Biological Names and Taxonomies on the Semantic Web – Managing the Change in Scientific Conception

Jouni Tuominen, Nina Laurenne and Eero Hyvönen

Semantic Computing Research Group (SeCo)
Aalto University School of Science and the University of Helsinki
http://www.seco.tkk.fi/, firstname.lastname@aalto.fi

Abstract. Biodiversity management requires the usage of heterogeneous biological information from multiple sources. Indexing, aggregating, and finding such information is based on names and taxonomic knowledge of organisms. However, taxonomies change in time due to new scientific findings, opinions of authorities, and changes in our conception about life forms. Furthermore, organism names and their meaning change in time, different authorities use different scientific names for the same taxon in different times, and various vernacular names are in use in different languages. This makes data integration and information retrieval difficult without detailed biological information. This paper introduces a meta-ontology for managing the names and taxonomies of organisms, and presents three applications for it: 1) publishing biological species lists as ontology services (ca. 20 taxonomies including more than 80,000 names), 2) collaborative management of the vernacular names of vascular plants (ca. 26,000 taxa), and 3) management of individual scientific name changes based on research results, covering a group of beetles. The applications are based on the databases of the Finnish Museum of Natural History and are used in a living lab environment on the web.

1 Introduction

Exploitation of natural resources, urbanisation, pollution, and climate changes accelerate the extinction of organisms on Earth which has raised a common concern about maintaining biodiversity. For this purpose, management of information about plants and animals is needed, a task requiring an efficient usage of heterogeneous, dynamic biological data from distributed sources, such as observational records, literature, and natural history collections. Central resources in biodiversity management are names and ontological taxonomies of organisms [1, 19, 20, 3, 4]. Animal ontologies are stereotypical examples in the semantic web text books, but in reality semantic web technologies have hardly been applied

to managing the real life taxonomies of biological organisms and biodiversity on the web. This paper tries to fill this gap.[1]

Managing taxonomies of organisms provides new challenges to semantic web ontology research. Firstly, although we know that lions are carnivores, a subclass of mammals that eat other animals, the notion of 'species' in the general case is actually very hard to define precisely. For example, some authors discuss as many as 22 different definitions of the notion of species [16]. Secondly, taxonomic knowledge changes and increases due to new research results. The number of new organism names in biology increases by 25,000 every year as new taxa to science are discovered [11]. At the same time, the rate of changes in existing names has accelerated by the implementation of molecular methods suggesting new positions to organisms in taxonomies. Thirdly, biological names are not stable or reliable identifiers for organisms as they or their meaning change in time. Fourthly, the same name can be used by different authors to refer to different taxa (units of classification that commonly have a rank in the hierarchy), and a taxon can have more than one name without a consensus about the preferred one.

As a result, biological texts are written, content is indexed in databases, and information is searched for using different names and terms from different times and authorities. In biological research, scientific names are used instead of common names, but in many applications vernacular names in different languages are used instead. Data fusion is challenging and information retrieval without deep biological knowledge is difficult.

We argue that a shared system for publishing and managing the scientific and vernacular names and underlying conceptions of organisms and taxonomies is needed. From a research viewpoint, such a system is needed to index research results and to find out whether a potential new species is already known under some name. Biological information needed by environmental authorities cannot be properly indexed, found or aggregated unless the organism names and identifiers are available and can be aligned. For amateur scientists and the public, aligning vernacular names to scientific names and taxonomies is often a prerequisite for successful information retrieval.

This paper presents a meta-ontology and its applications addressing these problems. Our research hypothesis is that semantic web technologies are useful in practise in modelling change in the scientific perception of biological names and taxonomies, for creating a platform for collaboratively managing scientific knowledge about taxonomies, and for publishing taxonomies as ontology services for indexing and information retrieval purposes in legacy systems.

In the following, biological classification systems are first discussed and a meta-ontology TaxMeOn for defining such systems is presented [13]. Three use case applications of the meta-ontology are then discussed: a system for managing vascular plant names collaboratively (26,000 species) based on the SAHA

---

[1] We discuss the taxonomies of contemporary species, not 'phylogenetic trees' that model evolutionary development of species, where humans are successors, e.g., of dinosaurs.

metadata editor [12], application of the ONKI ontology service [25] for publishing taxonomic species lists on the semantic web (over 80,000 taxa of mammals, birds, butterflies, wasps, etc.), and a more focused application for managing the names and scientific findings of the Afro-tropical beetle family Eucnemidae. Finally, contributions of our work are summarised, related work discussed, and directions for further research are outlined.

2   Biological Names and Taxonomies

The scientific name system is based on the Linnean binomial name system where the basic unit is a species. Every species belongs to some genus and every genus belongs to a higher taxon. A scientific name often has a reference to the original publication where it was first published. For example, the scientific name of the bumblebee, *Apis mellifera* Linnaeus, 1758, means that Linnaeus published the description of the bumblebee in 1758 (in Systema Naturae 10th edition) and that bumblebee belongs to the genus *Apis*. The upper levels of the taxonomic hierarchy do not show in a scientific name. A confusing feature of scientific names is that the meaning of the name may change although the name remains the same. Taxon boundaries may vary according to different studies, and there may be multiple simultaneous views of taxon limits of the same organism group. For example, a genus may be delimited in three ways and according to each view different sets of species are included in the genus as illustrated in Fig. 1. These differing views are *taxonomic concepts*. The usage of the correct name is not enough, and Berendsohn [1] suggested that taxonomic concepts should be referred to by an abbreviation *sec* (*secundum*) after the authors name to indicate in which meaning the name is used.
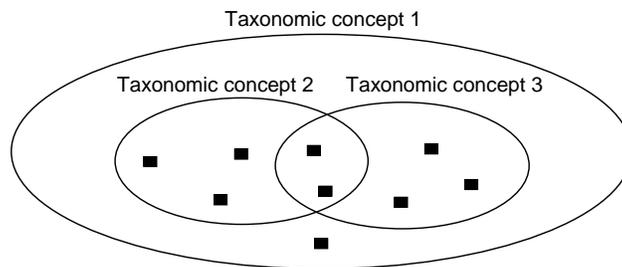
Fig. 1. A genus is delimited in three different ways according to three different studies. Black squares indicate species.

The nature of a biological name system is a change, as there is no single interpretation of the evolution. Typically there is no agreement if the variation observed in an organism is taxon-specific or shared by more than one taxon,

which makes the name system dynamic. For example, the fruit fly *Drosophila melanogaster* was shifted into the genus *Sophophora*, resulting in a new name combination *Sophonophora melanogaster* [7]. The most common taxonomic changes and their implications to the scientific names are the following: 1) A species has been shifted to another genus - the genus name changes. 2) One species turns out to be several species - new species are described and named, and the old name remains the same with a narrower interpretation. 3) Several species are found to be just one species - the oldest name is valid and the other names become its synonyms.

Species lists catalogue organisms occurring in a certain geographical area, which may vary from a small region to global. Often species lists contain valid taxon names with author information and synonyms of the valid names. They are snapshots of time and used especially by environmental authorities. The problem with species lists is that not all organism groups are catalogued and changes are not necessarily recorded in the species lists. Traditionally printed lists tend to be more detailed than online lists and their status is higher.

Species lists often follow different hierarchies and species may be associated with different genera according to the person who published the list. The hierarchy in a species list is a compromise that combines several studies, and the author can subjectively emphasise a view that he/she wishes. A taxon may also have different taxonomic ranks in literature, for example the same taxon can occur both as a species and a subspecies.

Common names tend to have regional variation and they do not indicate hierarchy unlike scientific names. Vernacular names have an important role in everyday language, but due to the variation and vagueness, they have little relevance in science. Vernacular names are used mainly in citizen science.

3   TaxMeOn – Meta-ontology of Biological Names

We have developed a meta-ontology for managing scientific and vernacular names. The ontology model consists of three parts that serve different purposes: 1) name collections, 2) species lists, and 3) name changes resulting from research. These parts are manageable separately, but associations between them are supported. Being a meta-ontology, TaxMeOn defines classes and properties that can be used to build ontologies. The ontologies can be used for creating semantic metadata for describing e.g. observational data or museum collections. TaxMeOn is based on RDF using some features of the OWL. The model contains 22 classes and 53 properties (61 including subproperties), of which ten classes and 15 properties are common to all the three parts of the model[2].

The core classes of TaxMeOn express a taxonomic concept, a scientific name, a taxonomic rank, a publication, an author, a vernacular name, and a status of a name. Taxonomic ranks are modelled as classes, and individual taxa are instances of them, for example the species forest fir *forrestii* (belongs to the genus *Abies*)

---

[2] The TaxMeOn schema is available at http://schema.onki.fi/taxmeon/

is an instance of the class *Species*. The model contains 61 taxonomic ranks, of which 60 are obtained from TDWG Taxon Rank LSID Ontology[3]. In order to simplify the management of subspecific ranks, an additional class that combines species and taxonomic levels below it was created.

References embody publications in a broad sense including other documented sources of information, for instance minutes of meetings. Bibliographic information can be associated to the reference according to the Dublin Core metadata standard. In biology, author names are often abbreviated when attached to taxon names. The TaxMeOn model supports the referring system that is typical to biology. Some of the properties used in TaxMeOn are part-specific as the uses of the parts differ from each other. For instance, the property that refers to a vernacular name is only available in the name collection part as it is not relevant in the other parts of the model.

The most distinctive feature of the research part [14] is that a scientific name and taxonomic concepts associated to it are separated, which allows detailed management of them both. In the name collection and species list parts, a name and its taxonomic concepts are treated as a unit. Different statuses can be associated to names, such as validity (accepted/synonym), a stage of a naming process (proposed/accepted) and spelling errors.

The model has a top-level hierarchy that is based on a rough classification, such as the division of organism *classes* and *orders*. Ontologies that are generated using TaxMeOn, can be hung on the top-level classification. A hierarchy is created using the transitive *isPartOfHigherTaxon* relation, e.g. to indicate that the species *forrestii* belongs to the genus *Abies*.

Taxon names that refer to the same taxon can occur as different names in the published species lists and different types of relations (see Table 1) can be set between the taxa. Similarly, research results of phylogenetic studies can be mapped using the same relations. The relations for mapping taxa are divided on the basis of attributes of taxa (*intensional*) or being a member of a group (*ostensive*). If it is known that two taxa have an association which is not specified, a class is provided for expressing incomplete information (see the empty ellipse in Fig. 2). This allows associations of taxa without detailed taxonomic knowledge, and especially between taxa originating from different sources.

Relation	Description
congruent with taxon	taxonomic concepts of two taxa are equal
is part of taxon	a taxonomic concept of a taxon is included in a taxonomic concept of another taxon
overlaps with taxon	taxonomic concepts of two taxa overlap

Table 1. Mapping relations used in species lists and research results. The three relations can be used as intensional and/or ostensive, using their subproperties.

---

[3] http://rs.tdwg.org/ontology/voc/TaxonRank
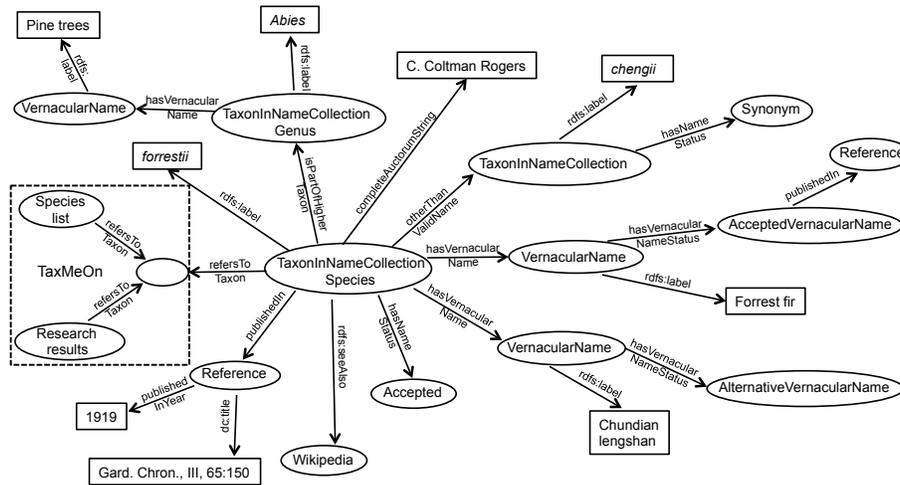
Fig. 2. An example of vernacular names in a name collection. The ellipses represent instances of TaxMeOn classes and literals are indicated as boxes. Other parts of the model are connected to the example taxon in the box with dotted line, in which the empty ellipse illustrates a general representation of a taxon.

In TaxMeOn, a reference (an author name and a publication year) to the original publication can be attached to a name. A complete scientific name is atomised into units that can be combined in applications by traversing the RDF graph by utilising the *isPartOfHigherTaxon* and *publishedIn* relations.

Name collections. Scientific names and their taxonomic concepts are treated as one unit in the name collection, because the scope is in vernacular names. The model supports the usage of multiple languages and dialects of common names. There may be several common names pointing to the same taxon, and typically one of them is recommended or has an official status. Alternative names are expressed defining the status using the class *VernacularNameStatus* and references related to the changes of a name status can be added. This allows the tracking the temporal order of the statuses. The model for vernacular names is illustrated in Fig. 2.

Species lists. Species lists have a single hierarchy and they seldom include vernacular names. Species lists have more relevance in science than name collections, but they lack information about name changes and a single list does not express the parallel or contradictory views of taxonomy which are crucial for researchers. Synonyms of taxa are typically presented and the taxonomic concept is included in a name like in a name collection. Taxa occurring in different species lists can be mapped to each other or to research results using the relations in Table 1. In addition, a general association without taxonomic details can be used (see Fig. 2).

Biological research results. In biological research results a key element is a taxonomic concept that can have multiple scientific names (and vice versa). Instead of names, taxonomic concepts are used for defining the relations between taxa. The same relations are applied here as in the speies list part (see Table 1). The latest research results often redefine taxon boundaries, for example a split of taxa narrows the original taxonomic concept and the meaning of the name changes although the name itself may remain the same. The new and the old concepts are connected into a temporal chain by instantiation of a change event. In Fig. 3 the concept of the beetle genus Galba is split into the concepts of the Balgus and Pterotarsus. The taxon names are shown inside the ellipses representing taxonomic concepts in order to simplify the presentation. Other change types are a lump of taxa, a change in taxon boundaries and a change in a hierarchy. These changes lead to the creation of a new instance of a taxonomic concept in order to maintain the traceable taxon history. An instantion of a new concept prevents evolving non-existing name combinations and artificial classifications. For instance, a species name is not associated with a genus name in which it has never been included.

The status of a scientific name may change in time as an accepted name may become a synonym. Multiple statuses can be attached to a name, but according to the nomenclatural rules only one of them is accepted at time. The temporal order of the statuses can be managed according to the same idea as in the name collections part.

4   Use Cases

We have applied the TaxMeOn ontology model to three use cases that are based on different needs. The datasets include a name collection of common names of vascular plants, several species lists of different animal groups and a collection of biological research results of Afro-tropical beetles. The use cases were selected on the basis of the active usage of the data (vernacular names), usefulness to the users (species lists), and the taxonomic challenges with available expertise (scientific names based on research results). The datasets used are depicted in Table 2.

4.1   Collaborative Management of Vascular Plants Names

The biological name collection includes 26,000 Finnish names for vascular plants that are organised into a single hierarchy. A deeply nested hierarchy is not necessary here as the classification used is robust, containing only three taxonomic ranks. The need is to maintain the collection of the common names and to manage the name acceptance process. The number of yearly updates exceeds 1,000. The typical users of the name collection are journalists, translators and other non-biologists who need to find a common name for a scientific name.
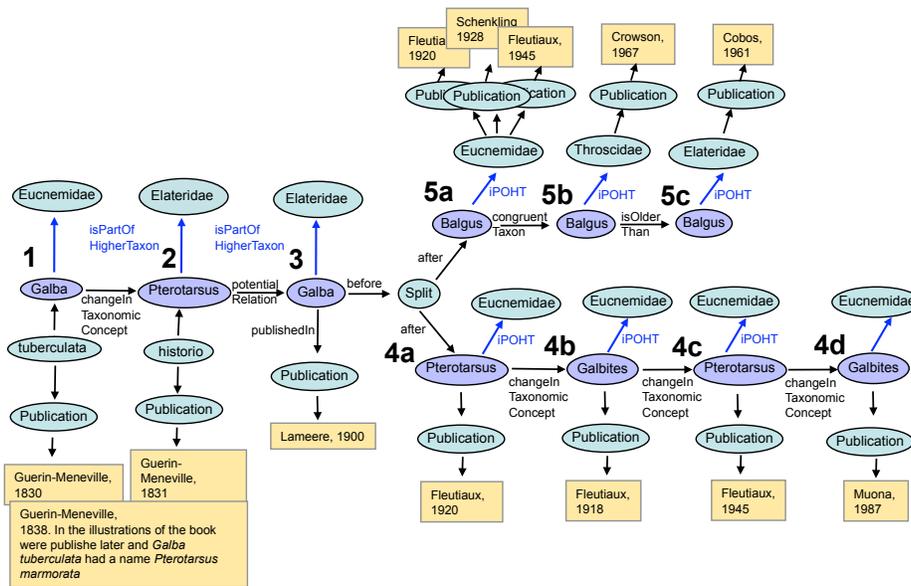
Fig. 3. An example of name changes and taxonomies of eucnemid beetles based on research results. The ellipses represent instances of TaxMeOn classes. Taxonomic hierarchies are expressed with the *isPartOfHigherTaxon* (iPOHT) relations, and the name change series of taxa are illustrated with a darker colour. The following abbreviations are used for the change types: S = Split of taxa, NC = Name change, TCC = Taxon concept change and CH = Change in hierarchy. The meaning of the numbers: 1) The species description of *Galba tuberculata* was originally published in 1830, but the illustrations of the book were published in 1838. However, in the illustrations *G. tuberculata* appeared with the name *Pterotarsus marmorata* (conflicting information). 2) Meanwhile, in 1831, the same taxon was independently described as *Pterotarsus historio* (independent events). 3) Lameere was confused by the two independently published works and changed the name to *Galba* in 1900 (uncertain relation between 2 and 3). 4a) Fleutiaux split the genus *Galba* into two genera. The name *Galba* was changed into *Pterotarsus* as there turned out to be a crustacean genus *Galba* (S, NC,TCC). 4b) Fleutiaux re-examined the genus and concluded that it is new to science and described it as *Galbites* (NC, TCC). 4c) Later Fleutiaux changed his mind and renamed the genus as *Pterotarsus* again (NC, TCC). 4d) Muona discovered that Fleutiaux was originally right and renamed the genus as *Galbites* (NC, TCC). 5a) When *Galba* was split, a part of its species were shifted into the genus *Balgus* that was described as new to science at the same time. *Balgus* was placed in the family Eucnemidae (CH). 5b) And changed into the family Throscidae (CH). This was originally published in a monthly magazine in the 1950's, but the magazines were published as a book in 1967 which is most commonly cited. 5c) *Balgus* was changed into the family Elateridae in 1961 (CH and conflict in publication years).

Taxon group	Region	Publ. years	# of taxa
Vascular plants	World	constantly updated	25726
Long-horn beetles (Coleoptera: Cerambycidae)	Scandinavia, Baltic countries	1939, 1960, 1979, 1992, 2004, 2010, 2010	205, 181, 247, 269, 300, 297, 1372
Butterflies and moths (Lepidoptera)	Scandianavia, North-West Russia, Estonia	1962, 1977, 1996, 2002, 2008	313, 256, 265, 4573, 12256, 3244, 3251, 3477
Thrips (Thysanoptera)	Finland	2008	219
Lacewings and scorpionflies (Neuroptera and Mecoptera)	Finland	2008	113
True bugs (Hemiptera)	Finland	2008	2690
Flies (Diptera: Brachycera)	Finland	2008	6373
Parasitic wasps (Hymenoptera: Ichneumoidae)	Finland	1995, 1999, 1999, 2000, 2003	282, 398, 919, 786, 733
Bees and wasps (Hymenoptera: Apoidea)	Finland	2010	1048
Mammals	World	2008	6062
Birds	World	2010	12125
False click beetles (Coleoptera: Eucnemidae)	Afrotropics	–	9 genera

Table 2. Datasets TaxMeOn has been applied to. Vascular plants are included in the name collection, the false click beetles are biological research results, and all other datasets are based on species lists.

The name collection of vascular plants is managed in SAHA[4] [12]. SAHA is a simple, powerful and scalable generic metadata editor for collaborative content creation. Annotation projects can be added into SAHA by creating the metadata schema for the content and loading it into SAHA. The user interface of SAHA adapts to the schema by providing suitable forms for producing the metadata. The values of the properties of the schema can be instances of classes defined in the schema, references to ontologies or literals. The annotations created using SAHA are stored in a database, from which they can be retrieved for use in semantic applications. SAHA also provides a SPARQL endpoint for making queries to the RDF data.

New scientific species names are added by creating a new instance of the *Species* class and then adding the other necessary information, such as their status. Similarly, a higher taxon can be created if it does not already exist, and the former is linked to the latter with the *isPartOfHigherTaxon* relation. SAHA has search facilities for querying the data, and a journalist writing a non-scientific article about a house plant, for example, can use the system for finding a common name for the plant.

---

[4] http://demo.seco.tkk.fi/saha/VascularPlants/index.shtml

4.2 Publishing Species Lists as Ontology Services

The users of species lists are ecologists, environmental authorities and amateurs searching for the correct scientific name occurring in a certain geographical area. In this use case ca. 20 published species lists obtained from the taxonomic database of the Finnish Museum of Natural History[5] containing more than 80,000 names were converted into TaxMeOn ontologies. In addition, seven regional lists of long-horn beetles (cerambycids) with 100 species are available from the years 1936–2010. The various names meaning the same taxon were mapped by an expert. The most common differences between the lists are a shift of a genus for a species, a change in a hierarchy and/or in a name status. Similarly, ca. 150 species of butterfly names from five lists were mapped.

Currently, the mapped beetle names are published as services for humans and machines in the ONKI Ontology Service[6] [25]. The ONKI Ontology Service is a general ontology library for publishing ontologies and providing functionalities for accessing them, using ready-to-use web widgets as well as APIs. ONKI supports content indexing, concept disambiguation, searching, and query expansion.

Fig. 4 depicts the user interface of the ONKI server [24]. The user is browsing the species lists of cerambycid beetles, and has made a query for taxon names starting with a string "ab". The selected species *abdominalis* has been described by Stephens in 1831, and it occurs in the species list Catalogue of Palaearctic Coleoptera, published in the year 2010 [15]. The species *abdominalis* belongs to the subgenus and genus *Grammoptera*. The taxonomy of the family Cerambycidae is visualised as a hierarchy tree. The same species also occurs in other species lists, which is indicated by *congruentWithTaxonOst* relation. Browsing the taxa reveals varying taxon names and classifications. For example, the *Grammoptera (Grammoptera) abdominalis* has a subgenus in this example, but the rank subgenus does not exist in the other lists of cerambycid. Also, the synonyms of the selected taxon are shown (*analis*, *femorata*, *nigrescens* and *variegata*).

The ONKI Ontology Services can be integrated into applications on the user interface level (in HTML) by utilising the ONKI Selector, a lightweight web widget providing functionalities for accessing ontologies. The ONKI API has been implemented in three ways: as an AJAX service, as a Web Service, and as a simple HTTP API.

The ONKI Ontology Service contains several ontologies covering different fields and is a part of the FinnONTO project [6] that aims to build a national ontology infrastructure. The Finnish Spatio-temporal Ontology (SAPO) [8], for example, can be used to disambiguate geographical information of observational data. Combining the usage of species ontologies and SAPO, extensive data harmonisation is avoided as both taxon names and geographical names change in time.

---

[5] http://taxon.luomus.fi/
[6] http://demo.seco.tkk.fi/onkiskos/cerambycids/

Fig. 4. The species of *abdominalis* shown in the ONKI Browser.

4.3 Management of Individual Scientific Names

The use case of scientific names is the Afro-tropical beetle family Eucnemidae, which consists of ca. nine genera that have gone through numerous taxonomic treatments. Also, mistakes and uncertain events are modelled if they are relevant to name changes. For example, the position of the species *Pterotarsus historio* in taxonomic classification has changed 22 times and at least eight taxonomic concepts are associated to the genus *Pterotarsus* [17]. Fig. 3 illustrates the problematic nature of the beetle group in a simplified example. A comparable comparable situation concerns most organism groups on Earth. Due to the numerous changes in scientific names, even researchers find it hard to remember them and this information can only be found in publications of taxonomy. The option of managing individual names is advantageous as it completes the species lists and allows the mapping of detailed taxonomic information to the species lists. For example, environmental authorities and most biologists prefer a simple representation of species lists instead of complicated change series.

5 Discussion

We have explored the applicability of the semantic web technologies for the management needs of biological names. Separating taxonomic concepts from scientific and vernacular names is justified due to the ambiguity of the names

referring to taxa. This also enables relating relevant attributes separately to a concept and to a name, although it is not always clear to which of these an attribute should be linked and subjective decisions have to made. The idea of the model is simplicity and practicality in real-world use cases.

The fruitfulness lays in the possibilities to link divergent data serving divergent purposes and in linking detailed information with more general information. For example, a common name of a house plant, a taxonomic concept that appears to be a species complex (a unit formed by several closely related species) and the geographical area can be linked.

The most complex use case is the management of scientific name changes of biological research results. The main goal is to maintain the temporal control of the name changes and classifications. The instantiation of taxon names and concepts lead to a situation in which they are hard to manage when they form a long chain. Every change increases the number of instances created. Protegé[7] was used for editing the ontologies, although managing names is quite inconvenient because they are shown as an alphabetically ordered flat list, not as a taxonomic hierarchy.

As Protegé is rather complicated for a non-expert user, the metadata editor SAHA was used for maintaining the continuous changes of common names of plants. The simplicity of SAHA makes it a suitable option for ordinary users who want to concentrate on the content. However, we noticed that some useful features are missing from SAHA. The visualisation of a nested hierarchy would help users to compare differing classifications.

In many biological ontologies the 'subclass of' relation is used for expressing the taxon hierarchies. However, in the TaxMeOn model we use the *isPartHigherTaxon* relation instead. If the 'subclass of' relation was used to express the taxonomic hierarchy, a taxon would incorrectly be an instance of the higher taxon ranks, e.g., a species would be an instance of the class *Genus*. This would lead to a situation in which queries for genera also return species.

5.1 Related Work

NCBO BioPortal[8] and OBO Foundry[9] have large collections of life science ontologies mainly concentrating on biomedicine and physiology. The absence of taxonomic ontologies is distinctive which may indicate the complexity of the biological name system. The portals contain only three taxonomic ontologies (Amphibian taxonomy, Fly taxonomy and Teleost taxonomy) and one broader classification (NCBI organismal classification). The taxonomic hierarchy is defined using the *rdfs:subClassOf* relation in the existing ontologies. Taxonconcept.org[10] provides Linked Open Data identifiers for species concepts and links data about them originating from different sources. All names are expressed using literals

---

[7] http://protege.stanford.edu/
[8] http://bioportal.bioontology.org/
[9] http://www.obofoundry.org/
[10] http://www.taxonconcept.org/

and the following taxonomic ranks are included: a combination of a species and a genus, a class and an order. Parallel hierarchies are not supported. Geospecies[11] uses the properties *skos:broaderTransitive* and *skos:narrowerTransitive* to express the hierarchy.

Page [19] discusses the importance of persistent identifiers for organism names and presents a solution for managing names and their synonyms on the semantic web. The taxon names from different sources referring to the same taxon are mapped using the *owl:sameAs* relation which is a strong statement. Hierarchy is expressed using two different methods in order to support efficient queries.

Schulz et al. [20] presented the first ontology model of biological taxa and its application to physical individuals. Taxa organised in a hierarchy is thoroughly discussed, but the model is static and based on a single unchangeable taxonomy. Despite recognising the dynamic nature of taxonomy and the name system, the model is not applicable in the management of biological names as such.

Franz and Peet [3] enlighten the problematic nature of the topic by describing how semantics can be applied in relating taxa to each other. They introduce two essentially important terms from philosophy to taxonomy to specify the way, in which differing classifications that include different sets of taxa can be compared. An ostensive relation is specified by being a member of a group and intensional relations are based on properties uniting the group. These two fundamentally different approaches can be used simultaneously, which increases the information content of the relation.

Franz and Thau [4] developed the model of scientific names further by evaluating the limitations of applying ontologies. They concluded that ontologies should focus either on a nomenclatural point of view or on strategies for aligning multiple taxonomies.

Tuominen et al. [23] model the taxonomic hierarchy using the *skos:broader* property, and preferred scientific and common names of the taxa are represented with the property *skos:prefLabel* and alternative names with *skos:altLabel*. The property *rdf:type* is used to indicate the taxonomic rank. This is applicable to relatively simple taxonomies such as species lists, but it does not support expressing more elaborate information (changes in a concept or a name).

The Darwin Core (DwC) [2] is a metadata schema developed for observation data by the TDWG (Biodiversity Information Standards). The goal of the DwC is to standardise the form of presenting biological information in order to enhance the usage of it. However, it lacks the semantic aspect and the terms related to biological names are restricted due to the wide and general scope of the DwC.

The scope of the related work presented above differs from our approach as our focus is on practical name management and retrieval of names.

Research on ontology versioning [10] and ontology evolution [18] has focused on finding mappings between different ontology versions, performing ontology refinements and other changes in the conceptualisation [9, 21], and in reasoning with multi-version ontologies [5]. There are similarities in our problem field, but our focus is to support multiple parallel ontologies interpreting the domain

---

[11] http://lod.geospecies.org/

differently, not in versioning or evolution of a specific ontology. For example, there is no single taxonomy of all organisms, but different views of how they should be organised into hierarchies.

A similar type of an approach for managing changes and parallel views of concepts has been proposed by Tennis and Sutton [22] in the context of SKOS vocabularies. However, TaxMeOn supports richer ways of expressing information, e.g. for managing changes of taxon names and concepts separately.

5.2   Future Work

The model will be tested using different datasets to ensure its applicability. Currently, the research results part covers animal names, but will be expanded to plant names as well. The lack of user-friendly tools is obvious and the metadata editor SAHA is planned to be expanded to respond to the needs. Describing evolutionary trees and their information content is a challenging application area as phylogenetics produces name changes.

References

1. Berendsohn, W.: The concept of "potential taxon" in databases. Taxon 44, 207–212 (1995)
2. Darwin Core Task Group: Darwin core. Tech. rep. (2009), http://www.tdwg.org/standards/450/
3. Franz, N., Peet, R.: Towards a language for mapping relationships among taxonomic concepts. Systematics and Biodiversity 7(1), 5–20 (2009)
4. Franz, N., Thau, D.: Biological taxonomy and ontology development: scope and limitations. Biodiversity Informatics 7, 45–66 (2010)
5. Huang, Z., Stuckenschmidt, H.: Reasoning with multi-version ontologies: A temporal logic approach. In: International Semantic Web Conference. pp. 398–412 (2005)
6. Hyvönen, E., Viljanen, K., Tuominen, J., Seppälä, K.: Building a national semantic web ontology and ontology service infrastructure—the FinnONTO approach. In: Proceedings of the ESWC 2008, Tenerife, Spain. Springer–Verlag (2008)
7. ICZN: Opinion 2245 (case 3407) drosophila fallén, 1823 (insecta, diptera): Drosophila funebris fabricius, 1787 is maintained as the type species. Bulletin of Zoological Nomenclature 67(1) (2010)
8. Kauppinen, T., Väätäinen, J., Hyvönen, E.: Creating and using geospatial ontology time series in a semantic cultural heritage portal. In: Proceedings of the ESWC 2008, Tenerife, Spain. pp. 110–123. Springer–Verlag (2008)

---

9. Klein, M.: Change Management for Distributed Ontologies. Ph.D. thesis, Vrije Universiteit Amsterdam (Aug 2004)

10. Klein, M., Fensel, D.: Ontology versioning on the Semantic Web. In: Proceedings of the International Semantic Web Working Symposium (SWWS). pp. 75–91. Stanford University, California, USA (Jul 30 – Aug 1, 2001)

11. Knapp, S., Polaszek, A., Watson, M.: Spreading the word. Nature 446, 261–262 (2007)

12. Kurki, J., Hyvönen, E.: Collaborative metadata editor integrated with ontology services and faceted portals. In: Workshop on Ontology Repositories and Editors for the Semantic Web (ORES 2010), the Extended Semantic Web Conference ESWC 2010, Heraklion, Greece. CEUR Workshop Proceedings, http://ceur-ws.org/ (June 2010)

13. Laurenne, N., Tuominen, J., Koho, M., Hyvönen, E.: Modeling and publishing biological names and classifications on the semantic web. In: TDWG 2010 Annual Conference of the Taxonomic Databases Working Group (September 2010), poster abstract

14. Laurenne, N., Tuominen, J., Koho, M., Hyvönen, E.: Taxon meta-ontology TaxMeOn – towards an ontology model for managing changing scientific names in time. In: TDWG 2010 Annual Conference of the Taxonomic Databases Working Group (September 2010), contributed abstract

15. Löbl, I., Smetana, A.: Catalogue of Palearctic Coleoptera. Volume 6. Chrysomeloidea. Apollo Books, Stenstrup (2010)

16. Mayden, R.L.: A hierarchy of species concepts: the denouement in the saga of the species problem. In: Claridge, M.F., Dawah, H.A., Wilson, M.R. (eds.) Species: The Units of Biodiversity, vol. Systematics Association Special Vol. 54, pp. 381–424. Chapman and Hall, London (1997)

17. Muona, J.: A revision of the indomalesian tribe galbitini new tribe (coleoptera, eucnemidae). Entomologica Scandinavica. Supplement 39, 1–67 (1991)

18. Noy, N., Klein, M.: Ontology evolution: Not the same as schema evolution. Knowledge and Information Systems 6(4) (2004)

19. Page, R.: Taxonomic names, metadata, and the semantic web. Biodiversity Informatics 3, 1–15 (2006)

20. Schulz, S., Stenzhorn, H., Boeker, M.: The ontology of biological taxa. Bioinformatics 24(13), 313–321 (2008)

21. Stojanovic, L.: Methods and Tools for Ontology Evolution. Ph.D. thesis, University of Karlsruhe, Germany (2004)

22. Tennis, J.T., Sutton, S.A.: Extending the simple knowledge organization system for concept management in vocabulary development applications. Journal of the American Society for Information Science and Technology 59(1), 25–37 (2008)

23. Tuominen, J., Frosterus, M., Laurenne, N., Hyvönen, E.: Publishing biological classifications as SKOS vocabulary services on the semantic web. In: TDWG 2010 Annual Conference of the Taxonomic Databases Working Group (September 2010), demonstration abstract

24. Tuominen, J., Frosterus, M., Viljanen, K., Hyvönen, E.: ONKI SKOS server for publishing and utilizing SKOS vocabularies and ontologies as services. In: Proceedings of the ESWC 2009, Heraklion, Greece. pp. 768–780. Springer–Verlag (2009)

25. Viljanen, K., Tuominen, J., Hyvönen, E.: Ontology libraries for production use: The Finnish ontology library service ONKI. In: Proceedings of the ESWC 2009, Heraklion, Greece. pp. 781–795. Springer–Verlag (2009)