A Biodiversity Observation Data Standard for the NatureServe Network

March 2020

NatureServe
Connecting Science With Conservation

NatureServe Network

This Standard was produced by the Observation Data Standard Work Group; all members are listed in Appendix I.

Recommended Citation:
NatureServe. 2020. A biodiversity observation data standard for the NatureServe Network. NatureServe: Arlington, VA.

Table of Contents

EXECUTIVE SUMMARY

NatureServe and the Network of natural heritage programs and conservation data centers are the recognized leaders for providing decision-quality distribution and abundance data for at-risk species and ecosystems. Our focus has been on Network-collected occurrence data that identifies key locations of demonstrable conservation value for those species and ecosystems. Now, with increasing pressures on biodiversity, there is an increasing need for rapid and more comprehensive biodiversity assessments, and there is an increasing volume of observation data available from citizen science and other data collection efforts across the globe. We have reached a critical moment where we must bring together as much relevant data as possible to address rising conservation challenges. To meet this need, we require a straightforward and flexible standard for collecting and compiling biodiversity observation data across the Network and from external sources.

To create the standard, we convened an Observation Data Standard Work Group (hereafter Work Group), comprised of individuals from multiple Network programs and NatureServe. It focused on two key objectives:

1) Develop a standard that contains a comprehensive set of fields that are relevant to most Network programs and meet individual program needs for collection and management of observation data

2) Identify a subset of core biodiversity observation data fields that can be used to aggregate data in support of key Network-wide products:
    - Spatial biodiversity distribution products (including element occurrences, range maps, hexagon grid observations, and habitat suitability models)
    - Conservation status assessments
    - Quality/condition assessments of at-risk species and ecosystems

The Work Group built on previous drafts of an observation data standard and consulted the Darwin Core, an international standard for sharing of information about biodiversity from varied and variable sources. We conducted a Network-wide survey of the draft list of fields, asking programs to identify fields that should be added, deleted, or modified in some way. The Work Group systematically reviewed all fields based on the 80 comments received from 28 programs, and we conducted a priority-setting exercise to establish a standard that a) effectively manages observation data at the individual program level, b) supports the key network products, and c) maximizes compatibility with Darwin Core to facilitate the use of data from other sources. We developed a series of apps, tools, and protocols based on the standard that will facilitate the collection and aggregation of observation data and the creation of key products.

The Biodiversity Observation Data Standard contains a comprehensive list of 166 fields for use by Network programs to collect and manage biodiversity observation data. Fields that are directly equivalent to Darwin Core are noted. The Standard identifies the subset of fields required to create a basic observation, a larger subset that will support the generation of core network products, and additional fields that are needed to import observation data from external sources. No single program is likely to use all 166 fields, but collectively they will support priority needs at the program level.

This standard places NatureServe and the Network in a better position to place the most current, complete, and consistent biodiversity data into the hands of researchers and decision-makers. It will leverage Network data by facilitating the creation of a wide range of products beyond element occurrences, including inputs to habitat suitability models, gridded occurrence or hotspot maps, and species and ecosystem conservation status and condition assessments. The standard will continue to evolve to meet the changing needs of the Network as we address pressing conservation challenges.

1. INTRODUCTION

1.1. The Need for a Biodiversity Observation Data Standard

Biodiversity conservation critically depends on understanding the distribution and condition of native ecosystems and species, both spatially and temporally. Scientists have devised numerous approaches for tracking the distributions of species and ecosystems, ranging from high-level satellite information to detailed ground surveys. NatureServe and its Network programs and partner organizations have chosen the path of more intensive ground-based surveys, and as a result are the recognized leaders in North America for providing decision-quality distribution data for at-risk species and ecosystems and exemplary locations of all native ecosystems (e.g., Stein et al. 2000, Hammerson et al. 2017, Rainer et al. 2017, Comer et al. in press). Now, with increasing pressures on biodiversity and the need for rapid and more comprehensive assessments of species and ecosystems, it is timely to expand our approach to producing these data.

To track most rare and at-risk elements of biodiversity, NatureServe developed an innovative spatial unit —the “element occurrence”—that depicts areas that, if conserved, can contribute to the persistence of the at-risk species or targeted ecosystem type at a site (NatureServe 2002). Element occurrences are based on reliable field-based identification and accurate measurement of locations. They can involve a labor-intensive and expert-based task of establishing well-defined polygon boundaries and synthesis of useful information about the last observation and condition of the population or ecosystem type at the site. The approach has worked well where species and ecosystem concepts are well-established and where at-risk elements often form discrete patches on the ground.

Over time, conservation scientists have identified situations in which the element occurrence concept becomes awkward or even unattainable. For example, a wide-ranging species like the northern goshawk (*Accipiter gentilis)* may be best approached by recognizing multiple locations where nesting pairs occur from year to year and foraging habitat that may extend over large areas. In addition, scientific methods and technological solutions have greatly advanced, enabling maps to be produced by correlating field observations with remotely sensed information, such as images provided by satellites. Simultaneously, the citizen science revolution has resulted in the availability of millions of potentially useful observations (McKinley et al. 2017). Consequently, some Network programs in the United States and Canada and all programs in Latin America and the Caribbean now emphasize collection of basic observation data—localities where there is evidence that a species or ecosystem is or was present. Observations are valuable because they can serve not only as building blocks of element occurrences, but also as input to habitat suitability models, range maps, and other methods for estimating or characterizing species distributions and abundances. Still, these data need to be of sufficiently high quality to meet the needs and products of our Network; that is, they need to be “standardized and structured” (Bayraktarov et al. 2019) and “taxonomically trustworthy” (Franz and Sterner 2018).

As the NatureServe Network and our partners increasingly collect observation data, it is critical to establish a standard that enables not only sharing of observation data across the Network, but also aligns our efforts with global initiatives (e.g., Global Biodiversity Information Facility [GBIF], Group on Earth Observations Biodiversity Observation Network [GEO BON]). For example, much of the rest of the world has coalesced around the Darwin Core standard for observations, and we need to utilize observations built on this standard (Wieczorek et al. 2012). There is also a growing set of observation data gathered through citizen science efforts, such as iNaturalist and eBird, that could be valuable to our work. To make the most of the data collected within the Network, we need to maximize compatibility with Darwin Core in order to generate geospatial biodiversity data products that support conservation and resource management decisions across boundaries and scales. At the same time, we need to maintain the high-quality data standards expected of our Network.

1.2. Goals and Objectives

NatureServe, in consultation with the Network Section Councils, formed an Observation Data Standard Work Group. The goal of the work group was to establish a straightforward and flexible standard for collecting and managing biodiversity observation data from multiple sources to support development of key knowledge products that support conservation and resource management decisions. To meet this goal, the work group focused on two main objectives:

1) Develop a standard that contains a comprehensive set of fields that are relevant to most Network programs and meet individual program needs for collection and management of observation data

2) Identify a subset of core biodiversity observation data fields that can be used to aggregate data in support of key Network-wide products:
   - Spatial biodiversity distribution products (including element occurrences, range maps, hexagon grid observations, and habitat suitability models)
   - Conservation status assessments
   - Quality/condition assessments of at-risk species and ecosystems

Although no program is expected to use every field in the comprehensive set of fields, this standard will help ensure compatible data collection and management for the data they do collect. It will leverage Network data by facilitating the creation of a wide range of products beyond that of element occurrences. The standard will continue to evolve as needed to meet the changing needs of the Network.

2. DEVELOPMENT OF THE STANDARD

2.1. Observation Data Standard Work Group

The NatureServe Network has been engaged in developing a biodiversity observation data standard since 2006, when the first draft was produced, but at the time there was insufficient momentum to implement the standard throughout the Network. The 2006 draft was reviewed in 2016 by the Network's Spatial Methodology Review Team (SMRT), which was tasked with scoping a variety of issues, of which an observation data standard was one. The work of SMRT laid the groundwork for the Work Group. Additionally, several individual Network programs have developed various solutions for handling observation data. In 2019, the Work Group, consisting of staff from NatureServe and Network programs (Appendix I), was established to review and build on those past efforts to create this standard.

2.2. Guiding Principles

The following principles were used to guide the development of the standard:

- <u>Keep it simple</u>: Make managing large numbers of observation records as simple as possible by keeping the number of core attributes small and focused on answering the basic questions of:
  - What (Taxon)
  - Where (Location)
  - When (Date)
  - Who (Observer)
- <u>Maximize compatibility with Darwin Core</u>: follow Darwin Core standards as much as possible, or at least make fields easily translatable; balance this need with the need to meet Network processes for managing spatial data (e.g., ensure compatibility with Biotics)
- <u>Provide options</u>: Recognize that compatible data, especially from external sources, can be collected in varying ways and still be useful. Our requirements should provide flexibility in order to take advantage of these data (e.g., accommodate several compatible ways to record location data).
- <u>Consider the need to query data</u>: use domain tables as appropriate to support data queries
- <u>Promote assessment of data quality</u>: facilitate assessment of the quality and confidence of observation records
- <u>Facilitate the gathering and sharing of large amounts of data</u>: interoperability is key to the use of observation data; enable aggregation of large observation data sets from multiple sources
- <u>Use quantitative data</u>: where possible, use observation attribute data that are quantitative

2.3. Work Group Process

The Work Group began by reviewing the 156 observation data fields drafted by the Spatial Methodology Review Team in 2016 and by cross walking them with the list of Darwin Core fields to identify opportunities for consistency with that standard. NatureServe Network programs reviewed the draft list of fields and associated documentation via a directed survey. Programs were asked to provide feedback about modifying, deleting, or adding fields in the initial list. Eighty comments were received from 28 programs.

The Work Group reviewed the survey results and reached consensus on a new draft set of fields through two rounds of priority setting. Individual Work Group members first identified the suite of fields that they felt were most important to their program. When compiled, this resulted in a comprehensive set of fields that could be widely used across the Network to manage data at the program level. Work Group members then identified the core fields that were most essential to supporting the production of the key Network products: a) spatial distribution products b) element occurrences, c) conservation status assessments, and d) quality and condition assessments. Results from Work Group members were tallied to create a draft list of core fields, which was then reviewed and adjusted by the entire Work Group. After additional rounds of review, Work Group members established the observation data standard that includes a basic, core, and comprehensive set of fields for collecting and managing observation data. It also provides a set of additional fields required to import data from external sources. Presentations were made to the Canadian and U.S. Section Councils to summarize the process and receive feedback. Consultation with the Latin American Section Council is ongoing.

The Biodiversity Observation Data Standard is meant to be a dynamic standard, given the diversity of Network program needs, the wide range of species and ecosystems that the Network tracks, and the diversity of conservation and management issues that we seek to address. The Work Group will reconvene after the standard has been field tested to revise as needed.

3. THE STANDARD

This Biodiversity Observation Data Standard includes a definition of an observation and a comprehensive set of fields that guide local programs in collecting and managing observation data for conservation purposes. It identifies the subset of basic fields required to create a minimally viable observation and a larger subset of core fields needed to support the development of key Network products.

3.1. Definition of an Observation

An observation is "*an occurrence, or documentation of lack of an occurrence, of an organism, a set of organisms, or an ecosystem type through a data collection event at a location at a given time by an observer(s).*" Individual observations can be linked through common characteristics such as time, place, protocol, and co-occurring organisms. Simply stated, an observation is the documentation of a) a species or ecosystem, b) at a location, c) at a given time, d) by a sensor (human or machine). Although only a small set of data fields may be needed to meet this basic observation definition, inevitably, additional information is needed to create specific conservation products and appropriately manage the observation data and judge its accuracy.

3.2. Description of Observation Data Fields

The full set of data fields of the Biodiversity Observation Data Standard are available online. Table 1 provides an explanation for the key attributes. All fields and their attributes are provided, organized by six basic categories:

- **What** (Taxon/Type): fields that document species taxa or ecosystem types found in a jurisdiction or region
- **Where** (Location): fields that document the location where a species or ecosystem is found
- **When** (Date): fields that document when a species or ecosystem was observed
- **Who (**Observer): fields that document the observer(s) name(s)
- **Details**: A variety of fields that document an observation, including measures of quantity, condition, area occupied, suitability for modeling, and field measures of abundance

- **Other**: A variety of database fields that can be used to manage observation data and contribute to metadata (e.g., observation grouping fields, dataset name, collection ID, sampling protocols, etc.)

Table 1. *Explanation of the key attributes for each of the fields in the observation data standard.*

Column Name*	Description
Category	Type of field used to document an observation; there are six basic categories: what (taxon), where (location), when (time), who (observer), details (e.g., quantitative data, survey methods), and other (e.g., references)
Group	Subcategories that organize fields within the basic categories
Field Label	A readable version of the field name
Definition	Definition of the field compiled from the Darwin Core Standard, Biotics, and other sources.
Darwin Core Equivalent	Darwin Core field name equivalent
Biotics Equivalent	NatureServe Biotics field name equivalent
Field Form/Device or Database Only	Distinguishes whether the field is used when the observation is recorded in the field (field form or field device) or is only used in a database after the observation is collected or imported from an external source
Basic Observation	Identifies the minimum fields required to document any observation. A value of “O” means it is required for a basic observation. A value of “E” means it is required if the observation comes from an external source (but see Conditional Requirements).
Conditional Requirement	Indicates options for collecting comparable fields (e.g., one can use a single date field or a start date/end date pair of fields when the observation is made over more than one day)
Core Products	Fields needed to create any of the primary products of the NatureServe Network
Distribution	Fields needed for distribution products such as habitat suitability models, hexagon grids, range maps
Element Occurrence	Fields needed to create an element occurrence from the observation
Conservation Status Assessment	Fields required to conduct conservation status assessments (global, national, subnational ranks)
Quality/Condition	Fields required to assess the quality or condition of an observation. The fields provide information relevant to population viability and ecological integrity and contribute to the element occurrence rank.

The observation data standard is presented in a series of steps (Fig. 1) based on the level of detail needed to support desired inputs from Network programs or external sources and desired products. The standard identifies a set of fields minimally *required* to document an observation, whether the data are collected directly by a Network program or from an external source. For data from an external source, several additional fields are *required* to establish standard taxonomy and provide appropriate record identifiers. An additional set of fields supports the creation of the core network products. In addition to the basic, external, and core observation fields, the standard provides 127 fields that support a wide variety of individual Network products.

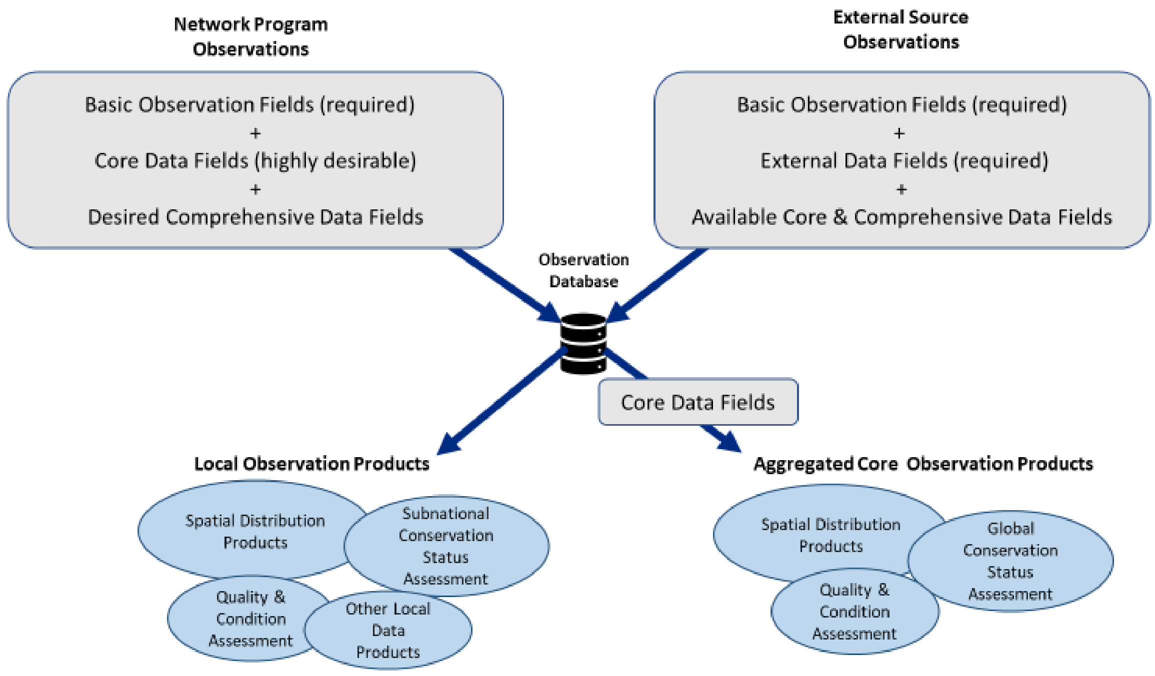

Figure 1. *The building of an observation from basic observations from the Network and external sources to support the creation of conservation data products.*

3.3. Basic Observation Fields

In its most basic form, the observation data standard only requires documentation of one field or a related set of fields in each of the What, Where, When, and Who categories. Users must document a) Taxon or Ecosystem Type, b) Location, c) Date/Time, and d) Observer (Table 2), but have some flexibility as to how they record them; i.e., for Taxon, use common or scientific name; for Location, use a point with latitude/longitude, a line, or a polygon; for Date/Time, use a single date or a date range; for Observer, only one field is available. Thus, there are 12 fields available to capture these four basic information types, with anywhere from 4 to 7 needed for any given observation. These fields are sufficient to document a basic field observation. All 12 fields are part of Darwin Core.

When these attributes are brought into a database, an additional set of seven fields are needed to manage the data, including: a) for point observations, the two fields for the coordinate system and geodectic datum, b) the date as recorded by the observer, c) a database assigned observation number, d) the date the record was created, e) who created the database record, and f) the higher classification category that a taxon or type belongs to (Table 2). The first three items are part of Darwin Core. The second three are not, but they are necessary for effective data management. This practice is not without precedent; the database maintained by GBIF contains many database fields not found in the Darwin Core in order to manage Darwin Core observation data.

Thus, to both collect a basic observation in the field and to store it in a database, some subset of 19 fields are needed. The fields used could be as few as nine; e.g., a point observation is made at a specific time on a specific day, using a standard geographic coordinate system, and is uploaded into a Network database.

Table 2. *Basic observation fields. Nineteen fields are provided in the standard. Some definitions are shortened here for brevity but are provided in full in the online version. The "Conditional Requirement" column provides options for documenting data in each category, only one or two of which are required. The "Field(FF)/Database(DO)" column distinguishes between fields that are collected on field forms or field devices (FF) versus those used in the database only (DO).*

Category	Group	Field Label (Displayed)	Definition	Darwin Core Equivalent	Field / Database (FF, DO)	Conditional Requirement
WHAT	Taxon	Scientific Name	The full scientific name, with authorship and date information if known. When forming part of an Identification, this should be the name in lowest level taxonomic rank that can be determined.	scientificName	FF	Select this field *or select* Common Name *or* Higher Classification *or* Taxon ID
WHAT	Taxon	Common Name	A nontechnical name for the element.	vernacularName	FF	Select this field *or select* Scientific Name *or* Higher Classification *or* Taxon ID
WHAT	Taxon	Higher Classification Level	Taxonomic or Ecosystem Type level(s) for the levels *above the one* represented by the Scientific Name	higherClassification kingdom, phylum, class, order, etc..	FF	Select this field *together with* Higher Classification Name *and* Unit ID *or select* Scientific Name *or* Common Name *or* Taxon ID
WHAT	Taxon	Higher Classification Name	Name of level(s) *above the one* represented by the Scientific Name.		FF	Select this field *together with* Higher Classification Level *and* Unit ID
WHAT	Taxon	Higher Classification Unit ID	Unique identifier for taxonomic levels *above the one* represented by the Scientific Name.		DO	Select this field *together with* Higher Classification Level *and* Name
WHERE	Location	Latitude (Decimal)	The latitude of the location from which the organism or observation was collected, expressed in decimal degrees.	decimalLatitude	FF	Select this field *together with* decimalLongitude *or select* Polygon *or* Line
WHERE	Location	Longitude (Decimal)	The longitude of the location from which the organism or observation was collected, expressed in decimal degrees.	decimalLongitude	FF	Select this field *together with* decimalLatitude *or select* Polygon *or* Line
WHERE	Location	Line	The location from which the organism or observation was collected, as represented by a line.		FF	Select this field *or select* Polygon *or* Lat / Long
WHERE	Location	Polygon	The location from which the organism or observation was collected, as represented by a polygon.		FF	Select this field *or select* Line *or* Lat / Long
WHERE	Location	Coordinate System	Coordinate system as it was originally recorded (geographic, UTM, etc.).	verbatimCoordinate System	DO	
WHERE	Location	Geodetic Datum	The geodetic datum to which the latitude and longitude refer. Datum should be selected from a picklist.	geodeticDatum	DO	

Category	Group	Field Label (Displayed)	Definition	Darwin Core Equivalent	Field / Database (FF, DO)	Conditional Requirement
WHEN	Date/Time	Observation Date	The date of the observation, consisting of year, month, and day.	eventDate	FF	Select this field *or select* Verbatim Date *or* Observation Start Date *and* Observation End Date
WHEN	Date/Time	Verbatim Date	The original representation of the date as recorded by the observer. (e.g., March 1998, or 17IV1999). *Note* – this allows for imprecise dates or idiosyncratic notation to be preserved.	verbatimEventDate	DO	Select this field *or select* Observation Date *or* Observation Start Date *and* Observation End Date
WHEN	Date/Time	Observation Start Date	The precise start date of a data collection event that spanned multiple days.	startDayOfYear	FF	Select this field *together with* Observation End Date *or select* Observation Date *or* Verbatim Date
WHEN	Date/Time	Observation End Date	The precise end date of a data collection event that spanned multiple days.	endDayOfYear	FF	Select this field *together with* Observation Start Date *or select* Observation Date *or* Verbatim Date
WHO	Observer	Observer	Full name of the person(s) who made or reported the observation or have other knowledge of it.	recordedBy	FF	
DETAILS	Observation	Observation ID	Unique identifier of the observation record in the local (i.e. NatureServe Network program) observation database.	occurrenceID	DO	
OTHER	Reference	Record Created By	Name of user who created the record within the database.		DO	
OTHER	Reference	Record Created Date	Date the record is created within the database.		DO	

3.4. External Source Observation Fields

When bringing in observation data from external sources, eight *additional* database fields are needed to a) document taxonomic concepts to ensure that we can match the observation to a standard taxonomy (five fields) and b) capture basic dataset identifiers (three fields) (Table 3). Five of these eight fields have Darwin Core equivalents.

Table 3. *External database fields required when bringing in observations from external sources. These eight fields are required in addition to the fields that describe the basic observation from Table 2. Some definitions are shortened here for brevity but are provided in full in the online version.*

Category	Group	Field Label (Displayed)	Definition	Darwin Core Equivalent	Field / Database Only (FF, DO)	Conditional Requirement
WHAT	Taxon	Concept Reference Citation ID	Unique ID of the citation for the reference that describes or points to the circumscription of the taxon or community	nameAccordingToID	DO	
WHAT	Taxon	Concept Reference Citation	Citation for the reference that describes or points to the circumscription of the taxon or community	nameAccordingTo	DO	
WHAT	Taxon	Name used in Concept Reference ID	Identification number (ID) of name used in concept reference		DO	
WHAT	Taxon	Name used in Concept Reference	Scientific name used in concept reference		DO	
WHAT	Taxon	Name Published in Year	The four-digit year in which the scientific name was published	NamePublishedInYear	DO	
OTHER	Data Source	Dataset ID	An identifier for the set of data. May be a global unique identifier or an identifier specific to a collection or institution.	datasetID	DO	Select this field *or* select Dataset Name *together with* Global ID
OTHER	Data Source	Dataset Name	The name identifying the data set from which the record was derived	datasetName	DO	Select this field *or* use Dataset ID *together with* Global ID
OTHER	Data Source	Global ID	A field of type UUID (Universal Unique Identifier) in which values are automatically assigned by the original source geodatabase. Should be used in combination with Dataset Name.		DO	Select this field *together with* DatasetID *or* select Dataset Name

3.5. Core Product Observation Fields

Building on the 19 basic observation fields (Table 2), the observation data standard specifies 12 additional fields needed to meet the needs of core Network products:

- Spatial biodiversity distribution products (including element occurrences, range maps, hexagon grid observations, and habitat suitability models)
- Conservation status assessments
- Quality/condition assessments of at-risk species and ecosystems

The additional fields are provided in Table 4 and largely fall in two categories: 1) additional specifics on where the observation was collected, and 2) additional details on species behavior (migratory), data sensitivity, and quality condition of the observation. Programs are encouraged to collect these core fields so that core products can be developed, but they are not required. Full documentation of these fields is available in the online Biodiversity Observation Data Standard, which can be easily filtered to provide only the core observation fields.

Table 4. *Additional fields needed for core products. These 12 fields are in addition to those required for the Basic Observation (Table 2). See Table 2 for further explanation of column headings. Some definitions are shortened here for brevity but are provided in full in the online version.*

Category	Group	Field Label (Displayed)	Definition	Darwin Core Equivalent	Field / Database Only (FF, DO)	Conditional Requirement
WHAT	Taxon	Taxon ID	An identifier for the set of taxon information (data associated with the Taxon class). May be a global unique identifier or an identifier specific to the data set.	taxonID	DO	
WHERE	Location	Locational Uncertainty Type	The type of inaccuracy in the mapped location of an observation compared with its actual on-the-ground location		FF	Use of this field *requires use of* Locational Uncertainty Distance/Unit
WHERE	Location	Locational Uncertainty Distance	Distance within which the location of the observation is believed to be captured. If location was mapped from GPS coordinates, accuracy recorded by GPS unit in meters.	coordinate Uncertaint yInMeters	FF	This field is required *if* Locational Uncertainty Type is Estimated
WHERE	Location	Locational Uncertainty Unit	Unit associated with the Location Uncertainty (meters)		FF	This field is required *if* Locational Uncertainty Type is Estimated
DETAILS	Observation	Data Sensitive Category	Value selected from a drop-down menu that best captures the category/reasoning for which this observation should be considered sensitive and should not be distributed without permission		FF	
DETAILS	Observation	Data Sensitive?	Is the locational information of this observation sensitive and hence, should be restricted from unsecured use?		FF	
DETAILS	Observation	Migratory Use	The descriptive label indicating which season or behavior (e.g., breeding, nonbreeding) is associated with an observation area for migratory animal species that utilize geographically and seasonally disjunct locations		FF	
DETAILS	Observation	Conceptual Feature Type	Indicates the cartographic feature that would result from mapping the underlying field data, based on the observed feature as compared to the size of the observed area with the minimum mapping unit (mmu) for the scale map used		FF	
DETAILS	Observation	Detected?	Was the element found? (Yes/No). 'No' indicates negative observation data.		FF	
DETAILS	Observation	Condition of Element Comment	Comments on the condition of the element at the location (such as alive or dead)		FF	
DETAILS	Observation	Suitable for EO?	Is this observation suitable for use in an Element occurrence (EO)? Considerations include whether this is a tracked element, the quality of the data, etc.		DO	
DETAILS	Observation	Suitable for Modeling?	Documents whether the record is suitable for use in habitat distribution modeling.		DO	

3.6. Comprehensive Observation Fields

There are many uses of observations beyond the core products developed by the Network, and many programs share the need to develop similar products. Other programs are looking to build biodiversity observation databases and are looking for guidance. The standard provides a set of 166 fields, including the ones already described: the 31 fields used to build core products (Table 2 and 4) plus the 8 additional fields required to bring in external observation data (Table 3). The additional 127 fields provide standards for a wide diversity of Network products (see the online Biodiversity Observation Data Standard for the complete list). We expect this part of the standard to be much more dynamic as programs apply it and learn from each other.

3.7. Comparison with Darwin Core

The standard is shaped by internationally recognized observation data standards in order to take advantage of the large amount of external data being gathered by agencies, organizations, and individuals. We focus on compatibility with Darwin Core, because it is the most widely recognized standard for collecting and distributing biological observation data (Wieczorek et al. 2012). Our goal is not to be 100% compliant, however, because NatureServe and Network products do more than is possible with only Darwin Core data. We have maximized compatibility wherever possible.

A comparison of the NatureServe Biodiversity Observation Data Standard fields with Darwin Core fields shows that most fields in the Basic Observation (Table 2) are part of Darwin Core. Fields that are not part of Darwin Core typically have a conditionally required field that is equivalent, allowing the Network to be compatible with other data sources.

Greater differences occur for fields that contribute to our core products (Table 4). In general, fields in the observation data standard that do not equate to Darwin Core fields are needed for NatureServe Network–specific workflows and products (e.g., Biotics data fields for Source Feature creation of data management or individual Network program needs). Examples of Biotics fields in these two categories that are not in Darwin Core include: Data Sensitivity, Migratory Use, Conceptual Feature Type, Observation Area, Threats, Primary Reference, and QC Status. Fields in these two categories that were recommended by Network programs to fulfill their local programmatic needs, and not in Darwin Core, include: Total Count, Evidence Comment, Condition of Element, Habitat Quality, Sampling Protocol Description, Permit, Suitability of Observation for various purposes (element occurrences, modeling), various fields related to Destination of the observation (into other databases), and grouping of observations. There is a larger divergence (approximately 60% similarity) between the comprehensive set of fields and Darwin Core.

As we gain experience in working with both Network and external observation data, we will better understand how the fields we are gathering may be used to improve observation data standards. Our goal is to work with Darwin Core developers to have them consider adopting some of these fields so that others may benefit from NatureServe Network methodology.

4. BIODIVERSITY OBSERVATION DATA FLOW

4.1. A Model for Collecting, Aggregating, and Managing Observation Data

Implementing the Biodiversity Observation Data Standard will facilitate the flow of observation data throughout the NatureServe Network and will support the creation of core biodiversity location knowledge products (Fig. 2). Adoption of the Biodiversity Observation Data Standard by Network programs and partners, or at least alignment of core database attributes and fields, will increase efficiency of observation aggregation and the data exchange process and will improve data currency, making spatial data more rapidly available to create our core products and support conservation applications.

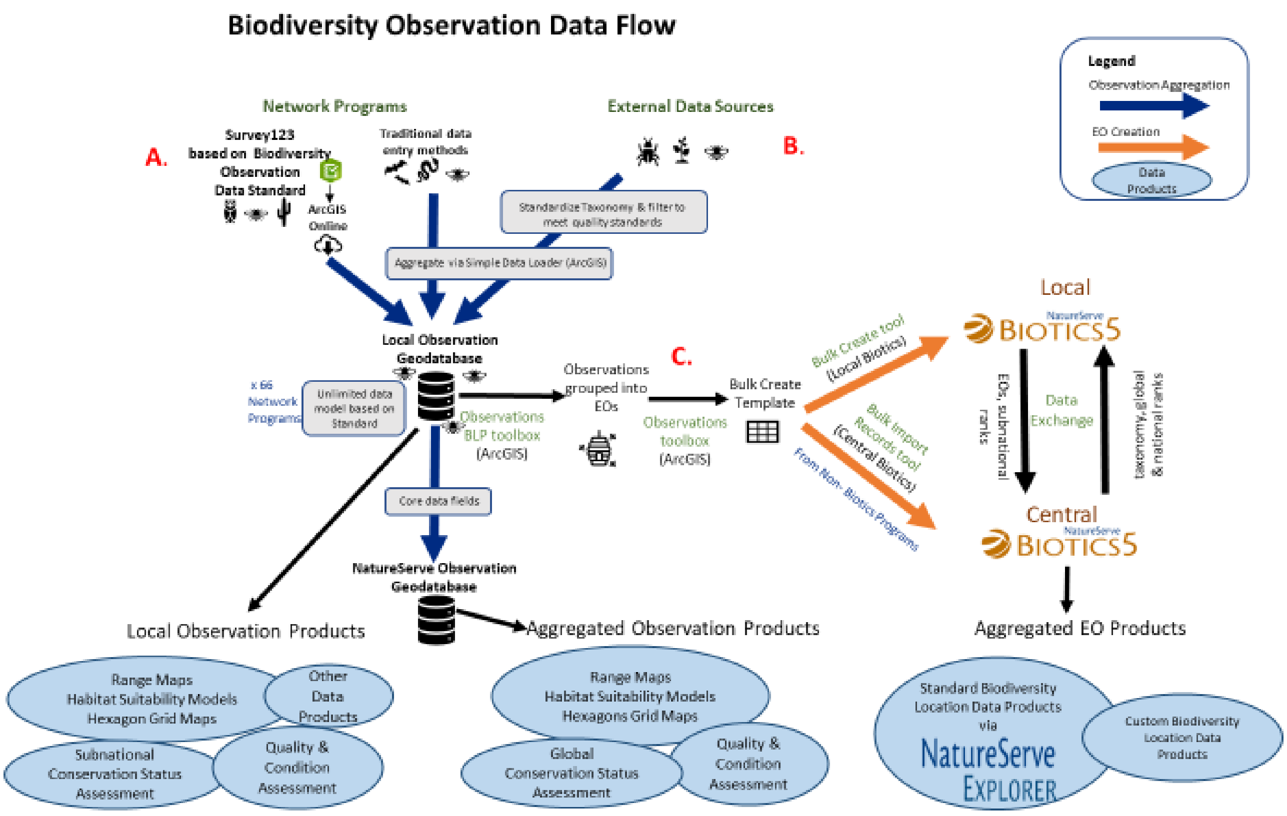

Figure 2. *Flow of observation data from Network programs and external sources to create conservation data products. Field data collected by Network programs or from external sources can be aggregated into local geodatabases to produce local observation products. Observations that have data for the core fields can be aggregated into a NatureServe geodatabase to create observation products from multiple programs. Observations that have the necessary fields can be used to create element occurrences with a suite of tools and processes that enforce NatureServe standard methodology. These EOs can then be integrated into Biotics to produce standard Biodiversity Location Data products via NatureServe Explorer or custom Biodiversity Location Data products.*

The solutions outlined in Figure 2 and described below offer a relatively quick and efficient conversion of observation data from a variety of sources into new or existing element occurrences following standard NatureServe methodology, which can then be aggregated into central Biotics. This greatly reduces data backlogs and data gaps and puts NatureServe and the Network in a better position to provide researchers and decision-makers the most current, complete, and consistent biodiversity data, including inputs to habitat suitability models, gridded occurrence or hotspot maps, conservation plans, and species and ecosystem conservation status and condition assessments.

4.2. Field Data Collection

The Biodiversity Observation Data Standard will be used as the basis for a survey template, created within Survey123 for ArcGIS, to enable collection of observation data in the field via a smartphone or tablet (Fig. 2, Letter A and Fig. 3). Data collected via the Observation Survey will be saved in ArcGIS Online and can then be downloaded into a Biodiversity Observation Data Standard geodatabase at the level of the Network program. Some Network programs have generated their own Survey123 for ArcGIS forms or other means of entering observation data. These can continue to be used to enter observation data into local biodiversity observation geodatabases.

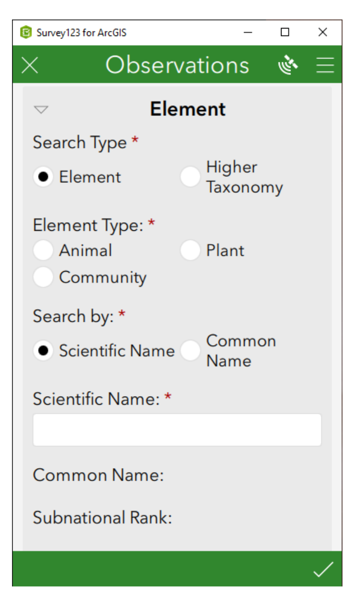

Figure 3. *First entry screen for Survey123 for ArcGIS Observation Survey Template.*

4.3. Integration of Data from External Sources

Observation data from external sources can be used to augment Network data collection activities. For example, Young et al. (2019) highlight how observation data were used to improve our element occurrence product for six at-risk bird species tracked by the New York Natural Heritage Program (Fig. 4). Now that we have a standard and a core set of fields that are compatible with the Darwin Core, and have identified the fields needed to establish standard taxonomy and track dataset identifiers, the next step is to develop streamlined ways to vet, filter, and integrate data from these sources to support the creation of our core products. Methods are needed to automate the vetting and aggregation of the massive number of observations collected by sources such as GBIF and GEO BON, other partner organizations and agencies, as well as citizen scientists (e.g., eBird, iNaturalist). These methods must include ways to evaluate the quality of those observations including locational accuracy, identification accuracy, date accuracy, reported habitat appropriateness, etc. Once vetted, those data can be added to Network programs' local geodatabases via the Simple Data Loader (ArcGIS), which facilitates cross walking of fields between varying data models (Fig. 2, Letter B).

These structured, high-quality observation data can become important contributions to larger efforts to conduct biodiversity research on all species and ecosystems.

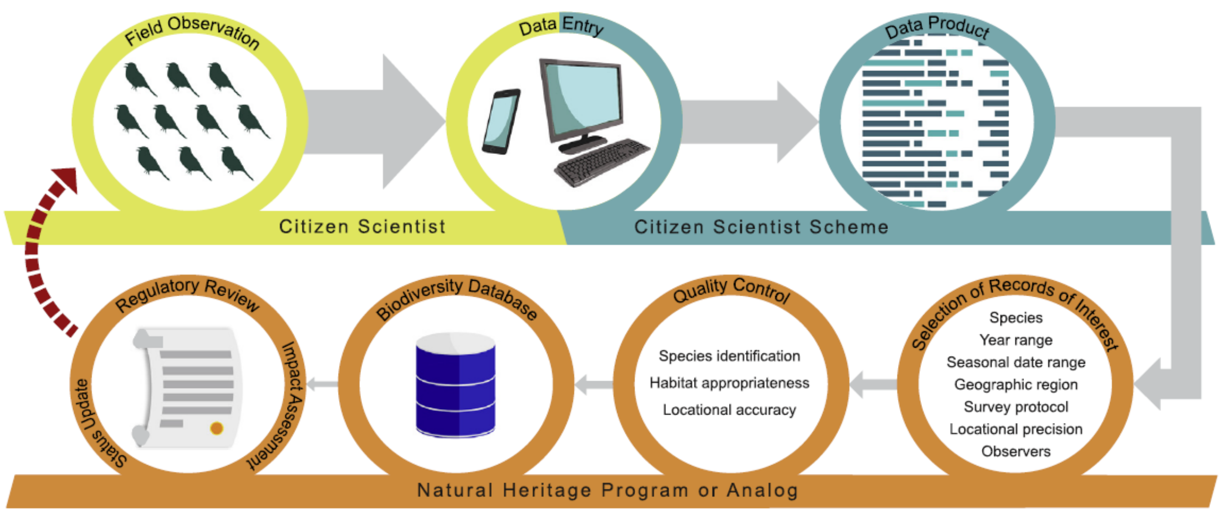

Figure 4. *Data pipeline from observation by citizen scientists to regulatory decisions, showing a pathway of increasing levels of data structure and data accuracy (e.g., taxonomic confidence of species assignment). Depth of gray arrows reflects volume of data flow. Significant filtering of records takes place at every step, based on data being entered into local databases, aggregated into large databases (e.g., Global Biodiversity Information Facility [GBIF] or iNaturalist), and then filtered based on selection of records of interest and data quality relevant to the mission of the Network or equivalent organization. From Young et al. 2019, reprinted with permission.*

4.4. Bulk Creation of Element Occurrences from Observations

NatureServe has developed a solution to efficiently aggregate observations to create element occurrences that follows standard methodology and facilitates data exchange. The solution is simple to apply, eliminating the need for technical expertise, while providing a repeatable, automated process for converting observations into EOs that can be imported into central Biotics (Fig. 1, Letter C). This process utilizes two toolboxes, both created in ArcGIS, to automate the steps to convert observation data into a format used to create EOs. First, the Observations Bulk Load Prep (BLP) Toolbox, designed by the Pennsylvania Natural Heritage Program, groups observations into new or existing EOs according to specified separation distances and in accordance with the EO standard methodology. Second, the Observations Toolbox formats and populates a geodatabase template, required for importing spatial data into Biotics. Finally, two Biotics tools, the Bulk Create tool (used by subnational Biotics instances) and Bulk Import Records tool (used to import data into central Biotics from programs which do not use Biotics) function to create source features and element occurrences from the data within the geodatabase template. Combining the two observations toolboxes with the Biotics Bulk Create and Bulk Import Records tools provides a streamlined, repeatable, and automated solution for efficiently importing observation data into central Biotics.

This solution will increase the speed and efficiency of converting observation data into source features and EOs, both by programs that use Biotics and those that do not. In turn, this speeds the rate and timing of data exchanges and ultimately the publication of those data to NatureServe Explorer and other biodiversity location data products.

5. CONCLUSIONS

In the United States and Canada, NatureServe Network programs have long maintained databases of locations of populations of at-risk species and ecosystems and exemplary locations of all ecosystems. These location data have been widely used for our core Network products and used for regulatory review processes to determine whether certain proposed activities might affect at-risk species or ecosystems (e.g., Faber-Langendoen et al. 2016). In the past, the location data used to create our products have largely been based on element occurrences (EOs), which require intensive field collection and data processing by program staff. Now, use of observation data will greatly improve our ability to meet our mission, because we can access a much wider set of data, such as those generated through citizen science, and apply the observation data standard filters needed to make them sufficiently rigorous for our products.

Thus, a streamlined process for collecting and aggregating observation data will help our work in the following ways:

- Ensure that more at-risk species and ecosystem observation data will be processed and kept current for use in environmental impact analysis, conservation planning, conservation status assessment, and resource management/restoration decisions
- Increase efficiency for modeling distributions of species that span multiple jurisdictions
- Ensure that exemplary ecosystem element occurrences and observations are compiled, thereby increasing development of reference locations for mitigation and restoration of ecosystems
- Provide a more effective use of the Biotics database to improve data standards and management, greatly increasing efficiency of data aggregation and data exchange across the network

6. REFERENCES

Bayraktarov, E., G. Ehmke, J. O'Connor, E.L. Burns, H.A. Nguyen, L. McRae, H.P. Possingham, and D.B. Lindenmayer. 2019. Do big unstructured biodiversity data mean more knowledge? Frontiers in Ecology and Evolution 6: 239.

Comer, P.J., J.C. Hak, C. Josse, and R. Smyth. In press. Long-term loss in extent and current protection of terrestrial ecosystem diversity in the temperate and tropical Americas. PLoS ONE (in press).

Faber-Langendoen, D., W. Nichols, J. Rocchio, K. Walz, J. Lemly, R. Smyth and K. Snow. 2016. Rating the condition of reference wetlands across states: NatureServe's Ecological Integrity Assessment method. National Wetlands Newsletter 38 (3):12-16.

Franz, N.M. and B.W. Sterner. 2017. To increase trust, change the social design behind aggregated biodiversity data. Database 2017: 1-12.

Hammerson, G.A., M. Kling, M. Harkness, M. Ormes, B.E. Young. 2017. Strong geographic and temporal patterns in conservation status of North American bats. Biological Conservation 212: 144–152.

McKinley, D.C., A J. Miller-Rushing, H.L. Ballard, R. Bonney, H. Brown, S C. Cook-Patton, D.M. Evans, R.A. French, J.K. Parrish, T.B. Phillips, S.F. Ryan, L.A. Shanley, J.L. Shirk, K.F. Stepenuck, J.F. Weltzin, A. Wiggins, O.D. Boyle, R.D. Briggs, S.F. Chapin, D.A. Hewitt, P.W. Preuss, M.A. Soukup. 2017. Citizen science can improve conservation science, natural resource management, and environmental protection. Biological Conservation 208: 15-28.

NatureServe. 2002. Element Occurrence Standard. NatureServe: Arlington, VA.

Rainer, R., B. Bennett, S. Blaney, A. Enns, P. Henry, E. Lofroth, & J. Mackenzie. 2017. On guard for them: Species of global conservation concern in Canada. NatureServe Canada: Ottawa, Ontario, Canada.

Stein, B.A., L.S. Kutner, and J.S. Adams. 2000. Precious Heritage: the status of biodiversity in the United States. Oxford University Press, New York.

Wieczorek J., D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, and D. Vieglais. 2012. Darwin Core: An evolving community-developed biodiversity data standard. PLoS ONE 7: e29715.

Young, B.E., N. Dodge, P.D. Hunt, M. Ormes, M.D. Schlesinger, and H.Y. Shaw. 2019. Using citizen science data to support conservation in environmental regulatory contexts. Biological Conservation 237: 57-62.

APPENDIX I - OBSERVATION DATA STANDARD WORK GROUP

Table 1. Members of the Observation Data Standard Work Group. Section column refers to either the NatureServe Section Councils or NatureServe organization, where CA = Canadian Section Council, NS = NatureServe, NSC = NatureServe Canada, US = U.S. Section Council.

Section	ORGANIZATION/ AGENCY	Last Name	First name	Position	Contact Information
CA	Ontario	Gemmiti	Sandy	Data Integration Analyst	sandy.gemmiti@ontario.ca
CA	Ontario	Sorrill	Peter	Natural Heritage Information Specialist	peter.sorrill@ontario.ca
CA	Quebec	Hebert	Nancy	Coordinator	Nancy.Hebert@environnement.gouv.qc.ca
CA	Saskatchewan	Benville	Andrea	Data Manager	andrea.benville@gov.sk.ca
US	California	Nelson	Misty	Lead Scientist	misty.nelson@wildlife.ca.gov
US	Kentucky	Salam	Nour	Geoprocessing Specialist	nour.salam@ky.gov
US	Michigan	Rogers	Rebecca	GIS Specialist and Database Administrator	RLR@msu.edu
US	Minnesota	Cieminski	Karen	Natural Heritage Information Manager	karen.cieminski@state.mn.us
US	Montana	Coleman*	Karen	Biological Data System Coordinator	kvwcoleman@mt.gov
US	Montana	Maxell	Bryce	Program Coordinator	BMaxell@mt.gov
US	Nevada	Johnson	Janel	Botanist/Webmaster	jdjohnson@heritage.nv.gov
US	New York	Buff	Matthew	Web/Database Manager	matthew.buff@dec.ny.gov
US	Oklahoma	Fagin	Todd	Database Manager	tfagin@ou.edu
US	Oregon	Gaines	Eleanor	Director	egaines@pdx.edu
US	Texas	Gottfried	Bob	Texas Natural Diversity Database Administrator	bob.gottfried@tpwd.texas.gov
NS	NatureServe	Faber-Langendoen**	Don	Senior Ecologist & Conservation Methods Coordinator	don_faber-langendoen@natureserve.org
NS	NatureServe	Fernandez	Miguel	Director, Latin America and the Caribbean Programs	miguel_fernandez@natureserve.org
NS	NatureServe	Goodin	Kathy	Vice-President, Data & Methods	kathy_goodin@natureserve.org
NS	NatureServe	Gratz	Allison	Director of Network Relations	allison_gratz@natureserve.org
NS	NatureServe	Weber	Whitney	Database Administrator & Product Support Specialist	whitney_weber@natureserve.org
NS	NatureServe	Young	Bruce	Chief Zoologist and Senior Conservation Scientist	Bruce_young@natureserve.org
NSC	NatureServe Canada	Enns	Amie	National Data Manager	aenns@natureserve.ca
NSC	NatureServe Canada	Siemens Worsley	Allison	National Data Support Biologist	aworsley@natureserve.ca

* Karen Coleman is now Systems Analyst with the Montana Department of Natural Resources.
**Don Faber-Langendoen was chair of the Work Group.