# GBIF Glossary of Terms used in Biodiversity

DRAFT scoping document prepared by Dag Endresen, David Remsen and Éamonn O Tuama, Global Biodiversity Information Facility (GBIF), January 2012.

## Summary

This draft document provides the scoping for a new tool to create, maintain and present terms for concepts used in biodiversity informatics.

[Question: do we need a new and dedicated tool for creating and maintaining new basic terms? Existing terms “owned” by term-vocabularies such as Darwin Core, Audubon Core, and Dublin Core are maintained today using different tools. We could perhaps use the Semantic Media Wiki platform to create new terms?]

## Background

Biodiversity standard vocabularies such as the Darwin Core[[1]](#footnote-1), Audubon Core[[2]](#footnote-2), Structured Descriptive Data (SDD)[[3]](#footnote-3) and Access to Biological Collection Data (ABCD)[[4]](#footnote-4) from the Biodiversity Information Standards (TDWG) sometimes include the same or similar concepts. A general and flat list of terms used in biodiversity informatics including their definition and a persistent identifier (PID) is here proposed to support (through improved overview of basic terms) the re-use of terms for common concepts (Catapano, 2011). The Dublin Core (DCMI) and the TDWG Darwin Core (DwC) could provide a general model for how to maintain and present terms.

## Management of new terms

The target tool should maintain a flat list of terms. The tool should support the minting of new terms including the description (definition) for these new terms. A persistent identifier (PID) should be assigned to identify each of the terms. If an existing PID is available then this new tool should support providing this PID when adding new terms.

[Question: Should the interface for this tool include only the “new” terms or also include existing terms e.g. from other domains?]

## GBIF Vocabulary Server

The GBIF Vocabulary Server[[5]](#footnote-5) can be used to create new terms, however these terms are created as part of a controlled term-vocabulary (“extension”) or as part of a controlled value-vocabulary (“vocabulary”). A new tool where terms are created free from this context could enhance the potential for re-use of terms. The GBIF Vocabularies is well suited to build controlled vocabularies (“extensions” and “vocabularies”), but could be limited to only re-use terms defined in other tools. The interface would be updated to make this limitation explicit.

[Question: are the features and capabilities of the GBIF Vocabularies to define new terms useful and appropriate?]

## GBIF Resources Registry

The present GBIF infrastructure uses the GBIF Resources Registry[[6]](#footnote-6) for the official version of the Darwin Core extensions and vocabularies that are created by tools such as the GBIF Vocabulary Server. The final version of the extensions and vocabularies is uploaded and cached by the GBIF Resources Registry. This registry can be used also for caching the official version of the recommended basic terms and provide access for re-using these terms when building new Darwin core extensions and vocabularies as well as when building new ontologies for biodiversity information resources.

## Properties for each term

Currently in use at the rs.gbif.org (dc = <http://purl.org/dc/terms/>)

* name
* namespace
* qualName [PID, namespace+name]
* dc:title
* dc:description
* dc:relation
* examples
* required
* group
* thesaurus
* columnLength

Darwin Core uses these properties for terms

* Identifier
* Class
* Definition
* Comment
* Details

ISOcat uses some additional properties

* Type
* Owner
* Scope
* Version
* Registration Status
* Administration Status
* Justification
* Origin
* Effective Date
* Creation Date
* Last Change Date
* Change Description

Proposed properties for describing basic terms

* rdf:type="skos:Concept…”
* rdfs:isDefinedBy = PID, URI
* skos:hasTopConcept = [term for describing the grouping]
* rdfs:range = [for terms with controlled values, or for numeric, string, …]
* dc:identifier = [CONCEPT\_PID/CONCEPT\_URI]
* rdfs:label = [concept name, no spaces]
* dc:title = [descriptive title for the concept]
* dc:description, skos:definition = [descriptive definition for the concept]
* skos:example = [examples of use]
* skos:note = [remarks to the use of the concept]

## Terms Used in Bionomenclature

The online user interface to the Terms Used in Bionomenclature[[7]](#footnote-7) (Hawksworth, 2010) was used as a model to build a more general glossary of terms for concepts used in biodiversity informatics[[8]](#footnote-8). GBIF and SilverBiology developed this glossary presentation tool using the Ext JS JavaScript framework[[9]](#footnote-9) and the source code was made available at Google Code[[10]](#footnote-10). This tool can be updated to provide a more generic presentation of “Terms Used in Biodiversity”. Including a cached version of the recommended basic terms to the GBIF Resources Registry would provide the content for the updated version of this tool.

## Annotation

As implemented for the “Terms Used in Bionomenclature” the new tool should support the keyword tagging of terms.

[Question: Should the keyword be terms with PIDs or should the tool support tagging using free keywords provided as strings?]

## Relationships

The glossary of terms provides a flat list of terms free from the eventual context provided by an ontology instance of the term. Providing the basic term will maximize the potential for re-use of terms.

[Question: Should this tool perhaps provide an interface to state very light and general (non-normative) relationships between terms?]

## Persistence, versioning of terms

The definition of a term can be updated and the meaning of a term may even change over time (semantic rot). A tool to maintain terms will therefore need a versioning of terms to track all modifications and updates to term definitions.

[Question: when different groups use the same term, but with a different definition (meaning) should the term be defined as a new term or somehow as a subclass of the parent term?]

## Discussion forum

The new tool should include a threaded discussion forum. Each discussion post or thread could be annotated using the PID for the relevant term. The discussions could sometimes be relevant to multiple terms. A PID should identify each discussion post and thread.

Perhaps the tool could be configured to receive email discussions from the TDWG mailing lists? The emails could be annotated with the relevant term(s) and be included in the threaded archive of discussions available for each term.

[Question: a discussion forum tool could be a separate tool from the glossary of terms tool to better support discussions around all terms including terms maintained elsewhere such as in other communities for geographic data and genomics.]

## References

Catapano, T., D. Hobern, H. Lapp, R.A. Morris, N. Morrison, N. Noy, M. Schildhauer, and D. Thau (2011). Recommendations for the use of knowledge organization systems by GBIF. Released on 4 February 2011. Global Biodiversity Information Facility (GBIF), Copenhagen. Available at <http://www.gbif.org/orc/?doc_id=2942&l=en>, verified 15 Nov 2011.

Hawksworth, D.L. (2010). Terms used in bionomenclature: The naming of organisms (and plant communities). Global Biodiversity Information Facility, Copenhagen. ISBN 87-92020-09-7. Available at <http://www.gbif.org/orc/?doc_id=2430&l=en>, verified 15 Nov 2011.

1. Darwin Core, <http://rs.tdwg.org/dwc/> [↑](#footnote-ref-1)
2. Audubon Core, <http://species-id.net/wiki/Audubon_Core> [↑](#footnote-ref-2)
3. Structured Descriptive Data (SDD), <http://www.tdwg.org/standards/116/> [↑](#footnote-ref-3)
4. Access to Biological Collection Data (ABCD), <http://www.tdwg.org/standards/115/> [↑](#footnote-ref-4)
5. GBIF Vocabulary Server, <http://vocabularies.gbif.org/> [↑](#footnote-ref-5)
6. GBIF Resources Registry, <http://rs.gbif.org> [↑](#footnote-ref-6)
7. Terms Used in Bionomenclature, <http://bionomenclature-glossary.gbif.org/> [↑](#footnote-ref-7)
8. GBIF RS browser, <http://tools.gbif.org/resource-browser/> [↑](#footnote-ref-8)
9. Ext JS, <http://www.sencha.com/products/extjs/> [↑](#footnote-ref-9)
10. Source Code, <http://code.google.com/p/terms-of-bionomenclature/> [↑](#footnote-ref-10)