Basics

One page for every species.

Each page to have the following:

- To be linked to others by an inbuilt taxonomic hierarchy.
- To allow for more than one nomenclature.
- To include specimen and species information as well as individual sightings/records.
- To include fields for multimedia and references (textual, digital).
- To allow for continuous changes/building of content; validation. To record changes in versions.
- To include information for more than interest group within a single field.
- To provide metadata on each version/change.
- To allow download of any one version with details of that particular version.
- To allow for information in more than one language to be included.
- To talk with the Map, Checklist features currently on IBP. Also, integrate with a Theme Feature, if built in future.
- To conform to globally accepted standards for exchange of biodiversity information.

Relevant Global Standards and Concepts

• Many organizations work towards developing and ratifying global standards for sharing information:

CODATA
Environmental Exchange Network
e-Learning Interoperability Standards (eSI)
The Global Grid Forum (GGF)
Institute of Electrical and Electronic Engineers (IEEE)
Internet Engineering Task Force (IETF)
Java Community process
International Organization of Standards (ISO)
Organization for the Advancement of Structural Information Standards (OASIS)
Web Services Interoperability Organization (WS-I)
World Wide Web Consortium (W3C)

Refer: http://www.tdwg.org/about-tdwg/faq/

• The most widely accepted group working towards establishing such standards for sharing of biodiversity information in particular is the Biodiversity Information Standards (BIS), more popularly known as the Taxonomic Data Working Group (TDWG).

URL: http://www.tdwg.org

TDWG in Brief

• Who?
Non-profit science and education association. Members are people in museums and natural history libraries including Smithsonian, GBIF etc from different continents.

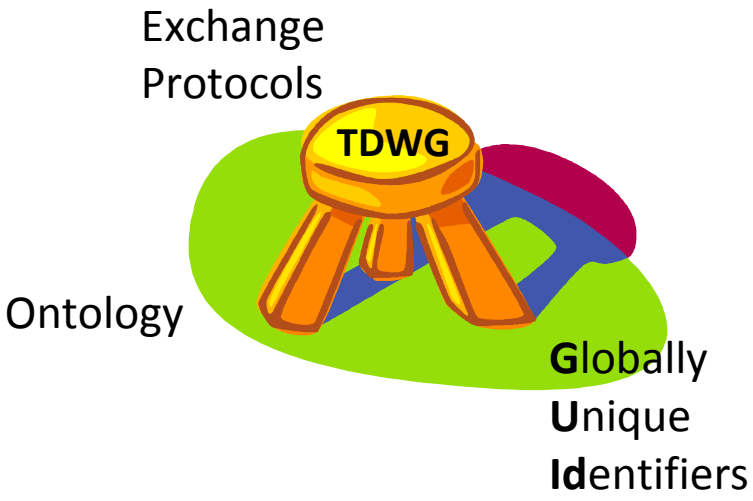
• What?
Develop, adopt and promote standards and guidelines for the recording and exchange of data about organisms.
Promote use of standards through the most effective and appropriate means.
Act as a forum for discussion through holding meetings and releasing publications.

• For?
Many organizations, institutions and individuals utilize, conserve and support the exchange of biological collection data over the Internet. These include:
*Archiving initiatives
*Biodiversity data holders (e.g. academia, museums, libraries, herbaria, botanical gardens, government agencies, researchers, students)
*Biodiversity network developers (e.g. BioCASE, GBIF, IABIN)
*Consumers of biodiversity data
*Developers of collections management systems (e.g. Specify, KEmu)
*GBIF Nodes
*TDWG sub groups and members.

All TDWG Standards

- ✓ LSID Applicability Statement (2008)
- ✓ NCD (2008)
- ✓ ABCD Extension for Geosciences (2007)
- ✓ TDWG Access Protocol for Information Retrieval—TAPIR (2006)
- ✓ Applicability Statement for the Deployment of Life Science Identifiers (LSIDs) in Bioinformatics (2006)
- ✓ Access to Biological Collection Data (ABCD) (2005)
- ✓ Natural Collections Descriptions (2005)
- ✓ Structure of Descriptive Data (2005)
- ✓ Taxonomic Concept Transfer Schema (2005)
- ✓ Delta 3 (2004)
- ✓ Darwin Core (2003)
- ✓ Distributed Generic Information Retrieval DiGIR—2 (2003)
- ✓ International Transfer Format for Botanic Garden Plant Records –2 (2003)
- ✓ HISPID3 - Herbarium Information Standards and Protocols for Interchange of Data (1996)
- ✓ Economic Botany Data Collection (1995)
- ✓ Plant Occurrence and Status Scheme (1995)
- ✓ Plant Names in Botanical Databases (1995)
- ✓ Authors of Plant Names (1992)
- ✓ World Geographical Scheme for Recording Plant Distributions (1992)
- ✓ Botanico-periodicum-huntianum/supplementum (1991)
- ✓ Index Herbariorum. Part I: The Herbaria of the World (1990)
- ✓ Users Guide to the DELTA System (1986)

TDWG Design Principles



Principle 1: **TDWG standards apply to shared data.**

- Enables interoperability of providers and consumers with radically different internal implementations
- Does not dictate internal structures on data providers or consumers
- 'Format' matters when data crosses boundaries.

Principle 2: **Biodiversity data will be modeled as graph of identifiable objects.**

- Objects are defined by an ontology: Understandable by humans and computers
- Requires globally unique identifiers to link objects across the network
- Requires a transport protocol to 'wrap' the biodiversity data for transport

Recommended TDWG Standards

- Most widely deployed formats for biodiversity **occurrence** data:

Darwin Core.

Access to Biological Collections Database (ABCD) .

- XML **exchange** protocol:

TDWG Access Protocol for Information Retrieval (TAPIR).

- **Tagging** Data with Identifiers:

Globally Unique Identifiers (GUID) especially Life Science Identifiers (LSID).

ABCD and Darwin Core

• Both are **complementary** standards for specimen based biodiversity information:
"conceived to facilitate the discovery, retrieval, and integration of information about modern biological specimens, their spatiotemporal occurrence, and their supporting evidence housed in collections (physical or digital). "

ABCD
Introduction: http://www.bgbm.org/TDWG/CODATA/Schema/default.htm
Schema:
http://www.bgbm.org/scripts/ASP/TDWG/frame.asp?config=0&configurl=http://www.bgbm.org/TDWG/CODATA/S

Darwin Core
Introduction: http://rs.tdwg.org/dwc/
Terms and Definitions: http://rs.tdwg.org/dwc/terms/index.htm#Taxon

• Comparison
The **Darwin Core 2 (DwC) is a smaller set of data element definitions** designed to support the sharing and integration of primary biodiversity data.

DwC has a flat structure of 44 elements, ABCD has a hierarchical structure that supports repeating elements and complex types. In order to accommodate detailed data, often according to multiple standards, as well as free text alternatives where detailed data cannot be provided, ABCD contains about 700 elements. While making the mapping process and the development of data portals more demanding than with DwC, this complexity widens the application of the schema from a minimum-common-denominator approach for resource discovery to a standard that can meet the demands where more detailed data are needed.

To accomplish interoperability, the development of both ABCD and DwC will strive to keep both standards compatible on the element level, so that a mutual translation of documents will be possible.
http://www.bgbm.org/TDWG/CODATA/Schema/Mappings/DwCAndExtensions.htm

TAPIR

What?
A computer protocol designed for discovery, search and retrieval of distributed data over the Internet. It consists of a specification that determines how client applications seeking information should communicate with server applications hosting data. (http://www.tdwg.org/activities/tapir/executive-summary/)

How?
• specifies a standardized, stateless, HTTP transmittable, XML-based request and response protocol for accessing structured data that may be stored on any number of distributed databases of varied physical and logical structure.
• combines and extends features of the BioCASe and DiGIR protocols to create a new and more generic means of communication between client applications and data providers using the Internet.

Where?
http://www.tdwg.org/standards/449/

Key Features
• Integrates well with the Web infrastructure because it is based on established international Web standards like HTTP, XML and XML Schema.
• Is independent of the data being exchanged and therefore can be used to access a wide range of data.
• Can return data in many different formats - search responses are flexible and customizable.
• Includes five operations to cover all necessary aspects for searching and retrieving data in a network: access to metadata (descriptive data) and capabilities of the provider, preliminary data discovery and data mining, searching, and monitoring the data provider service.
• Allows all operations to be invoked through simple Web addresses.
• Supports complex requests.

GUID

What?
Unique Identifiers that are used to tag digital objects and other resources on the internet within a database for quick and accurate retrieval.

Principles (http://www.tdwg.org/about-tdwg/news/article/lsid-and-guid-applicability-statements-approved/)

Important -
*7.A GUID technology **should** be chosen from the list of recommended GUID types.*

HTTP URI (this technology is used as a basis for some of the following options). A URI scheme whose identifiers are prefixed with “http://”. An HTTP URI can be used to locate network resources via the HTTP protocol, and therefore supports the Linked Data practices well.

LSID — Life Science Identifier. A particular kind of actionable identifier, recommended for use by TDWG, which enables global uniqueness by including an Internet domain name, which is itself subject to rules and procedures ensuring uniqueness, and uses the domain name system to locate a resolution service that enables a user to find out more about the entity to which an LSID refers.

DOI — Digital Object Identifier. A digitally managed system for persistent identification of entities. The term "DOI" is understood to mean "digital identifier of an object", rather than "identifier of a digital object". As well as identifying content items such as digital files and digital media manifestations of intellectual property, DOI names can also identify physical objects, performances and abstract works.

PURL — Permanent URL. An HTTP Uniform Resource Identifier (URI) (i.e. location-based Uniform Resource Identifier or URI) with a redirect mechanism. It does not directly describe the location of the resource to be retrieved but instead describes an intermediate (more persistent) location which, when retrieved, results in redirection (e.g. via a 302 HTTP status code [http://www.w3c.org/Protocols/rfc2616/rfc2616-sec10.html]) to the current location of the final resource.

UUID — Universally Unique Identifier. A GUID created by an algorithm that virtually guarantees that no two identical UUIDs will ever be generated at any time or place. This avoids the need to check for identical GUID values when generating identifiers, and ensures that any search application will only return a single interpretation.

Handle System. A technology specification for assigning, managing, and resolving persistent identifiers for digital objects and other resources on the Internet. The protocols specified enable a distributed computer system to store identifiers (names, or handles), of digital resources and resolve those handles into the information necessary to locate, access, and otherwise make use of the resources. That information can be changed as needed to reflect the current state and/or location of the identified resource without changing the handle.

*2. Providers **must** assign at most one GUID to any particular object.*

It makes sense to give physical objects, such as specimens, a single identifier that everyone reuses, but for abstract
objects, this becomes harder to enforce when most data holders have their own versions of the abstract concepts.

The task of integrating all data holder concepts and identifiers is currently considered an impossible task. It therefore follows that separate identifiers will exist for possibly the "same" object. Following the linked data model for globally integrated data, the preferred method to handle this situation is to ensure that links are maintained between the various versions of the same concept.

Examples of objects in the biodiversity domain that **should** be assigned GUIDs—
•scientific names;
•taxonomic concepts;
•taxon name usages;
•observations;
•individual organisms and observations;
•published and unpublished reference citations (e.g., literature);
•specimens;
•collections;
•images, videos, and sound recordings.

Others -

3. HTTP GET resolution **must** be provided for non-self-resolving GUIDs.

4. Only one globally unique identifier should be assigned to each object.

5. Providers should only assign GUIDs to objects for which they are the authority.

6. Aggregators should assign new GUIDs to derived objects.

7. GUIDs should be resolvable.

8. Information systems should use existing GUIDs when available to refer to external objects.

9. Aggregators should use GUIDs and the Dublin Core metadata term 'source' to link derived objects to their sources.

10. The default metadata response format **should** be RDF serialized as XML.

11. Objects in the biodiversity domain that are identified by a GUID should be typed using the TDWG ontology or other well-known vocabularies in accordance with the TDWG common architecture.

12. Working Groups and Use Case Engineers should determine the degree of change required for GUID reassignment.

13. GUID authorities should use appropriate metadata properties to represent relationships between revisions of an object.

14. Clients must not try to infer relationships between objects based on any part of a GUID. Instead, clients must dereference the GUID and retrieve any assertions about revisions from the returned metadata.

Species Profile Model (SPM)

What, Why? (http://wiki.tdwg.org/SPM)

TDWG has established/ratified more than one standard - ABCD Schema, Dublin Core, Darwin Core etc - schema to define, categorize and share specimen based biodiversity information online.

But, groups like the EoL would like to create portals to aggregate species information - which means, information for species that we may/may not have specimens in any natural history museum/ university for; information that goes beyond just the classification and location records of a specimen and includes everything known in scientific and other communities about behaviour, ecology, habit, reproduction, folklore etc; and allows for records of occurrence of the species in the wild/in national parks or zoos, in cities, villages and homes, even. GBIF also has plans to expand it's site to include this kind of information.

Now, the TDWG has to also start building/reviewing standards for Species pages that are applicable globally and are on par with those for Specimen pages. Since not much work has been done on developing a schema for Species Pages, TDWG simply calls their draft schema SPM or Species Profile Model (http://wiki.tdwg.org/SPM). This is loosely based on what EOL has built to categorize all information on it's site.

A variety of people are working on the TDWG Wiki for SPM, including their inputs on what they think the schema should contain. It is under this project that the developers of the Plinian Core, one of whom is from GBIF, have submitted their format for review. As of now, the Plinian Core is still under review as an accepted TDWG standard for SPM.

Plinian Core

What?

A set of concepts that defines the necessary basic attributes to integrate and recover the information about species of organisms required by users specialized in biodiversity as well as users of other areas.

Why?
The current version compiles a list of high-priority concepts selected by different groups of people interested in having apt mechanisms to access integrated information of species using Internet.

Who?
Development led by the Costa Rican and the Spanish nodes (CRBio and GBIF.ES) of the Global Biodiversity Information Facility (GBIF) network.

Organizations involved:
GBIF.CR
GBIF.ES
Herbarium of the University of Granada
Species and Specimen Thematic Network

Validation? (http://www.pliniancore.org/en/antecedentes.htm)
Species and Specimens Thematic Network of IABIN (SSTN) selected the Plinian Core as the standard to share species information inside the network; SSTN is part of the project and is developing compatible tools using it.

Where?
http://www.pliniancore.org/en/inicio.htm

Objectives

•To specify the necessary basic concepts to integrate and to recover the information about species of organisms that is distributed in data bases that are managed by institutions around the world.

•The model must be simple but with the adequate degree of detail to gather the information required by the target audience.

Target Audience

•Decision makers.
•Biodiversity researchers (Scientists, taxonomists, biologists, technicians, biochemical, health professionals , university professors and students, among others);
•Teachers, students and opinion makers (teachers, journalists, religious, among others);
•Producers that make a sustainable use of the resource;
•Professionals outside of biodiversity (literary authors, historians, ethnographers , among others) and graphic designers (Illustrators, painters, among others);

General Principles of Design

•The model objective does not include being used as a database structure to develop systems to capture information.
•The model must reuse the existing international standards.
•Must support the combination of data from multiple database.
•Must allow sharing structured information although it is not normalized properly.
•Must give support to information in multiple languages.
•Must allow the Integration of data of relatively common use: Concepts used in multiple disciplines, Concepts that are commonly in relevant databases, Concepts useful for the selected target audience.
•Must allow managing species records versions.
• The design must be structured using a core and extensions.