyche	Morse and Holzenthal, 1996	Hydropsyche
Hydropsyche	Morse and Holzenthal, 2008	Hydropsyche
Hydropsyche	Wiggins, 1996	Ceratopsyche/Hydropsyche
Hydropsyche	Schuster, 1984	Hydropsyche



So What Do We Do?

A lot of mapping

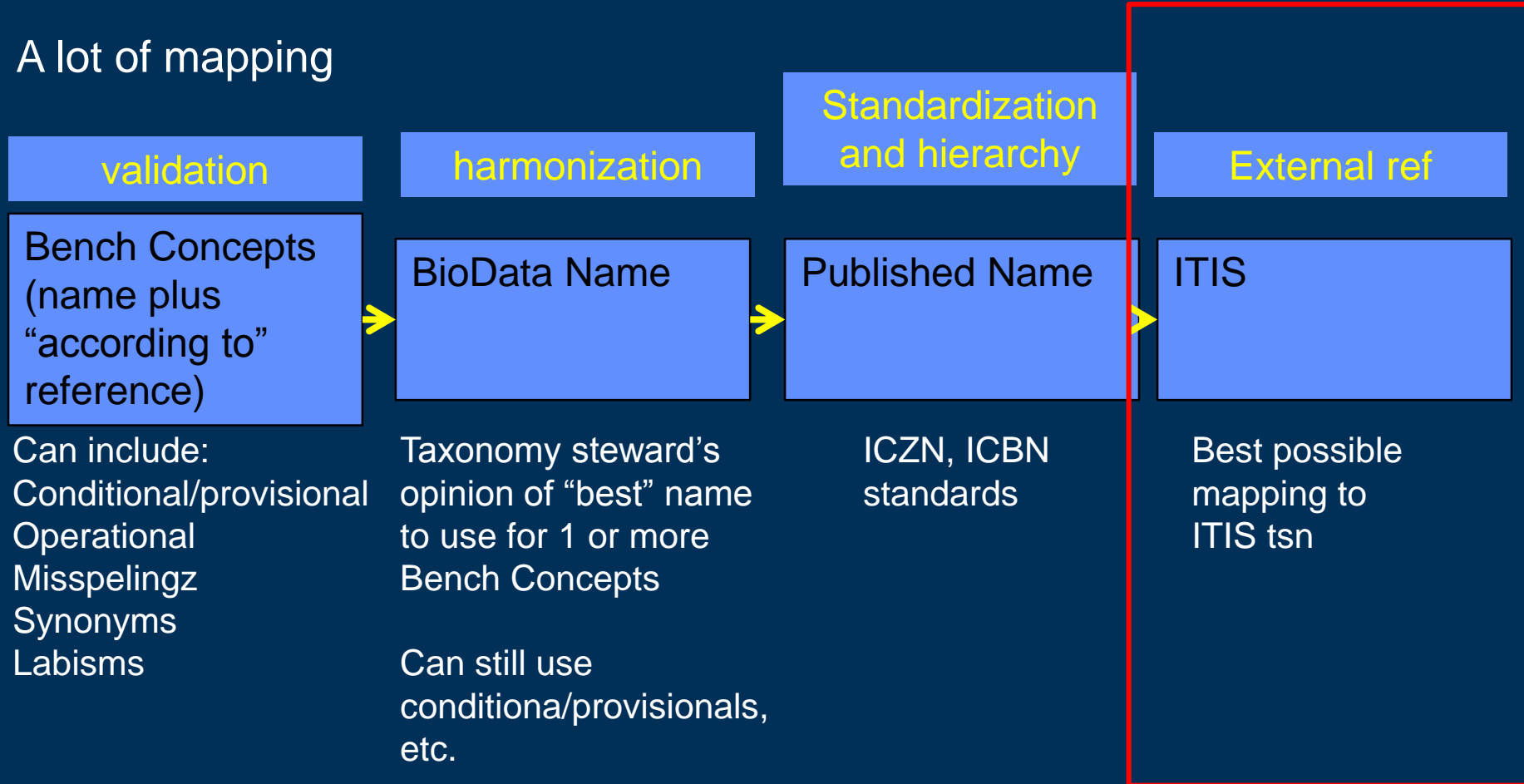


Published Names Hierarchy

- Validly published, conform to international standards for publication
- Curated by taxonomy stewards (we pick which published names and hierarchical relations to use)
- Can be edited, deleted and re-arranged, but not without publishing a new version of the BioData Taxonomy
- Phylogenetic ordering is controlled, and data can be reported that way – key user requirement
- Ragged hierarchy (can skip levels)
- Linked to ITIS
 - Exact matches
 - Inexact matches (sometimes have to link to higher rank in ITIS)

So What Do We Do?

A lot of mapping



Linking to ITIS

Bench Name	Reference	BioData Name	Published Name	ITIS tsn	ITIS Match Code
Hydropsyche		Hydropsyche	Hydropsyche	115453	Automated Exact Match
Hydropsyche	Geraci and others, 2010	Ceratopsyche/Hydropsyche	Hydropsychinae	598191	Automated Exact Match
Hydropsyche	Morse and Holzenthal, 1996	Hydropsyche	Hydropsyche	115453	Automated Exact Match
Hydropsyche	Morse and Holzenthal, 2008	Hydropsyche	Hydropsyche	115453	Automated Exact Match
Hydropsyche	Schuster, 1984	Hydropsyche	Hydropsyche	115453	Automated Exact Match
Hydropsyche	Wiggins, 1996	Ceratopsyche/Hydropsyche	Hydropsychinae	598191	Automated Exact Match

Managing It All

- **Dedicated stewards for each group (with decision-making authority)**
 - **Fish – AFS list, adjust using literature in between editions**
 - **Algae – coordinate with Diatoms of N. America**
 - **Inverts – primary literature and coordination with major government and private-sector labs**

Versioning

- Deliver snapshot with data retrieval
- Control and track versions
 - Version number is delivered with the count data
- All versions are archived and available

Versioning

- Two-part numbering scheme
- XX.YY
 - XX – major version
 - YY – minor version
- Major version – changes that can affect ongoing analysis
- Minor version – changes that will not affect ongoing analysis

Major Version Criteria

- Deletion, change (e.g. spelling correction), or Movement (e.g. associate a Family with a different Order) of a Published Name
- Deletion, change (e.g. spelling correction), or Movement (e.g. associate a BioData Name with a different Published Name) of a BioData Name
- Reassigning a Bench Concept to a different BioData Name.
- Reassigning a BioData Name to a different Published Name.
- Deletion, change or addition of an ITIS TSN
- Deletion, change, or addition of a BioData Common Name
- Addition or deletion of an ITIS TSN Match Code - since that would require the addition or deletion of the TSN
- Toggling of the Hybrid Flag in either a BioData Name or Bench Concept
- Deletion, change or addition of a BU_ID or NADEDID
- Changing the Rank (e.g. change from Species to Subspecies) of a Published Name
- Changing the Algae Group (e.g. green algae, red algae)

Minor Version Criteria

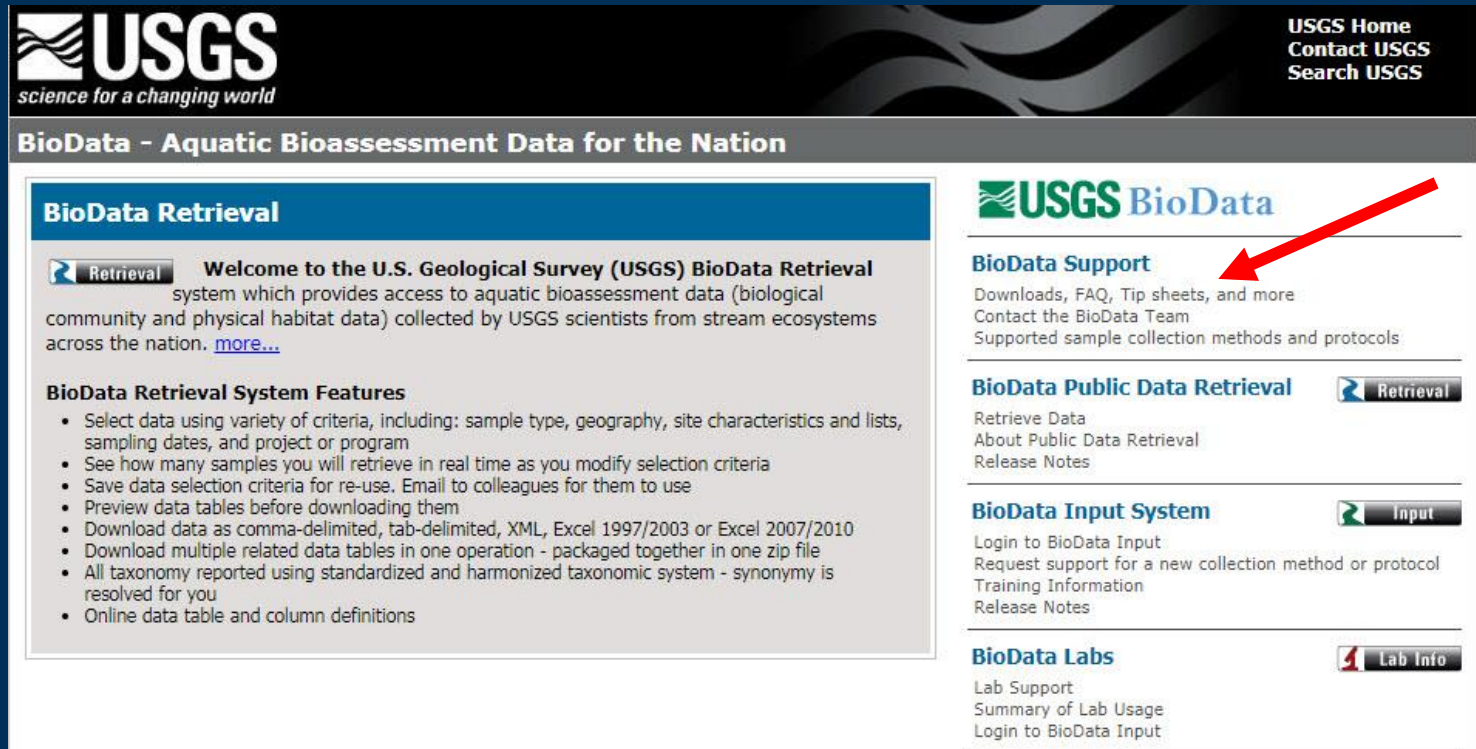
- Addition of a totally NEW Bench Concept, BioData Name, or Published Name
- Simple State change (e.g. deprecated to retired)
- Deletion, change, or addition of a BioData Short Name
- Deletion, change, or addition of an Authority to a Published Name
- Toggling of the Provisional/Conditional Flag in either a BioData Name or Bench Concept
- Change of an ITIS TSN Match Code
- Deletion, change, or addition of the Comment associated with a Bench Concept

Volatility

- 36 invert versions since 2011 (11 major)
- 49 algae versions (13 major)
- 15 fish versions (3 major)

Getting Our Taxonomy

From home page at <https://aquatic.biodat.usgs.gov>



USGS
science for a changing world

USGS Home
Contact USGS
Search USGS

BioData - Aquatic Bioassessment Data for the Nation

BioData Retrieval

Retrieval Welcome to the U.S. Geological Survey (USGS) BioData Retrieval system which provides access to aquatic bioassessment data (biological community and physical habitat data) collected by USGS scientists from stream ecosystems across the nation. [more...](#)

BioData Retrieval System Features

- Select data using variety of criteria, including: sample type, geography, site characteristics and lists, sampling dates, and project or program
- See how many samples you will retrieve in real time as you modify selection criteria
- Save data selection criteria for re-use. Email to colleagues for them to use
- Preview data tables before downloading them
- Download data as comma-delimited, tab-delimited, XML, Excel 1997/2003 or Excel 2007/2010
- Download multiple related data tables in one operation - packaged together in one zip file
- All taxonomy reported using standardized and harmonized taxonomic system - synonymy is resolved for you
- Online data table and column definitions

USGS BioData

BioData Support
Downloads, FAQ, Tip sheets, and more
Contact the BioData Team
Supported sample collection methods and protocols

BioData Public Data Retrieval **Retrieval**
Retrieve Data
About Public Data Retrieval
Release Notes

BioData Input System **Input**
Login to BioData Input
Request support for a new collection method or protocol
Training Information
Release Notes

BioData Labs **Lab Info**
Lab Support
Summary of Lab Usage
Login to BioData Input

Search for “BioData Support”




BioData Support - MyUSGS

<https://my.usgs.gov/confluence/display/biodata/BioData+Support> ▾

A description for this result is not available because of this site's robots.txt – learn more.

Getting Our Taxonomy

[Retrieve Data](#)[About](#)[Support Home](#)[Contact Us](#)[Login](#)[USGS Home](#)[Contact USGS](#)[Search USGS](#)

Pages

BioData Support

Created by Peterson, Sibert H., last modified by Ruhl, Peter M. on Aug 23, 2017

Getting Started

- [What will BioData do for me?](#)
- [Getting started checklist](#)
- [What sampling methods are supported?](#)
- [Next steps](#)

About BioData

- [Fact Sheet](#)
- [Release Notes](#)
 - [Data Retrieval System](#)
 - [Data Entry and Management System](#)

Taxonomy

- [Taxonomic System Overview](#)
- [Taxonomy Downloads](#)
- [Request a NEW Taxon](#)
- [Naming Conventions for Conditional and Provisional Taxa](#)

BioData Taxonomy Downloads

Created by Unknown User (mwstern@usgs.gov), last modified by Grotheer, Scott A. on Apr 27, 2017

Overview of the BioData Taxonomic System

BioData Taxonomic System

Current Versions of the BioData Taxonomy

Community	CSV	Tab-delimited (TSV)	Excel 97/03	Excel 07	XML
Fish	CSV	TSV	Excel 97/03	Excel 07	XML
Invertebrate	CSV	TSV	Excel 97/03	Excel 07	XML
Algae	CSV	TSV	Excel 97/03	Excel 07	XML

[Complete BioData Taxonomy](#) (includes current, deprecated, and retired taxa)

[Previous Version](#) of the BioData Taxonomy ([Fish](#) | [Invertebrate](#) | [Algae](#))

[Naming Conventions for Conditional and Provisional Taxa](#)

Citations used in BioData Taxonomy

This download includes all the Bench Taxon Name Reference Codes and their complete citations

Community	CSV	Tab-delimited (TSV)	Excel 97/03	Excel 07	XML
Invertebrates	CSV	TSV	Excel 97/03	Excel 07	XML

Background

There are three important attributes (columns) that establish a taxon's identity (the "Taxon Concept") and whether the taxon is acceptable to the system for a data upload.

1. BenchTaxonName
2. BenchTaxonNameReferenceCode
3. TaxonConceptState

BioData uses the combination of **BenchTaxonName** and the "according to" **BenchTaxonNameReferenceCode** to establish the identity of a taxon. Each combination of name and reference code defines a distinct **Taxon Concept**. A concept that does not have a reference code is a "nominal" concept. BioData accepts many of these types of concepts, but the long-term value of data can be significantly enhanced by proper usage of a referenced concept.

Taxon Concepts can have one of three states (**TaxonConceptState**)

1. **Current** - meaning the concept is considered up-to-date and will be accepted in data
2. **Deprecated** - meaning the concept will be accepted in data but may be rejected in the near future
3. **Retired** - meaning the concept has been retired from use and will be rejected if used for incoming data

For more information please see the [BioData Taxonomic System](#) page.

Download format definitions

- CSV = Comma-separated values or comma delimited
- Tab-delimited (TSV) = Tab-separated values or tab delimited
- Excel 97/03 = Format used in Microsoft Excel 97- Excel 2003 workbooks, extension is .xls
- Excel 07 = Format used in Microsoft Excel 2007 - present workbooks, extension is .xlsx
- XML = Extensible Markup Language

Previous Versions of the BioData Fish Taxonomy

Created by Bunch, Aubrey R., last modified by Grotheer, Scott A. on Dec 19, 2017

Versioning of the BioData Taxonomic System

Fish Taxonomy Version	Excel File	Production Date	Comments
1.0	Fish_1.0.xlsx	March 2, 2011	From BioTDB March 2, 2011.
1.1	Fish_1.1.xlsx	November 28, 2011	Saved making no changes; differences are new sort order and authority added to hybrids
1.2	Fish_1.2.xlsx	November 28, 2011	NO changes. Moving this version to check on - 'clear cache' call which is the last step
1.2.1	Fish_1.2.1.xlsx	November 29, 2011	NO changes. Moving this version to check on - 'clear cache' call which is the last step
1.3	Fish_1.3.xlsx	December 7, 2011	Added Blue ridge sculpin, Umatilla dace
1.4	Fish_1.4.xlsx	October 2, 2012	adding 5 taxa for Amy Gill and reviewed with Mike Meador; one conditional/provisional taxa added

Requesting Revisions

- **Contact us links**
- **Request new taxon (lab interface for logged-in labs, in fish data entry GUI)**
- **Lab upload data GUI**



BioData

<http://aquatic.biodata.usgs.gov>

For more information:

Pete Ruhl	pmruhl@usgs.gov	703-648-6841
Scott Grotheer	sagrothe@usgs.gov	303-236-3486