

November 21, 2000

# **BIOLOGICAL TAXONOMY DATA STANDARD BUSINESS RULES**

**United States Environmental Protection Agency  
Office of Environmental Information  
1200 Pennsylvania Avenue, NW.  
Washington, DC 20460**

## BIOLOGICAL TAXONOMY DATA STANDARD BUSINESS RULES

### 1. THE STANDARD

- a. The Standard specifies the key data elements necessary to constitute consistent and unambiguous identification of biological organisms of interest to the Environmental Protection Agency (EPA). The standard data elements and their definitions are presented in Appendices A and B and are derived from EPA's Environmental Data Registry (EDR) at <http://www.epa.gov/edr/>
- b. The purpose of the standard is to ensure uniformity and comparability in the identification of biological organisms in the collection, analysis, and exchange of environmental data.
- c. The Biological Taxonomy Data Standard for EPA provides for the use of the *Integrated Taxonomic Information System* (ITIS) as the primary source of identification of biota of interest to EPA. Since ITIS does not contain identification information on all organisms of interest to EPA, the standard also provides for the use of information from other sources (e.g., the International Committee on Taxonomy of Viruses database, or ICTVdB).
- d. The Standard provides for the use of a unique Biological Identification Number (BIN), assigned by EPA, for the identification of all biological organisms of interest. The Standard also provides for the use of an ITIS Taxonomic Serial Number (TSN), where available, for the identification of biota in information management systems maintained by the EPA for the collection and reporting of data under the authority of Federal environmental legislation. The Standard also provides for the use of other identification numbers, such as the ICTVdB Taxon Identifier for the identification of viruses of interest.

### 2. DEFINITIONS

- a. *Biological organisms* for this standard include biota, groups, and viruses.
- b. *Biological Registry System* (BioRS) is used in the context of this document to represent the EPA central registry of biological organisms of interest to EPA. The BioRS will assign a BIN to biological organisms of interest to EPA.

- c. *Biological taxonomy* is the science that describes, identifies, names, and classifies organisms.
- d. *Biota* refers to the biological entities that comprise the taxonomic and nomenclatural standard reference known as the Integrated Taxonomic Information System. For the purpose of this standard, biota is an ecological term that refers to animals, plants, fungi, and other organisms. Biota, in the EPA biological taxonomy standard, does not refer to viruses. (See *Pollutants* definition.)
- e. *Biological Taxonomy Data Steward* is the person or organization responsible for managing the set of data resources for biological taxonomy.
- f. *Environmental Data Registry* (EDR) serves as the major tool in supporting EPA's standard-setting process, by recording and disseminating Agency standards, and ultimately facilitating data sharing between organizations and users. The EDR is a comprehensive, authoritative source of environmental metadata.
- g. *EPA national program systems* are those systems that exchange data with stakeholders or the public.
- h. *Groups* of biological organisms are those biota that are regulated as an aggregate of organisms that are related (e.g., macro-invertebrates, minnows, and total coliforms).
- i. *ICTVdB* is a database that supports virus classification and nomenclature maintained by the International Committee on Taxonomy of Viruses (ICTV). The ICTVdB provides easy access to records describing an individual virus, as well as providing information on its relatives. It uses a decimal system to number uniquely identified viruses. The user can access the ICTVdB at <http://life.anu.edu.au/viruses/Ictv/index/htm>
- j. *Integrated Taxonomic Information System (ITIS)* is a partnership to create a taxonomic and nomenclatural standard reference for biota that promotes scientific excellence and is fully supported by the world taxonomic community. The database is accessible over the World Wide Web (WWW) at <http://www.itis.usda.gov/>. ITIS classifications conform to the International Code of Botanical Nomenclature and the International Code of Zoological Nomenclature.
- k. *Name Context* is the context in which the name of a biological entity is used, e.g., Registry Name (the unique name assigned to the data element by the EDR). An organism may have more than one name depending on the context in which it is used.

- l. *Pollutants* are any substances, such as chemicals and waste products, that render the air, soil, or other natural resource harmful or unsuitable for a specific purpose. Biota, for the purpose of the EPA standard, might be an affected resource (e.g., trout in an affected stream) or, in some cases (e.g., *Escherichia coli* in a drinking water source), might be a pollutant.
- m. *Program System Data Stewards* for biological taxonomy are the individuals responsible for managing biological taxonomy data for EPA national program systems.
- n. *Reinventing Environmental Information (REI)* is an EPA initiative, in partnership with the states, to implement core data standards and to make electronic reporting universally available in EPA's national systems.
- o. *Taxon (taxa)* is a general term applied to any taxonomic element, irrespective of its classification level, e.g., species, class, or order.
- p. *Viruses* are ultra-lightmicroscopic, metabolically inert infectious agents that replicate only within the cells of living hosts, mainly bacteria, plants, and animals.

### 3. **APPLICABILITY**

- a. The EPA Biological Taxonomy Data Standard will apply where information about biological organisms is currently being collected and stored in the Reinventing Environmental Information (REI) national program systems. The REI program identifies 13 national program systems that will implement data standards. The REI national program systems are:
  - Permit Compliance System (PCS).
  - National Compliance Database (NCDB).
  - OECA Docket (Docket).
  - RCRA Information System (RCRIS).
  - Aerometric Information Retrieval System/Air Quality Subsystem (AIRS/AQS).
  - AIRS/Air Facility Subsystem (AFS).
  - Biennial Reporting System (BRS).
  - CAA 112(r) Risk Management Plan Information System (RMP\*Info).
  - CERCLA Information System (CERCLIS 3).
  - Safe Drinking Water Information System (SDWIS).
  - Toxics Release Inventory System (TRIS).
  - STORAGE and RETRIEVAL of Water Quality Data (STORET).
  - Envirofacts (EF) Data Warehouse.

- b. This standard is applicable to new EPA data collections and information systems and to reengineered information systems.
- c. This standard is applicable to regulations, rules, and Federal Register Notices created by EPA staff and contractors.
- d. The use of this standard is encouraged in any instances where the naming of biota may be ambiguous.
- e. This standard does not apply to retired systems not presently in use or to EPA documents and publications that were created prior to the adoption of this standard.
- f. State and Tribal partners and other stakeholders are encouraged to adopt this standard on a voluntary basis. This standard is applicable to state partners and other stakeholders only when the work plan of the Environmental Performance Partnership Agreement, or the appropriate Grant program agreements, includes mechanisms for adopting the standard. These agreements are typically reached between the regional offices or the program offices and the states or other stakeholders.

#### **4. DATA REQUIREMENTS**

- a. EPA Program Systems
  - 1) Mandatory data elements include: one number (ITIS Taxonomic Serial Number, EPA Biological Identification Number, or ICTVdB Taxon Identifier) and one name (Biological Systematic Name, Biological Vernacular Name, or Biological Group Name) and an associated name context (Biological Systematic Context Name, Biological Vernacular Name Context Name, or Biological Group Context Name). The data elements are described in detail in Appendix A.
- b. Biological Registry System (BioRS)
  - 1) Mandatory data elements for the BioRS are outlined in Appendix A. They include one number (ITIS Taxonomic Serial Number, EPA Biological Identification Number, or ICTVdB Taxon Identifier) and one name (Biological Systematic Name, Biological Vernacular Name, or Biological Group Name) and an associated name context (Biological Systematic Context Name, Biological Vernacular Name Context Name, or Biological Group Context Name).

- 2) Optional data elements for the BioRS are outlined in Appendix B. These data elements provide the ability to support versioning, tracking, and functionality inherent in ITIS.

## **5. PROCESSING**

The following section describes the methodology for implementing the requirements of the standard with respect to use of identifiers in the BioRS.

- a. EPA will maintain a central registry system for biological taxonomy, the BioRS, that will contain verified information identifying each registered biological organism. The BioRS will be available on the Internet.
- b. Every applicable information management system will store the ITIS TSN, EPA BIN, or ICTVdB Taxon Identifier when biological organisms are identified.
- c. Any EPA office or data administrator or State representative will satisfy the following procedures to register biological organisms in the BioRS.

To obtain an ITIS TSN for biota:

- 1) Perform a search of ITIS to obtain identification information including a TSN. This TSN will be submitted to the Biological Taxonomy Data Steward, with a name and name context for assignment of a BIN.
- 2) Whenever there are ambiguous search results and the unique identification of an organism is in question, contact the Biological Taxonomy Data Steward for assistance.
- 3) Where name information submitted to ITIS does not yield a TSN, a submitter must request and obtain a BIN.

To obtain an ICTVdB number for viruses:

- 4) Perform a search of the ICTVdB <http://life.anu.edu.au/viruses/Ictv/index.html> to obtain an ICTVdB Taxon Identifier for viruses. This number will be submitted to the Biological Taxonomy Data Steward, with a name and name context for assignment of a BIN.

To obtain a BIN for all organisms and groups of organisms without a TSN or ICTVdB Taxon Identifier:

- 5) The office requesting a BIN for organisms without a TSN must supply the Biological Taxonomy Data Steward with a name and name context.
  - 6) The Biological Taxonomy Data Steward will research ITIS, ICTVdB, and the BioRS and validate that the biological organism is not represented in any of the systems.
  - 7) If the organism is not represented in the BioRS, the Biological Taxonomy Data Steward will initiate registration into the BioRS and provide the submitter with a BIN, where appropriate. The data steward will also provide a request to ITIS for assignment of a TSN, where appropriate.
  - 8) When a TSN is assigned by ITIS, the Biological Taxonomy Data Steward will notify the submitter. When TSNs are assigned, program systems are encouraged to incorporate the storage of TSNs through their normal reengineering cycle and to use TSNs for reporting purposes. If a TSN is not assigned, the data steward will notify the submitter and the organism will continue to be reported through use of its BIN.
- d. Program Systems can interact with the BioRS with either the ITIS TSN or the EPA BIN, depending on which identifier the Program System chooses to store. All biological entities of interest to EPA will have an EPA BIN number assigned by the BioRS; ITIS-registered biota will also have an ITIS TSN.

## **6. ROLES AND RESPONSIBILITIES**

- a. The EPA Chief Information Officer (CIO) will:
- 1) Ensure adherence to these business rules and will be responsible for resolving conflicts and issues relating to these business rules, including applicability and waiver.
  - 2) Provide guidance and technical assistance to program offices and the regulated community in meeting the requirements of this standard.

- 3) Ensure public availability of tools for determining the proper ITIS or ICTVdB identifiers.
  - 4) Ensure the availability and maintenance of a central Biological Registry System (BioRS) where biological organisms can be uniformly identified with standard data elements and assigned a unique identifier.
  - 5) Ensure the appointment of a Biological Taxonomy Data Steward who will be responsible for the accuracy, reliability, and currency of the data standard.
  - 6) Establish an agency process and operating procedures to review Agency documents, information systems, and publications for compliance with this standard.
  - 7) Be responsible for issuing waivers from compliance with this standard in accordance with the procedures identified in Section 8 of the business rules.
- b. EPA Program Senior Information Resources Management Officers (SIRMO) and Regional Information Resource Management (IRM) Branch Chiefs will:
- 1) Promote compliance with this standard.
  - 2) Approve applications for waiver from this standard and submit them to the CIO.
  - 3) Ensure that new regulations, documents, and information systems identify biological organisms according to this standard.
- c. National System Program Managers will:
- 1) Ensure that this standard is implemented in their systems, as applicable.
  - 2) Work collaboratively with the CIO on continuing standards development and implementation.
  - 3) Identify and bring forward potential conflicts between these business rules, the underlying standards, and program systems needs.
  - 4) Submit waiver to the CIO for approval when a standard cannot be implemented.
- d. The Biological Taxonomy Data Steward, or an authorized delegate, will:



- 1) Work in collaboration with data stewards from states and national systems to develop and maintain identification information for biological organisms of interest to the Agency.
  - 2) Manage the reconciliation process between ITIS, ICTVdB, and the BioRS.
  - 3) Inform program system manager, program system data steward, or submitter of ITIS TSNs when new biota of interest to the Agency are registered in ITIS and assigned TSNs.
  - 4) Be responsible for the ongoing maintenance and update of the central database of biological identification data (BioRS).
  - 5) Be responsible for all agency communication with ITIS, including searches of ITIS, in connection with obtaining TSNs and BINs.
  - 6) Report assignment of a BIN:
    - Within 3 working days provided a submission is received with equivalent information to that needed by the Biological Taxonomy Data Steward to make the determination described in section 5.c.6. or
    - Within reasonable time for all other submissions, that time being dependent on the complexity of the review required by the Biological Taxonomy Data Steward.
  - 7) Validate information from submitters.
- e. Program System Data Stewards for biological taxonomy will:
- 1) Manage biological taxonomy data for EPA program systems.
  - 2) Be responsible for searching ITIS and ICTVdB for identification information prior to request of a BIN from the Biological Taxonomy Data Steward.
  - 3) Serve as the point of contact for notification of assignment of TSNs by the Biological Taxonomy Data Steward.

## **7. IMPLEMENTATION**

- a. Pursuant to REI mandates, Agency support functions will be developed, including guidance, data values, and appropriate tools to facilitate access to the ITIS database by the second quarter of Fiscal Year 2001 (FY01). Agency support functions will also include development of a BioRS to store identifying information about biological organisms and to assign and maintain EPA BINs. There will be continual maintenance of the support functions.
- b. EPA's REI national systems (as identified in section 3.a) will implement this biological identification data standard no later than March 31, 2003. REI national program system managers will include delegated state partners in the implementation planning process for the standard.
- c. This standard will be incorporated into the development cycle of all new Agency information collection and data management systems.
- d. This standard will be implemented into non-REI information systems as part of the normal reengineering cycle.

## **8. PROVISION FOR WAIVER**

- a. The Agency's CIO may grant waivers for sufficient reasons.
- b. Applications for a waiver shall contain:
  - 1) An outline of the reasons the data standard should not be implemented.
  - 2) A risk assessment and cost-effectiveness evaluation of continued noncompliant operation, if mandated by other Agency requirements.
  - 3) The approval of the decision officials in the requesting office, as defined by EPA's Information Resources Management Policy Manual, and by the organizations = SIRMO.
- c. The CIO shall notify the applying office in writing of the disposition of the waiver within 60 days of receipt.

**9. MAINTENANCE**

- a. Upon implementation of the standard, the Biological Taxonomy Data Steward shall review the standard and associated business rules annually for a period of five years and thereafter on a 3-year cycle.
- b. Any changes to the standard and business rules will be regularly updated and published in the EDR.

**10. REFERENCES**

- a. *Biological Data Profile of Content Standard for Digital Geospatial Metadata*, Federal Geographic Data Committee (FGDC).
- b. EPA Directive 2100, Information Resources Management Policy Manual, August 1, 1997.
- c. *EPA Environmental Data Registry (EDR)* (<http://www.epa.gov/edr/>).
- d. *Integrated Taxonomic Information System (ITIS)*, last updated June 26, 1997. (<http://biology.usgs.gov/cbi2/programs/itis.html> or <http://www.itis.usda.gov/>).
- e. *National Biological Information Infrastructure (NBII) Home Page*, (<http://www.nbio.gov/>).
- f. *Research School of Biological Sciences, Australian National University, Universal Virus Database (ICTVdB), International Committee on Taxonomy of Viruses (ICTV), Virology Division, International Union of Microbiology.* <http://life.anu.edu.au/viruses/Ictv/index.html>.

November 21, 2000

## **APPENDIX A**

### **Mandatory Biological Taxonomy Interim Standard Data Elements**

**APPENDIX A**

## Mandatory Biological Taxonomy Interim Standard Data Elements

<b>REQUIRED</b>		<b>NAME</b>	<b>DEFINITION</b>	<b>EDR DE ID</b>
<b>One Number</b>	This <b>OR</b>	ITIS Taxonomic Serial Number	The unique identification number assigned to an Integrated Taxonomic Information System biological entity.	20727:1
	This <b>OR</b>	ICTVdB Taxon Identifier	The unique identification number assigned to a virus by the International Committee on Taxonomy of Viruses (ICTV).	TBD
	This	EPA Biological Identification Number	The unique non-intelligent identification number assigned to a biological entity by the Environmental Protection Agency.	20728:1

**AND**

<b>One Name and Name Context</b>	These	Biological Systematic Name	The name assigned to a biological entity by a classification identification system.	20729:1
		Biological Systematic Context Name	The name of the classification system used to assign a systematic name to a biological entity.	20743:1
	<b>OR</b>			
	These	Biological Vernacular Name	The vernacular name associated with an occurrence of a biological entity.	20744:1

		Biological Vernacular Name Context Name	The name of the source of the vernacular name for a biological entity.	20745:1
	<b>OR</b>			
	These	Biological Group Name	The name of a collection of biological entities that are related.	20897:1
		Biological Group Context Name	The name of the source of the name of a group of biological organisms.	TBD

Note: "EDR DE ID in column 4 means "Environmental Data Registry Data Element Identification Number."

November 21, 2000

## **APPENDIX B**

Optional Biological Taxonomy Interim Standard  
Supplemental Standard Data Elements

**APPENDIX B**

Optional Biological Taxonomy Interim Standard  
Supplemental Standard Data Elements

NAME	DEFINITION	EDR DE ID
	NOTE DESCRIPTION	
Biological Kingdom Name	The systematic name that represents a biological Kingdom.	20730:1
	A biological Kingdom is a taxonomic category of the highest rank, grouping together all forms of life having certain fundamental characteristics in common.	
Biological Division/Phylum Name	The systematic name that represents the biological Division/Phylum.	20731:1
	A Division/Phylum is the primary subdivision of a taxonomic kingdom, grouping together all classes of organisms that have the same body plan.	
Biological Subdivision/ Subphylum Name	The systematic name that represents the biological Subdivision/Subphylum.	20732:1
	A Subdivision/Subphylum is a category of related classes within a Division/Phylum.	
Biological Class Name	The systematic name that represents the biological Class.	20733:1
	A Class is a major subdivision of a Phylum or Division, usually consisting of several Orders.	
Biological Subclass Name	The systematic name that represents the biological Subclass.	20734:1
	A Subclass is a category of related orders within a Class.	
Biological Infraclass Name	The systematic name that represents the biological Infraclass.	20735:1
	An Infraclass is a category of related classes within a Class.	



Biological Order Name	The systematic name that represents the biological Order.	20736:1
	An Order is a major subdivision of a Class or Subclass that consists of several families.	
Biological Suborder Name	The systematic name that represents the biological Suborder.	20737:1
	A Suborder is a category of related families within an Order.	
Biological Family Name	The systematic name that represents the biological Family.	20738:1
	A Family is a major subdivision of an Order or Suborder usually consisting of several genera.	
Biological Subfamily Name	The systematic name that represents the biological Subfamily.	20739:1
	A Subfamily is a category of related genera within a Family.	
Biological Genus Name	The systematic name that represents the biological Genus.	20740:1
	A Genus is a major subdivision of a family or subfamily, usually consisting of more than one Species.	
Biological Species Name	The systematic name that represents the biological Species.	20741:1
	A Species is a major subdivision of a genus or subgenus and is regarded as the basic division of a genus or subgenus. A Species is composed of related individuals that resemble one another and are able to breed among themselves, but are not able to breed with members of another Species.	
Biological Subspecies Name	The systematic name that represents the biological Subspecies.	20742:1
	A Subspecies is a subdivision within a Species.	
Biological Synonymous Name	The name that represents one of two or more names applied to a single biological entity.	20746:1
	None	
Biological Synonymous	The name of the source of a synonymous name for a biological entity.	20747:1

Biological Order Name	The systematic name that represents the biological Order.	20736:1
	An Order is a major subdivision of a Class or Subclass that consists of several families.	
Name Context	None	
Program System Biological Name	The name applied to a biological entity in a program system.	20748:1
	None	
Biological Group Description Text	The text describing the relationship of the biological entities in a group.	20750:1
	None	
Program System Abbreviated Name	An abbreviated name that represents the name of an information management system for an environmental program.	5712:1
	None	
Program System Name	The name of an information management system for an environmental program.	5714:1
	None	
Citation Reference Code	The code that represents a reference to an official printed copy of an environmental regulation.	20061:1
	None	