Version 1.0 Date: September 30, 2006

Observational Data Standard, Version 1.0

Table of Contents

I. Purpose of the Standard Development Effort and Definitions

a. Goals / Purpose

Observational data are recognized to be fundamental for scientific inventory, conservation planning, habitat management, invasive-species assessments, predictive range modeling, monitoring, and much more.

The goal of this project has been to collaboratively develop a provisional standard for observational data that can be broadly applicable regardless of data or survey type. The standard is intended to benefit the research and conservation communities by facilitating data aggregation and sharing within and between organizations, such as data discovery through global search portals, and by fostering interoperability and collaboration.

To achieve these broader goals, NatureServe is a co-sponsor of the Taxonomic Databases Working Group (TDWG) Observational Data Subgroup. NatureServe will be offering this provisional standard as an input for that group's work towards an international observational data standard.

b. Working definitions

An **observation** is a set of information that describes the presence or absence of an organism or assemblage of organisms through a data collection event at a location. Observations can be

documented by physical specimens or vouchers, as well as through the collection of different attribute data including but not limited to location, abundance, distribution, reproductive status or phenology, ecological associations, and environmental conditions. Observations are not necessarily independent, but can be linked via characteristics such as time, place, protocol, and co-occurring organisms.

Although there are numerous methods used to collect these different types of observational data, the methodologies can be categorized into two general classes:

> **Place-based protocols** – The survey effort is focused on characterization of a defined locality, with the resulting observations typically consisting of a checklist of species observed and a measure of their abundance. In this context, individual taxa can be considered attributes of the observation record.
>
> **Taxon-based protocols** – The survey documents the individual occurrences of single or multiple species or ecological elements in the environment, and results in a collection of records describing their location, distribution, and condition. In this context, place (i.e., location and extent) can be considered an attribute of the observation record. Such inventories may focus on characterizing the set of biological elements within a specified site (e.g., a nature preserve, small watershed, or county), with the result being a set of discrete observation records in which the taxa are the objects, rather than attributes, of the observation.

Monitoring is an activity that results in a set of observational data gathered during repeated visits over a period of time using consistent or comparable methodology. Generally, the overall focus in monitoring is the detection and documentation of change and trends over time .

II. Review Process

Two review cycles were used in developing version 1.0 of the Observational Data Standard. Reviewers were from the Observation Working Group, NatureServe and member network programs, as well as external partners such as the U.S. National Park Service, the Canadian Wildlife Service, the University of North Carolina, and the Cornell Laboratory of Ornithology. The first review was initiated in March 2006, while the second began in May 2006. Feedback from reviewers used to revise and refine the Observational Data Standard. In addition to these reviews, version 1.0 of the standard will be submitted for consideration by the TDWG Observational Data Subgroup in the development of their standards work.

III. Requirements

The following requirements were identified through meetings of the Observation Working Group, input received from participants at regional NatureServe conferences, discussions among the project team, and review by external reviewers. The requirements list should not be considered complete, as further suggestions are expected with broader review.

Requirements – The Observational Data Standard shall:
1. Be independent of any particular software implementation in order to be adaptable to a variety of current or future applications.
2. Initially focus on "core" attributes that are as universal as possible and apply to observations of all taxonomic groups and classification types, regardless of the origin (nonindigenous vs. native) and of whether the data represent negative survey results.
3. Identify a smaller subset of required attributes.
4. Incorporate and be compatible with existing standards where appropriate. Some examples could include: Darwin Core, Taxon Concept, IUCN habitats, IUCN/CMP threats, and IUCN conservation actions.
5. Accommodate data from both taxon-based (focused on individual occurrences of single or multiple species or ecological elements) and place-based (focused on characterization of a defined locality, such as a plot) surveys.
6. Include both species and collections of species (e.g., ecological communities, animal assemblages) as data types.
7. Permit linkages between place-based and taxon-based observations.
8. For place-based observations, allow users to record a species list.
9. For place-based observations, allow users to capture certain minimum information for each taxon on a species list, without the need to create a full observation record for each.
10. Accommodate negative data—explicit documentation of locations where a species or ecological community was not observed to be present at a specific moment in time.
11. Accommodate monitoring data, a set of observational data gathered during repeated visits over a period of time using consistent methodology, usually for the purpose of detecting and documenting changes and trends.
12. Allow users to group observations by any user-specified criteria to enable various types of data analyses.
13. Allow multiple observations to be linked as part of the same survey. The survey would serve as an umbrella for a group of related observations.
14. Allow multiple surveys to be linked as part of the same project. The project would serve as an umbrella for a group of related surveys.
15. Allow users to record information on conditions or actions at a particular moment in time: quantity observed, management data, management actions, conditions, events (natural and anthropogenic events such as flood, burn, landslide, chemicals applied, etc).
16. Include nonindigenous species as a data type, with attributes to track invasiveness.
17. Include attributes for assessment of data quality and validation. Because it may not be possible to verify an observation after it was originally made, there are differing levels of quality and/or confidence/reliability associated with observational data (e.g., whether a taxon or ecological community observed was correctly identified, confidence in various biological attributes, confidence / accuracy of locational information, etc.). Validation could include information about: whether or not the information was collected at a location, on a specific date/time, and under conditions appropriate for the targeted taxa by a person familiar with the organism(s).
18. Biological data – allow extensibility to make the Observational Data Standard the most useful to specific communities of users. (NOTE: Ideally a community of users [such as the bird monitoring scientists] would develop a standard extension of user-defined fields to the core standard.)

IV. Required Fields

The following fields are proposed as required for a minimally complete observation record.

The definition of a set of required fields will be critically important for data sharing and aggregation, either among partners or through international data portals such as the Global Biodiversity Information Facility (GBIF).

For field definitions and additional details, please see the data dictionary (Appendix B) and Section VI "Major Entities and Subject Areas."

Field	Comments
Unique identifier(s) of the observation record and its source	For data sharing, a way to uniquely identify a record and its database of origin is required.
Formal name (scientific name)	Can be provided at the most precise level of identification available (e.g., at the genus or family level for a taxon with uncertain identification); use TDWG Taxon Concept standard.
Where found (GIS shape, coordinates, and/or a text description)	Some type of mappable location information is required, whether or not the location is actually mapped in a GIS.
Name of observer(s)	Full name of the person(s) who made the observation, to the amount of detail possible. If the observer is not known (such as for historical data), "unknown" is an acceptable value.
Date of observation, including: • visit start date • visit end date • imprecise (verbatim) dates (such as "2001-Spring")	Enter date to the most specific / detailed level possible; at least a four-digit year or range of years is required.

VI. Major Entities and Subject Areas

The Observation Working Group defined a set of core concepts and data for an Observation record during its workshop in June 2005 and subsequent reviews. A subset of the group[1] then developed a draft data standard by grouping the data into entities and establishing the numeric relationships or "cardinality" between the entities. "Entity" is a data modeling term for a logical grouping of pieces of information, or "attributes."

[1] NatureServe staff Lynn Kutner, Jennifer Nichols, and Donna Reynolds, with NatureServe data modeler Carol Fogelsong.

This section gives a description of the major parts of the Observational Data Standard and how they relate to each other. For each major entity, a list of attributes with draft definitions is included in a data dictionary, which is attached as an Excel spreadsheet (Appendix B). The data dictionary also indicates which attributes are required, which are considered core to the standard, and, among core fields, which should receive highest priority.

A conceptual data model is attached in the Appendix to help readers visualize the relationships between the entities.

a. Observation

This entity contains the information that defines an observation: what was observed, where, when, and by whom. It also includes attributes for recording biological data about the taxon or community and information on the process used to make and document the observation.

i. Identification: To identify the object of the observation (taxon or community), the Observation entity contains the required attributes scientific name and name type, as well as nonrequired but priority core data such as scientific name author and a taxonomic concept reference. The latter, along with the scientific name used in the concept reference, is essential for data sharing, as these attributes allow both parties to be sure they agree on the underlying circumscription of the taxon or community, even if they call it by a different name. Also included are attributes for indicating the degree of confidence in the identification, and for supplying a secondary designation for communities (where the observation represents a mixed type) or for taxa (where the identification is not certain). Because it may not be possible to identify a taxon or community to the finest taxonomic level prior to analyzing the observational data, an Observation record may be created based on preliminary identification or using a scientific name at a higher level of the classification hierarchy and indicating that the name assigned is provisional. After the identification is narrowed, it is possible to preserve the earlier interpretation (if desired) by recording it as the "verbatim name." For communities, there is a need to track changes in the identity of the community over time, as succession occurs. This can be done using an “observation grouping” (see section IV.a.vii) to link together Observation records with different identifications, indicating that the difference is due to succession.

The environmental conditions and the biological characteristics of the element[2] at the time and place of the observation can also be recorded. These include whether the element is native or nonindigenous, whether it is invasive, its abundance and density, reproductive status, and associated species. See the data dictionary for a full list of biological and environmental attributes.

ii. Location: Ideally, the location of the observation will be represented as a GIS feature (point, line, or polygon). However, recognizing that historical observational data sometimes lack precise geographic information (e.g., a plant observation with only a county name), and that some users

[2] The term "element" is used in the NatureServe network to mean any unit of natural biological diversity. Elements may be species (or infraspecific taxa), natural communities, ecological systems, or other nontaxonomic biological entities (e.g., migratory species aggregation areas).

of this proposed standard may not use GIS, the relationship between an observation record and a GIS feature is optional. Instead, location can **either** be mapped in a GIS **or** described as precisely as possible by using coordinates and datum and/or by filling out one or more fields in the Country – State/Province – County hierarchy and/or by filling out the Location Description text field. Therefore, a valid observation record must be mappable in the broad sense but does not necessarily have to be mapped. In the example mentioned above of a plant with only a county as the location, the observation could be mapped as a polygon whose boundaries coincide with the county in which the plant was recorded, but it may not be desirable to map it as a large polygon in a GIS. In this case, the county name would be put in the County field, and the state and country names would also need to be entered.

iii. Date: In order to accommodate different levels of date precision, the standard contains several different date fields, but only the applicable field(s) would be filled in for any given observation. For example, if a precise year-month-day is known, only the Observation Date field is filled in. The minimum requirement is to enter an Observation Year. Dates should be compliant with ISO standard 8601.(Implementation note: An application could be programmed so that the user will not have to choose between the array of multiple date attributes shown in the data dictionary but would simply enter the available date information and have it parsed into the proper fields "behind the scenes.")

iv. Observer: Observer is a required attribute, but "Unknown" will be an allowed value for cases where the observer's name is not available. The data standard allows for multiple observers per Observation record, each of which can be assigned a role. Information such as address, phone number(s), email, and organization may be included if available. In order that such information be kept up-to-date as easily as possible and also to provide for consistency in the spelling of personal names, which is important for searchability and reporting, it is recommended that data about observers be stored in a Contacts database and linked to the Observation record, rather than having one or more text fields where observer names and contact information are entered.

In the Observation record itself, a "verbatim observer name" text field can be used to record the name in its original format in the observational data if it differs from the name in the Contact record or if the full name is unknown but some identifying information (like initials) is available. A "verbatim affiliation" field can be used to record the institutional affiliation of the primary observer at the time the observational data were collected. (Implementation note: It should be simple to review existing Contact records to prevent multiple records from being created for the same person but with different spellings.)

v. Observation Methods and Evidence: The Observation record can be linked to information about the protocol used to collect data. Search adequacy, level of effort, and the need for further inventory work may also be recorded. Observation Evidence Type should be recorded to indicate the basis of the Observation record.

vi. Negative Observations: A "negative observation" is indicated through a simple found/not found attribute. An associated confidence field allows for a rating of confidence that the search results confirm either presence or absence. (Examples: species not found, but believed to exist in the area—low confidence in negative result; species found as remains of dead individual in

unlikely habitat—low confidence in positive result.) Comments about the adequacy of the search and the reasons for the confidence rating may be added, regardless of whether the element was found or not found. The search area polygon, line, or coordinates would be the mapped representation of the negative observation, but users may vary in their need or desire to map negative observations.

vii. Monitoring: Monitoring efforts are represented by multiple observations made over time that can be linked together through common attributes, such as the identity of the element and the search area. The proposed standard actually goes a step further in that it allows the grouping of Observations records to be done for any reason specified by the user. The entity Observation Grouping represents one or more Observations that a user has chosen to associate together. For example, a user could choose to create Observation records for the constituent species of a community as well as for the community itself and link them all together. Observation Grouping is also related to the Protocol entity, in order to allow a group of observations to be associated with a specific protocol.

b. Survey

A Survey is defined as a coordinated effort to gather information on biological entities. Such an effort may result in numerous Observation records. Survey is not represented by a spatial feature, but both the search area and the observations resulting from the survey can be mapped.

In the Observational Data Standard, the Survey entity is more important as an organizing concept than as a repository for information. It includes attributes for descriptive comments about the survey, as well as information on how it was performed (through a link to the protocol(s) used) and why (through a link to a project). The "where," "what," "when," and "who" information is stored in the Observation and Search Area records.

Surveys can be either taxon-based (the focus is to record data on a particular taxon or taxa) or location-based (the focus is to characterize the biological elements present in a defined location). One specialized type of location-based survey is a vegetation plot, but a location-based survey may also be performed in all or part of an ecologically or geopolitically defined area (examples: a swamp, a state park, a county).

For a plot type of location-based survey, one or more Observation records may be created for the ecological community(ies) represented in the plot, but it is also possible to:

- List the species found as well as create a community Observation record;
- List the species found **and** create Observation records for these constituent species, in addition to the Observation record created for the community in which they occur.

This approach was taken to provide maximum flexibility.

c. Search Area

Information about the place where a survey took place is managed through the Search Area entity and its associated GIS shape. It may be a plot, protected area, ecoregion, or some other defined type of location. A Search Area record may be associated with a Survey (which in turn may be linked to several Observations) or with an individual Observation record. Therefore, it is not necessary to map a Search Area for every Observation made within a defined area, but it is possible to do so if desired. Search Areas may be nested within other Search Areas. They can be linked to one or more references or to identifiers of externally maintained information about the plot or other location type. A Search Area may be mapped as a point (e.g., coordinates), line, single polygon, or multiple polygons.

d. Species List

When the search area is a vegetation plot, the outcome of the survey may be a list of species found in the plot, with data such as cover class and stratum for each species. These data are stored in the Species List entity. The Species List is linked to the Observation entity, which allows date and observer information to be recorded for a plot survey.

Putting a species on a Species List does not make it the object of an Observation record. The associated Observation record would be about the *community(ies)* that occur in the plot, not the constituent species. However, it is recommended that the software in which the standard is implemented allow users to easily create selected Observation records for species in the Species List. In addition, there should be a way to tell when Observation records have been created from the Species List. If Observation records generated from the Species List are mapped as polygons, they should be flagged automatically to indicate that the mapped area of the species' occurrence was constrained by the size of the plot. Unless the decision is made to create Observation records for species on the SpeciesList, they are not mapped as individual polygons.

A Species List is not the same thing as a list of Associated Elements. The data standard allows for both types of information to be recorded. Associate Elements are any other taxa or communities co-occuring in the vicinity of an Observation of a different taxon or community. For an Associated Element, the observer cannot record data other than its relationship to the observed element and comments. If an observer does want to record other data, such as metrics on its abundance, he or she should create an Observation record.

e. Documentation

Observation, Survey, Search Area, Protocol, and Project records may all be linked to metadata about external documentation and/or to an internal reference database. A link to documentation (external or internal) is not required for any of these record types, but each of them may be linked to multiple sources of documentation.

The External Documentation entity provides a way to link to information that is managed by other institutions and data that is beyond the scope of the Observation record. An example is information associated with voucher specimens. The specimen ID and museum or herbarium

name are stored in the External Documentation record, so that original specimen data may be located if desired, without the need to enter and managed it in multiple places.

f. Protocol

The protocol is the plan or procedure used to collect data. The standard assumes that detailed information about a protocol will not be managed as part of the Observation record. Therefore, the Protocol entity contains minimal information (protocol name, description, and type), but it may be linked to detailed documentation through the External Documentation entity or a Reference database.

There is a link between a Protocol and an Observation. More than one Protocol may also be specified for a Survey.

g. Project

A project is a formal plan of work during a defined time period. The standard assumes that detailed information about a project will not be managed as part of the Observation record, although it may be housed in the same database. Therefore, the Project entity contains minimal information, but it may to linked to documentation through the External Documentation entity or a reference database.

Linkage of a Survey to Project information is not required.

h. Cardinality of Relationships between Entities

The following relationships are illustrated in the Conceptual Data Model, attached as Appendix A.

Observation

- An Observation may be related to zero or one Survey.
- An Observation may be related to zero or one Protocol.
- An Observation may be related to zero or one GIS Shape.
- An Observation may be related to zero, one, or many References.
- An Observation may be related to zero, one, or many External Documentation records.
- An Observation may be related to zero or one Element List.
- An Observation may be related to zero, one, or many Observation Groupings.
- An Observation may be related to zero or one Search Area.

Protocol

- A Protocol may be related to zero, one, or many Observations.
- A Protocol may be related to zero, one, or many Surveys.
- A Protocol may be related to zero, one, or many Observation Groupings.
- A Protocol may be related to zero, one, or many References.
- A Protocol may be related to zero, one, or many External Documentation records.

Observation Grouping

- An Observation Grouping may be related to zero or one Protocol.
- An Observation Grouping may be related to zero, one, or many Observations.

Survey

- A Survey may be related to zero, one, or many Observations.
- A Survey may be related to zero, one, or many Protocols.
- A Survey may be related to zero, one, or many Search Areas.
- A Survey may be related to zero or one Project.
- A Survey may be related to zero, one, or many References.
- A Survey may be related to zero, one, or many External Documentation records.

Project

- A Project may be related to zero, one, or many Surveys.
- A Project may be related to zero, one, or many References.
- A Project may be related to zero, one, or many External Documentation records.

Search Area

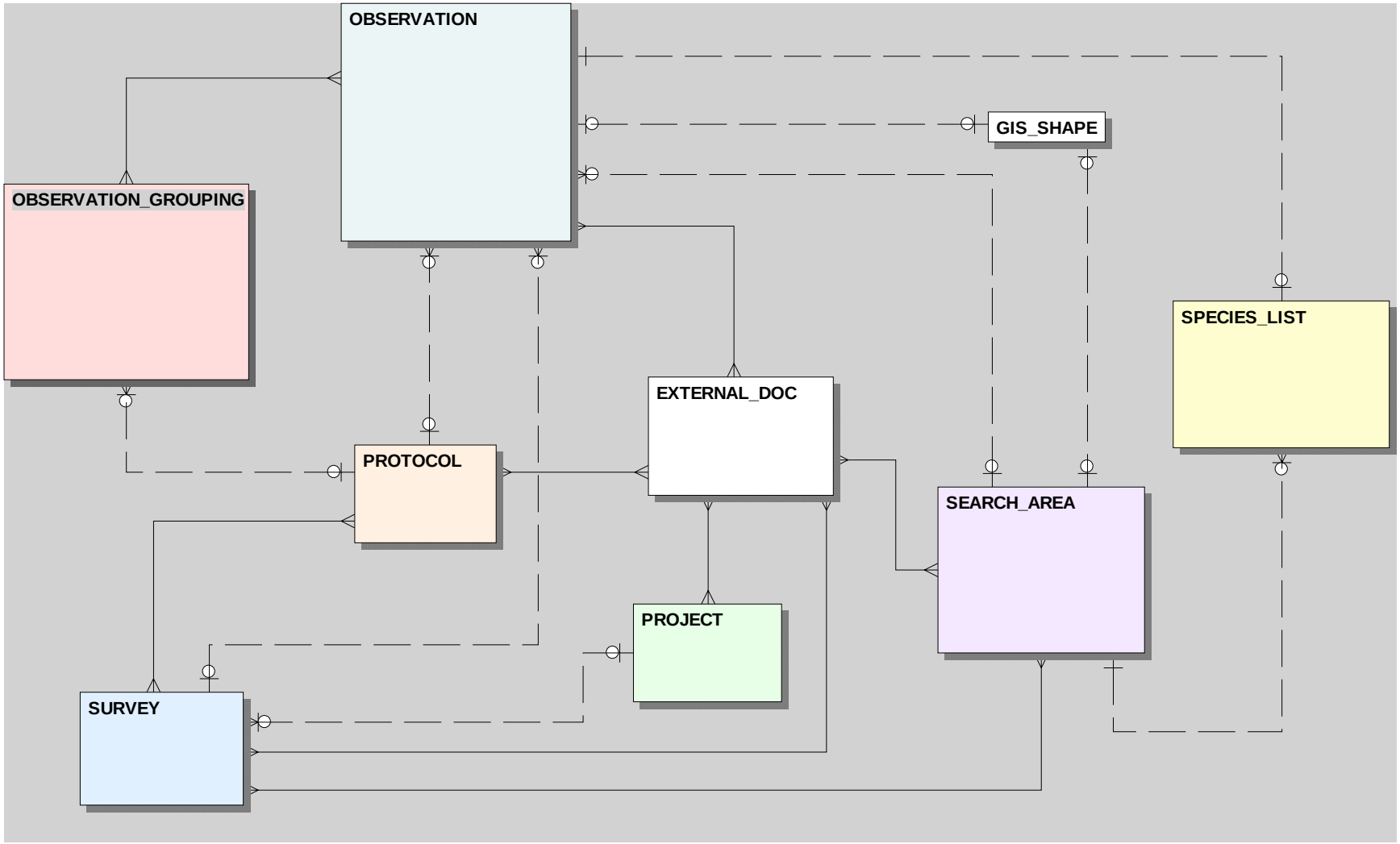
- A Search Area may be related to zero, one or many Surveys.
- A Search Area may be related to zero, one, or many Observations.
- A Search Area may be related to zero or one GIS Shape.
- A Search Area may be related to zero, one, or many Element Lists.
- A Search Area may be related to zero, one, or many References.
- A Search Area may be related to zero, one, or many External Documentation records.

Species List

- A Species List must be related to exactly one Search Area.
- A Species List must be related to exactly one Observation.

VII. Appendices

A. Conceptual Model of Observational Data Standard Entities and Relationships



KEY

Zero or one instance of Entity A may be related to zero or one instance of Entity B

Zero, one, or many instances of Entity A may be related to zero, one, or many instances of Entity B

One instance of Entity A is related to zero, one, or many instances of Entity B

Zero or one instance of Entity A is related to zero, one, or many instances of Entity B

B. Observation Data Dictionary

See separate Excel document.