Technical paper N° 8/2015

Dyntaxa taxon concept administration and how to handle information related to taxa

Oskar Kindvall

in collaboration with
Sabine Roscher, Jérôme Bailly-Maître, Želmíra Šípková-Gaudillat

01.12.2015

Authors' affiliation:

Oskar Kindvall, Swedish University of Agricultural Sciences (SE)
Sabine Roscher, Muséum national d'Histoire naturelle (FR)
Jérôme Bailly Maitre, Muséum national d'Histoire naturelle (FR)
Želmíra Šípková-Gaudillat, Muséum national d'Histoire naturelle (FR)

EEA project manager:

Mette Palitzsch Lund

ETC/BD production support:

Muriel Vincent, Muséum national d'Histoire naturelle (FR)

Context:

The Topic Centre has prepared this Technical paper in collaboration with the European Environment Agency (EEA) under its 2015 work programmes as a contribution to the EEA's work on the biodiversity data center.

Citation:

Please cite this report as
Kindvall, O., in collaboration with Roscher, S., Bailly-Maître, J. and Šípková-Gaudillat, Ž., 2015. Dyntaxa taxon concept administration and how to handle information related to taxa. ETC/BD report to the EEA.

Disclaimer:

ETC/BD Technical paper N° 8/2015
European Topic Centre on Biological Diversity
c/o Muséum national d'Histoire naturelle
57 rue Cuvier
75231 Paris cedex, France
Phone: + 33 1 40 79 38 70
E-mail: etc.biodiversity@mnhn.fr
Website: http://bd.eionet.europa.eu/

Contents

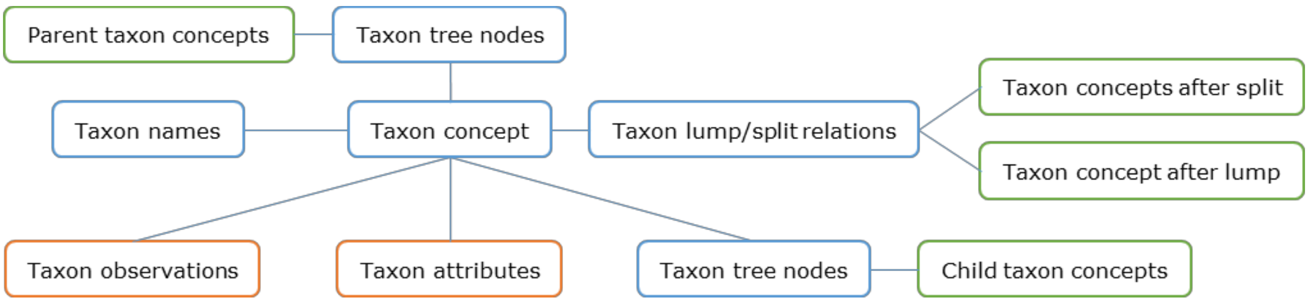
1 Taxon information model

1.1 Overview

The Dyntaxa web application (www.dyntaxa.se) and its underlying main web service (TaxonService) use a rather simple and pragmatic information model (Figure 1.1). This model supports effective listing of taxa, searches for taxon concepts and matching processes between different taxon name lists. Furthermore, it supports administration of the changes of the taxonomic information in a way that keeps track of the changes and making the dynamic history of taxa traceable.

The taxonomic information handled by Dyntaxa (Figure 1.1) can be categorized in merely four distinct types of objects, i.e. *Taxon names*, *Taxon concepts*, *Hierarchical relations* and *Changes*. The most central object in this information model is the Taxon Object, i.e. the Taxon Concept, to which an infinite number of Taxon Name Objects can be related. Each taxon object has at least one parent taxon related to it via a Taxon Tree Node Object which describes that particular pairwise hierarchical relation. An exception from this is the 'Root'-taxon concept named Biota which represent the taxon tree root for all taxa handled in the system (https://www.dyntaxa.se/Taxon/Info/0). Each taxon object may also have one or several objects describing the exact lump/split relations with other taxa that has occurred as result from changes of the taxon concepts.

Figure 1.1 Overview of the different information object types described in this report and their relations to each other.



Note: Blue boxes indicate pieces of information (objects) representing a single taxon; Green boxes represent examples of related other taxa; Orange boxes represent pieces of information that have been linked to the taxon.
Source: ArtDatabankenSOA, 2015.

Taxon specific information of all kinds can of course be linked to the taxonomic information in various ways. In many systems handling such information only the taxon name or the GUID (Globally Unique Identifier) for that name is used to label taxon related information. However, in order to make it more robust and less sensitive to problematic effects of taxonomic changes it is better to link information about taxa to the taxon concept instead.

At the Swedish Species Information Centre all information about a particular species, or any other taxon category, are packaged as handy taxon attribute objects (sometimes referred to as 'Species Facts') which are linked to the taxon concepts via their TaxonIds, i.e. the Ids of the Taxon Concepts (Figure 1.1). These objects can hold any sort of information which constitute some statement, text or measurement that is representative for the specific taxon or at least for a recognizable group of individuals of that taxon. The general data model of taxon attributes is described in more detail in section 1.6.

Another type of information that preferably should be linked to the taxon concepts via their TaxonIds is observations (Figure 1.1). All types of observations of taxa can be handled as Taxon observation objects. There are useful standards on how to organize this type of data that link to the taxon information model presented here (TDWG, 2015). As individual records of taxon observations are not directly handled in the article 12 and 17 reports there is no need to describe this information type any further in this report.

All object types described in this report do have stable identifiers plus a set of general properties (Table 1.1) that handles update information (Dates and User Information) relevant to trace editing history of each object. Besides, all object types hold information about the user that currently has retrieved it from the web service, i.e. the Data context. This information makes it possible to adjust the content to the users' selected role and language selection.

The stable identifiers representing each object are expressed both as an Id and as a GUID. The Id is an integer value that is unique within its object type in the domain. These Ids are often referred to by names constructed by combining the information type with Id. The Id of a Taxon name will thereby be called 'TaxonNameId' or to be more specific in general situations 'Dyntaxa TaxonNameId'. When referring to the ids of the Taxon concepts in Dyntaxa 'TaxonIds' and 'Dyntaxa TaxonIds' are commonly used. Sometimes the 'TaxonIds' are also referred to as the 'Taxon Concept Ids'. The GUID property of each data type in the information model described here is always constructed as an LSID (Life Sciences Identifier). LSIDs are constructed according to a specific standard (Object Management Group, 2004).

Table 1.1 List of general properties which all taxonomic objects in the described information model do have.

Property	Description
CreatedBy	Id of user that created the object.
CreatedDate	The object was created at the date.
DataContext	Data context with meta data about the content of the object such as current language of text strings.
Guid	GUID (Globally Unique Identifier) for the object. It is constructed for each object as an LSID, i.e., Life Sciences Identifier (Object Management Group, 2004).
IsInRevision	Indicates if the object is currently checked out for revision and thereby can be modified. Objects where this property is set to true are only handled by users authorized to work with a particular revision in Dyntaxa.
IsPublished	Indicates whether or not the information in the object has been published. If set to false only users authorized to work with a particular revision in Dyntaxa can retrieve it.
ModifiedBy	The user id of the user who modified the object content.
ModifiedByPerson	Name of the person who last modified the object content.
ModifiedDate	Date the object content was modified. Set by database when a revision is checked in.
ValidFromDate	Date object content is valid from. Is set to date created by default.
ValidToDate	Date the object is valid to.

Source: ArtDatabankenSOA, 2015

There is often a need of linking pieces of information to References and Illustrations. Besides, it is often valuable to annotate individual information objects by adding different types of comments. Generally all this issues are best solved by linking these pieces of information as GUID to GUID triplets which can be easily retrieved when needed. How this is handled in detail in the context of Dyntaxa taxon information model is beyond the scope of this report.

1.2 Taxon Names

Normal search for information related to taxa includes the use of taxon names or some type of identifier associated with the requested taxon. The very same taxon may often have several names or identifiers associated to it. All accepted taxa has at least one scientific name each which was given to them by the taxonomist who originally described it, i.e. the author of each taxon. However, most taxa have several scientific names associated. Usually, the reason for this is that taxonomists discover that entities described as separate taxa actually are the same thing, or because of species that are moved between genera. Consequently, names become synonymized with each other. Many taxa also have one or several common names which of course differ among languages. Furthermore, most taxa tend to have a whole set of different identifiers (Id numbers or GUID:s) stemming from various usages in various databases.

In the Dyntaxa information model all mentioned kinds of names are handled as taxon name objects (TaxonName). Table 1.1 summarizes the properties that are unique for the Taxon Name Object. Among the listed properties the Name Category needs further explanation as its usage together with the other properties will help a lot when presenting the taxonomic information to a reader, or when dealing with cross mapping and advanced taxon search where it can be helpful to discriminate between different types of names. The three types of name categories are described separately below.

The Dyntaxa taxon information page for the bird species *Regulus ignicapilla* (https://www.dyntaxa.se/Taxon/Info/Regulus%20ignicapilla) can illustrate the usage of the name categorisation in combination with all the other name properties listed in Table 1.2. The very same species is also exemplifying the name shifting issue described in section 2.1 (Figure 2.1.2). To read more information about each Taxon name one can click in the presentation directly on the names. Try also to click on the green i-symbol which will open a box with a detailed information summary of the taxon concept.

Table 1.2 List of properties specific for taxon names. A short description is given for each property.

Property	Description
Name	The name string.
Author	A string representing the author or authors of the taxon name if relevant for the specific category of the name.
NameStatus (Nomenclature)	The nomenclatural status of the taxon name. Possible values are e.g. 'Correct', Incorrect, Misspelled, 'Preliminary suggestion', Incorrect citation' and 'Removed'.
NameCategory	All names have a category. Examples of name categories are Scientific Name, Swedish name and Norwegian name. The number of possible name categories is infinite. See text for more information.
IsOriginalName	Indicates whether or not the name is the original name for its taxon.
IsRecommended	Indicates whether or not the name is recommended. For each name category one name is set to recommend. There can only be one name per name category where this property is set to true.
IsUnique	Indicates whether or not the name is unique. If the same name string exists among several taxa then the name is not unique.
NameUsage	Name usage of the taxon name. Possible values are 'Accepted', 'Synonym', 'Homotypic', Heterotypic, 'Pro Parte Synonym', and 'Misapplied Author Name'.
Description	In case needed this field can be used to add clarification about the name usage or to comment on problems associated with the name.
Taxon	The taxon object which the taxon name belongs to.

Source: ArtDatabankenSOA, 2015

1.2.1 Scientific names

The basic name category is the ‘Scientific Name’. Every Taxon concept object requires at least one ‘Scientific Name’. Otherwise published lists based on the information contained by the taxon concept object will lack a name which makes it unreadable. However, most established taxon concepts belonging to the taxon category ‘Species’ do have several scientific names associated to it. Some of them are considered to be correctly used for that taxon. When the nomenclatural status of these names is ‘Correct’ they are true ‘Synonyms’. Beside these scientific names there could be names that are miss spellings or names that do not correctly follow the taxonomical codex (ICZN, 2000, IAPT, 2011). It could be due to incorrect Latin grammar. As species are moved from one genus to another, the species epithet must sometime be modified to fit properly to the new genus name according to the rules of spelling.

One name among those categorized as ‘Scientific name’ could be pointed out as being the original one using the IsOriginalName-property (Table 1.2).

When presenting taxon information or any set of data that are associated with taxa in tabular list format or in tree views it is often not feasible to present all the names referring to each taxon. Instead, there is definitively a need for selection just one of them. Dyntaxa administrators therefore set one of the most widely used and ‘correct’ scientific names as recommended.

1.2.2 Common names

Beside the Scientific names it is often very useful to handle the common names which of course are often very different for different languages and sometime also between regions within countries. In Dyntaxa each language or dialect can have its own Taxon Name Category, which corresponds to the globalization locale. This opens for the possibility to handle many alternative common names or spelling alternatives for each of the languages. As with the scientific names one of the common names in each name category is selected to be the recommended one.

1.2.3 Identifiers

A third group of taxon name categories is different kinds of identifiers. In many older observatory systems different codes or practical short names have been adopted within certain organism groups. If these are needed by some users it could be practical to have them associated with the taxon concept. Similarly, Ids of different kinds that are used in different other taxon systems are also very practical to store together with the taxon concept. In Dyntaxa there are several examples of name categories that correspond to Id in other systems.

One of these identifier categories is just named ‘GUID’. The purpose of this name category is to store all GUID:s that have been used for the particular taxon concept. By default the GUID of the Dyntaxa Taxon Concept as it is written in the GUID property of the taxon object is added to the list of taxon names of the name category labelled ‘GUID’ and set to the recommended one among GUID:s. If the recommended scientific name suggest a perfect match with a taxon in PESI then the GUID stored by PESI is added to the list of GUID name categories. Furthermore, if the taxonomic expert think that the match to PESI is correct then the PESI GUID is set to the recommended one. By following this procedure more and more concepts stored in Dyntaxa will be linked in a robust way to the taxa in PESI as editing of the content proceeds.

1.3 Taxon concepts

A taxon concept is very much defined by the list of names associated to it and all the data that are added to it. When a new species has been described for the first time, the concept only has one scientific name and a paper describing it in relation to other closely related taxa. As times goes more and more observations are often registered and thereby knowledge are gradually growing. After a while a lot of different pieces of information have been stored in databases that refer to the original name. It could be information about habitat preferences, generalization of its geographical distribution, conservation status, Redlist Categories, identification characteristics and all other kinds of taxon attributes.

Having observed a taxon concept over time within a region there is often no obvious needs for changing the taxon concept associated with the data. When the taxonomic knowledge grows and taxon names become synonymized or when revisions move species to another genus it is often not practical to move the data to another concept bearing the new name. Dyntaxa is based on the idea that as long as the data describe a taxonomically interrelated group of individuals or populations there should be no shift of taxon concept. Instead the name shift is treated as a shift of the recommended scientific name for the very same concept.

Taxonomic research may of course sometimes also reveal that all populations that have been observed within a certain region is actually not belonging to the very same taxon concept. In such cases one needs to split the former concept. There may also be situations where two well established concepts prove not to be genetically differentiated and taxonomists therefore suggest to merge the two separate taxa. Taxonomic changes reflected as changes of the taxonomic concepts or successive replacements of concepts are described in more detail in section 1.5.

In Dyntaxa the taxon concept is constructed as a rather light weighted object with a limited set of properties described in Table 1.3. If one needs to describe the concept with all relevant taxonomic information including all associated names and their usage, its definition, history of taxon concept changes, list of parent and/or child taxa and historic hierarchical taxon relations then most of this information has to be retrieved using various specific methods available at the Taxon Service. The reason for this object architecture is that it is by far the most efficient way to retrieve the data. When searching for taxon information starting from a name or just a part of the name the common first step will be to present all taxon names that match for the user in a list. In many cases it is necessary to present basic taxon concept information beside each matched name string such as the recommended Scientific and common name and also the taxon category (rank). The reason for this is that many names or similar spellings of names occur on several alternative taxon concepts. To be able to pick the one that was actually searched for this additional information is sometimes crucial. Also when taxon searches are done using taxon tree views the light weighted concept object which basically holds the recommended name and the taxon category is just what is needed.

Table 1.3 List of taxon properties specific for taxon concepts. A short description is given for each property.

Property	Description
ScientificName	The currently recommended scientific name.
Author	Author string of the currently recommended scientific name.
CommonName	The currently recommended common name. Not all taxa has a

	recommended common name.
TaxonCategory	Category that the taxon belongs to. Taxon categories corresponds basically to the taxonomic ranks or levels, e.g. 'Kingdom', 'Phylum', 'Class', 'Order', 'Family', 'Genus', 'Species' etc. The number of taxon categories handled by Dyntaxa may increase over time. A complete list of taxon categories is obtainable from Dyntaxa.se using the following link which summarize the number of child taxa for the 'Root' taxon Biota: https://www.dyntaxa.se/Taxon/Info/0
ChangeStatus	Change status for the taxon. Indicates if this taxon has been lumped, split or deleted. For details read section 1.5.
IsValid	Indicates if the taxon is 'Accepted' or not (='Not accepted').
AlertStatus	Alert status for the taxon. A classification of the need for communication of problems related to the taxon status and recognition. Might be used to decide if the concept definition text should be displayed as warning to end users. Examples of taxon definition texts are shown in Chapter 2.
PartOfConceptDefinition	Part of concept definition for the taxon. The concept definition of the taxon is a text which is retrieved on demand using a specific method. The complete taxon definition is automatically generated based on information about lump/split events if changes have affected the taxon. The PartOfConceptDefinition-property stores hand written descriptions of the taxon concepts and its usage which will be incorporated in the full Definition text.
SortOrder	Sorting order for the taxon. This sorting order can be used when sorting lists of taxa according to current taxonomic sorting. The value is in relation to all other taxa following the structure of the main tree. In order to get localized sorting relevant for pragmatic organism groups or other unaccepted hierarchies it is recommended to use the SortOrder values held by the specific taxon tree nodes, i.e. the actual hierarchical relations between pairs of taxa.

Source: ArtDatabankenSOA, 2015

In Dyntaxa, the taxonomic rank of a taxon concept is expressed by its taxon category (Table 1.3). Taxon category is an object holding information on both the name of the rank and its sort order in relation to all other currently occurring categories. This solution makes it possible to handle an infinite number of categories. This is practical as the selection of accepted taxonomic ranks and their names changes over time and differs between various parts of the taxonomic tree.

The property called alert status (Table 1.3) could have one of three possible values: 'Green', 'Yellow' and 'Red'. When 'Green', then the taxon concept is accepted and no particular problems exists concerning its classification and recognition. On the other hand, when 'Yellow' the taxon concept is problematic as it could potentially be confused with other taxa. The concept is still considered accepted according to scientific consensus. However, when alert status is 'Red' then the taxon concept is no longer accepted and it has been either deleted or replaced by another taxon concept. The alert status property helps communicating possible problems associated with the particular taxon concept. In any graphical information system the alert status can be used to trigger popup text boxes which can inform the reader about important aspects of the taxon concept. In Dyntaxa an i-symbol coloured in Red, Yellow or Green is used for this. This symbol is shown in both list views and in the top right of each taxon information page (Pictures 2.3.1, 2.3.2, 2.4.1, 2.4.2).

1.4 Hierarchical relations between taxa

Current and historical hierarchical relations between taxon concepts can be described in terms of Taxon Tree Node objects. It is basically an object that holds the link between each pair of hierarchal related taxa (Table 1.4) plus the general properties related to the editing of the relation (Table 1.1).

Table 1.4 List of taxon properties specific for hierarchical relations between taxa. A short description is given for each property.

Property	Description
ParentTaxon	The parent taxon in the relation.
ChildTaxon	The child taxon in the hierarchical relation.
IsMainRelation	Indicates whether or not this taxon relation is the main relation or not. This property support the need for adding infinite numbers of extra parent relations which is useful when one needs to create pragmatic non-taxonomic groups of any kind or just alternative taxonomic hierarchies. The main relations are used to form the currently accepted taxon tree.
SortOrder	Sort order for the taxon relation amongst other taxon hierarchical relations chairing the same parent taxon. This parameter is enables sorting of taxonomic lists and tree views according to some taxonomic sorting often makes it easier to find a particular taxon compared to the alphabetic order.

Source: ArtDatabankenSOA, 2015

1.5 Taxon changes

The name 'Dyntaxa', refers to the web application (www.dyntaxa.se) used for the basic presentation and administration of the content in the Swedish Taxonomic Database. The name stems from its original name which was 'Dynamic Taxa'. The notion that taxonomic information is rather dynamic and change as research progress was the primary reason for constructing the first version of Dyntaxa in 2004. It was done merely to solve the problems associated with taxonomic changes that occurred between each period of Red list evaluations in Sweden, which were exactly the same type of problems now being identified on the European level by ETC, and that are described in this report (Section 2).

There are practically only very few types of change events associated with taxon concepts: 1) *name shifts*, 2) *category shifts*, 3) *movements*, 4) *lumps*, 5) *splits* and 6) *deletes*.

Name shifts are often linked to the other types of changes that happen over time with a taxon concept. A category shift, implies that the taxon rank is altered from e.g. the 'Sub species' level to the 'Species' level. When this happens the Recommended Scientific Name has to be adjusted to the rank dependent naming convention. Movements of taxa within the hierarchical taxonomic tree is another change type that for species or most sub ordinate ranks gives rise to a name shift. This is because the scientific names of species and their subtaxa include the genus name as a part of the name string. When a species is moved to another genus it will get a new recommended name. In case of movements of taxon concepts where the category constitute of a higher rank than 'Species', like 'Family' or 'Orders', the recommended name seldom changes as a consequence. The same applies for category shifts of taxon concepts representing taxon groups.

While name and taxon category shifts per se do not affect the interpretation of the taxon attribute or observation data that are linked to the particular taxon concept the process of taxon movements may do so. When the focal taxon concept is moved from one parent taxon to another it will not cause any problems. However, when the child taxa of the focal taxon concept are moved to another taxon parent, then the change of the taxon concept may have implications on the representativeness of the linked information which in some cases has to be considered. Consequently, it would be useful to have a specific property that indicates whether or not the inclusiveness of the particular taxon concept that represents a taxon group has changed significantly. This aspect is not handled by the current version of Dyntaxa.

Among the different types of taxon changes only lumps and deletes make the taxon concept totally invalid. In those cases, all data that has been linked to the taxon also become invalid or need to be relinked to another replacing taxon concept. A taxon that becomes split is also invalid taxonomically. However, if a taxon that has become taxonomically invalid due to a split event at the same time shifts its taxon category to a 'Collective taxon' (see section 2.4) it can be used as a valid pragmatic taxon concept which still can be very useful as something to link data to. If this is done there is no need to relink old data that originally were associated to it. In order to make it clearer for the reader of data linked to the collective taxon that resulted from a split it is a good practice to shift the recommended name of the concept to its former recommended name plus 's. lat.', i.e. sensu lato.

As lumps, splits and deletes have important implications on how related taxon attribute or observation data should be interpreted and managed, every taxon concept object in Dyntaxa has a Change Status which specify its current state in relation to these change types (Table 1.5.1).

Table 1.5.1 List change statuses possible for a taxon concept. A short description is given for each status value.

Taxon change status	Description
InvalidDueToSplit	The taxon concept has split into several taxa. The taxon concept is no longer valid in taxonomic terms. It may still be valid as a pragmatic collective taxon that hierarchically owns all the taxa that result from the split events which then becomes the child taxa of the collective taxon.
InvalidDueToLump	The taxon concept has been lumped with another taxon concept and replaced by a new accepted taxon concept. The lumped taxon concept is no longer accepted.
InvalidDueToDelete	The taxon concept has been deleted such that it is no longer accepted. In this case no replacing taxon is exists. This state should be avoided. It is only need when erroneous taxon concepts accidentally are created by the taxon editor.
Unchanged	Unchanged taxon concept. The initial default value.
ValidAfterLump	The taxon concept which resulted from a lump of at least two taxa.
ValidAfterSplit	One of the new taxon concepts that is a result from a split of another taxon.

Source: ArtDatabankenSOA, 2015

When the taxon change status indicates that the concept has been involved in a lump or split event it is possible to get information on which taxon concepts have replaced a lumped or a split taxon. It is of course also possible to list taxa that has been replaced by the changes. To support this, the information of the lump/split relation between taxa is stored as relation objects that specify each pair of taxa that was involved in the particular change event (Table 1.5.2).

Table 1.5.2 List of taxon properties specific for taxon change relations. A short description is given for each property.

Property	Description
EventType	Indicates which type of change the relation is a result of.
TaxonId	The Id of the taxon concept that was split or lumped.
TaxonIdAfter	The Id of the taxon concept that replace the taxon concept that was split or lumped.

Source: ArtDatabankenSOA, 2015

1.6 Taxon attributes

At the Swedish Species Information Centre there is a long tradition of handling digitalized information about taxa. This class of information is made up of everything that constitute a generalization or statement representative for a particular taxon or a pair of taxa (e.g. herbivore/host relations). This characteristic of Taxon attributes make this kind of information distinctly different from Taxon Observations as that type of information rather constitute the raw data underlying the generalisations about distributions, habitat preferences or evaluations of conservation status which are examples of taxon attributes.

For a given combination of Taxon (=Taxon Concept), Factor (=Trait/Attribute), Individual Category, Period (=Reporting round) and Host (optionally related taxon) it is possible to specify a value, or actually a closely related set of values set in up to five different data fields (Table 1.6). Depending on the factor the values could be describing texts, factorial classifications using a specific well defined categories (enum values), boolean statements or numbers representing counts or different types of measurements. Using the up to five possible data fields for a single factor data type it is easy to handle measurements with mean, min and max values within the scope of a single Factor, e.g. 'body length' or 'Estimated Area of Occurrence".

Table 1.6 List of properties specific for taxon attributes (also known as 'Species Fact objects'). A short description is given for each property.

Property	Description
Taxon	The taxon concept for which the specified field values are representative.
Factor	This property holds information on what type of information about the taxon this object contains. The factor object has a name (e.g. 'Redlist category', Conservation Status according to Article 17' or 'Number of reproductive individuals'), a definition and information about the actual data types and data field labels.
IndividualCategory	If the data values only is valid for a subset of individuals, e.g. a specific regional population, life stage, sex or colour morph, then this property is useful. The default category is 'General' which imply that the data values presented in the fields are representative for all individuals of the focal taxon.
Period	If the associated Factor is something that are evaluated at repeated reporting rounds it could be classified as 'periodic'. When Factor.IsPeriodic is true then the actual period is specified using this property.

Host	Depending on the type of Factor it is possible to specify a related taxon. In the most common cases these extra taxa constitute the host taxon. If the Factor is 'Green leaves' then the Host will specify which kind of leaves in terms of its species or group of species.
Fields	This property holds a list of data fields relevant for the specific Factor Data Type. These fields holds the actual data values. By reading the information contained by the Factor object is possible to understand the content of each data field. Values of each field could be either text, boolean, a categorical enum value, integers or decimal numbers.
Quality	A categorical property which is used for declaration of the quality of the information contained in the specific taxon attribute object.

Source: ArtDatabankenSOA, 2015

1.7 Linking current EUNIS terminology to the information model

The different taxonomic terms used in EUNIS (currently under revision) are listed in Table 1.7 together with information on how these terms are handled by the Dyntaxa taxon information model.

Table 1.7 List of taxonomic terms used by EUNIS and their correspondence in the taxon information model presented in this report.

EUNIS term	Correspondence in the taxon information model
Species	One of the taxon categories which correspond to a specific taxonomically accepted type of taxon rank. Taxon category is a property of the taxon concept object.
Synonym	In general, all taxon names associated with a taxon concept that belongs to the same taxon name category and are not the recommended name. In more strict sense it is the scientific names beside the recommended scientific name, which are of correct both usage and nomenclatural status.
Subspecies	One of the taxon categories which correspond to a specific taxonomically accepted type of taxon rank. Taxon category is a property of the taxon concept object.
Synonym (variety)	Dyntaxa do not handle synonyms of sub species differently from synonym names associated with taxon concepts of any taxon category.
Questionable	Has no correspondence in Dyntaxa.
Hybrid	One of the taxon categories which correspond to a specific taxonomically accepted type of taxon rank. Taxon category is a property of the taxon concept object.
Variety	One of the taxon categories which correspond to a specific taxonomically accepted type of taxon rank. Taxon category is a property of the taxon concept object.
Genus	One of the taxon categories which correspond to a specific taxonomically accepted type of taxon rank. Taxon category is a property of the taxon concept object.
Group	Is called 'Organism group' in Dyntaxa. It is one of the taxon categories which is used for various practical groups of taxa which do not correspond to a particular rank in the taxonomic hierarchy. Taxon category is a property of the taxon concept object.
Misspelling	One possible value of 'Nomenclature' which is a property of the taxon name object.
Synonym (subspecies)	Dyntaxa do not handle synonyms of sub species differently from synonym names associated with taxon concepts of any taxon category.
Form	One of the taxon categories which correspond to a specific taxonomically

	accepted type of taxon rank. Taxon category is a property of the taxon concept object.
Invalid name/author combination	Corresponds to the term 'Incorrect citation' which is one possible value of the taxon name property 'Nomenclature'.
Cultivar	One of the taxon categories. Taxon category is a property of the taxon concept object.
status unclear	Has no correspondence in Dyntaxa.
Expression in Habitats Directive	Has no correspondence in Dyntaxa but could easily be added as a separate Taxon Name category.
Unresolved	Has no correspondence in Dyntaxa.
Subgenus	One of the taxon categories which correspond to a specific taxonomically accepted type of taxon rank. Taxon category is a property of the taxon concept object.
Species complex	One of the taxon categories which correspond to a specific taxonomically accepted type of taxon rank. Taxon category is a property of the taxon concept object.
Doubtful	Has no correspondence in Dyntaxa.

Source: EUNIS Task #27963, https://taskman.eionet.europa.eu/issues/27963

2 Issues and solutions

2.1 *How to handle name shifts*

Name shifts occur very often among taxa in several groups of organisms. Figures 2.1.1 and 2.1.2 give two examples from the articles 12 and 17 reporting which arises due to quite different reasons. Not knowing that the shift has happened may cause great problems for the reader who wants to compare lists of reported results between rounds. The two synonymous names will appear as if they represent two separate taxa. If the reader of the list is provided with a column containing the established common name for that taxon it is often not a problem as taxon concepts very seldom change the common name compared to how often the scientific name shifts for species and their subtaxa due to revisions based on phylogeny.

Figure 2.1.1 An example where a species reported previous reporting round (orange) has been moved to another genus and consequently will have a new name in the next reporting round (green).

Note: ETC/BD Species codes are shown in brackets.
Source: ETC/BD, 2015.

Figure 2.1.2 An example where an incorrect name spelling was used for a species when reported previously (orange). Next round it will be reported using the corrected name (green).

Note: ETC/BD Species codes are shown in brackets.
Source: ETC/BD, 2015.

The pragmatic solution to the name shifting problem is simple if one use the Dyntaxa taxon information model, as the Taxon concept, and thereby is identifier (TaxonId or GUID) will stay the same between reporting rounds. The reader of the date will then be able to compare reports from different reporting rounds. When searching for data it will be possible to use any of the names. When listing old data the new name, which has become the recommended one for that taxon concept, can be shown in a column labelled 'Scientific name' while the old name can be put in a separate column. Notice that the 'Species codes' used in Figures 2.1.1 and 2.1.2 are not equivalents to the TaxonIds (or the GUID:s of the Taxon concept). Using the terminology from the Dyntaxa taxon information model the so called 'Species codes' of ETC/BD is instead equivalents to the TaxonNameIds (or the GUID:s of the Taxon names).

Depending on the nomenclatural status of the scientific name that is no longer recommended, it may be beneficial to use separate columns to put it in. 'Synonyms' is adequate if the non-recommended name is still considered to be correct as in the case of the example in Figure 2.1.1. If the nomenclature is not classified as 'Correct' as would be the case shown in Figure 2.1.2, then it may be more appropriate to put the name in a column labelled 'Other names'.

When linking data to taxon concepts rather than directly to the scientific names, presentations of data will automatically change over time. This is of course very powerful in terms of comparability of data within the domain that uses the same taxonomic data system as its backbone. However, it is important to realize that there will always be different opinions about taxonomy and consequently different taxonomic systems will define concepts and select recommended names differently. Furthermore, most taxonomic systems do not separate names from concepts as Dyntaxa does. This implies that normally there is no way to trace the history of changes automatically. As a consequence of the heterogeneous taxonomic views it is wise to save the name for which data was originally reported separate from the TaxonId and the currently recommended name, associated synonyms and other names. This name should be stored in a data field labelled e.g. 'VerbatimScientificName' together with the report. The verbatim name is not a part of the strict taxonomic information model but a valuable property of each record of taxon attributes or taxon observations.

2.2 Reports on different taxonomic levels

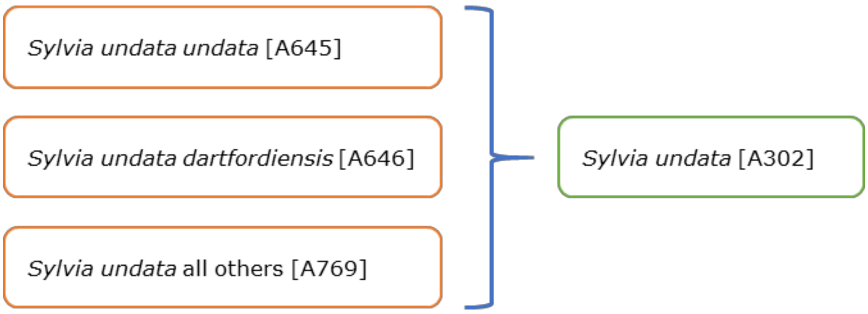
In previous reporting round of Article 12 and 17, some information was reported on the subspecies level which in the next reporting round will be reported on the species level. Figure 2.1 gives an actual example from ETC/BD 2015. When this is the case obvious problems arises dealing with comparability of the actual information reported for the involved taxa between subsequent reporting events. The Dyntaxa information model does not cover all aspects of this problem. Additional solutions are needed and at The Swedish Species Information centre we experienced the same problem several times during the course of repeated rounds of Red List compilations.

Direct comparisons of data values between subsequent reporting events are in most cases not possible when the reports are done on different taxonomic levels. Of course one could think of calculating some kind of summary statistics, selecting minimum or maximum values based on the individual reports for the subordinate taxa in order to obtain single metrics on the species level. However, although it would be technically very simple to perform such calculations automatically based on the hierarchical relationships obtained from Dyntaxa or similar systems, it is often not statistically justified to do so.

Nevertheless, having all the hierarchical relations at hand the presentation of the reports can be much improved by adding routines that search for report results for all hierarchically related taxa. Then the reader can make own conclusions by reading all the relevant reports related to the requested taxon and trace the history of reporting routines. In order to facilitate this it is possible to construct specific factors which directly points out the relationship between taxa with comparable information between reporting events. In the Swedish Taxon attribute Service one such factor is called 'Red listed as' (see Table 1.6) and that factor points out which taxon that are related to the target taxon that has the Red List information in the current period. The general rule in the Swedish Red listing process has been to assess Red list category on the species level. If the species fulfils the criteria for the Red list then the sub taxa is not listed separately. However, if the species then becomes classified as LC (Least Concern), which implies that it is not Red listed, and any of its subtaxa fulfils the criteria for being Red listed, then the sub taxa should be listed.

If ETC/BD has a 'Reported as'-factor at hand for each taxon relevant for the reporting, it is very simple to present the relevant reporting results to whatever taxon the user happens to search for. If it is one of the subtaxa which has no report for the selected period then the 'Reported as'-factor may point out information reported for the species.

Figure 2.2 An example where information on taxa was reported on the subspecies level in the previous reporting round (orange) but where the next reporting round (green) will report on the species level.



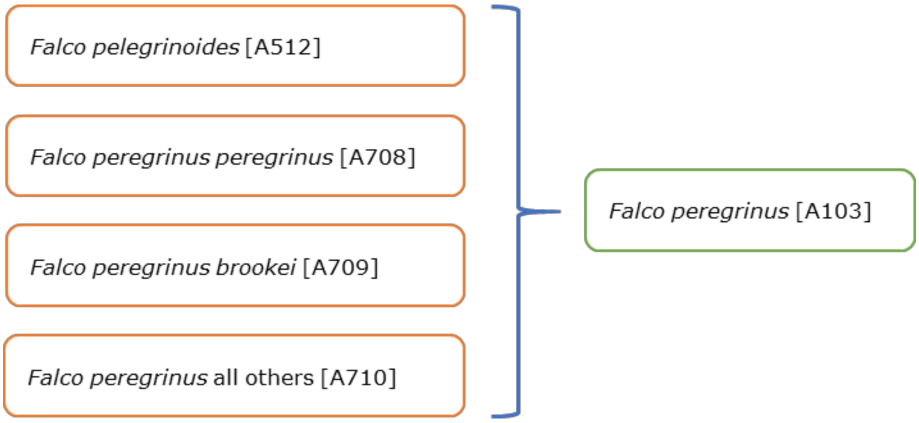
Note: ETC/BD Species codes are shown in brackets.
Source: ETC/BD, 2015.

If it is really important to analyse trends over subsequent reporting rounds then it may actually require a complete reassessment based on the same taxon concepts as was used for the last assessment. This is of course a costly approach but in some cases it may be necessary. In the Red listing field this is actually done when applying the Red List Index on a group of taxa in order to understand whether or not actual changes in the risk of extinction has occurred.

2.3 When taxa become merged between reporting rounds

A very common type of taxonomic change that can cause great problems for data management and comparisons between reporting rounds is the lumping, where two or several previously separate taxon become merged into a single replacing taxon. Figure 2.3 gives an example from the articles 12 and 17 reporting. When this happens between reporting rounds it causes very similar problems as the example where reporting strategy shifted from subspecies to only species level (Figure 2.2), i.e. comparisons of actual data values will in most cases be difficult without reassessing the data. When it comes to presentation of data for those who just need to read information for all the taxon concepts involved in the successive reporting rounds it is of course in principle as easy as in the case described in section 2.2. However, while it is usually straight forward to retrieve the list of subtaxa for any species using most taxonomic data systems available today, it is usually not possible to retrieve the list of taxa interrelated as a consequence of a taxon lump. Here Dyntaxa and its underlying web service makes a big difference. As that system keeps track of all lumps it is always possible to retrieve information about which taxon concept replaced those that have become unaccepted due to the merging process. Similarly, it is always possible to retrieve a list of the taxa that were accepted before the lump.

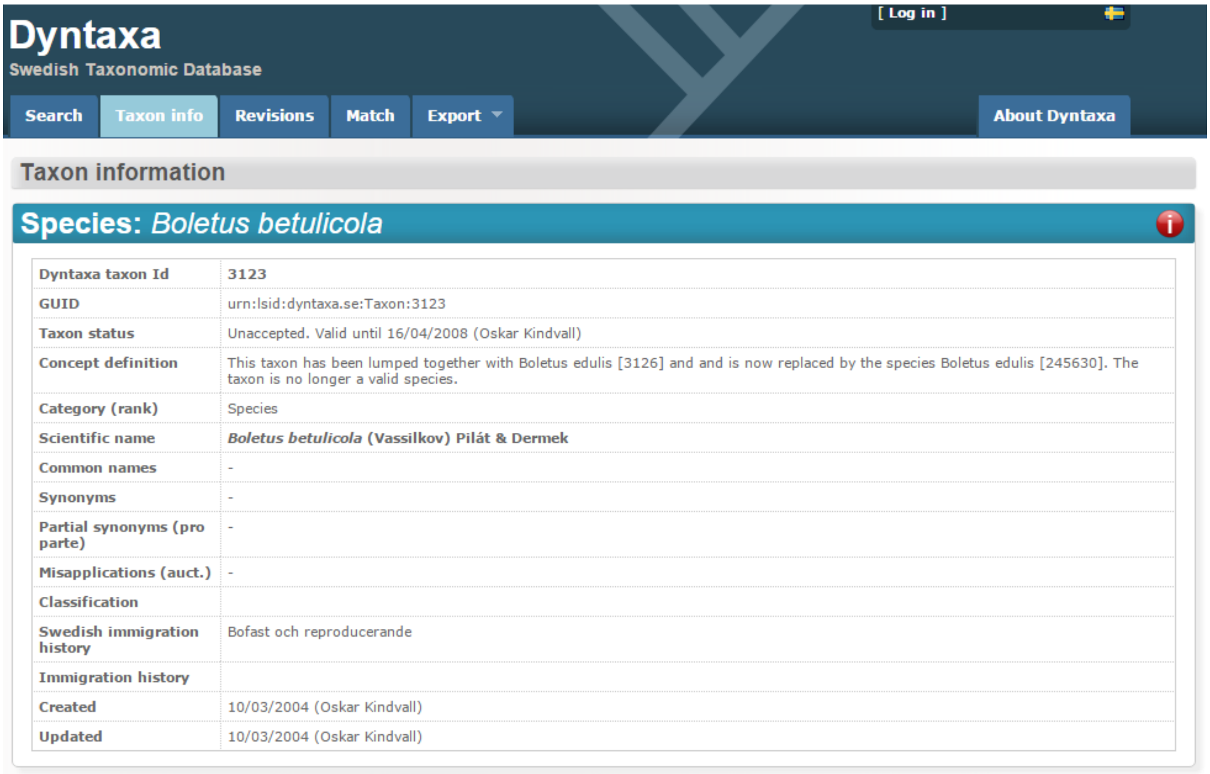
Figure 2.3 An example where a set of taxa becomes merged since the previous reporting round (orange) and replaced by a single taxon before the subsequent reporting round (green).



Note: ETC/BD Species codes are shown in brackets.
Source: ETC/BD, 2015.

Having the information model of Dyntaxa at hand, it is easy to present the reports from subsequent rounds together no matter which of the involved taxon names the end user searched for. Furthermore, it is also possible to dynamically produce text that explains to the reader what has happened with the involved taxon concepts.

Picture 2.3.1 An example of a taxon information page in Dyntaxa presenting a species that has been lumped with another species and thereby has become invalid (Taxon status = 'Unaccepted'). Notice that the text presented for the field 'Concept definition' has been automatically generated based on available information on lump/split relations.


Dyntaxa
Swedish Taxonomic Database
[Log in]
Search | Taxon info | Revisions | Match | Export | About Dyntaxa

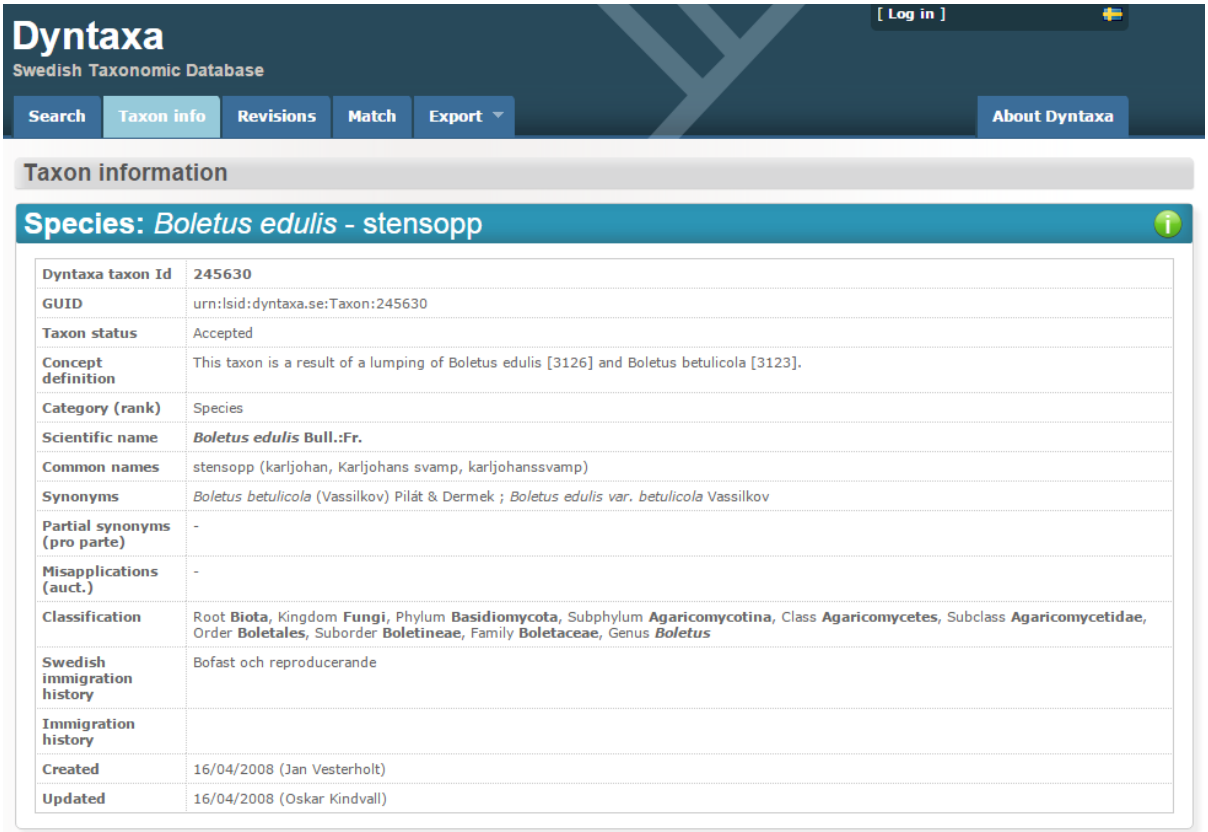
Taxon information

Species: *Boletus betulicola*

Dyntaxa taxon Id	3123
GUID	urn:lsid:dyntaxa.se:Taxon:3123
Taxon status	Unaccepted. Valid until 16/04/2008 (Oskar Kindvall)
Concept definition	This taxon has been lumped together with Boletus edulis [3126] and and is now replaced by the species Boletus edulis [245630]. The taxon is no longer a valid species.
Category (rank)	Species
Scientific name	***Boletus betulicola*** **(Vassilkov) Pilát & Dermek**
Common names	-
Synonyms	-
Partial synonyms (pro parte)	-
Misapplications (auct.)	-
Classification	
Swedish immigration history	Bofast och reproducerande
Immigration history	
Created	10/03/2004 (Oskar Kindvall)
Updated	10/03/2004 (Oskar Kindvall)

Note: Dyntaxa taxon concept ids are shown in brackets. The red i-symbol signals that the taxon is no longer accepted according to scientific consensus, i.e. taxon status equals 'Unaccepted'. Note that as the GUI of Dyntaxa currently works the information about Date for last modification, i.e. 'Updated' do not correspond to editing changes of the objects that holds the change information. This is of course not meant to be so. If name, hierarchies or change status change then the 'ModifiedDate'-property of the taxon concept should reflect this important changes.
Source: www.Dyntaxa.se, 2015: https://www.dyntaxa.se/Taxon/Info/3123

Picture 2.3.2 An example of a taxon information page in Dyntaxa presenting the taxon concept replacing the species shown in Picture 2.3.1 after it became lumped together with another former species. Notice that the text presented for the field 'Concept definition' has been automatically generated based on available information on lump/split relations.



Dyntaxa
Swedish Taxonomic Database
[Log in]
Search | Taxon info | Revisions | Match | Export | About Dyntaxa

Taxon information

Species: *Boletus edulis* - stensopp

Dyntaxa taxon Id	245630
GUID	urn:lsid:dyntaxa.se:Taxon:245630
Taxon status	Accepted
Concept definition	This taxon is a result of a lumping of Boletus edulis [3126] and Boletus betulicola [3123].
Category (rank)	Species
Scientific name	***Boletus edulis*** **Bull.:Fr.**
Common names	stensopp (karljohan, Karljohans svamp, karljohanssvamp)
Synonyms	*Boletus betulicola* (Vassilkov) Pilát & Dermek ; *Boletus edulis var. betulicola* Vassilkov
Partial synonyms (pro parte)	-
Misapplications (auct.)	-
Classification	Root **Biota**, Kingdom **Fungi**, Phylum **Basidiomycota**, Subphylum **Agaricomycotina**, Class **Agaricomycetes**, Subclass **Agaricomycetidae**, Order **Boletales**, Suborder **Boletineae**, Family **Boletaceae**, Genus ***Boletus***
Swedish immigration history	Bofast och reproducerande
Immigration history	
Created	16/04/2008 (Jan Vesterholt)
Updated	16/04/2008 (Oskar Kindvall)

Note: Dyntaxa taxon concept ids are shown in brackets. The green i-symbol signals that the taxon is accepted according to scientific consensus.
Source: www.Dyntaxa.se, 2015: https://www.dyntaxa.se/Taxon/Info/245630

In the presentation of reports related to taxon concepts that are no longer accepted due to a lump event, it is important to point out that the data must be interpreted with care as evaluations or measurements do not actually represent a subgroup of individuals that differ from any of the other groups of individuals belonging to the new replacing taxon concept.

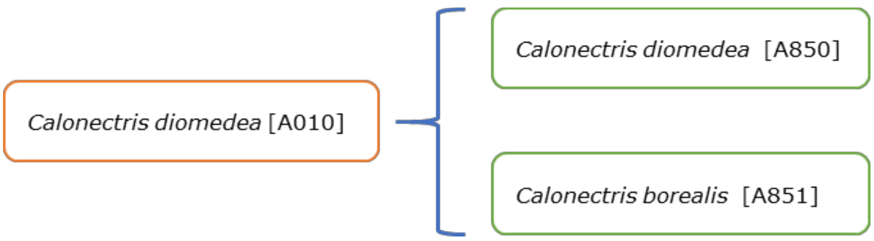
When listing report results from several taxa together it would be best to use the recommended scientific name for the currently valid taxon concepts only in the column labelled 'Scientific name' for all reports including old ones where the report may stem from lumped taxa which are no longer accepted. A column labelled 'Verbatim scientific name' could be used to inform the reader about the origin of data and making it clearer why several somewhat different reports occurred for the same taxon for some earlier reporting rounds.

2.4 When species become 'split up' between reporting rounds

As with the merging or lumping of taxa, the reverse process, i.e. when a single taxon becomes split up into two or more new taxa (Figure 2.4), may result in severe problems when trying to compare data from subsequent reporting rounds. In practice the problems and solutions for comparisons and presentations is very similar to the ones discussed in sections

2.2 and 2.3. Generally, most exiting taxonomic information systems do not support an automatic solution helping developers to construct interfaces that in a simple way can summarize the data from all involved taxa. To do this one needs information about which taxa replaced the old one and which taxon was replaced.

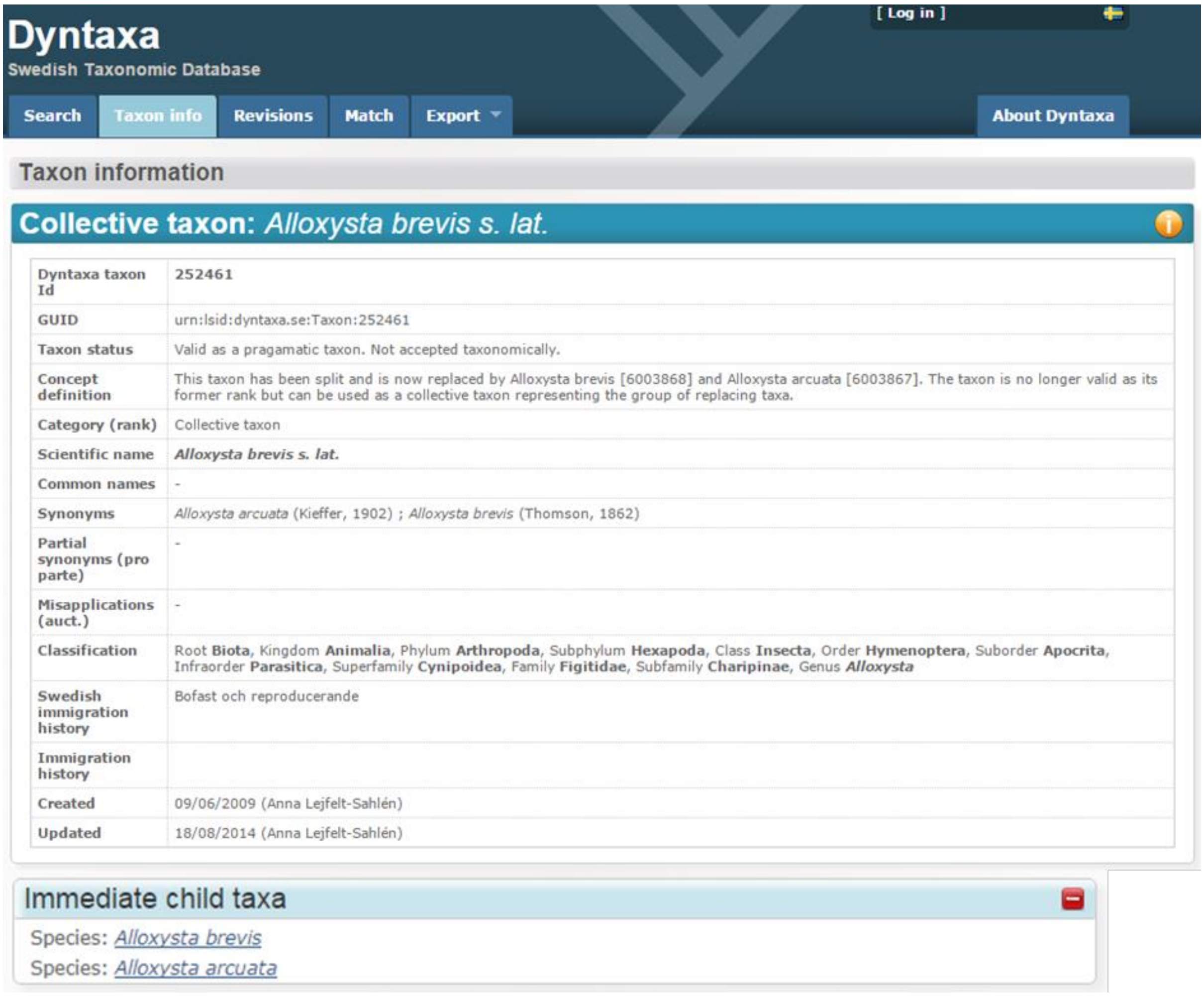
Figure 2.4 An example where a taxon has been split. In the previous reporting round it was considered as a single species (orange). In the next round the reporting has to be done for two species separately (green).



Note: ETC/BD Species codes are shown in brackets.
Source: ETC/BD, 2015.

In Dyntaxa, the full understanding of what has happened as a result of a split event can be presented for the reader (Picture 2.4.1 and 2.4.2) in a similar way as for the lump events (Pictures 2.3.1 and 2.3.2). However, there is a significant difference between the two cases as it is handled by default in Dyntaxa. Instead of making the replaced taxon totally invalid, as is the case with replaced taxa after a lump event, the taxon that has been split up is still considered an accepted taxon concept. Although, the split taxon has to change taxon category. If it was considered an accepted species before the split, its rank is now automatically changed to the informal category ‘Collective taxon’.

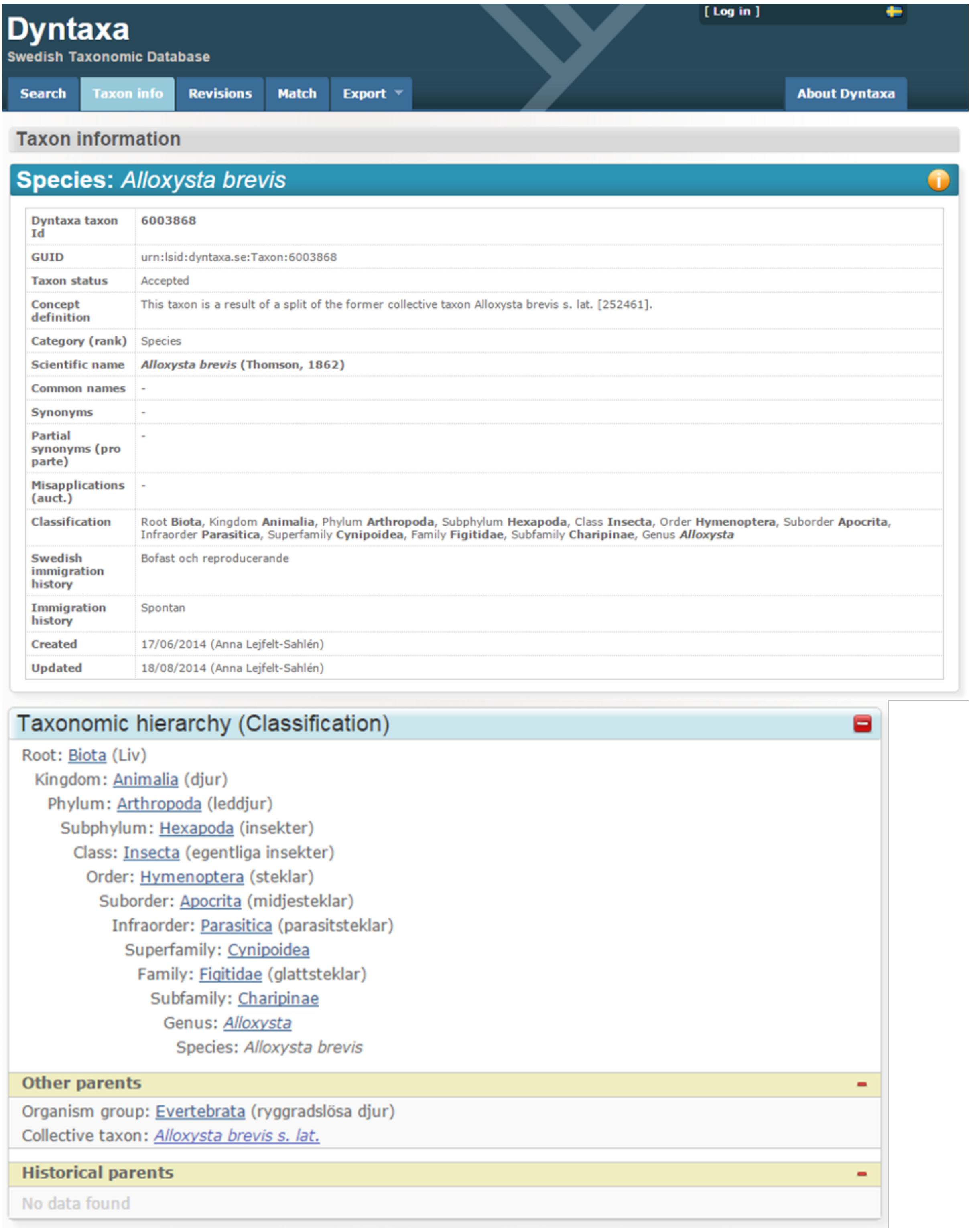
Picture 2.4.1 An example of a taxon information page in Dyntaxa for a former species that has been split into two other species. Notice that the text presented for the field 'Concept definition' has been automatically generated based on available information on lump/split relations.



Dyntaxa taxon Id	252461
GUID	urn:lsid:dyntaxa.se:Taxon:252461
Taxon status	Valid as a pragamatic taxon. Not accepted taxonomically.
Concept definition	This taxon has been split and is now replaced by Alloxysta brevis [6003868] and Alloxysta arcuata [6003867]. The taxon is no longer valid as its former rank but can be used as a collective taxon representing the group of replacing taxa.
Category (rank)	Collective taxon
Scientific name	***Alloxysta brevis s. lat.***
Common names	-
Synonyms	*Alloxysta arcuata* (Kieffer, 1902) ; *Alloxysta brevis* (Thomson, 1862)
Partial synonyms (pro parte)	-
Misapplications (auct.)	-
Classification	Root **Biota**, Kingdom **Animalia**, Phylum **Arthropoda**, Subphylum **Hexapoda**, Class **Insecta**, Order **Hymenoptera**, Suborder **Apocrita**, Infraorder **Parasitica**, Superfamily **Cynipoidea**, Family **Figitidae**, Subfamily **Charipinae**, Genus ***Alloxysta***
Swedish immigration history	Bofast och reproducerande
Immigration history	
Created	09/06/2009 (Anna Lejfelt-Sahlén)
Updated	18/08/2014 (Anna Lejfelt-Sahlén)

Note: Dyntaxa taxon concept ids are shown in brackets. The yellow i-symbol which refers to current Alert Status of the taxon signalizes that the taxon is somewhat problematic. As it has been split the name usage may cause problems as the name in itself is ambiguous.
Source: www.Dyntaxa.se, 2015: https://www.dyntaxa.se/Taxon/Info/252461

Picture 2.4.2 An example of a taxon information page in Dyntaxa for one of two related species that arose from a split of the species presented in Picture 2.4.1. Notice that the text presented for the field 'Concept definition' has been automatically generated based on available information on lump/split relations.


Dyntaxa
Swedish Taxonomic Database
[Log in]
Search | Taxon info | Revisions | Match | Export | About Dyntaxa

Taxon information

Species: *Alloxysta brevis*

Dyntaxa taxon Id	6003868
GUID	urn:lsid:dyntaxa.se:Taxon:6003868
Taxon status	Accepted
Concept definition	This taxon is a result of a split of the former collective taxon Alloxysta brevis s. lat. [252461].
Category (rank)	Species
Scientific name	***Alloxysta brevis*** **(Thomson, 1862)**
Common names	-
Synonyms	-
Partial synonyms (pro parte)	-
Misapplications (auct.)	-
Classification	Root **Biota**, Kingdom **Animalia**, Phylum **Arthropoda**, Subphylum **Hexapoda**, Class **Insecta**, Order **Hymenoptera**, Suborder **Apocrita**, Infraorder **Parasitica**, Superfamily **Cynipoidea**, Family **Figitidae**, Subfamily **Charipinae**, Genus ***Alloxysta***
Swedish immigration history	Bofast och reproducerande
Immigration history	Spontan
Created	17/06/2014 (Anna Lejfelt-Sahlén)
Updated	18/08/2014 (Anna Lejfelt-Sahlén)

Taxonomic hierarchy (Classification)

Root: Biota (Liv)
Kingdom: Animalia (djur)
Phylum: Arthropoda (leddjur)
Subphylum: Hexapoda (insekter)
Class: Insecta (egentliga insekter)
Order: Hymenoptera (steklar)
Suborder: Apocrita (midjesteklar)
Infraorder: Parasitica (parasitsteklar)
Superfamily: Cynipoidea
Family: Figitidae (glattsteklar)
Subfamily: Charipinae
Genus: *Alloxysta*
Species: *Alloxysta brevis*

Other parents

Organism group: Evertebrata (ryggradslösa djur)
Collective taxon: *Alloxysta brevis s. lat.*

Historical parents

No data found

Note: Dyntaxa taxon concept ids are shown in brackets. The yellow i-symbol which refers to current Alert Status of the taxon signalizes that the taxon is somewhat problematic. As it is the result from a split of a former species the name usage may cause problems as the name in itself is ambiguous.

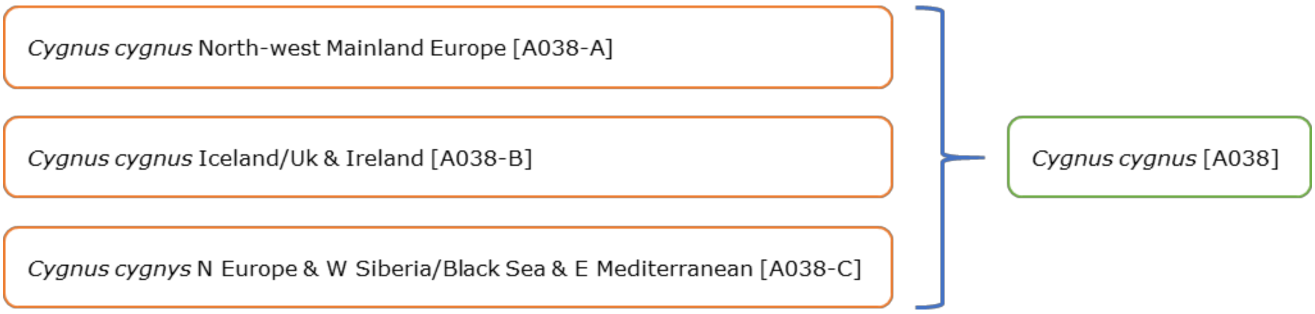
Source: www.Dyntaxa.se, 2015: https://www.dyntaxa.se/Taxon/Info/6003868

2.5 How to handle information related to populations of the same taxon

'Flyway' is a concept used by BirdLife and others, e.g. African-Eurasian Migratory Waterbird Agreement (AEWA) which separates groups of individuals (populations) belonging to the same species as different pragmatic taxa. It is in principle no problem to treat these entities as taxon concepts although they do not correspond to any taxonomic entities that are formally named. Similar concepts has been dealt with in the process of compiling the Red List of Sweden. Examples include the Baltic Sea population of the seal *Phoca vitulina* and the wintering population of the Long-tailed Duck (*Clangula hyemalis*). In these cases the taxon category is set to 'Population' which also would be an appropriate solution for the example from the Article 12 reporting shown in Figure 2.5.

Having the taxon information model of Dyntaxa at hand, each 'Flyway' type of taxa in Figure 2.5 could be treated as an 'Accepted pragmatic taxon' where the taxon category is set to 'Population'. All these taxa could then also be hierarchically linked as child taxa to their corresponding species taxon concept.

Figure 2.5 An example where the reporting was made separately for different distinct groups of individuals based on flyways of the same species in the previous reporting round (orange) while the next round will report only for the species (green).



Note: ETC/BD Species codes are shown in brackets.
Source: ETC/BD, 2015.

2.6 When legislation lists a whole genus

In some cases legislation points out a whole group of taxa belonging to a specified organism group. One example of that from the Article 17 reporting is the genus *Cladonia* (subgenus *Cladina*). This is of course from one point of view very practical as it reduces the number of items which require reporting. Still when compiling the assessments of conservation status practical problems sometimes arise when trying to gather all relevant raw data. The problem is that it is not always clear which species belong to a particular genus and without knowing the list of species and other subordinate taxa it is sometimes hard to search for observations and other types of relevant information.

Another problem with having legislation associated to the genus level instead of being more specific is that species are often moved from one genus to another during taxonomic revisions. These changes of the taxon concepts may impose juridical uncertainty regarding whether or not the legislation still applies for those subtaxa that do no longer belong to the original genus. However, no matter the legal aspects of this issue there is a need for taxonomic systems that can support listing of all taxa that belong to a certain genus both currently and historically. Furthermore, if the systems that store observations needed for assessment during reporting have not linked their observation records to the Dyntaxa type of taxon concepts, then one also require a way of obtaining all relevant names of all the relevant subtaxa that belong to the genus. On the other hand, if the observatory system is based on the type of information model outlined in Figure 1.1 none of these issues should cause any problem. In the observation systems connected to the Swedish LifeWatch infrastructure all relevant observation data can be retrieved by just searching for the desired genus name and using its Taxon Concept Id as the search parameter.

2.7 How to handle arbitrary groups and 'all others'

In the more strict taxonomic data systems like e.g. Catalogue of Life and PESI it is probably not suitable to include pragmatic taxa of any kind. Concepts that constitute non-phylogenetic arbitrary groups of taxa are not accepted taxonomically and thereby do not fit into the scope of most taxonomic database systems. In the checklists used for Article 12 and 17 the term 'all others' occur in order to collectively cover a group of subtaxa that are related to a certain species. One such example is the name '*Troglodytes troglodytes* all others' [A676] (ETC/BD, 2015) which is used to cover both the species and all its subordinate taxa occurring in the geographical scope of the Nature Directives except *Troglodytes trogoldytes fridanensis* EUNIS cannot fully lean on external and globally accepted taxonomic databases such as Catalogue of life if the reporting is linked to name strings of this pragmatic type.

When dealing only with taxon name objects as in the current checklist for reporting under the Nature Directives, names with the extension 'all others' is of course one solution that works for reporting. However, using the hierarchical structure of related taxa implemented by the Dyntaxa taxon information model names like the ones with 'all others' would not be needed. All subordinate taxa are always included in any concept. So if a report would have been linked to the taxon concept of the species level, instead of to a name object as in the case of the checklist example, it would automatically imply that all its subtaxa also are included in the report. This would of course help when trying to link the data to other taxonomic systems as the taxonomically valid taxon concepts of Dyntaxa normally would be easy to match with the accepted names listed in e.g. the Catalogue of Life (see section 2.8).

As the mechanism for creating hierarchical relations between taxa in the Dyntaxa taxon information model includes the property 'IsMainRelation' (Table 1.4), it enables editors to connect taxon objects to several alternative parent taxa. This implies that it is possible to add pragmatic groups of taxa which do not follow the nomenclatural rules without destroying the tree structure for those users who prefer a more strict taxonomic view. Arbitrary groups are quite common in Dyntaxa today. Such groups are set up as valid taxon concepts where taxon category is set to e.g. 'Organism group', Species complex' or 'Collective taxon'. All the taxon concepts which are members of such a specific group are then connected to this group hierarchically as child taxa.

2.8 How to link information across separate taxonomic systems

Both PESI and Dyntaxa provide graphical web interfaces supporting the process of matching taxon lists against the content in the underlying databases. The matching can of course also be done programmatically in relation to these two sources by means of their web services. Programmatically, it is also possible to match whole taxon lists against Catalogue of Life.

Several algorithms useful for cross-mapping has been developed during recent years within the community related to Catalogue of Life and GBIF (Culham *et al*., 2013).
When new taxon concepts are added to Dyntaxa by any of the editors the PESI Web Service is used in order to automatically check whether or not the concept names match a name in the PESI Database. If there is a match then its GUID is stored as an extra taxon name of the type 'GUID'. By doing this a direct link between the taxonomic information held by the two systems is created. A taxon expert may thereafter validate the linkage and set the GUID to the recommended GUID for that taxon concept, if appropriate. For all taxon concepts that do not match to PESI the Dyntaxa taxon concepts GUID will still be the recommended one. If applying this approach iteratively during the progress of editing a taxonomic database, preferably also using several alternative sources, hopefully most taxon concepts will eventually become interlinked with the global systems.

From now on also Catalogue of Life (www.catalogueoflife.org) will record IDs stable for names of species and taxa subordinate the species level. Their new identifiers are constructed as hashed strings (GNA, 2015) that are unique depending on both the name itself and its context in terms of its current parent taxa. This procedure aims to exclude the possibility that the same name used for e.g. a plant and an animal will become ambiguously mixed. Furthermore, the procedure will give rise to identifiers that are globally unique and may therefore work as GUID:s which opens up for the possible use together with other GUID:s stemming from other sources. This definitively helps the process of interlinking and homogenizing the taxonomic systems around the world. Once a match was done properly with Catalogue of Life it will be very easy to continuously check the current status of names using one of the available methods in the web service for Catalogue of Life.

Catalogue of Life does not provide any identifiers for the taxon concepts as defined by the information model described in this report (Jones *et al*., 2011). Nevertheless, as it provides the best source for GUID:s of scientific names of species in the world, it can be used as the core name list which all other taxonomic systems can link to.

In the current version of Dyntaxa there is no property of the taxon name objects called CatalogueOfLifeId. However, imagine that all taxon name objects that prove to match a name in Catalogue of Life bares its unique identifier. Then the taxon concepts provided by Dyntaxa can be described in terms of how generally accepted they are. If the name usage fits with the current view presented by Catalogue of Life the concept could be described as 'Established'. On the other hand, if none of the names associated with the taxon concept links to any name in Catalogue of Life then the Dyntaxa taxon concept must be considered as being 'Not established'. In some cases, the taxon concept will clearly appear as being 'Ambiguous' as it may be associated with two or more accepted names in Catalogue of Life, implying the underlying idea of the concept definition differs.

References

ArtDatabankenSOA, 2015, ‘Code Documentation of the Service Orientated Architechture Developed by Swedish Species Information Centre’ (https://lampetra2-1.artdata.slu.se/ArtDatabankenSOA/client/) accessed 24 August 2015.

Culham, A., Sitko, M., Roskov, Y., Didziulis, V., Cheung, K., Kunze, T, Schalk, P., Addink, W., Döring, M., Cochrane, G., Riviere, S., Robert, V., Bogdanowicz, W., Hilton-Taylor, C., Berendsohn, W., Güntch, A., Jones, A., White, R., Bourgoin, T., 2013, ‘i4Life: Standardising the World's Biodiversity Catalogue’, Conference: IST-Africa, At Nairobi, Kenya, Volume: 2013.

ETC/BD, 2015, ‘ETC/BD Working Document’ (Issues_art12_next_reporting_20150824.docx) accessed 24 August 2015.

GNA, 2015, ‘New UUID v5 Generation Tool -- gn_uuid v0.5.0’ (http://globalnames.org/news/2015/05/31/gn-uuid-0-5-0/).

IAPT, 2012, ‘International Code of Nomenclature for algae, fungi, and plants’ (http://www.iapt-taxon.org/nomen/main.php).

ICZN, 2000, ‘International code of zoological nomenclature’ (http://www.iapt-taxon.org/nomen/main.php).

Jones, A. C, White, R. J. and Orme, E. R., 2011, ‘Identifying and relating biological concepts in the Catalogue of Life’, *Journal of Biomedical Semantics* 2:7.

Object Management Group, 2004, ‘Life Sciences Identifiers Specification, OMG Final Adopted Specification’, (http://www.omg.org/cgi-bin/doc?dtc/04-05-01.pdf).

TDWG, 2015, ‘Darwin Core’ (http://rs.tdwg.org/dwc) accessed 4 September 2015.