Taxon Concept Schema - User Guide

Roger Hyam (editor) – 2005-05-31 - Version 1.0 – First Version.

Table of Contents

1 Credits

This document is part of an general effort by a international community of individuals to improve the way taxonomic information is exchanged between organisations. It is the result of the work of the many people who have responded to direct requests for help or contributed to email lists, wiki sites and verbal discussions. It is not possible to produce a comprehensive list of all those involved and a partial list would not do justice to those who were omitted. The names of a few contributors appear in the text where their examples have been used more or less verbatim. Errors and inaccuracies are, of course, attributable to the editor.

The compilation of the guide has been made possible by the financial support of:

- The Global Biodiversity Information Facility (www.gbif.org) who fully funding the work of the editor and partially funded the development of TCS by the Napier University team under Professor Jessie Kennedy – who made time available to help with the guide.

- The Science Environment for Ecological Knowledge initiative (seek.ecoinformatics.org) who partially funded the development of TCS by the Napier University team.

2 Comments, Corrections and Suggestions

Any comments, corrections or suggestions as to how this guide could be improved should be emailed to roger@hyam.net. Messages titles should start with "TCS User Guide: ". All feedback is most welcome.

3 Introduction

This guide is meant to act as a readable introduction to the Taxon Concept Schema (TCS). It is aimed at both decision makers and implementers of systems and so the content varies from being of a general nature, assuming little prior knowledge of taxonomy or XML technologies, to being quite detailed concerning the technicalities of biological nomenclature and XML schemas. Generally it becomes more specific as it goes on.

This document is not a step by step manual for mapping a data source to a TCS document. Each data source is likely to be different in structure and so this is not a feasible exercise. What it attempts to do is provided the information needed for an implementer to make intelligent decisions as to how their data source should be mapped to the schema so that they can publish and/or receive data from other systems with as little confusion as possible.

4 Conventions

- When XML elements are referred to in the text they are surrounded by angled brackets like this <> to make it clear that we are referring to an actual element that could occur in an instance document rather than a hypothetical notion.
- When XML attributes are referred to in the text they are preceded by an @ to make it clear we are referring to XML attributes of instance documents not hypothetical notions.
- When the names of XML Schema types appear they are surrounded by curled brackets like this {}. This differentiates them from elements that can be instantiated.
- When XML examples are given they contain only the elements necessary to illustrate the current point. They are not valid XML documents but should validate if included correctly in valid TCS instance documents. XML in actual instance documents is likely to be far more complete.

5 History

The need for a common mechanism for the providers of taxonomic information to exchange data with other providers and users of varying expertise in taxonomy was recognized at TDWG Lisbon 2003. As a result J Kennedy was appointed Chair of the Taxonomic Names Subgroup to lead the development of a Taxonomic Names/Concepts Exchange Schema. This process began with a series of consultation meetings covering the range of Taxonomic data providers and users specifically Vegbank, Taxonomer, SEEK, Berlin Model, GBIF, Nomencurator, Species 2000, IT IS and IPNI and informal discussion with APNI. From these meetings a Strawman Schema (TCS) was developed and presented at a meeting on Taxonomic Exchange Standards held at the eScience Centre in Edinburgh in May 2004 to which all interested parties were invited to attend and contribute. This schema focussed on concepts and inherited the Names subschema from the ABCD Schema. The TCS was favourably received although some changes or extensions were suggested which were taken into account in preparation for presenting the draft schema at TDWG Christchurch 2004. In May 2004 Jerry Cooper proposed an alternative schema (Linnean Core) to TCS to focus more specifically on Names issues particularly as seen by Nomenclators which was initially distributed only to GBIF ECAT members, but was not planned to be developed further by Jerry. The draft schema was presented at TDWG Christchurch and received a mixed reception. The main problem appeared to be that Names were being treated as part of a Concept and some users wanted to be able to deal with names independently. In addition it was felt that not all name issues had been addressed. A Linnean core sub group of the Names sub group was formed to investigate in depth the name issues which were currently represented in the Name element and through Nominal concepts in the TCS. As a result Linnean Core was modified and extended and much discussion took place over the relationship between the two. The current Schema takes what we consider to be the best work from both Linnean core and TCS combined into a coherent schema which can satisfy the needs of all taxonomic name and concept users. The major difference from the schema presented at Christchurch being that Names have been elevated to top level elements and specify specific valid relationships that can be held between names while still trying to ensure that names are not used where concepts are meant. This schema will be presented at TDWG St Petersburg 2005 for voting as a TDWG standard by the members.

6 Rational for a Taxon Concept Schema.

The availability, exchange and interpretation of taxonomic information (such as species check lists, distribution and identification data) is of critical importance to taxonomists, ecologists and other biologists and legislators, amongst others. This information is provided by a number of global and local taxonomic database services. These databases hold records, often based on valid scientific names for species, according to their own models of what constitutes a taxonomic 'entity' or concept (i.e. a species or higher level taxon). Databases typically model a single view of taxonomy, whilst making some attempt to relate their concepts to synonymous names or concepts.

The need for a common mechanism for the providers of taxonomic information to exchange information with other providers and users of varying expertise in taxonomy was recognized at TDWG Lisbon 2003. Such a mechanism must adequately represent the data as modelled by the owners of the data, whilst facilitating integration with data provided according to different models of taxonomy.

The development of an abstract model for a taxonomic concept, which can capture the various models represented and understood by the various data providers is clearly required. The Taxon Concept Schema was developed to meet this need. The model is presented as an XML schema document that is proposed as a standard to allow exchange of data between different data models. It aims to capture data as understood by the data owners without distortion, and facilitate the query of different data resources according to the common schema model.

7 What are Taxon Concepts and what are they for?

A taxon is a unit of biodiversity. The species *Narcissus pseudonarcissus* L. (daffodil) is a taxon just as the genus *Ratus* and the kingdom *Animalia* are taxa. The scientific names that are used to refer to taxa are controlled by a series of legalistic codes that govern their use. These codes state that names must be published in a specific way and that each name should have a voucher or 'type' specimen. A type specimen is usually an individual or small group of individuals that anchor a name in the literature to the 'real world' of biodiversity.

As taxonomists work to improve the classification of organisms they produce definitions or circumscriptions for taxa. They do this by writing descriptions or keys or providing lists of specimens or other data. Once they have delimited a taxon a decision is taken as to what it should be called. This is done by considering all the type specimens which fall within the circumscription of the taxon and using the oldest (first published) name associated with these specimens. Only if no type specimens occur within the delimitation is a new name created and a type nominated from within the new taxon. (This is a gross simplification of what are complex rule sets but is enough to illustrate the points below.).

When a non-taxonomist uses a name they use it to refer to a taxon. The problem is that there may have been more than one taxon defined that uses that name because the type specimen of the name may have fallen into several different circumscriptions. There may be several different **Taxon Concepts** or taxon delimitations associated with that name.

To use a name precisely a worker should use a fully qualified scientific name plus an indication of which concept of that name they are using. Within the schema this is done using the "AccordingTo" element. Traditionally the Latin *sensu* or *sec.* (for *secundum* – according to) have been used. Data that is marked up in this way can be compared in a far richer way than data that isn't.

Unfortunately most data is not available with <AccordingTo> statements. Typically workers just refer to the names without indicating which particular concept they are using. In the Taxon Concept Schema these are referred to as **Nominal Concepts**. They are uses of a name to indicate a circumscribed taxon without indicating which particular circumscription. Clearly if there is only one possible circumscription as with a recently published species then this is not as much of a problem as for a taxon that is part of a heavily revised group. When there were fewer names in use and fewer standard texts nominal taxon concepts were satisfactory. With increasing amounts of data and increasing levels of interconnectivity properly defined concepts are needed.

To summarise: A Taxon Concept is a name plus a description of a taxon. To refer to a Taxon Concept one must use a name plus an "AccordingTo" which gives the location of the circumscription.

8 Scientific Names

From the above discussion it can be seen that scientific names are separate from taxon concepts. One name may be used by several concepts. Names are governed by rules that are separate from concepts (such as whether they have been lectotypified, their basionym relationships etc.). The TCS therefore treats names as separate <NameObject> elements that are referenced by <TaxonConcept> elements.

<TaxonConcept> elements do not need to use <NameObject> elements. They have an element that can contain an arbitrary name string that can be a scientific name or a vernacular. <NameObject> elements are only used to expand on this name string and provide details of the scientific name that are relevant for the nomenclatural codes.

9 What the Taxon Concept Schema is NOT for?

The TCS schema was conceived to allow the representation of taxonomic concepts as defined in published taxonomic classifications, revisions and databases etc.

The TCS schema is not designed to facilitate the exchange or documentation of information about Taxon Concepts where this information is not part of a taxonomic revision creating new concepts. The amount and variety of (additional) information that can be potentially assigned to concepts is outside the scope of a taxonomic concept transfer schema, but development of domain specific models that use or extend this schema is encouraged. XML supports this flexibility by allowing the use of different name spaces.

For example, whilst a TCS taxon concept definition may include details of specimen circumscription (i.e. list specimens that are asserted to define the taxon concept) datasets that merely include observations identifying specimens as being examples of a taxon concept would reference a defined Taxon Concept, not constitute a new or modified concept. That is, TCS documents are for transferring the definitions of taxon concepts, not for detailing observations of these defined concepts.

Examples of observational datasets that could refer to defined taxon concepts might include

- ecological field records, which record abundance data on a variety taxa
- descriptive datasets that include descriptions of representative specimens or taxa
- specimen records for biological collections

The observations recorded in these would **not** constitute definitions of the concepts, but document instances of the concept. (However, observation datasets might themselves be included as part of a taxon concept definition, e.g. as specimen or character circumscriptions - if this is what the 'creator' of the concept intended).

10 What is an XML Schema?

The Taxon Concept Schema is an XML Schema. An XML Schema is a means of specifying how an XML document should be structured - which elements are permitted where, which elements are optional or required and what the elements and their attributes can contain.

A **TCS Document** is an XML document containing taxonomic data that conforms to the XML Schema called the Taxon Concept Schema.

TCS is a data exchange format i.e. A means of marking up the taxonomic data that is to be communicated. It is **not** a data exchange protocol. A TCS document does not 'know' anything about moving between computers. It relies on an exchange protocol to achieve this.

A common mistake is to think of TCS as a database schema. It is not a database schema. TCS documents should be thought of ephemeral software objects that pass between applications that make use of the their contents. It is presumed that the applications on either end of this process will have their own internal ways of representing and processing the data passed. All that is required of the schema is that the two applications can map their own data onto it unambiguously.

Because TCS is expressed as an XML Schema any documents that are passed from one machine to another that claim to be TCS documents can be validated using TCS to check that they conform to the standard. This is very useful for maintaining data integrity.

The fact that TCS is an XML Schema also greatly increases the power of tools such as XMLSpy, OxygenXML and many others to manipulated TCS documents.

11 What about GUIDs?

Globally Unique Identifiers (GUIDs) have been much discussed during the development of TCS. The importance of them can be understood when we consider how Taxon Concepts might be referenced. How can we tell if data from two or more sources is referring to the same concept? There are three types of keys that could be provided.

1. **A scientific name alone.** This only resolves to a nominal concept. We can not know *which* delimitation of the taxon the user is referring to. We can only know that they are referring to *a*

circumscription that has used this name. Even the resolution of one scientific name to another can be complex. Users do not need to supply ranks, authors or publications in any standard form and so software algorithms are needed to try and match names. These will make errors. Scientific names are '**partial user keys**' to concepts.

2. **A scientific name plus an "AccordingTo" publication.** This is much more useful for reasoning as we now know which taxon circumscription has been used but it suffers from the same problems as 1 above. Firstly the scientific name needs to be resolved and then the "AccordingTo" authors and publication need to be resolved. Neither of these is trivial and may be difficult to automate. Names plus AccordingTos are '**composite user keys**'.
3. **A globally unique identifier (GUID).** This makes the problem technically simple if there is a single authority who issues codes. It becomes more complex if there are multiple issuing authorities as the codes then need to be synonymised. Of course using GUIDs to mark-up data is not 'user-friendly', and does not provide a human readable record of the concept. Users may prefer to use or be provided with composite user keys for concepts (even alongside GUIDs where these are available).

Despite their importance to the usefulness of TCS and the future of taxonomic data exchange the whole subject is actually orthogonal to the structure of TCS itself. At any point in the schema where an object is referenced a GUID could be used or an ID that is local to the document or some other kind of ID. Further discussion of GUIDs in this context is therefore left to other works. The resolution of partial and composite user keys to taxon concepts is within the scope of the TCS however and is discussed in the section on **Signatures** later in the guide.

11.1 Stability of externally referenced objects

Linked to the issue of GUIDs is the notion of object stability. If an organisation issues an ID for an object (whether part of a globally unique system or just a institution specific identifier) they need to make clear the contract associated with that ID and how the object it refers to might change in the future. An example would be a ID for a NameObject. It may be acceptable for a NameObject to be changed by the addition of a lectotypification event. It may not be acceptable for the spelling of the actual name to be changed. Such a change would affect the meaning of any TaxonConcepts that referenced this NameObject. Changes such as spelling corrections should instead result in the creation of new NameObjects. This is an area that needs further discussion in the context of GUIDs.

12 Summary of Schema Structure

The XML Schema definition language has the ability to contain it's own annotations. There are tools available to generate documents in different formats from these internal annotations. TCS is well commented internally and documentation is generated from these comments for each release so there is little point in discussing each and every element in this guide. What is given here is a broad overview of the main sections that might occur in a TCS Document and what they are for.

12.1 {ReferenceType} Complex Type

This is not a structural element within the schema but is an important complex type that is used multiple times in different places. Understanding the way that elements of this type can be used is central in understanding how the schema can be used.

In XML Schema a complex type is an abstract definition of a element. This means that a number of different elements within a schema can have the same centrally defined structure although they may have different names. It is also possible for elements to 'extend' a type definition and add to or override some of the structure.

All the elements based on {ReferenceType} have at least two optional attributes: @ref and @linkType and they have optional string content. The @ref attribute contains the id of a resource the @linkType attribute contains the location/kind of resource. The content of the element contains a string representation of the resource. There are three basic ways the {ReferenceType} elements can be used: as a non-reference, as an internal reference and as an external reference - here illustrated with the <PublishedIn> element that occurs throughout the schema.

12.1.1 Non-reference reference.

```
<PublishedIn>Clapham, Tutin &amp; Moore (1987)</PublishedIn>
```

If both the attributes are missing from the tag then the instance of {ReferenceType} is not acting as a reference to another resource but actually contains the data. This is especially useful if the data being published is not atomised beyond the level of the reference. If, for example, the database contains a single string field to hold the place of publication for names then the value of that field can be contained within the <PublishedIn> element and there is no need to create separate <Publication> elements. Likewise with voucher specimens and other links including those in <TaxonConcepts><Relationships>.

12.1.2 Internal reference

```
<PublishedIn ref=”123”>Clapham, Tutin &amp; Moore (1987)</PublishedIn>
<PublishedIn ref=”123” />
```

If a @ref attribute is present but there is no @linkType attribute the reference is presumed to be the id of an element within the current document. In the case of the <PublishedIn> element this would be a <Publication> element. In the case of a <Name> element it would be a <NameObject>.

If the reference has text content then the text is taken to be a summary of the content of the element referenced. By convention this should be the same as the content of the <Simple> element in the target element if it has one or if the reference is pointing to a <TaxonConcept> it should be the name plus the content of the <Simple> element from the <AccordingTo> element.

12.1.3 External reference

```
<PublishedIn linkType=”external” ref=”123”>
      Clapham, Tutin &amp; Moore (1987)
</PublishedIn>
<PublishedIn linkType=”external” ref=”123” />
```

If there is a @linkType attribute present in the element and it has a value of anything other than 'local' then the reference is to a resource outside the current document. This is what the possible values of @linkType mean:

- **external** = A link to a resource in another document. It is presumed that the target will be returned as a valid TCS instance.
- **local** = This is the id of another element within this document. The default.
- **other** = A link to another resource external to the current document that is not in TCS format. e.g. a jpeg, html page, pdf, DarwinCore or ABCD.

If the element has text content then it is taken as having the same meaning as above i.e. It is a summary of the resource pointed at.

12.1.4 Simple instance document design

By using a value of 'external' in @linkType it is possible to avoid ever publishing more than a single type of information within a TCS document i.e. Only one of <Publications>, <Vouchers>, <NameObjects> or <TaxonConcepts>. All other element types can be referred to using external references with text summaries and can be fetched only if required by the consuming application. This method has it's drawbacks but maybe a useful approach for some data providers.

12.2 {PlaceholderType} Complex Type

Like the {ReferenceType} complex type {PlaceholderType} is integral to the way TCS works.

The schema's main role is to communicate the existence of taxon concepts and how they are related. It does not contain the ability to actually describe concepts beyond listing specimens and other concepts. Should character, text or other descriptive data need to be communicated then TCS allows for instances of foreign schemas to be embedded within the <TaxonConcept><CharacterCircumscription> element. Neither does TCS provide detailed break downs of meta data or voucher specimens but allows for foreign schemas to be included. There is also an element to allow this to occur within <Publication>, <Voucher> and <Metadata>.

The {PlaceholderType} complex type provides the mechanism for inclusion of these external schemas. It is a wrapper around the XMLSchema 'any' element type and 'anyAttribute' attribute type. The example below shows the use of the <PublicationDetailed> element to pass a piece of valid XML in the TCS publication format. Clearly the provider of the data has to be sure that the consumer will know how to handle data in the schema included.

```
<Publication id="123">
      <Simple>Clapham, Tutin &amp; Moore (1989) Flora of the British Isles</Simple>
      <PublicationDetailed >
            <Publication
            xmlns="http://www.tdwg.org/schemas/tcs/publication/1.00"
            xsi:schemaLocation="http://www.tdwg.org/schemas/tcs/publication/1.00
            publication.xsd"
            >
                  <Author>Clapham, A.R., Tutin, T.G. &amp; Moore, D.M.</Author>
                  <Title>Flora of the British Isles</Title>
                  <Edition>3</Edition>
                  <DatePublished>1989</DatePublished>
            </Publication>
      </PublicationDetailed>
</Publication>
```

12.3 MetaData

This is a container for meta data that applies to the whole of the current document. There is a {PlaceHolderType} element called <MetadataDetailed> that allows for arbitrary extension of the metadata. A schema is supplied for use in this place holder. It's name space is http://www.tdwg.org/schemas/tcs/metadata/1.00 but other schemas could be used here.

12.4 Publications

This is a container for <Publication> elements. Each <Publication> has a unique @id attribute that allows it to be referenced from elsewhere in the document. References to publications may occur in multiple places in <TaxonConcept> and <NameObject> elements. <Publication> elements can be extended using the <PublicationDetails> element which is of {PlaceholderType}. A schema is supplied for use in this place holder. It's name space is http://www.tdwg.org/schemas/tcs/publication/1.00 but other schemas could be used at this point.

12.5 Vouchers

This is a container for <Voucher> elements. Each <Voucher> represents a specimen or illustration. <Vouchers> have unique @id attributes and are referenced from the <Typification> section of the <NameObject> and from the <SpecimenCircumscription> section of <TaxonConcept>. <Voucher> can be extended using the <Specimen> element that extends {PlaceholderType}.

12.6 NameObjects

This is a container for <NameObject> elements. <NameObject> is a data structure for holding nomenclatural data. Each <NameObject> has a unique @id attribute. <NameObject> elements have references to <Publication>, <Voucher> and other <NameObject> elements but not <TaxonConcept> elements.

12.7 TaxonConcepts

A container for <TaxonConcept> elements. These elements contain the description of taxon concepts only. These descriptions may include lists of specimens, links to other concepts and character data. Each <TaxonConcept> has a unique @id attribute and may have a reference to a single <NameObject> element and multiple <Voucher>, <Publication> and <TaxonConcept> elements.

12.8 RelationshipAssertions

A container for <RelationshipAssertion> elements. These elements are used to express relationships between concepts without actually presenting new concepts. If, for example, a data provider wanted to arrange an arbitrary set of published concepts according to a taxonomic hierarchy for a regional check list but didn't want to publish a new classification then they would use relationship assertions. Each <RelationshipAssertion> has a unique @id attribute and can point to multiple <TaxonConcepts>

13 How to use the TCS

This section contains some general guidelines on how to approach encoding data in TCS. The next section contains more detailed examples.

13.1 When to use <NameObjects> and when <TaxonConcepts>.

- <TaxonConcept> elements are used to represent real world taxa as published. They are the basic unit of taxonomic data exchange. Generally, whenever a scientific name is used a TaxonConcept is implied. All taxonomic **opinion** can be expressed using <TaxonConcept> elements and the relationships between them.
- <NameObject> elements do not represent taxa. They serve only as abstract nomenclatural data structures that encapsulate the core rules of the different nomenclatural codes. Their purpose is to prevent nomenclatural statements becoming confused with statements about the circumscription of, and relationships between, different taxon concepts.
- No taxonomic opinion can be expressed using <NameObject> elements in TCS.
- As a rule of thumb if you are dealing with anything beyond a type specimen and references to it you are talking about a TaxonConcept of some form.

13.2 Linking between <NameObject> elements.

<NameObject> elements can be linked to each other in various ways to signify different kinds of nomenclatural relationship. Most of these relationships are apparent from the names of the elements involved or are documented in the 'Specific Usage Examples' section below. <NameObject> linking should never be used to express taxonomic relationships only nomenclatural facts.

There are situations where two <NameObject> elements appear to have more than one link joining them. If, for example, a name is conserved against its earlier homonym then it could have a <LaterHomonymOf> link and a <ConservedAgainst> link to the earlier name but **it is highly recommended that two <NameObject> elements should only ever have one link joining them.** In the case of a conserved name the <ConservedAgainst> link is enough. Other relationships are implied. There are several combinations of links between <NameObject> elements that would be contradictory if they occurred together for example <Basionym> and <LaterHomonymOf>. There is no way in XML schema to prevent these links occurring together and there are no combinations of links between two <NameObjects> that say more than is said with a single link. This also reduces the cost of implementing applications that map between internal data structures and TCS documents.

Links between <NameObject> elements always point backwards in time. It may be useful within an application data base to have links pointing forward in time or even both ways but it would only add complexity to a transport schema like TCS to incorporate this redundancy and add to the cost of implementing applications that make use of TCS documents.

13.3 When to use <RelationshipAssertion> and <TaxonConcept> <Relationships>

TCS allows taxonomic experts to express relationships between concepts in two distinct places: (1) as an integral part of the <TaxonConcept> structure, and (2) separate from TaxonConcepts under <RelationshipAssertions>. Sometimes the latter are called "Third-Party-Relationships."

If an expert asserts relationships between two concepts that were both authored at an earlier time, then by default these assessments are allocated to place (2), i.e. outside the <TaxonConcepts> section. Example: *Brachycerinae sec.* Kuschel (1995) “includes” *Brachycerinae sec.* Marvaldi & Morrone (2000), asserted according to Franz (2005; the external third party).

The TCS allows more choices if an expert simultaneously authors concepts AND asserts concept relationships. In that case, he or she has the option of placing the relationships inside the newly authored <TaxonConcepts> OR outside in the <RelationshipAssertions>.

The former solution (1) should be used if the author wishes to make the relationship to another concept an integral part of the new concept definiton. This includes all relationship types not just parent-child relationships. Earlier, external concepts may be used as well to nail down the meaning of a newly published concept. Example: *Ranunculus abortivus* L. *sec.* Kartesz (2004) “is congruent to” *Ranunculus abortivus* L. sec. FNA (1997), where Kartesz (2004) places this relationship into the <TaxonConcept> definition to indicate that the newly published 2004 concept is defined by its relationship to the FNA 1997 concept.

The latter solution (2) should be used if the author does not think of the relationship as an integral part of his or her new concept definition. Perhaps other data (diagnosis, included concepts, etc.) are sufficient to specify the present meaning. Or the relationship to earlier concepts is not so clear as to nail down the new definition exactly how the author wants it to be. Example: *Equisetum hyemale* L. ssp. *affine* (Engelmann) Calder & R.L. Taylor *sec.* Weakley (2005) “includes” *E. hyemale* var *affine* (Engelmann) A.A. Eaton *sec.* Radford, Ahles & Bell (1968), as asserted by Weakley (2005). Here the latter author does not wish to define his new concept via its (inclusive) relationship to the 1968 concept. Rather, Weakley intends to provide readers with a "guide" to understanding the taxonomic legacy. The precise definition of his new concept lies in the diagnosis and explanatory comments.

No matter what option an author chooses at the time of authoring a concept, the possibility of authoring another (now by default) third-party relationship at a later time remains. Such a reassessment would "coexist" with the earlier relationship.

As a general principle it is best to avoid asserting multiple relationships between nominal concepts. This provides little more information than linking a series of names. In cases where this is occurring it may be worth considering whether a the data source should actually be publishing their own concepts – perhaps versioned by date. Data consumers can then make use of these concepts.

13.4 Verbatim representation of original publications

TCS does not attempt the verbatim representation of the contents of taxonomic sources. This role is fulfilled by other standards such as TaXMLit. TCS represents the taxa and names in those taxonomic sources in a way that can be handled intelligently by different software applications.

14 Specific usage examples.

14.1 Basionyms and New Combinations - ICZN vs ICBN

One of the most notable differences between ICBN and ICZN is the way names are cited when a species is placed in a different genus from the one it was originally published in (a *comb. nov.*). Botanists have a convention of always citing the authors of the original combination in brackets followed by the names of the authors of the combination. Zoologists don't follow this convention, they simply place the author of the original combination in brackets for the new combination and don't cite the authors who were first to make the new combination. This difference is cosmetic. Indeed ICZN Recommendation 51G is that new combinations in zoology should be quoted in a similar way to the way they are quoted in botany.

'Basionym' is a well established term in the ICBN where it is defined as “name-bringing or epithet-bringing synonym” (ICBN Art. 33.3). It is the orginal combination of a name as viewed from a new combination. The basionym is therefore always relative to a new combination. A name can't be a basionym in it's own right only relative to another name. ICZN does not mention the term basionym but the notion is clearly present in zoological nomenclature as zoologists also have the concept of the new recombination of a name. TCS therefore follows the botanical terminology and allows <Basionym> links between names and the specification of <BasionymAuthors> in recombined names.

14.2 Classifications (Taxonomic Hierarchies)

The building of hierarchical classifications involves the linking of taxon concepts together with “is child of” and “has child” relationships. Taxonomic hierarchies reflect taxonomic opinion and so are expressed with <TaxonConcept> elements not <NameObject> elements. There are, however, links between <NameObject> elements at different ranks and this may lead to confusion. The example below illustrates the differences between the two kinds of linking.

This example shows the placement of *Dianthus gratianopolitanus* within the genus *Dianthus* in the family Caryophyllaceae. This hierarchy is indicated by the linking between <TaxonConcept> elements with Ids 986, 987 and 988. Note that only two of these concepts have links to <NameObject> elements. The family was considered well known and so a break down of the scientific name has not been given. It is represented by a nominal concept. Looking at the <NameObject> for *D. gratianopolitanus* (id = 124) it has a link to the <NameObject> for *Dianthus* (id = 123). This is present only to indicate the construction of the name and no taxonomic conclusions should be drawn from this link. Looking at the <NameObject> for *D. caesius* it can be seen that this name also has a link to the <NameObject> for the genus but its associated <TaxonConcept> is of @type nominal and has no upward placement in a higher genus. We are not recognising this as a taxon concept that is part of the classification we are presenting. It is merely a synonym but its <NameObject> is still linked to the genus <NameObject> because that is an attribute of its name.

```
<NameObjects>

	<NameObject id="123" nomenclaturalCode="Botanical">
		<Simple>Dianthus</Simple>
		<Rank code="gen">genus</Rank>
	</NameObject>

	<NameObject id="124" nomenclaturalCode="Botanical">
		<Simple>Dianthus gratianopolitanus Vill.</Simple>
		<Rank code="sp">species</Rank>
		<CanonicalName>
			<Simple>Dianthus gratianopolitanus</Simple>
			<Genus ref="123">Dianthus</Genus>
			<SpecificEpithet>gratianopolitanus</SpecificEpithet>
		</CanonicalName>
	</NameObject>

	<NameObject id="125" nomenclaturalCode="Botanical">
		<Simple>Dianthus caesius Sm.</Simple>
		<CanonicalName>
			<Simple>Dianthus caesius</Simple>
			<Genus ref="123">Dianthus</Genus>
			<SpecificEpithet>caesius</SpecificEpithet>
		</CanonicalName>
	</NameObject>

</NameObjects>

<TaxonConcepts>

	<TaxonConcept type="nominal" id="986">
		<Name scientific="true">Caryophyllaceae</Name>
		<Rank code="fam">family</Rank>
	</TaxonConcept>

	<TaxonConcept id="987">
		<Name scientific="true" ref="123">Dianthus L.</Name>
		<AccordingTo>
			<AccordingToSimple>
				Clapham, Tutin &amp; Moore (1987)
			</AccordingToSimple>
		</AccordingTo>
		<Relationships>
			<Relationship type="is child of">
				<ToTaxonConcept ref="986"/>
			</Relationship>
			<Relationship type="is parent of">
				<ToTaxonConcept ref="988"/>
			</Relationship>
		</Relationships>
		<CharacterCircumscription>...</CharacterCircumscription>
	</TaxonConcept>

	<TaxonConcept id="988">
		<Name scientific="true" ref="124">Dianthus gratianopolitanus Vill.</Name>
		<AccordingTo>
```

```
                    <AccordingToSimple>
                          Clapham, Tutin &amp; Moore (1987)
                    </AccordingToSimple>
             </AccordingTo>
             <Relationships>
                    <Relationship type="is child of">
                          <ToTaxonConcept ref="987"/>
                    </Relationship>
                    <Relationship type="has synonym">
                          <ToTaxonConcept ref="989"/>
                    </Relationship>
                    <Relationship type="has vernacular">
                          <ToTaxonConcept ref="990"/>
                    </Relationship>
             </Relationships>
             <CharacterCircumscription>...</CharacterCircumscription>
      </TaxonConcept>

      <TaxonConcept type="nominal" id="989">
             <Name scientific="true" ref="125">Dianthus caesius</Name>
      </TaxonConcept>

      <TaxonConcept type="nominal" id="990">
             <Name scientific="false" language="en">Cheddar Pink</Name>
      </TaxonConcept>

</TaxonConcepts>
```

For completeness the vernacular name mentioned in the concept definition is also included.

14.3 Homonyms

Strictly speaking homonyms are two names that are spelled exactly alike. The ICBN and ICZN do, however, have rules that say that names that are spelled so similarly that they could be confused should also be considered homonyms (ICBN Art. 53.3 & ICZN Art. 58). ICZN Art 52 deals with zoological homonyms. For the purposes of TCS homonyms can be divided into four kinds: Real, Correction, Publication and Combination. It is hoped that the different approaches taken by different workers under the different codes can all be accommodated using this approach.

14.3.1 Type I - 'Real' Homonyms

These occur when two or more authors use the same name for different taxa (taxon concepts) and base their versions of the name on separate types (Occasionally a single author might publish two identical names base on different types). These are termed 'real' because using the name without the author string can lead to real confusion. From the name alone it is not possible to tell which type is being referred to and so which set of taxon concepts. *Pedicularis inconspicua* is an example of a real homonym (see below under publication homonyms)

Real homonyms should be marked up in TCS as two separate <NameObjects>. The later homonym should have a <LaterHomonymOf> reference to the earlier one. If the later homonym is conserved under ICBN Art. 14 or suppressed under ICZN Art. 81 then it should have a <ConservedAgainst> reference to the earlier name but **not** a <LaterHomonymOf> link. <TaxonConcept> elements should then be constructed that link to the name actually used in the concept.

'Real' homonyms are characterised by having different types and usually different authors.

14.3.2 Type II -'Correction' Homonyms

Sometimes an author makes an error when publishing a name so the name is not validly published. The name may still slip into common usage though. Later the same author or another may publish the name correctly. In botany this is usually shown by using *ex* in the author string although this isn't required.

An example from the ICBN Art. 15 - Seemann (1865) published *Gossypium tomentosum* "Nutt. mss.", followed by a validating description not ascribed to Nuttall; the name may be cited as *G. tomentosum* Nutt. ex Seem. or *G. tomentosum* Seem.

Correction homonyms should be marked up in TCS as two separate names. The validating name should have a <BasedOn> link to the incorrectly published name. The validating name should **not** have a <LaterHomonymOf> link to the incorrectly published name. Published taxon concepts should then be linked to the version of the name that they actually use. It is not necessary to produce a <NameObject> for the incorrectly published name unless it is used by a <TaxonConcept> element.

```
<NameObject id="123" nomenclaturalCode="Botanical">
    <Simple>Gossypium tomentosum Nutt. ex Seem.</Simple>
    <CanonicalAuthorship>
        <Simple>Nutt. ex Seem.</Simple>
        <Authorship>
            <Simple>Nutt. ex Seem.</Simple>
            <Authors>
                <Name>Nuttall</Name>
                <Name>ex</Name>
                <Name>Seem</Name>
            </Authors>
        </Authorship>
    </CanonicalAuthorship>
    <BasedOn>
        <RelatedName ref="124"/>
    </BasedOn>
</NameObject>
<NameObject id="124" nomenclaturalCode="Botanical">
    <Simple>Gossypium tomentosum Nutt.</Simple>
</NameObject>
```

Note that, as with the authorship of sanctioned names (see below) the “ex” in the author strings is treated as if it were an author.

Correction homonyms are characterised by one of the names not being validly published. They are usually both based on the same type if a type was specified in the original publication.

14.3.3 Type III - 'Publication' homonyms

Publication homonyms arise when an error has occurred in the order in which publications appear or there is confusion over publications in different languages.

The situation occurs in zoology in past when an author has written a paper (1) describing a new species. The same author then writes a second paper (2) and mentions the species. The author assumes that paper 1 will be published before paper 2 but in fact paper 2 is comes out before paper 1. Most workers catalogue the new taxa and name from paper 2 (even though in that paper the taxon and name are not marked as new) and simply ignore paper 1. In most cases, there are no real taxonomic consequences as the name in 1 is merely a later homonym of the name in 2 which is an objective (homotypic) synonym because the types are the same. However there may be cases where the type series are different. That is, if the second reference does not include a reference to the first, but it appear first then the type series is determined by the material mentioned in the “wrong” paper. The ICZN was changed in the 4th edition to prevent this. To be available after

1999, taxonomic names must be clearly marked as new and must include type designations. (Christian Thompson pers. com. 2005).

An example of a publication homonym and a real homonym from botany is

1. *Pedicularis inconspicua* P.C. Tsoong, Acta Phytotax. Sin. 3: 292 & 323, Jan. 1955
2. *Pedicularis inconspicua* Vved., Fl. URSS 22: 811, 18 Jun. 1955.
3. *Pedicularis inconspicua* P.C. Tsoong, Bull. Brit. Mus. (Nat. Hist.) 2:17, Nov. 1955.

Names 1 and 3 relate to the same taxon from Bhutan and represent double publication of the same name in Chinese and Western journals. This is one of many such names published in the same two papers by Tsoong; all the names in Acta Phytotax. Sin. have priority over their publication in the British Museum journal. Between the two papers published by Tsoong, the Russian botanist Vvedensky published a real homonym *P. inconspicua* Vved. for a totally different species from Uzbekistan. (Robert Mill 2005 pers. comm.)

To express publication homonyms in TCS create two <NameObjects> and link the later one to the earlier one with a <LaterHomonymOf> link.

```
<NameObject id="123" nomenclaturalCode="Botanical">
    <Simple>Pedicularis inconspicua P.C. Tsoong</Simple>
    <PublishedIn>Acta Phytotax. Sin. 3: 292 &amp; 323, Jan. 1955</PublishedIn>
</NameObject>

<NameObject id="124" nomenclaturalCode="Botanical">
    <Simple>Pedicularis inconspicua Vved.</Simple>
    <PublishedIn>Fl. URSS 22: 811, 18 Jun. 1955.</PublishedIn>
    <LaterHomonymOf>
        <RelatedName ref="123"/>
    </LaterHomonymOf>
</NameObject>

<NameObject id="125" nomenclaturalCode="Botanical">
    <Simple>Pedicularis inconspicua P.C. Tsoong</Simple>
    <PublishedIn>Bull. Brit. Mus. (Nat. Hist.) 2:17, Nov. 1955</PublishedIn>
    <LaterHomonymOf>
        <RelatedName ref="123"/>
    </LaterHomonymOf>
</NameObject>
```

Publication homonyms are characterised by having the same authors and same types.

14.3.4 Type IV - 'Combination' Homonyms

Sometimes a new combination is made more than once thus creating a homonym. An example is *Trillium texanum* Buckley which was recombined as a variety of *Trillium pusillum* twice (Susan Farmer 2005, pers. comm.):

1. *Trillium pusillum* var. *texanum* (Buckley) J.L.Reveal & C.R.Broome in Castanea, 46(1): 56 (1981)
2. *Trillium pusillum* var. *texanum* (Buckley) C.F.Reed in Phytologia, 50(4): 279, 283 (1982)

The combination made by C.F. Reed is a later homonym and so invalid. In TCS these two names should be represented by two <NameObjects>. The C.F. Reed name should have a <LaterHomonymOf> link to the earlier combinations. If a <NameObject> is presented for *Trillium texanum* then both combinations should have a <Basionym> link to it. <TaxonConcept> elements should be created to represent the different concepts the two different authors have of the names (not shown in the example below).

```
<NameObject id="123" nomenclaturalCode="Botanical">
```

```
    <Simple>Trillium texanum Buckley</Simple>
    <PublishedIn>Proc. Acad. Nat. Sci. Philadelphia 1860: 443. 1861</PublishedIn>
</NameObject>

<NameObject id="124" nomenclaturalCode="Botanical">
  <Simple>Trillium pusillum Michx. var. texanum J.L.Reveal &amp; C.R.Broome</Simple>
  <PublishedIn>Castanea, 46(1): 56 (1981)</PublishedIn>
  <Basionym>
      <RelatedName ref="123"/>
   </Basionym>
</NameObject>

<NameObject id="125" nomenclaturalCode="Botanical">
    <Simple>Trillium pusillum Michx. var. texanum (Buckley) C.F.Reed</Simple>
    <PublishedIn>Phytologia, 50(4): 279, 283 (1982)</PublishedIn>
    <Basionym>
        <RelatedName ref="123"/>
    </Basionym>
    <LaterHomonymOf>
        <RelatedName ref="124"/>
    </LaterHomonymOf>
</NameObject>
```

If a name collision occurs as a result of two taxa with the same epithet but based on different types both being recombined into the same genus then the homonym is a Type I homonym rather than a Type IV and should be treated as such.

Combination homonyms are characterised by having the same type and therefore basionym but different publications and combination authors (in botany).

14.3.5 Junior and Senior homonyms

The ICZN uses the terms 'Junior homonym' and 'Senior homonym'. These are concisely defined in the ICZN glossary:

- **Junior Homonym**: “Of two homonyms: the later established, or in the case of simultaneous establishment the one not given precedence under Article 24.”
- **Senior Homonym:** “Of two homonyms: the first established, or in the case of simultaneous establishment the one given precedence under Article 24.”

A decision has been taken to avoid the use of these terms in favour of earlier and later homonyms. This is for two reasons:

1. If the senior homonym is suppressed then some workers refer to the name published before as the junior homonym and the one published after as the senior homonym. This is incorrect but does occur and does cause confusion.
2. It is clearer, within the context of TCS, to consistently use the terms “Later Homonym” and “Earlier Homonym” where earlier is always chronologically before and later is always chronologically after. When acts of conservation or suppression have occurred then the terms “Conserved Later Homonym” and “Rejected Earlier Homonym” should be used. These also reflect the relationships between the <NameObjects> in the TCS more clearly.

The disadvantage of taking this approach is that it does not take account of the 'First Reviser' rule (ICZN Art. 24.2). In cases of simultaneous publication then the first author to used the name acts as the 'First Reviser' and resolves the dispute. These cases should be marked up in TCS using the <ConservedAgainst> link in <NameObject> and entering the appropriate rule in <RuleConsidered> and publication in <PublishedIn>.

14.3.6 Primary and Secondary homonyms

The ICZN uses the terms 'Primary homonym' and 'Secondary homonym'. These are concisely defined in the ICZN glossary:

- **Primary Homonym**: “Each of two or more identical specific or subspecific names established for different nominal taxa and originally combined with the same generic name [Art. 57.2]. For variant spellings deemed to be identical see Article 58.”
- **Secondary Homonym: “**Each of two or more identical specific or subspecific names established for different nominal taxa and originally combined with different generic names but subsequently combined with the same generic name [Art. 57.3]. For variant spellings deemed to be identical see Article 58.”

From these definitions it can be seen that primary and secondary homonyms fall within the definition of Type I homonyms described above (as they are based on different types) but are restricted to taxa below the rank of genus. Whether two names are fall into these categories can be told from their context and so there is no special feature in TCS to identify them.

14.3.7 Earlier Homonyms of Sanctioned Names

The ICBN has the notion of sanctioned names for some fungal names. These are dealt with separately below under Sanctioned Names.

14.4 Synonyms

From the point of view of TCS there are four basic types of synonym

14.4.1 Type I Synonyms - Objective/Homotypic

Homotypic synonyms are two names that share the same type. Under ICZN they are termed 'Objective Synonyms'. The ICZN glossary defines the term 'Objective Synonym' as:

> “*Each of two or more synonyms that denote nominal taxa with the same name-bearing type, or (in the cases of family-group and genus-group taxa) that denote nominal taxa with name-bearing types whose own names are themselves objectively synonymous.*”

Within TCS homotypic synonyms are handled by links between <NameObjects> unless <NameObjects> are not being used in which case they can be treated in the same way as heterotypic synonyms (see below). The <NameObject> links that imply homotypic synonyms are <Basionym>, <BasedOn>, <SpellingCorrectionOf> and <NomenNovumFor>.

```
<NameObject id="325" nomenclaturalCode="Botanical">
    <Simple>Fallopia dumetorum (L.) J. Holub</Simple>
    <Basionym>
        <RelatedName ref="326"/>
    </Basionym>
</NameObject>

<NameObject id="326" nomenclaturalCode="Botanical">
    <Simple>Polygonum dumetorum L.</Simple>
</NameObject>
```

14.4.2 Type II Synonyms - Subjective/Heterotypic

Frequently authors of taxon concepts provide a list of synonyms in addition to the homotypic synonyms dealt with above. These lists are usually just lists of names and not references to concepts with a '*sensu*', or 'according to' authority (such synonyms are dealt with in the next section). These are heterotypic synonyms - names based on different types. Under ICZN they are termed 'Subjective Synonyms' and defined in the glossary as:

> *“Each of two or more names whose synonymy is only a matter of individual opinion, i.e. it is not objective. See also Article 61.3.1.”*

In order to appreciate how subjective/heterotypic synonyms should be handled in TCS it is worth examining what such a statement of synonymy could mean in terms of taxon concepts.

1. The author includes only the type specimen of the synonymous name in their definition of the taxon. This is the most literal interpretation of what is presented but is unlikely to convey what the author actually meant by making the statement. The author will have known of other concepts associated with the name and is implying that they are included in the concept being described. Certainly this is how many workers interpret these synonyms.
2. The author has their own concept of a taxon associated with the synonymous name but includes all the members of it within their circumscription of the concept being published. This is theoretically appealing but actually of little use as the author does not describe the concept associated with the synonymous name and from their action does not accept it as being a 'good' taxon worthy of a proper name.
3. The author accepts some one else's concept associated with this name but hasn't indicated whose. As with 2 above this is appealing but of little use.

Bearing the above discussion in mind the way subjective/heterotypic synonyms of this type should be handled in TCS is that the <TaxonConcept> for the concept being described should have a <Relationship> reference of @type 'has synonym' to the nominal concept for the synonymous name. This can only be interpreted as the accepted taxon concept including the concept for a synonymous name whose description can never be known.

14.4.3 Type III Synonyms – Concept relationships

Sometimes authors of taxon concepts provide a list of synonyms that include some names that are *sensu* or *secundum* an authority. In these cases the authors are giving some indication as to the concept associated with the synonym. In such cases a separate <TaxonConcept> should be created to represent the concept defined by the synonym and the accepted concept should have a <Relationship> reference of type 'includes' to the concept for the synonym.

14.4.4 Type IV Synonyms - Pro parte relationships.

If a synonym is cited as being *pro. parte* (in part) then the <TaxonConcept> being defined should have a <Relationship> of type 'overlaps' to the nominal concept for the synonymous name unless, of course, the synonymous name has a *sensu/secundum* (as in the Misapplied Names example belowMisspell) in which case the <Relationship> should point at a proper concept for the synonymous name according to the cited authorities.

Some workers use *pro. parte* in a stricter sense to only refer to taxa where type specimen is now thought to be a mixed gathering of more than one taxon. This is by far the less common usage of the term. It can be expressed in TCS by the associated lectotypification event that is required to split the type (ICBN Art. 9.9).

14.5 Typification

Typification is a nomenclatural process and is dealt with entirely within the <NameObject>. Although it is common to talk in terms of the type of a species or the type of a genus this is not entirely accurate as the circumscription of a taxon beyond its type specimen has no effect on its typification unless, of course, it should change to include other types in which case the taxon may have a different name – but the name would never have a different type.

There is a <Typification> element in the NameObject element to express matters concerning the typification of a the name. The <Typification> element can contain either a <TypeName> element or a series of <TypeVoucher> elements. This is to allow for the different rules governing names at and below species from those above species level.

14.5.1 Names at and below species level

Names at species level and below are typified by vouchers specimens. Here the term specimen is used loosely as illustrations can also serve as vouchers. Strictly only a single type specimen is of importance – the holotype - but it is customary to list other specimens such as isotypes, paratypes and so the schema allows for this. (See also Lectotypification below)

14.5.2 Names above species level.

Names above the species level are typified by other names. Although nomenclatural codes refer to names being typified by taxa none of the rules are affected by the circumscription of the taxa beyond the type so taxa in this context means name. The <TypeName> element holds a reference to another <NameObject> for this purpose.

14.5.3 Lectotypification

Elements are provided within the <Typification> element to indicate if the typification is the result of a later publication.

14.6 Misapplied Names

Misapplied names occur when there is an accepted taxon which has a valid name but this name has been consistently and erroneously used for some other taxon or part of another taxon. There are three or more concepts involved in such situations.

1. The accepted taxon that the name applies to.
2. The rejected concept to which the name is erroneously applied by some other group of workers.
3. The accepted concept or concepts to which specimens and observations in concept 2 above should belong.

The schema does not provide a specific mechanism for flagging misapplied names because, in terms of TaxonConcepts, they are not special cases. Concept 2 above is merely a concept that is not accepted. Here is an example:

1. *Acarospora discreta* (Ach.) Arnold Is the name of an accepted taxon of lichen according to Index Fungorum.
2. The concept of *Acarospora discreta* (Ach.) Arnold as used by British authors (sensu auct. brit.) is rejected by Index Fungorum.
3. *Acarospora veronensis* A. Massal. Is the name of an accepted taxon of lichen to which specimens in the concept in 2 above should be placed i.e. *A. veronensis* has a 'contains' relationship to A. discreta sensu auct. brit.

This is the most simple example of the misapplication of a name. A more complex example is given by *Acrospora admissa* (Nyl.) Kullh. Which is the name of an accepted taxon but which has been used *sensu auct brit.* for specimens that should be placed in three other taxon concepts, *A.*

veronensis A. Massal., *A. badiofusca* (Nyl.) Th. Fr and *A. impressula* Th. Fr. The *pro parte* relationships between *Acrospora admissa* (Nyl.) Kullh. *sensu auct. brit.* and the three accepted taxon concepts can be expressed in terms of 'overlaps' relationships within the schema.

In summary a “misapplied name” is simply a rejected taxon concept that bears the same name as an accepted concept but can be differentiated by the contents of the <AccordingTo> element.

14.7 Misspelled Names

Sometimes names are misspelled. By misspelling here we mean orthographic and typographic variants of all forms including having the wrong gender for an epithet. This can happen for a number of reasons. The original author of a name can spell it incorrectly and the name can then be corrected under the code (e.g. see ICBN Article 60 for examples.), authors of revision concepts can make mistakes in interpreting the code or typographically errors can be made. In TCS if a <NameObject> is known to represent the corrected spelling of value in another <NameObject> then a <SpellingCorrectionOf> link can be created from the correct spelling. A correctly spelled <NameObject> can link to multiple incorrectly spelled <NameObjects>.

```
<NameObject id="225" nomenclaturalCode="Botanical">
  <Simple>Persicaria segetum (Kunth) Small (1903)</Simple>
  <SpellingCorrectionOf>
      <RuleConsidered>23.5</RuleConsidered>
       <Note>Correction of epithet gender</Note>
      <RelatedName ref="226">Persicaria segeta (Kunth) Small (1903)</RelatedName>
  </SpellingCorrectionOf>
</NameObject>

<NameObject id="226" nomenclaturalCode="Botanical">
  <Simple>Persicaria segeta (Kunth) Small (1903)</Simple>
</NameObject>
```

The <RuleConsidered> and <Note> elements are not required but included here for illustrative purposes.

Note that name spelling here refers only to the spelling of the uninomial, binomial or trinomial name not to the author name string or any other part of the <NameObject>.

14.8 Vernacular Names

TCS allows the mark up of vernacular names. Vernacular names are dealt with entirely within the <TaxonConcepts> part of the schema. The <NameObjects> area of the schema is only for marking up names that are governed by one of the nomenclatural codes. Vernacular names are stored in the <Name> element within the <TaxonConcept> element. The fact that the <Name> contains a vernacular name should be indicated by setting the @scientific attribute to false. There is an optional @language attribute to indicate which language the name is in. Two usages patterns are envisaged.

14.8.1 Vernacular names mentioned in concept definitions

Authors of taxon concepts for scientific names will often provide one or more vernacular names associated with the concept. From the point of view of TCS this is logically the same as providing heterotypic synonyms as discussed above. We can not be sure what is meant by the author who supplies the vernacular name and we have no description for the concept that is associated with that vernacular name. A nominal <TaxonConcept> for the vernacular name should therefore be created and the <TaxonConcept> containing the circumscription should have a <Relationship> of @type 'has vernacular' to it. This is envisaged as being by far the most common way of marking up vernacular names.

```
<TaxonConcept id="123">
    <Name scientific="true" ref="345">Lactuca sativa L.</Name>
    <AccordingTo>
        <AccordingToSimple>Clapham, A.R., Tutin, T.G. &amp; Moore, D.M. (1987)
</AccordingToSimple>
    </AccordingTo>
    <Relationships>
        <Relationship type="has vernacular">
            <ToTaxonConcept ref="456"/>
        </Relationship>
    </Relationships>
    <CharacterCircumscription>
        ...
    </CharacterCircumscription>
</TaxonConcept>

<TaxonConcept id="456">
    <Name scientific="false" language="en">Garden Lettuce</Name>
</TaxonConcept>
```

14.8.2 Vernacular concepts

These are envisaged as being a far rarer usage pattern and is only likely to be used in studies with a particular focus on vernacular name usage. If an author has a specific definition of a vernacular name then a <TaxonConcept> should be created to hold this concept description. If, as part of the description, the author mentions scientific names without indicating which concepts of those names are referred to then the <TaxonConcept> element for the vernacular should have a <Relationship> of @type 'is vernacular for' to a nominal <TaxonConcept> for the scientific name.

```
<TaxonConcept id="123">
      <Name scientific="false" language="en">Prickley Lettuce</Name>
      <AccordingTo>
            <AccordingToSimple>Hyam, R.D. (1999)</AccordingToSimple>
      </AccordingTo>
      <Relationships>
            <Relationship type="is vernacular for">
                  <ToTaxonConcept ref="124"/>
            </Relationship>
      </Relationships>
      <CharacterCircumscription>
            ...
      </CharacterCircumscription>
</TaxonConcept>

<TaxonConcept id="124">
      <Name scientific="true" ref="987">Lactuca serriola L.</Name>
</TaxonConcept>
```

14.9 Nomen Novum/Replacement Names

Sometimes authors wish to recognise taxa whose names are homonyms. In such cases both the ICBN (Art. 7.3) and the ICZN allow for *nomen novum* or replacement name to be created. The ICZN glossary defines *nomen novum* as:

> *"A name established expressly to replace an already established name. A nominal taxon denoted by a new replacement name (nomen novum) has the same name-bearing type as the nominal taxon denoted by the replaced name [Arts. 67.8, 72.7]."*

In TCS a *nomen novum* <NameObject> should have a <NomenNovumFor> link to the illegitimate homonym name it replaces. The illegitimate homonym should have a <LaterHomonymOf> link to the legitimate homonym of the name.

```
<NameObject id="123" nomenclaturalCode="Botanical">
  <Simple> Myrcia lucida McVaugh (1969)</Simple>
  <Typification>
      <TypeVoucher typeOfType="holo">
          <VoucherReference>Spruce 3502</VoucherReference>
      </TypeVoucher>
  </Typification>
  <NomenNovumFor>
      <RelatedName ref="124"/>
  </NomenNovumFor>
</NameObject>

<NameObject id="124" nomenclaturalCode="Botanical">
  <Simple>Myrcia laevis O. Berg (1862)</Simple>
  <Typification>
      <TypeVoucher typeOfType="holo">
          <VoucherReference>Spruce 3502</VoucherReference>
      </TypeVoucher>
  </Typification>
  <LaterHomonymOf>
      <RelatedName ref="125"/>
  </LaterHomonymOf>
</NameObject>

<NameObject id="125" nomenclaturalCode="Botanical">
  <Simple>Myrcia laevis G. Don (1832)</Simple>
  <Typification>
      <TypeVoucher typeOfType="holo">
          <VoucherReference>Some other type</VoucherReference>
      </TypeVoucher>
  </Typification>
</NameObject>
```

14.10 Hybrids

In ICBN it is possible for an author to publish a name as a hybrid and another author to come along and revise it as not being a hybrid or as having different parents. None of this affects nomenclature. One simply adds an X (multiplication symbol) if one believes the taxon concept is of hybrid origin the name remains unchanged.

Bearing this in mind the way hybrids are handled in TCS is that the <NameObject> has no notion of hybridisation. This is because a single <NameObject> could be used by two different concepts one of which is a hybrid and the other isn't.

<TaxonConcept> elements do 'care' about hybrids. They signify this by having 'is hybrid child of' relationships with other TCs. There is also a @form attribute in the <TaxonConcept> that can be set to the value 'hybrid' to indicate that a taxon is of hybrid origin when it's parentage is not known. A data consumer can make decisions on how to render a <TaxonConcept> on the basis of whether this flag is present or a relationship of the appropriate type. It is also possible for a data supplier to include an X or multiplication sign within the <TaxonConcept><Name> element.

Hybrids are excluded from the ICZN under Art. 1.3.1.

14.11 Autonyms

Autonyms are names that are automatically created under the ICBN to refer to the part of a taxon that is not included in a subdivision of that taxon at a particular rank. Autonyms are only created below the rank of genus. See ICBN Articles 22 and 26. ICZN does not have the notion of autonyms.

The schema does not take a position on autonyms. Should a data supplier wish to provide TaxonConcepts that circumscribe autonyms then they can although some workers would question whether this is theoretically possible.

If need be NameObjects can be created that provide the autonym name in canonical form with details of the publication which triggered their creation. (See ICBN Article 32.6).

14.12 Anamorphs and Teleomorphs

The ICBN allows for the naming of mitotic, asexual morphs of fungi (ICBN Art 59.). Some definitions may be useful at this point.

- **Anamorphic Taxon**: A taxon containing only the asexual part of the life cycle of a fungus. Represented in TCS by a <TaxonConcept>.
- **Teleomorph Taxon**: A taxon containing the sexual part of the life cycle of a fungus. Represented in TCS by a <TaxonConcept>.
- **Holomorphic Taxon**: A taxon containing both sexual and asexual forms of a fungus. Represented in TCS by a <TaxonConcept>.
- **Anamorphic Name:** A name based on an asexual fungal type. These names can only be used for anamorphic taxa because they are always of lower precedence than teleomorphic names when they occur in the same holomorphic taxon (Art.59.4). Represented by a <NameObject> in TCS. There is an attributed @isAnamorphic in the <NameObject> that should be set to “true” to indicate that the type of the name is asexual.
- **Teleomorphic Name:** A name based on a sexual fungal type. These names can be used for both teleomorphic taxa and holomorphic taxa. Represented by a <NameObject> in TCS. The @isAnamorphic flag could be set to false to indicate that this is a based on fertile type material.

Relationship types are available for both <TaxonConcept> and <RelationshipAssertion> to allow concepts to be related in terms of whether they are anomorphic-teleomorphic forms or not. The @form attribute of <TaxonConcept> can be set to 'anamorphic' to indicate a taxon is anamorphic when it isn't known to be in an 'is anamorph of' relationship with another taxon.

14.13 Sanctioned Names

The names of fungi presented in certain publications have been sanctioned by ICBN (Art. 13.1(d)). This means that these names are treated as if conserved against earlier homonyms and competing synonyms (ICBN Art. 15). The earlier names are not illegitimate as would be the case if these works had been set as the starting date for fungal nomenclature but are available for use in different combinations.

In TCS sanctioned names should be indicated using the <Sanctioned> element. This element can either be left empty or populated with <PublishedIn> and <MicroReference> elements giving the location of the sanctioning event. A sanctioned name may be conserved against more than one other names and so may contain more than one <ConserveredAgainst> elements with references to the other names.

```
<NameObject id="123" nomenclaturalCode="Botanical">
 <Simple>Agaricus personatus Bolton : Fr</Simple>
 <PublishedIn>Hist. fung. Halifax 2: 58 (1788)</PublishedIn>
 <ConservedAgainst><RelatedName ref="124"/></ConservedAgainst>
 <ConservedAgainst><RelatedName ref="125"/></ConservedAgainst>
 <ConservedAgainst><RelatedName ref="126"/></ConservedAgainst>
 <Sanctioned>
     <PublishedIn>Syst. mycol. 1: 126 (1821)</PublishedIn>
 </Sanctioned>
</NameObject>

<NameObject id="124" nomenclaturalCode="Botanical">
  <Simple>Agaricus peronatus Valenti</Simple>
```

```
</NameObject>
<NameObject id="125" nomenclaturalCode="Botanical">
  <Simple>Agaricus peronatus Lasch</Simple>
</NameObject>
<NameObject id="126" nomenclaturalCode="Botanical">
  <Simple>Agaricus peronatus With.</Simple>
</NameObject>
```

14.13.1 <CanonicalAuthorship> in Sanctioned names

The authority strings of sanctioned names contain the authors of the name followed by a colon and the sanctioning authority. This should be represented in <CanonicalAuthorship> elements with the colon as if it were an author name. This is the same method as used with “ex” in botanical names.

15 Notes on Cultivated Plants

The International Code of Nomenclature for Cultivated Plants (ICNCP) can be viewed as an extension of the ICBN. It principally introduces the ranks of cultivar and cultivar Group and the concept of a denomination classes as well as mechanisms for handling graft chimeras.

A cultivar is a taxon consisting of plants of cultivated origin or a selection from a 'wild' taxon brought into cultivation. Cultivar Groups are collections of cultivars. Graft chimeras are combinations of plants in which the phenotypes have become mixed without full genetic hybridisation having occurred. Most of the rules concerning cultivar names also applies to cultivar Groups and chimeras.

All Cultivars and Groups must be registered with the appropriate International Cultivar Registration Authoritity (ICRA). These authorities are appointed through the Commission for Nomenclature and Cultivar Registration of the International Society for Horticultural Science (ISHS). (http://www.ishs.org/icra/).

15.1 Cultivar and Group names in NameObjects

The name of a Cultivar or Group consists of the name of the genus or lower taxonomic unit to which it is assigned together with a cultivar or Group epithet (ICNCP Art. 7.1). The minimum needed to name a cultivar unambiguously is the genus name plus the cultivar epithet so long as the denomination class for the taxa concerned does not cause confusion.

The denomination class is the taxonomic scope a cultivar epithet must be unique within. Denomination classes are typically genera. For example the hybrid epithet 'Buff Beauty' must be unique within the genus *Rosa*. Not all denomination classes are congruent to genera however. Some may consist of several genera. An example would be the more common cereal genera (*Avena*, *Hordeum*, *Secale*, *Triticale* and *Triticum*) which form a single denomination class and so force all cereal cultivar epithets to be unique. Other denomination classes are at lower taxonomic level than genus. Examples would be the common sunflower (*Helianthus annuus*) and Jerusalem artichoke (*Helianthus tuberosus*) which both have their own denomination classes.

The result of the denomination classes and naming rules are that a Cultivar epithet or a cultivar Group epithet or the name of a graft chimera can be attached to the end of a ICBN governed name of genus rank and below. This is accomplished in the schema by the use of the optional <CultivarNameGroup> element and the appropriate ranks in the {TaxonomicRanksEnumeration} simple type.

It should be noted that although the ICNCP stipulates that a cultivar name should be surrounded by single quotes, that a cultivar Group should be followed by the word “Group” and that graft chimeras should be preceded by a "+" sign none of these things should be included within the <Published <CultivarNameGroup > element. The consuming software should render the name appropriately (using the correct language for "Group" etc) on the basis of the value of the @code attribute in the <RANK> element of the <NameObject> element.

15.1.1 Vernacular alternatives to genus names.

Article 19.2 example 2 of the ICNCP shows that it is acceptable to use vernacular equivalents for genera or denomination groups. e.g. Apple 'James Grieve' being the equivalent of *Malus domestica* 'James Grieve'. The schema could pass this form of cultivar name in the <NameObject> by storing the vernacular part of the name in the <Uninomial> element. It is strongly recommend that data for general consumption **not** be published in this way. It assumes a single, language independent, meaning of the vernacular term. It can not be guaranteed that the consumer of the document will understand the vernacular name to be synonymous with the genus/denomination group name. The consumer has no way of telling that the vernacular is not a genus name and so is likely to render the name as a Latin genus name.

A better way to handle vernacular-cultivar names is as regular vernacular names. In the example above a TaxonConcept should be created with the <Name> element containing "Apple 'James Grieve'" and an appropriate @language attribute but no @ref to a NameObject. The relationships of this TaxonConcept to other concepts can then be expressed in the normal way.

15.2 Cultivars and Groups in TaxonConcepts

Cultivars and Groups are like any other taxa in that they can be described and reasoned about in different ways and therefore can be handled using TaxonConcepts although each cultivar should have a standard description. Denomination classes can also be described using TaxonConcepts and there is a rank in the {TaxonomicRankEnumeration} to allow for this.

16 Signatures

Until there is a globally accepted system of GUIDs for taxon concepts it will be necessary for systems to attempt to resolve concept identities on the basis of a combination of names and <AccordingTo> data. Even after GUIDs are available it may be necessary to de-duplicate or synonymise objects on the basis of their contents. For this purpose TCS has the notion of standardised signature fields. These are specified ways of citing existing information in order to help the process of name and concept resolution. The following points should be noted.

- Signature fields do not offer a guarantee of being globally unique. There will be collisions but these collisions should be very rare.
- Signatures are designed to be possible to create from most taxonomic databases and computationally cheap to compare using simple string comparison algorithms. They are functional not ideal.
- The signature fields do not include the name authorship. This is because of the vast range of ways in which author names have been quoted and stored in the past. A data consuming application may well choose to compare objects on the basis of matching signature fields and only then look at the name authorship fields.
- The completion of signature fields as indicated here is optional but their use would significantly enhance the ability of receiving applications to process the data.

The following elements within a TCS instance document are considered to be part of the signature.

16.1 <NameObject> Signature fields.

16.1.1 NameObject/@nomenclaturalCode

This is a controlled vocabulary to indicate which code governs the use of the name.

16.1.2 NameObject/Rank/@code

This is a controlled vocabulary to indicate the rank of the name.

16.1.3 NameObject/CanonicalName/Simple

This field should contain only the words that form the name. It should not contain rank, authorship or any other qualifiers. For scientific names it will contain one, two or three words. For cultivated plant names it may contain more but any special characters, including the quotes round the cultivar epithet, should be omitted.

16.1.4 NameObject/Year

A four digit representation of the year in which this name was published. This will be the same as the date of the publication in the <PublishedIn> element. Note that this is the year the combination was published not the year the basionym was published if this is a *comb nov*. This applies even in zoology where the year of basionym publication is often quoted even when the epithet is combined differently. There are fields in the <CanonicalAuthorship> element for quoting basionym dates correctly for animals.

16.2 <TaxonConcept> Signature fields

16.2.1 TaxonConcept/AccordingTo/AuthorTeam/Simple

This field should contain the authors of the concept. Two situations are envisaged, one where concepts appear in printed publications and the other where concepts are published on-line.

When representing a printed concept the field should contain the unabbreviated surnames of the authors in the order they appear in the publication separated by spaces. Initials and any punctuation marks should be omitted. If there are more than three authors only the first two author names should be included and they should be followed by the words “et al”. The full authorship of the concept will always be available via the <PublishedIn> element. Transliteration of names should be avoided unless they can't be represented in UTF-8 encoding.

If the concept is being published on line, and does not exist in a paper form, then the DNS name of the institution publishing the concept should be used. A policy should be formulated for how many sub-domains should be cited and this should be stuck to. It is recommended that the “www” sub-domain should not be used (e.g. ipni.org not www.ipni.org). These DNS names are not expected to resolve to anything now or in future and so artificial sub-domains could be created to represent publishing authorities within larger organisation if required. If the concept is version sensitive then the DNS name should be followed by a space and then the versioning information.

16.2.2 TaxonConcept/AccordingTo/AuthorTeam/Year

A four digit representation of the year in which the concept was published. If this concept appears in a paper publication then it will be the year of publication. If this concept is being published as part of an on-line data set then a decision will have to be taken as to when the concept was created. If the data is highly dynamic then it may just be the current year and version information will be present in the <AuthorTeam> element – see above.

16.3 Recommended Minimum Fields for Nomenclatural Data Sources.

If you are publishing data that you intend to be nomenclaturally precise then you should include following fields in addition to the signature fields.

16.3.1 NameObject/CanonicalAuthorship/Simple

This should contain the full authorship of the name. If abbreviations are used then an attempt should be made to follow a recognised standard such as Brummitt and Powell (1992) *Authors of Plant Names*.

16.3.2 NameObject/PublishedIn

The text of this element should contain a citation to the place of publication - whether or not the @ref attribute points to a full breakdown of the publication. This is because not all applications will be able to follow the reference or necessarily understand the schema the reference is presented in.

16.3.3 NameObject/Typification/Simple

This should contain a summary of the typification information if it is available.