International Union for Biological Sciences
Taxonomic Databases Working Group
http://www.tdwg.org

Authors: Lee Belbin and Roger Hyam
TDWG Infrastructure Project
Last modified: 2006-01-24

A Documentation and Supporting Software Strategy for TDWG

Index

1. Introduction

This report is a strategy for producing the first normative documents. This document will feed into the subsequent process document and the construction of TDWG's collaborative environment.

2. Definitions

2.1. Documentation

Documentation is the recording of information to define and support standards in a permanent format.

2.2. Supporting Software

Supporting software includes reference implementations, test suites and code libraries - specific tools for specific standards. It is separate from the collaboration environment which is used to develop the standards.

3. Current Best Practice

Study of other standards bodies (see Appendix B) indicates the following is best practice in relation to documentation:

a. The organisation uses documents as primary outputs.
b. The organisation has clearly specified its documentation process.
c. The specification of documentation is included within the standards process itself to allow for controlled evolution.
d. Clear documentation templates and style guidelines are provided.
e. Clear IP and copyright policies are used.

4. Recommendations for TDWG Documentation and Supporting Software

These recommendations are based on current TDWG standards (Appendix A) and best current practice (Appendix B). The recommendations govern the three broad categories of TDWG standards identified:

a. Administrative standards to control the TDWG standards process (see 4.9).
b. Applicability statements on the use of existing TDWG and non TDWG standards.
c. *De novo* standards for data modelling and data exchange.

4.1. All TDWG standards must be represented as a folio of documents

At a minimum, each standard should contain:

a. The normative (prescriptive) form of the standard itself. (e.g. XML Schema);
b. A 'Cover Page' document that summarises the content and context of the standard;
c. A 'Motivations' document that describes the reasons for the standards existence;
d. A 'Rationale' document that describes why the standard takes the form it does and
e. A 'Change History' document that describes how this version has changed since the last version.

4.1.1. Justification

It is important to preserve the context of the original standard documentation, particularly when standards take the form of XML Schema documents. If a normative form of a

specification is difficult to read, a companion document will enhance the accessibility of the standard.

a. A standard must be defined as an archival document.
b. For standards to be treated uniformly there must be a document containing metadata in a consistent, machine processable form.
c. The potential implementer of a standard must know why the standard exists and what functions the standard is intended to support. This provides the justification for its adoption.
d. The revisers of a standard cannot understand the intent of original creators without knowledge of the rationale behind the design decisions. Without a rationale document they are likely to introduce errors.
e. Implementers of a standard need to know how a standard differs from previous versions of the same standard so they can adapt their implementation.

4.2. Each TDWG standard must be accompanied by a ladder of documentation that gives easy access to those making implementation decisions at all levels

A task team charter is pivotal in defining the documents that will be produced as part of a standard. A charter accepted by the Executive Committee must define a list of documents that will be created as part of the work plan of the team. The Executive Committee should ensure that documentation is created for the complete array of potential clients, including managers, biodiversity scientists, data managers and technology experts.

4.2.1. Justification

It is unlikely that a third party will see a developing TDWG standard and write introductory literature, tutorials and criticisms as would happen with a W3C Draft Recommendation. It is therefore necessary for TDWG standards to conform to a minimum level of documentation. Without this documentation, standards are less likely to become widely adopted or maintained and TDWG will fail in its mission.

4.3. TDWG must attach clear Copyright and Intellectual Property Rights statements to all standards documents

4.3.1. Justification

Copyright and Intellectual Property statements must be unambiguous. Public sector and not-for-profit organisations are becoming increasingly aware of the value of the intellectual property they possess and expect clear terms on its release. Commercial organisations are unlikely to be involved in development and implementation of standards if ownership is ambiguous. Less scrupulous organisations may try to gain ownership of standards through copyright and patenting if they are produced in a legal vacuum.

4.4. A TDWG standard must have online documentation

4.4.1. Justification

The role of a standards body is to hold and distribute standards in an easily accessible way. If the standards are not readily available online, TDWG is not fulfilling its role as a standards development organization.

4.5. All TDWG standards documentation must be structured and in a limited range of open, archival formats

Archives of meetings, instant messaging conversations, email lists and unstructured wikis should not be considered documentation for the purposes of the TDWG standards process.

4.5.1. Justification

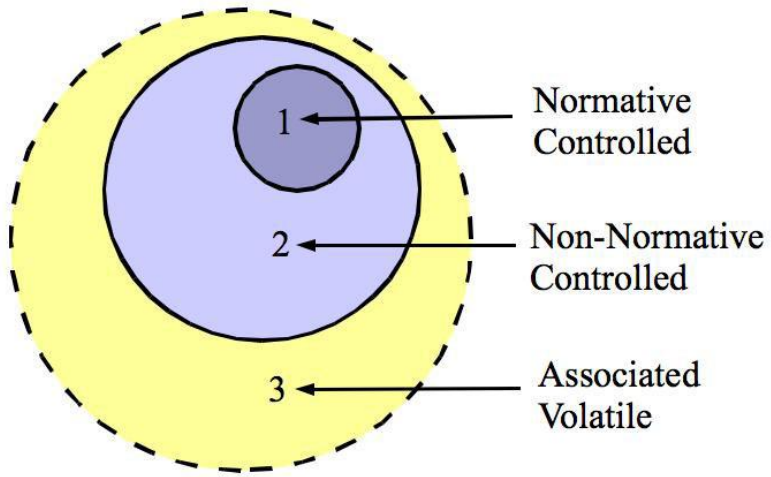
Potential users of standards find it hard to get information on both managerial and technical aspects of the standards. Simplified, uniform documentation procedures will solve this problem. When TDWG ratifies a standard, it is making a commitment of resources to host that standard, and to migrate it into any future repository. In order to create and maintain a functional repository, TDWG needs to specify the metadata elements, document structure, and file formats for documentation.

4.6. There should be 3 types of documentation

Type 1 documents are the normative parts of a standard. Type 2 documents are part of the standard that are non-normative. Type 3 documents are not part of the standard and will not be controlled by the TDWG process, but will provide help and support to people working with the standard. Type 3 documentation may contain examples, tutorials, introductory overviews, etc.

The three different types are illustrated in Diagram 1 and Table 1.

4.6.1. Diagram 1: Documentation Types



4.6.2. Table 1: Enumeration of document Types.

	Type 1	**Type 2**	**Type 3**
Normative	Yes	No	No
Part of a standard	Yes	Yes	No
Function	Defines	Explains and justifies	Helps and Supports
Versioned with Standard	Yes	Yes	No
Controlled by TDWG Process	Yes	Yes	No
Document format	Tightly Controlled	Tightly Controlled	Not Controlled
Document content	Tightly Controlled	Loosely Controlled	Not Controlled
Example formats	XML, RTF, PDF, XSD	XML, RTF, PDF	Plain-Text, HTML, Word, PDF

Maintained in repository	Yes	Yes	Sometimes or possibly only as a link.
Language	US English	US English + translations	Any

4.6.3. Justification

Authors and consumers of standards must know the status of the document they are producing or consuming. The three types recommended here provide the simplest system for indicating document status. Human readable type 2 and 3 documents are particularly useful when the normative parts of standards are machine readable.

4.7. All Type 1 documents must be in English using US spellings and grammatical constructs.

Translations of Type 1 documents may be supplied, but the translations will be treated as either Type 2 or Type 3 documents. The Executive Committee may require the production of Type 2 translations as part of the standard. All Type 2 documents should be in US English possibly with translations as decided by the Executive Committee. Type 3 documents may be produced in any language with no requirement for translation.

4.7.1. Justification

Normative documents in a single language avoid differences in translation. Requiring translation of non-normative documents would be a burden on standards authors but the availability of documentation in multiple languages may encourage adoption and so should be encouraged where it is appropriate.

4.8. All standards must be treated the same

Even if a standard is in the form of a web service, a folio of documents describing the interface to the services and data curation methods followed by the data provider should be put through the TDWG standards process.

4.8.1. Justification

Any variation would impose a burden on curation and use of the standard.

4.9. Six administrative standards must be created to initialise the TDWG process

Documentation within TDWG should be controlled by standards defined by the standards process. The first standards required in a new process must be those that govern documentation. These documentation standards are interlinked so they need to be created together as a group. Six standards are identified as being the minimum required to bootstrap the process. These will be included in milestone #16 of the TDWG Infrastructure Project.

a. File Formats: The file formats, naming and versioning conventions used in all TDWG standards.
b. Cover Page Specification: The format of the cover page document that should accompany every standard. This will be an XML Schema document that the Cover Page produced in 4.1 must validate against.
c. Layout Template: A specification of how human readable documents (see 4.1) should be laid out.
d. Process Document: A document specifying how the TDWG process track is administered. This document will include a list of standards maturity levels.
e. Copyright Notice: The text of a copyright notice to be used on TDWG standards documents along with notes on how it may or may not be modified.

f. Intellectual Property Statement: Guidelines for inclusion of Intellectual Property Rights statements in TDWG standards documents.

These standards must be treated like any other standards (see 4.8) and must contain the minimum documents required i.e. Standard, Cover Page, Motivation, Rational, Change History (see 4.1).

4.9.1. Justification

The documentation process will be controlled by the standards process. To initialise the system the first documents through the standards process must be those that control the documentation. These are the minimum documents required to initialise the standards track.

4.10. Task team charters must describe how systems can test whether they comply with the proposed standard

Testing will usually take the form of producing a compliance test suite. In the case of XML Schemas, the schema itself could be considered the test suite. In the case of exchange protocols the test suites are likely to require *Supporting Software*. Two independent teams must develop the systems that are subsequently used to test each other.

4.10.1. Justification

A compliance test suite is necessary to test whether a system meets a particular standard. Most TDWG standards have had intrinsic compliance testing. The paper-based data standards are lists of facts and so compliance is trivial. If the abbreviation is not in the list then it isn't compliant. The newer XML Schema-based standards are test suites. By definition, if an XML document validates against a schema then it is compliant with the standard. As more complex standards (such as exchange protocols) are ratified, the need for compliance testing will increase. Compliance is central to other standards organisations like OGC (http://www.opengeospatial.org/resources/?page=testing).

4.11. Where it is appropriate, a reference implementation must be part of the task team charter

Compliance test suites should be required alongside reference implementations to demonstrate conformance to the standard. In some situations (such as data exchange protocols and schemas) it may be possible to have two reference implementations that act as compliance tests for each other.

4.11.1. Justification

A reference implementation is an example system where software is used to demonstrate a standard. There are three specific ways reference implementations can help the standards process-

a. The most powerful barrier to reaching consensus on a standard is the criticism “it will never work”. An operational reference implementation is the most effective way to counter this argument.
b. A reference implementation is perhaps the most effective Level 3 documentation. “A standard is much easier to understand with a working example in hand.” (http://en.wikipedia.org/wiki/Reference_implementation)
c. Multiple reference implementations can demonstrate the independence of the standard from an implementation..

4.12. Dynamically served standards must be documented like any other standard:

a. Documents must describe the interface to the data and how the data will be curated. These documents must go through the TDWG standards track.
b. Software must be provided to serve the data in the way specified in the documentation. This software may be hosted by TDWG or a third party.
c. Test software must regularly monitor that data is being served in a way that is compliant with the documentation. This has to be hosted by TDWG or on TDWG's behalf.
d. Client software must demonstrate how to use the service.

4.12.1. Justification

Botanical author abbreviations and herbarium codes are examples of TDWG data standards that are now databases. If these databases are to be ratified as TDWG standards, either static versioned snapshots need to be released as regular standard documents, or the dynamic data has to be served in a controlled way. There appears to be little interest in formalising older TDWG standards, but this situation could soon change with the introduction of Globally Unique Identifiers. TDWG may need to ratify standard lists such as herbaria or entomological collections. The TDWG Process should handle these cases.

Appendix A: Review of Current Standards Documentation

1. Mature Standards:

- Authors of Plant Names: A book listing plant name authors and approved abbreviations published in 1992. Now integrated into International Plant Names Index. Help text associated with on-line searches.
- Botanical Periodicals: A list of botanical periodicals. Published 1968 with supplement 1991. Currently out of print.
- Economic Botany Data Collection Standard: A book published in 1995. Still in print.
- XDF: A language for the definition and exchange of biological data sets.
- Floristic Regions of the World: A section of a book published in 1986. Now out of print.
- Geographical Codes: A book published in 1992 and updated in 2001(not clear if 2001 version is ratified by TDWG). Polygons of listed regions have been made available by Kew but not been put forward for ratification by TDWG. (http://www.tdwg.org/geo2.htm)
- Herbarium Index Codes: A book published in 1990. Now maintained as an on-line database. Not clear if on-line version is TDWG standard. Help text associated with on-line searches. (http://sciweb.nybg.org/science2/IndexHerbariorum.asp)
- Herbarium Information Standards and Protocols for Interchange of Data. Available as a document on-line or in print. Latest version available as a searchable database. (http://plantnet.rbgsyd.nsw.gov.au/HISCOM/HISPID/HISPID3/hispidright.html)
- Plant Names in Botanical Databases: A 1995 publication available on-line (http://www.tdwg.org/plants.html)
- Publication Abbreviations: A book series started in 1976.
- Plant occurrence and status scheme: A document published in 1995 and available on-line (http://www.tdwg.org/poss_standard.html).
- DELTA (Description Language for Taxonomy): User guide published in 1986 with subsequent updates. Much available on line though linked to a particular implementation (http://delta-intkey.com/)

2. Newly Adopted Standards:

- ABCD: XML Schema, schema annotations, misc pdf documentation, database of 'concepts' used within the schema. (http://www.bgbm.org/TDWG/CODATA/ABCD-Versions.htm & http://ww3.bgbm.org/abcddocs/AbcdIntroduction)
- SDD: Extensive Wiki documentation and discussion including an initial draft of a 'Primer' document to get people started. Authors have difficulty keeping documentation up to date. (http://bdei.cs.umb.edu/twiki/bin/view/SDD/WebHome)
- Taxon Concept Schema: XML Schema, schema annotations exported as html, user guide as pdf, development wiki. (http://www.soc.napier.ac.uk/tdwg/index.php)

Appendix B: Documentation Best Current Practice in Other Similar Standards Organisations

1. Summary

Best current practice in the quality of documentation appears to be between W3C and IETF. Other bodies tend to copy these two organisations, but each organisation has useful lessons for TDWG. It is regrettable that a number of other standards organisations do not make their documentation freely available.

The GGF strategy is notable and appealing -

> *Rationale and specification of documentation + open document formats = best practice.*

2. Global Grid Forum

- http://www.ggf.org/index.php
- First document in series sets out documentation requirements (http://www.ggf.org/documents/GFD.1.pdf)
- Authors strongly encouraged to follow IETF Internet Drafts format where possible
- Documents must contain:
    - Document type: GWD-X or GFD-X, where X is one of several types including I(informational), E (experimental), P (Community Practice), or R (Recommendations track).
    - Author name(s), affiliation(s), and contact information
    - Date of the document (original date and revised date).
    - Name of working group or research group (where applicable)
    - Title of document
    - Document URL
    - 1-2 paragraph abstract
    - a summary of security considerations.

Institute of Electrical and Electronics Engineers

- http://www.ieee.org/portal/site
- Difficult to find out much about as this group; appear largely commercial.

Internet Engineering Task Force

- http://www.ietf.org/
- Detailed document describing document formatting (http://www.ietf.org/ietf/1id-guidelines.html)
    - ASCII text only for standard documents.
    - May have accompanying (but non-normative documents).
    - Documents are the standards.
    - Strong on IP issues.

International Multimedia Telecommunications Consortium

- http://imtc.org/

- Does not produce standards of its own but works with other bodies notably IETF.

International Organisation for Standardization

- http://www.iso.org/iso/en/ISOOnline.frontpage
- Over 14,000 standards and long history.
- Documents are the standards.
- Detailed documentation for docs (http://isotc.iso.org/livelink/livelink/fetch/2000/2122/3146825/4229629/texts_list.htm)

Organization for the Advancement of Structured Information Standards

- http://www.oasis-open.org/home/index.php
- "A specification may be composed of any number of files of different types, though any such multi-part specification must have a single specification name and version number. Irrespective of the number and status of the constituent parts, the specification as a whole must be approved by a single TC ballot. "
- "All documents and other files produced by the TC, including specifications at any level of approval, must use the OASIS file naming scheme, and must include the OASIS copyright notice. All document files must also use the OASIS document templates." (http://docs.oasis-open.org/templates/)
- Templates are provided for working in MS Word, OpenOffice? ? and XHTML.

Open Geospatial Consortium

- http://www.opengeospatial.org/
- Released as PDF (and white papers as Word docs).
- Appears to be a clear format/template used.
- Can't find documentation on website (may be members only)

Web Services Interoperability Organisation

- http://www.ws-i.org/
- Issue documents are PDF
- Standards ('deliverables') include software.
- Supply MS Word templates for creation of documents.
- Document format is very similar to W3C format.
- Details may be in members only area.

World Wide Web Consortium

- http://www.w3.org/
- Well defined document rules (http://www.w3.org/2005/07/pubrules)
- Well defined style (http://www.w3.org/Provider/Style/)
- Well defined help on authoring documents (http://www.w3.org/2001/06/manual/)
- Some guides only available to members (http://www.w3.org/Guide/Reports)