

RESEARCH

Open Access



# OpenBiodiv-O: ontology of the OpenBiodiv knowledge management system

Viktor Senderov<sup>1,2\*</sup> , Kiril Simov<sup>3</sup>, Nico Franz<sup>4</sup>, Pavel Stoev<sup>1,7</sup>, Terry Catapano<sup>5</sup>, Donat Agosti<sup>5</sup>, Guido Sautter<sup>5</sup>, Robert A. Morris<sup>6</sup> and Lyubomir Penev<sup>1,2</sup>

## Abstract

**Background:** The biodiversity domain, and in particular biological taxonomy, is moving in the direction of semantization of its research outputs. The present work introduces OpenBiodiv-O, the ontology that serves as the basis of the OpenBiodiv Knowledge Management System. Our intent is to provide an ontology that fills the gaps between ontologies for biodiversity resources, such as DarwinCore-based ontologies, and semantic publishing ontologies, such as the SPAR Ontologies. We bridge this gap by providing an ontology focusing on biological taxonomy.

**Results:** OpenBiodiv-O introduces classes, properties, and axioms in the domains of scholarly biodiversity publishing and biological taxonomy and aligns them with several important domain ontologies (FaBiO, DoCO, DwC, Darwin-SW, NOMEN, ENVO). By doing so, it bridges the ontological gap across scholarly biodiversity publishing and biological taxonomy and allows for the creation of a Linked Open Dataset (LOD) of biodiversity information (a biodiversity knowledge graph) and enables the creation of the OpenBiodiv Knowledge Management System.

A key feature of the ontology is that it is an ontology of the scientific process of biological taxonomy and not of any particular state of knowledge. This feature allows it to express a multiplicity of scientific opinions. The resulting OpenBiodiv knowledge system may gain a high level of trust in the scientific community as it does not force a scientific opinion on its users (e.g. practicing taxonomists, library researchers, etc.), but rather provides the tools for experts to encode different views as science progresses.

**Conclusions:** OpenBiodiv-O provides a conceptual model of the structure of a biodiversity publication and the development of related taxonomic concepts. It also serves as the basis for the OpenBiodiv Knowledge Management System.

**Keywords:** Biodiversity, Biodiversity informatics, Semantic web, Semantic publishing, Ontology, Knowledge management, Linked open data, RDF, OWL, Taxonomy, Concept taxonomy, Biological systematics, Data modeling

## Background

The desire for an integrated information system serving the needs of the biodiversity community dates at least as far back as 1985 when the Taxonomy Database Working Group (TDWG)—later renamed to Biodiversity Informatics Standards—was established [1]. In 1999, the Global Biodiversity Information Facility (GBIF) was created after the Organization for Economic Cooperation and Development (OECD) had arrived at the conclusion that

“an international mechanism is needed to make biodiversity data and information accessible worldwide” [2]. The Bouchout declaration [3] crowned the results of the pro-iBiosphere project (2012 - 2014) [4] dedicated to the task of creating an integrated biodiversity information system. The Bouchout declaration proposes to make scholarly biodiversity knowledge freely available as Linked Open Data. A parallel process in the U.S.A. started even earlier with the establishment of the Global Names Architecture [5, 6].

The specification and design of a semantic system, the Open Biodiversity Knowledge Management System (OBKMS, later simply OpenBiodiv), implementing the objectives of the Bouchout Declaration by focusing

\*Correspondence: vsenderov@gmail.com

<sup>1</sup>Pensoft Publishers, Prof. Georgi Zlatarski 12, 1700 Sofia, Bulgaria

<sup>2</sup>Institute of Biodiversity and Ecosystems Research, Bulgarian Academy of Sciences, Sofia, Bulgaria

Full list of author information is available at the end of the article

on knowledge extraction from academic journals and research databases, were outlined amongst others in [7, 8]. In this publication we present the OpenBiodiv Ontology (OpenBiodiv-O)—the knowledge and inferencing model of OpenBiodiv [9]. OpenBiodiv-O provides a conceptual model of the structure of a biodiversity publication and the development of related taxonomic concepts.

### Previous work

In the biomedical domain there are well-established efforts to extract information and discover knowledge from literature [10–12]. The biodiversity domain, and in particular biological systematics and taxonomy (from here on in this paper referred to as *taxonomy*), is also moving in the direction of semantization of its research outputs [13–15]. The publishing domain has been modeled through the Semantic Publishing and Referencing Ontologies (SPAR Ontologies) [16]. The SPAR Ontologies are a collection of ontologies incorporating—amongst others—FaBiO, the FRBR-aligned Bibliographic Ontology [17], and DoCO, the Document Component Ontology [18]. The SPAR Ontologies provide a set of classes and properties for the description of general-purpose journal articles, their components, and related publishing resources. Taxonomic articles and their components, on the other hand, have been modeled through the TaxPub XML Document Type Definition (DTD) (also referred to loosely as XML schema) and the Treatment Ontologies [19, 20]. While TaxPub is the XML-schema of taxonomic publishing for several important taxonomic journals (e.g. ZooKeys, Biodiversity Data Journal), the Treatment Ontologies are still in development and have served as a conceptual template for OpenBiodiv-O. In fact, they share many of the same authors.

Taxonomic nomenclature is a discipline with a very long tradition. It transitioned to its modern form with the publication of the Linnaean System [21]. Already by the beginning of the last century, there were hundreds of vocabulary terms (e.g. *types*) [22]. At present the naming of organismal groups is governed by the International Code of Zoological Nomenclature (ICZN) [23] and by the International Code of Nomenclature for algae, fungi, and plants (Melbourne Code) [24]. Due to their complexity (e.g. ICZN has 18 chapters and 3 appendices), it proved challenging to create a top-down ontology of biological nomenclature. Example attempts include the relatively complete NOMEN ontology [25] and the somewhat less complete Taxonomic Nomenclatural Status Terms (TNSS) [26].

There are several projects that are aimed at modeling the broader biodiversity domain conceptually. Darwin Semantic Web (Darwin-SW) [27] adapts the previously existing Darwin Core (DwC) terms [28] as RDF. These models deal primarily with organismal occurrence data.

Modeling and formalization of the strictly taxonomic domain has been discussed by Berendsohn [29] and later, e.g., in [30, 31]. Noteworthy efforts are the XML-based Taxonomic Concept Transfer Schema [32] and a now defunct Taxon Concept ontology [33].

### Aims

The present work introduces OpenBiodiv-O, which serves as the basis of OpenBiodiv. By developing an ontology focusing on biological taxonomy, our intent is to provide an ontology that fills in the gaps between ontologies for biodiversity resources such as Darwin-SW and semantic publishing ontologies such as the ontologies comprising the SPAR Ontologies. Moreover, we take the view that it is advantageous to model the taxonomic process itself rather than any particular state of knowledge.

OpenBiodiv [8] “lifts” biodiversity information from scholarly publications and academic databases into a computable semantic form. The implementation of the system will be treated in future works. In this contribution, we discuss OpenBiodiv-O by first introducing the modeled domain conceptually and then formalizing it in “Results” section.

### Domain description

Biological taxonomy is a very old discipline dating back possibly to Aristotle, whose fundamental insight was to group living things in a hierarchy [34]. The discipline took its modern form after Carl Linnaeus (1707 - 1778) [34]. In his *Systema Naturae* Linnaeus proposed to group organisms into *kingdoms, classes, orders, genera, and species* bearing latinized scientific names with a strictly prescribed syntax. Linnaeus listed possible alternative names and gave a characteristic description of the groups [21]. These groups are called *taxa*, which is a Greek word for *arrangement*. The hierarchy that taxa form is called *taxonomy*. The etymology of the word is Greek and roughly translates to *method of arranging*. Note the polysemy here: the science of biological taxonomy is called *taxonomy* as is the arrangement of taxa itself. We believe, however, that it is sufficiently clear from context what is meant by “taxonomy” in any particular usage throughout this paper.

Even though Linnaeus and his colleagues may have hoped to describe life on Earth during their lifetimes, we now know that there are millions of species still undiscovered and undescribed [35]. On the other hand, our understanding of species and higher-rank taxonomic concepts changes as evolutionary biology advances [36]. Therefore, an accurate and evolutionarily reliable description of life on Earth is a perpetual process and cannot be completed with a single project that can be converted into an ontology. Thus, our aim is not to create an ontology capturing a fixed view of biological taxonomy, but to create an ontology of the taxonomic process. The ongoing use of this ontology will enable the formal description of taxonomic

biodiversity knowledge at any given point in time. In the following paragraphs, we introduce what the taxonomic process entails and reflect on the resources that need modeling.

An examination of the taxonomic process reveals that taxonomy works by employing the scientific method: researchers examine specimens and, based on the phenotypic and genetic variation that they observe, form a hypothesis [37]. This hypothesis may be called a taxonomic concept, a potential taxon, a species hypothesis [29], or an operational taxonomic unit (OTU) [38] in the case of a numerically delimited taxon.

A taxonomic concept describes the allowable phenotypic, genomic, or other variation within a taxon by designating type specimens and describing characters explicitly. It is a valid falsifiable scientific claim as it needs to fulfill certain verifiable evolutionary requirements. For example, a species-rank taxonomic hypothesis needs to fit our current understanding of species (species concept, [36]). More generally, the aspiration is that species concepts are adequate and give certain tangible criteria for species delimitation. However, valid scientific discussions continue about concept adequacy. The discussions are nuanced because they often draw on different conceptions of the relative weight of certain evolutionary phenomena. This leads to having quite a few different species concepts—morphological, ecological, phylogenetic, genomic, biological, etc. [36]. Nevertheless, if we fix a species concept—let us say we take the biological species concept—we can falsify any given species-rank taxonomic hypothesis against our fixed species concept.

Similarly, hypotheses of higher rank (representing upper levels of the taxonomic hierarchy) also need to fulfill certain evolutionary requirements. For example, a modern genus concept requires all species assigned to it to be descendants of a separate lineage and to form a monophyletic clade.

The ranks (taxonomy hierarchy levels) are not completely fixed. The usage of lower ranks (species, genus, family, order) is governed by international Codes [23, 24]. In the example of Linnaeus' ranks, each organism is first a member of its species, then genus, then order, then class, and finally kingdom. Which specific ranks a given taxonomic study employs is dependent on the field (e.g. botany vs. zoology), on the particular author, on the level of taxonomic resolution required, as well as on the history of classifying in that particular group.

Once the researchers have formed their concept, it must be published in a scientific outlet (journal or book). The biological Codes put some requirements and recommendations aimed at ensuring the quality of published research but ultimately it is a democratic process guaranteeing that everyone may publish taxonomic concepts provided they follow the rules of the Codes. This means

that in order to create a knowledge base of biodiversity, we need to be able to mine taxonomic papers from legacy and modern journals and books.

As a first good approximation, a taxonomic concept is based on a number of specimens or occurrences that are listed in a section usually called "Materials Examined". In general terms, we can say that a sighting of a living thing, i.e. an organism, at a given location and at a given time is referred to as an occurrence, and a voucher for this occurrence (e.g. the sampling of the organism itself) is referred to as a specimen [27]. Moreover, a taxonomic article may include other specialized sections such as the Checklist section, where one may list all taxa (in fact: the taxonomic concepts for those taxa) for organisms observed in a given region.

Typically, the information content of a treatment consists of several units. First, we have the aforementioned nomenclatural information that pertains to the scientific name—its authorship, etymology, related names, etc. Then, we have the taxonomic concept information that can be considered to have two components, as well: the first one is the intensional component of the taxonomic concept made up mostly of *traits* or *characters*. Traits are an explicit definition of the allowable variation (e.g. phenotypic, genomic, or ecological) of the organisms that make up the taxon. For example, we can define the order of spiders, Araneae, to be the class of organisms that have specialized appendages used for sperm transfer called pedipalps [39]. Knowledge of this kind is found in the Diagnosis, Description, Distribution and other subsections of the treatment.

Non-traditionally delimited taxonomic hypotheses are called *operational taxonomic units* (OTU's). In the case of genomic delimitation, sometimes the concepts are published directly as database entries and not as Code-compliant taxonomic articles [40]. A genomic delimitation can, for example, be based on a barcode sequence and on a statistical clustering algorithm specifying the allowable sequence variability that an organism can possess in order to be considered part of the barcode sequence-bearing operational taxonomic unit. However, as, in the general case, we don't have a Linnaean name or a morphological description for an operational taxonomic unit, we refer to it as a *dark taxon* [40]. The term "dark" is, however, usually reserved for concepts at lower ranks. Operational taxonomic units are published, for example, in the form of *barcode identification numbers* (BIN's) in the Barcode of Life Data Systems (BOLD) [41], or as *species hypotheses* in Unified system for the DNA based fungal species linked to the classification (UNITE) [42].

The second part of the information content of a taxonomic concept is the ostensive component: a listing of some (but not necessarily all) of the organisms that belong to the taxonomic concept. This information

is found in the Materials Examined subsection of the treatment.

Finally, the relationships between taxonomic concepts—simple hierarchical (*is a*) or more fine-grained Region Connection Calculus 5 (RCC-5) [30, 43]—can be both intensionally defined in the nomenclature section or ostensibly inferred from the Materials Examined. However, given the customary idiosyncrasies of biological descriptions, providing an initial set of RCC-5 relationships for a machine reasoner to work with often requires expert assessment and cannot be easily lifted from the text.

Thus, in order to model the taxonomic process, our ontology models scholarly taxonomic papers, database entries, agents responsible for their creation, treatments, taxonomic concepts, scientific names, occurrence and specimen information, other entities (e.g. ecological, geographical) part-taking in the taxonomic process, as well as relationships among these.

## Methods

OpenBiodiv-O is expressed in Resource Description Framework (RDF). At the onset of the project [8], a consideration was made to use RDF in favor of a more complex data model such as Neo4J's. The choice of RDF was made in order to be able to incorporate the multitude of existing domain ontologies into the overall model.

To develop the conceptualization of the taxonomic process and then the ontology we utilized the following process: (1) domain analysis and identification of important resources and their relationships; (2) analysis of existing data models and ontologies and identification of missing classes and properties for the successful formalization of the domain.

The formal structure of the ontology is specified by employing the RDF Schema (RDFS) and the Web Ontology Language (OWL). It is encoded as a part of a literate programming [44] document titled “OpenBiodiv Ontology and Guide” [45]. The structure has been extracted from that file via *knitr* and provided here as Additional file 1. It is also possible to request the ontology via Curl from the endpoint with the indication of content-type: application/rdf+xml. The vocabularies can be found as more additional files: Taxonomic Statuses (Additional file 2) and RCC-5 (Additional file 3), on the website [9], and on the GitHub page [46] (under *ontology/*).

A partial dataset from Pensoft's journals has been generated with OpenBiodiv-O and can be found at the SPARQL Endpoint <<http://graph.openbiodiv.net/>>, select repository *obkms\_i6*. The endpoint is also accessible from the website, <<http://openbiodiv.net/>>, under “SPARQL Endpoint”. Demos are available as “Saved Queries” from the workbench.

## Results

We understand OpenBiodiv-O to be the *shared formal specification of the conceptualization* [47–49] that we have introduced in Background. OpenBiodiv-O describes the structure of this conceptualization, not any particular state of it.

There are several domains in which the modeled resources fall. The first one is the scholarly biodiversity publishing domain. The second domain is that of taxonomic nomenclature. The third domain is that of broader taxonomic (biodiversity) resources (e.g. taxonomic concepts and their relationships, species occurrences, traits). To combine such disparate resources together we rely on SKOS [50]. Unless otherwise noted, the default namespace of the classes and properties for this paper is <<http://openbiodiv.net/>>. The prefixes discussed in this paper are listed in Additional file 1, at the beginning of the ontology.

### Semantic modeling of the biodiversity publishing domain

An article as such may be represented by a set of metadata, while its content consists of article components such as sections, tables, figures and so on [51].

To accommodate the specific needs of scholarly biodiversity publishing, we introduce a new class for taxonomic articles, Taxonomic Article (`:TaxonomicArticle`), new classes for specific subsections of the taxonomic article such as Taxonomic Treatment, Taxonomic Key, and Taxonomic Checklist, and a new class, Taxonomic Name Usage (`:TaxonomicNameUsage`), for the mentioning of a taxonomic name (see next subsection) in an article. These new classes are summarized in Table 1.

The classes from this subsection are based on the Tax-Pub XML Document Type Definition (DTD) [19] (also

**Table 1** New biodiversity publishing classes introduced

Class QName	Comment
<code>:Treatment</code>	Section of a taxonomic article
<code>:NomenclatureSection</code>	Subsection of Treatment
<code>:NomenclatureHeading</code>	Contains a nomenclatural act
<code>:NomenclatureCitationList</code>	List of citations of related concepts
<code>:MaterialsExamined</code>	List of examined specimens
<code>:BiologySection</code>	Subsection of Treatment
<code>:DescriptionSection</code>	Subsection of Treatment
<code>:TaxonomicKey</code>	Section with an identification key
<code>:TaxonomicChecklist</code>	Section with a list of taxa for a region
<code>:TaxonomicNameUsage</code>	Mention of a taxonomic name

referred to loosely as XML schema), on the structure of Biodiversity Data Journal’s taxonomic paper [52], and on the Treatment Ontologies [20].

Furthermore, we introduce two properties: *contains* (:contains) and *mentions* (:mentions). *Contains* is used to link parts of the article together and *mentions* links parts of the article to other concepts.

A graphical representation of the relationships between instances of the publishing-related classes that OpenBiodiv introduces is to be found in the diagram in Fig. 1.

**Semantics, alignment, and usage**

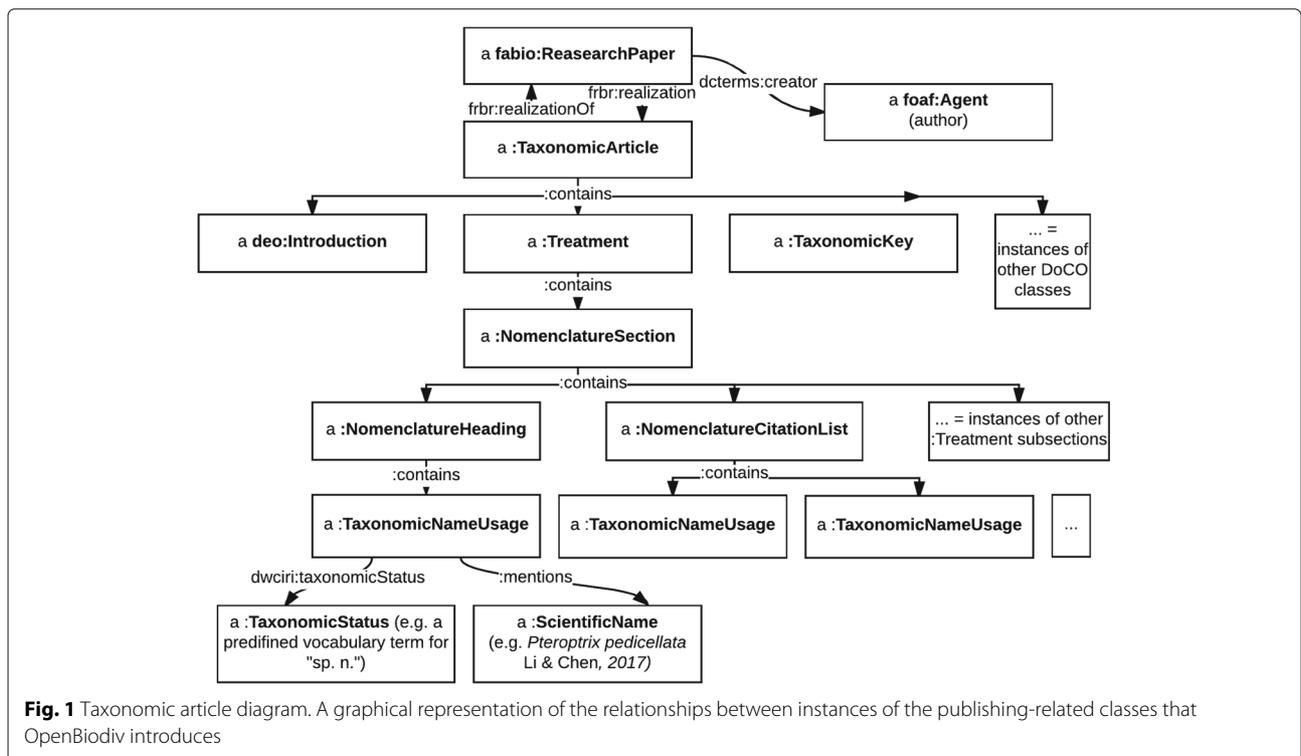
Our bibliographic model has the Semantic Publishing and Referencing Ontologies (SPAR Ontologies) at its core with a few extensions that we have written to accommodate for taxonomic elements. The SPAR Ontologies’ FRBR-aligned Bibliographic Ontology (FaBiO) uses the Functional Requirements for Bibliographic Records (FRBR) [53] model to separate publishable items into less or more abstract classes. We deal primarily with the Work class, i.e. the conceptual idea behind a publishable item (e.g. the story of “War and Peace” as thought up by Leo Tolstoy), and the Expression class, i.e. a version of record of a Work (e.g. “War and Peace,” paperback edition by Wordsworth Classics).

Taxonomic Article is a subclass of FaBiO’s Journal Article. Furthermore Journal Article is a FRBR Expression. This implies that taxonomic articles are FRBR expressions as well. This has important implications later on

when discussing taxonomic concept labels. Also, it means that we separate the abstract properties of an article (in a FaBiO Research Paper instance, which is a Work) from the version of record (in a Taxonomic Article, an Expression).

The taxonomic-specific section and subsection classes are introduced as subclasses of Discourse Element Ontology’s (DEO) Discourse Element (deo:DiscourseElement, [18]). So is the class Mention (:Mention), meant to represent an area of a document that can be considered a mention of something. This class, and the corresponding property, *mentions*, are inspired by pext:Mention and its corresponding property from PROTON [54]. The redefinition is necessary by the fact in OpenBiodiv-O they possess a slightly different semantics and a different placement in the upper-level hierarchy. We then introduce Taxonomic Name Usage as a subclass of Mention.

This placement of the document component classes that we’ve introduced in Discourse Element means that they ought to be used exactly in the same way as one would use the other discourse elements from DEO and DoCO (analogous to e.g. deo:Introduction). Note: DEO is imported by DoCO. Figures 2 and 3 give example usage in Turtle illustrating these ideas. A caveat here is that while the SPAR Ontologies use po:contains in their examples, we use *contains*, which is a subproperty of po:contains with the additional property of being transitive. We believe this definition is sensible as surely a sub-subcomponent is contained in a component. All other



**Fig. 1** Taxonomic article diagram. A graphical representation of the relationships between instances of the publishing-related classes that OpenBiodiv introduces

```

:biodiversity-data-journal rdf:type fabio:Journal ;
  skos:prefLabel "Biodiversity Data Journal"@en ;
  skos:altLabel "BDJ"@en ;
  fabio:issn "1314-2836" ;
  fabio:eIssn "1314-2828" ;
  frbr:part :b90f6933-ab5e-4ce1-9379-12de9ef4eaa6 .

<http://dx.doi.org/10.3897/BDJ.1.e953> rdf:type fabio:TaxonomicArticle ;
  skos:prefLabel "10.3897/BDJ.1.e953" ;
  dc:title "Casuarinicola australis Taylor, 2010
(Hemiptera: Triozidae), newly recorded from New Zealand"@en ;
  prism:doi "10.3897/BDJ.1.e953" ;
  dcelements:publisher "Pensoft Publishers"@en ;
  fabio:hasPublicationYear "2013"^^xsd:gYear ;
  prism:publicationDate "2013-9-16"^^xsd:date ;
  dcterms:publisher :pensoft-publishers ;
  frbr:realizationOf :thorpe-2013 .

:thorpe-2013 rdf:type :ResearchPaper ;
  skos:prefLabel "Thorpe 2013"
  skos:altLabel "paper10.3897/BDJ.1.e953" ;
  dcterms:creator :stephen-e-thorpe ;
  prism:keywords "Casuarinicola australis"@en ;
  fabio:hasSubjectTerm :a2ee4929-90dd-4a7a-aa5c-08836f49d549 .

:pensoft-publishers rdf:type :Publisher ;
  skos:prefLabel "Pensoft Publishers"@en .

:stephen-e-thorpe rdf:type foaf:Person ;
  skos:prefLabel "Stephen E. Thorpe" ;
  foaf:firstName "Stephen E." ;
  foaf:surname "Thorpe" ;
  foaf:mbox "stephen_thorpe@yahoo.co.nz" ;
  :affiliation "School of Biological Sciences (Tamaki Campus),
  University of Auckland, Auckland, New Zealand"@en .

:a2ee4929-90dd-4a7a-aa5c-08836f49d549 rdf:type fabio:SubjectTerm ;
  rdfs:label "Casuarinicola australis"@en ;
  skos:inScheme :openbiodiv-subject-terms .

```

**Fig. 2** Example article metadata. This example shows how to express the metadata of a taxonomic article with the SPAR Ontologies' model and the classes that OpenBiodiv defines. The code is in Turtle

aspects of expressing a taxonomic article in RDF according to OpenBiodiv-O are exactly the same as according to the SPAR Ontologies.

### Semantic modeling of biological nomenclature

While NOMEN and TNSS (introduced in subsection "Previous work") take a top-down approach of modeling the nomenclatural Codes, OpenBiodiv-O takes a bottom-up approach of modeling the use of taxonomic names in articles. Where possible we align OpenBiodiv-O classes to NOMEN.

Based on the need to accommodate taxonomic concepts, we have defined the class hierarchy of taxonomic names found in Fig. 4. Furthermore, we have introduced the class Taxonomic Name Usage (:TaxonomicNameUsage). Taxonomic name usages have been discussed widely in the community (e.g. in [55]); however, the meaning of term remains vague. The

abbreviation TNU is used interchangeably for "taxon name usage" and for "taxonomic name usage." In OpenBiodiv-O, a taxonomic name usage is the mentioning of a taxonomic name in the text, optionally followed by a taxonomic status.

For example, "*Heser stoevi* Deltschev 2016, sp. n." is a taxonomic name usage. The cursive text followed by the author and year of the original species description is the latinized scientific name. The abbreviation "sp. n." stands for the Latin *species novum*, indicating the discovery of a new taxon.

We also introduce the class Taxonomic Concept Label (:TaxonomicConceptLabel). A taxonomic concept label (TCL) is a Linnaean name plus a reference to a publication, where the discussed taxon is circumscribed. The link is via the keyword "sec." (Latin for *secundum*) [29]. An example would be "*Andropogon virginicus* var. *tenuispathus* sec. Blomquist (1948)". Here, Blomquist (1948) is

```

<http://dx.doi.org/10.3897/BDJ.1.e953>
  :contains :abstract, :casuarinicola-australis-treatment .

:introduction rdf:type deo:Introduction, doco:Section ;
  c4o:hasContent "Casuarinicola australis Taylor, 2010 was described from
  Australia, where it is the most common and widespread member of its
  genus, being widely distributed in New South Wales, Queensland,
  South Australia, Victoria and Western Australia. "

:casuarinicola-australis-treatment rdf:type doco:Section, :Treatment ;
  :contains :casuarinicola-australis-nomenclature ,
    :casuarinicola-australis-materials ,
    :casuarinicola-australis-description ,
    :figure-box-1 ,
    :figure-box-2 .

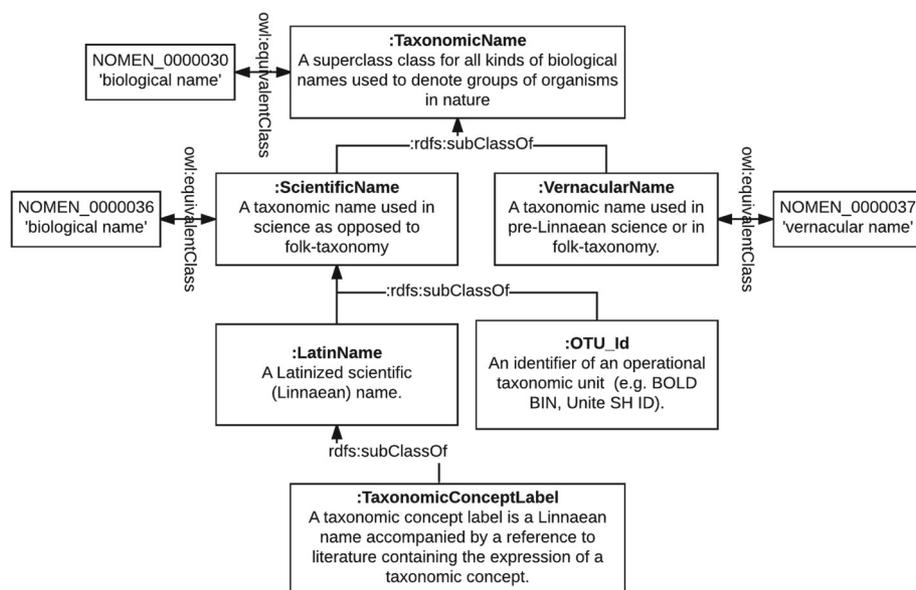
:casuarinicola-australis-nomenclature rdf:type :NomenclatureSection ;
  :contains :casuarinicola-australis-nomenclature-heading .

:casuarinicola-australis-nomenclature-heading a :NomenclatureHeading ;
  cnt:chars "Casuarinicola australis Taylor, 2010" .

:casuarinicola-australis-materials rdf:type :MaterialsExamined ;
  c4o:hasContent "country: New Zealand;
  verbatimLocality: Mechanics Bay, Auckland City;
  verbatimElevation: 0-5 m;
  verbatimLatitude: 36.8474938105S ;
  verbatimLongitude: 174.7869624545E ;
  eventDate: 6 January 2013;
  sex: 1 male, 1 female;
  recordedBy: Stephen Thorpe;
  institutionCode: Auckland Museum" .

:casuarinicola-australis-description rdf:type :DescriptionSection ;
  c4o:hasContent "On 6 Jan 2013, I examined some Casuarina glauca trees growing
  in the vicinity of Ports of Auckland at Mechanics Bay." .
  
```

**Fig. 3** Example article structure. This examples shows how to express the article structure with the help of :contains. The code is in Turtle



**Fig. 4** Taxonomic name class hierarchy diagram. We created this class hierarchy to accommodate both traditional taxonomic name usages and the usage of taxonomic concept labels and operational taxonomic units

**Table 2** OpenBiodiv taxonomic status vocabulary

Vocabulary instance QName	Example abbrev	Comment
:TaxonomicUncertainty	<i>incertae sedis</i>	Taxonomic uncertainty
:TaxonDiscovery	<i>sp. n.</i>	Taxonomic discovery
:ReplacementName	<i>comb. n.</i>	Replacement name
:UnavailableName	<i>nomen dubium</i>	Unavailable name
:AvailableName	<i>stat. rev.</i>	Available name
:TypeSpecimenDesignation	<i>lectotype designation</i>	Type specimen designation
:TypeSpeciesDesignation	<i>type species</i>	Type species designation
:NewOccurrenceRecord	<i>new country record</i>	New occurrence record (for region)

a reference to [56], the publication where the concept is circumscribed.

We extracted taxonomic status abbreviations from about 4000 articles across four taxonomic journals (ZooKeys, Biodiversity Data Journal, PhytoKeys, and MycoKeys) in order to create a taxonomic status vocabulary (Additional file 2) that covers the eight most common cases (Table 2). The Latin abbreviations that have been classified into these classes can be found on the OpenBiodiv GitHub page [46] (See “Methods” section for more details).

Based on our analysis of taxonomic statuses, we have identified two Code-compliant patterns of relationship between latinized scientific names (Fig. 5). The pattern *replacement name*, implemented via the property `:replacementName`, indicates that a certain Linnaean name should be used instead of another Linnaean name. It covers a wide variety of cases in the Codes, such as, for example, the placement of one species taxon in a new genus (“comb. n.”), the correction of a name for nomenclatural reasons (“nomen novum”), or the application of the Principle of Priority for the discovery of synonyms (“syn. nov.”) [23].

The other pattern is that of *related names* (`:relatedName`). It is a broader pattern, indicating that two names are somehow related. For example, they may be synonyms, with one replacing the other, or they may point to taxonomically related taxonomic concepts. For example, *Harmonia manillana* (Mulsant, 1866) is related to *Caria manillana* Mulsant 1866 since, as per

[57], a name-bearing type (lectotype) of *Harmonia manillana* (Mulsant, 1866) sec. Poorani [57] is named *Caria manillana* Mulsant 1866.

### Semantics, alignment and usage

As evident from Fig. 4, OpenBiodiv-O taxonomic names are aligned to NOMEN names.

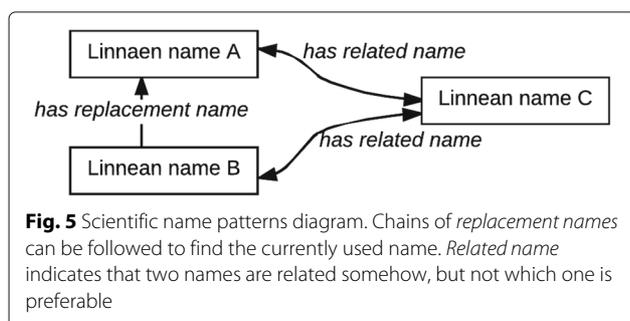
The linking between text and taxonomic names must pass through the intermediary class Taxonomic Name Usage. As parts of the manuscript, taxonomic name usages link document components to taxonomic names. Taxonomic name usages are *contained* in sections such as Treatment, and *mention* a taxonomic name as illustrated in the example in Fig. 6.

### Semantic modeling of the taxonomic concepts

In OpenBiodiv-O taxonomic names are not the carriers of semantic information about taxa. This task is accomplished by a new class, Taxonomic Concept (`:TaxonomicConcept`). A taxonomic concept is the theory that a taxonomist forms about a taxon in a scholarly biological taxonomic publication and thus always has a taxonomic concept label. We also introduce a more general class, Operational Taxonomic Unit (`:OperationalTaxonomicUnit`) that can be used for all kinds of taxonomic hypotheses, including ones that don't have a proper taxonomic concept label. The class hierarchy has been illustrated in Fig. 7.

Taxonomic concepts are related to taxonomic names—including taxonomic concept labels—via the property *has taxonomic name* (`:taxonomicName`) and its sub-properties mimicking in their range the hierarchy of taxonomic names that we introduced earlier. We have defined a property specifically to link taxonomic concepts to taxonomic concept labels, *has taxonomic concept label* (`:taxonomicConceptLabel`). The property hierarchy diagram is shown in Fig. 8.

There are two ways to relate taxonomic concepts to each other (Fig. 9). As we pointed out earlier, historically taxonomic concepts form the hierarchy known as biological taxonomy. To express such simple semantic



```

:casuarinicola-australis-nomenclature-heading
  po:contains :casuarinicola-australis-TNU .

:casuarinicola-australis-TNU a :TaxonomicNameUsage ;
  dc:date "2013-9-16"^^:xsd:date ;
  cnt:chars "Casuarinicola australis Taylor, 2010" ;
  dwc:genus "Casuarinicola" ;
  dwc:specificEpithet "australis" ;
  dwc:scientificNameAuthorship "Taylor, 2010" ;
  # we can infer the following because we are in the treatment heading
  dwc:nameAccordingToId "doi: 10.3897/BDJ.1.e953" ;
  pkm:mentions :casuarinicola-australis-taylor,
               :casuarinicola-australis-taylor-sec-thorpe-2013 .

:casuarinicola-australis-taylor a :ScientificName ;
  rdfs:label "Casuarinicola australis Taylor, 2010" ;
  dwc:genus "Casuarinicola" ;
  dwc:specificEpithet "australis" ;
  dwc:scientificNameAuthorship "Taylor, 2010" .

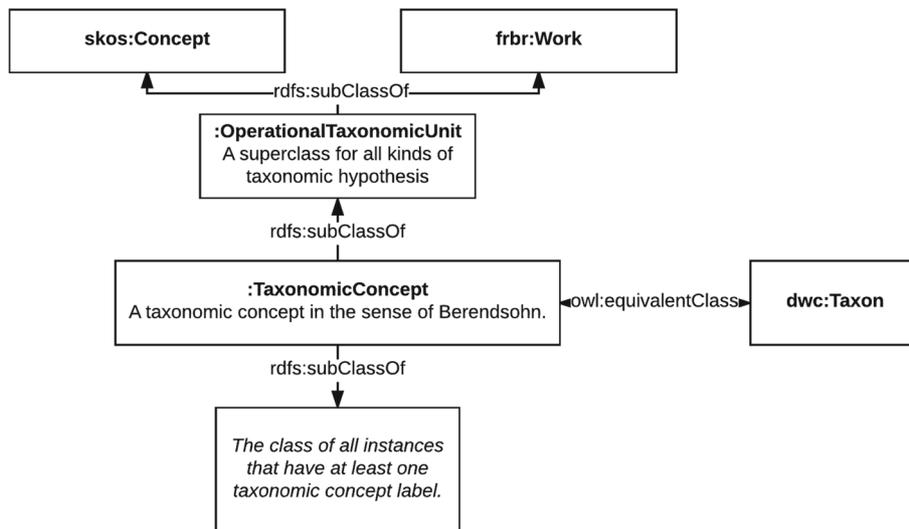
:casuarinicola-australis-taylor-sec-thorpe-2013 a :TaxonomicConceptLabel ;
  rdfs:label "Casuarinicola australis Taylor, 2010 sec. Thorpe 2013" ;
  dwc:genus "Casuarinicola" ;
  dwc:specificEpithet "australis" ;
  dwc:scientificNameAuthorship "Taylor, 2010" .
  dwc:nameAccordingToId "doi: 10.3897/BDJ.1.e953" ;
  :nameAccordingTo <http://dx.doi.org/10.3897/BDJ.1.e953> .
    
```

**Fig. 6** Example taxonomic name usage. This examples shows how taxonomic name usages link document components to taxonomic names. The code is in Turtle

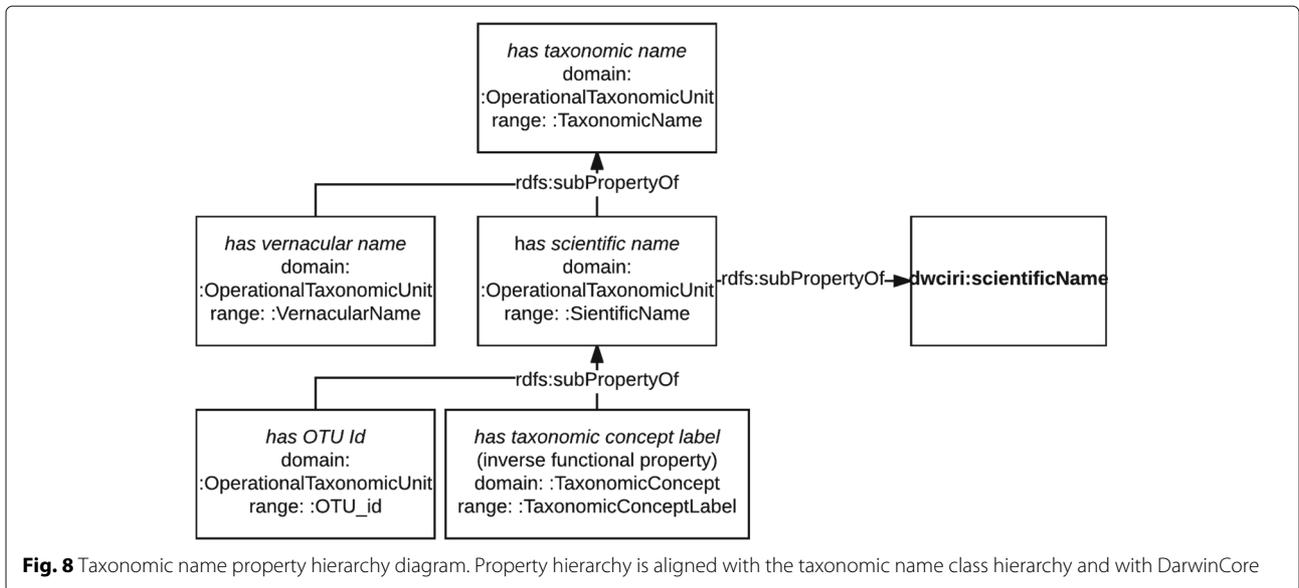
relations, it is fully sufficient to use the SKOS semantic vocabulary [50].

However, these simple relationships are not well suited for machine reasoning. This is why Franz and Peet [30] suggested, building on previous work by e.g. [58], to use the RCC-5 language to express relationships between

taxonomic concepts. Furthermore, the Euler [59] program was developed, which uses Answer Set Programming (ASP) to reason over RCC-5 taxonomic relationships. An answer set reasoner is not part of OpenBiodiv as this task can be accomplished by Euler; however, we have provided an RCC-5 dictionary class (:RCC5Dictionary), an



**Fig. 7** Taxonomic concept diagram. A taxonomic concept is a skos : Concept, a frbr : Work, a dwc : Taxon and has at least one taxonomic concept label

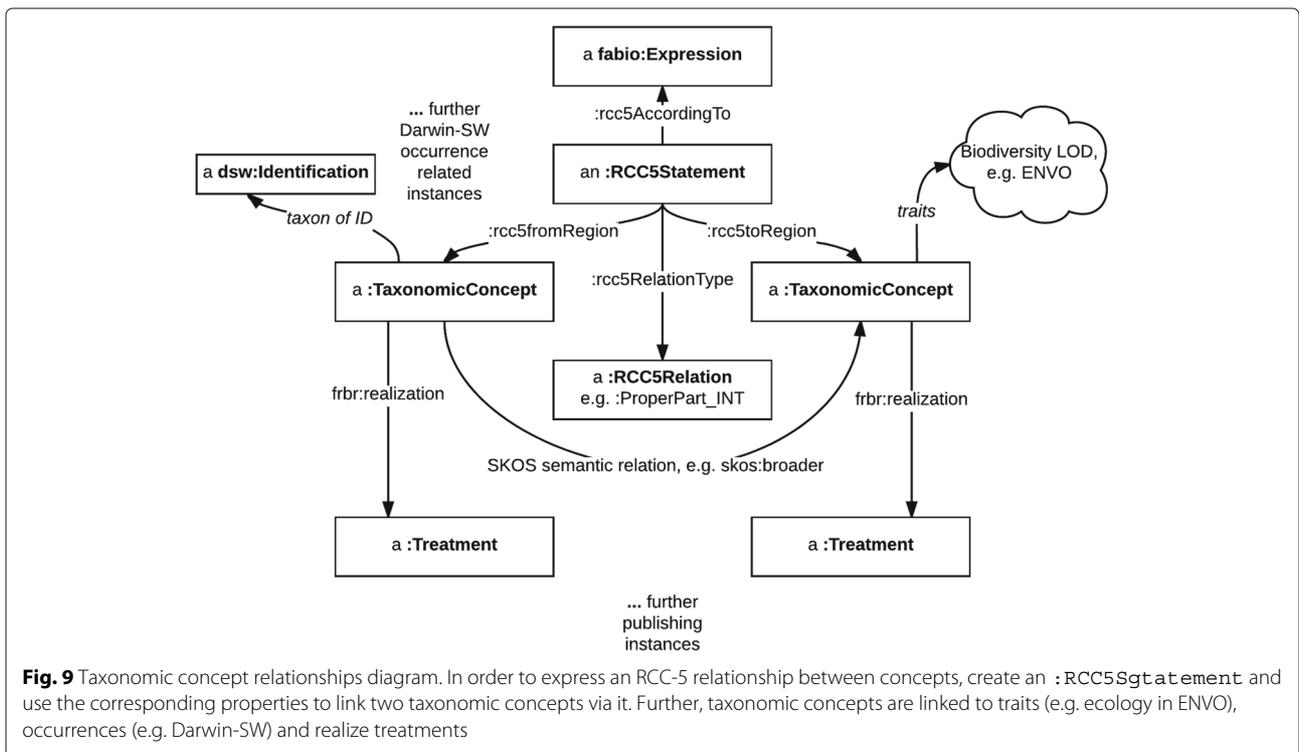


RCC-5 relation term class (`:RCC5Relation`), a vocabulary of such terms to express the RCC-5 relationships in RDF (Additional file 3), as well as a class and properties to express RCC-5 statements (`:RCC5Statement`, `:rcc5Property`, and subproperties).

**Semantics and alignment**

We introduce Taxonomic Concept as equivalent (`owl:equivalentClass`) to the DwC term Taxon

(`dwc:Taxon`) [60]. However, by including “concept” in the class’ name, we highlight the fact that the semantics it carries reflect the scientific theory of a given author about a taxon in nature. As we mentioned earlier, our ontology models the ongoing still unfinished process of taxonomic discovery. For this reason, we also derive Taxonomic Concept from Work. This derivation fits the definition of Work in FRBR/FaBiO, which is “a distinct intellectual or artistic creation.” Finally, as we use SKOS to connect



taxonomic concepts to each other, we derive Taxonomic Concept from SKOS Concept.

As with other semantic publishing-related aspects of the ontology, the creation of the RCC-5 vocabulary follows the SPAR Ontologies' model. Thus OpenBiodiv RCC-5 Vocabulary (:RCC5RelationshipTerms) is a SKOS concept scheme and every RCC-5 Relation is a SKOS concept. This allows to seamlessly share this vocabulary with other publishers of biodiversity information that also follow the SPAR Ontologies' model.

It is important to note that we have aligned the subproperty of *has taxonomic name*, *has scientific name* (:scientificName), to the DwC property *dwciri:scientificName*. The difference is that while the DwC property is unbound and provides more flexibility, the OpenBiodiv-O property has the domain Taxonomic Concept and the range Scientific Name and provides for inference. Furthermore, *has taxonomic concept label* is an inverse-functional property with the domain Taxonomic Concept. This means that a given taxonomic concept label uniquely determines its taxonomic concept. This is accomplished by a minimum cardinality restriction on the property.

Together with the declaration of *has taxonomic concept label* to be an inverse functional property, we can now list what types of relationships between names and taxonomic concepts are allowed: (1) The relationship between a taxonomic concept and a name that is not a taxonomic concept label is many-to-many—i.e. one Linnaean name can be a mention of multiple taxonomic concepts, and one taxonomic concept may have multiple Linnaean names. (2) The relationship between a taxonomic concept and a taxonomic concept label is one-to-many: while a taxonomic concept may have more than one (at least one is needed) labels, every label uniquely identifies a concept. These logical restrictions make taxonomic concept labels into unique identifiers to taxonomic concepts, something that Linnaean names are not.

### Usage

For an example of linking two taxonomic concepts to each other, let us look at the species-rank concept *Casuarinicola australis* Taylor, 2010 sec. Thorpe [61]. It is a narrower concept than the genus-rank concept of *Casuarinicola* Taylor, 2010 sec. Taylor [62]. As we have

aligned our concepts to SKOS, we can use its vocabulary to express this statement as seen in the example in Fig. 10. A further example of how to utilize the OpenBiodiv RCC-5 vocabulary is found in Fig. 11.

Furthermore, thanks to the alignment to DwC, we treat instances of our class Taxonomic Concept as functionally equivalent to DwC Taxa. This makes linking to other biodiversity ontologies possible. For example, the Open Biomedical Ontologies' (OBO) Population and Community Ontology (PCO) [63] has a class "collection of organisms" ([http://purl.obolibrary.org/obo/PCO\\_0000000](http://purl.obolibrary.org/obo/PCO_0000000)) that can be considered a superclass of DwC Taxon. Therefore, every taxonomic concept is a collection of organisms and the application of OBO properties on it is allowed.

In the paper that inspired our *Casuarinicola* example [61], we read: "On 26 February 2013, the species was found to be fairly common on *Casuarina* trees at Thomas Bloodworth Park, Auckland." This statement can be interpreted (in ENVO) as meaning that the taxonomic concept that the author formulated implies that it includes the habitat "forest biome" ([http://purl.obolibrary.org/obo/RO\\_0002303](http://purl.obolibrary.org/obo/RO_0002303)). The RDF example is shown in Fig. 12.

As we pointed out earlier, taxonomic concepts have an intensional component (traits or characters) and an ostensive component (a list of occurrences belonging to the concept). The ostensive component can be expressed by linking occurrences to the taxonomic concepts via Darwin-SW. This is possible as we have aligned the Taxon Concept class to DwC Taxon used by Darwin-SW. For an example refer to [27].

Lastly, describing traits is an active area of ontological research [64]. Due to the very complex language used to describe morphological characteristics, the Ontology Term Organizer (OTO) [64] software was developed to allow for user-created vocabularies. We will rely on such external efforts for expressing traits and trait equivalences (in the taxonomic sense) during the population of OpenBiodiv with triples. We are tightly working with the developers of OTO to integrate their efforts into OpenBiodiv [65].

Further, the interpretation of Taxonomic Concepts as Work means that they are realized by taxonomic treatments (e.g. Fig. 13).

```

:concept-casuarinicola-australis-thorpe rdf:type :TaxonomicConcept ;
:taxonomicConceptLabel :casuarinicola-australis-taylor-sec-thorpe-2013 .

:concept-casuarinicola-taylor rdf:type :TaxonomicConcept ;
skos:broader concept-thorpe .

```

**Fig. 10** Example simple taxonomic concept relationships. We can use SKOS semantic properties to illustrate simple relationships between taxonomic concepts

```

:statement rdf:type :RCC5Statement ;
:rcc5FromRegion :concept-casuarinicola-australis-thorpe ;
:rcc5ToRegion :concept-casuarinicola-taylor ;
:rcc5AccordingTo <http://dx.doi.org/10.3897/BDJ.1.e953> ;
:rcc5RelationType :ProperPart_INT .

```

**Fig. 11** Example of RCC-5 taxonomic concept relationships. In order to express an RCC-5 relationship between concepts, create an `:RCC5Statement` and use the corresponding properties to link two taxonomic concepts via it. SKOS relations relate concepts directly

## Discussion

OpenBiodiv-O is—together with the Treatment Ontologies [20]—the first effort to model taxonomic articles as RDF. It introduces classes and properties in the domains of biodiversity publishing and biological taxonomy and aligns them with the SPAR Ontologies, the Treatment Ontologies, the Open Biomedical Ontologies (OBO), Tax-Pub, NOMEN, and DarwinCore. We believe this introduction bridges the ontological gap that we had outlined in our aims and allows for the creation of a Linked Open Dataset (LOD) of biodiversity information (biodiversity knowledge graph [8, 66]).

Furthermore, this biodiversity knowledge graph, together with this ontology, additional semantic rules, and user software will form the OpenBiodiv Knowledge Management System. This system, as any taxonomic information system should, has taxonomic names as a key building block. For any given taxonomic name, the user will be able to rely on two patterns—*replacement name* and *related name*—to get answers to two questions of high importance to the working taxonomist. First: what is the current and historical usage of any given Linnaean name? Second: given a particular name, what other related names ought to be considered in a taxonomic discussion?

Both may be useful in building semantic search applications and the latter, in particular, is actively being researched by a group at the National Center for Text Mining in the UK (NaCTeM) [67]. OpenBiodiv-O proper does not include a mechanism for inferring replacement names and related names; however, such mechanisms are part of the OpenBiodiv knowledge system via SPARQL rules using information encoded in the document structure (Nomenclature section). Another way to infer related names is via a machine learning approach to obtain feature vectors of taxonomic names. Note that the ontology can describe related names independent of the process of

their generation and will enable the comparison of both approaches in a future work.

On the other hand, by using OpenBiodiv-O, a knowledge-based system does not have to have a backbone name-based taxonomy. A backbone taxonomy is a single, monolithic hierarchy in which any and all conflicts or ambiguities have been pragmatically (socially, algorithmically) resolved, even if there is no clear consensus in the greater taxonomic domain. Such backbone taxonomies are used in systems that rely solely on taxonomic names (and not concepts) as bearers of information. They are needed as it is impossible, in such a system, to express two different sets of statements for a single name.

In OpenBiodiv, however, multiple hierarchies of taxonomic concepts may exist. For example, large synthetic taxonomies such as GBIF's backbone taxonomy [68] or Catalogue of Life [69] may not agree or may have some issues [70]. With OpenBiodiv-O, we may, in fact, incorporate both these taxonomies at the same time! It is possible according to the ontology to have two sets of taxonomic concepts (even with the same taxonomic names) with a different hierarchical arrangement. By allowing this, we leave some room for human interpretation as an additional architectural layer. Thus, we delay the decision of which hierarchy to use to the user of the system (e.g. a practicing taxonomist) and not to the system's architect. Due to this design feature, it is likely that our system stands a better chance to be trusted as a science process-enabling platform as the system architects don't force a taxonomic opinion on the practicing taxonomist.

It should be noted that a successful concept-based system exists for the taxonomic order Aves (birds) [71]. The main issue that we will face is to develop tools to enable expert users to annotate taxonomic concepts with the proper relationships as only recently individual articles utilizing concept taxonomy in addition to nomenclature have been published [43, 72, 73]. We do believe that their

```

:australian-casuarina-forest rdf:type <http://purl.obolibrary.org/obo/ENVO_01000174> .
:hasHabitat owl:sameAs <http://purl.obolibrary.org/obo/RO_0002303> .
:concept-casuarinicola-australis-thorpe :hasHabitat :australian-casuarina-forest .

```

**Fig. 12** Example of combining ENVO with OpenBiodiv-O. We create a shortcut for *has habitat* and instance of the “forest biome” and link them to our taxonomic concept in order to express the fact that specimens of it have been found to live in *Casuarina* trees

```
:casuarinicola-australis-treatment frbr:realizationOf :concept-casuarinicola-australis-thorpe.
```

**Fig. 13** Example connection between a treatment and a taxonomic concept. A treatment is the realization of a taxonomic concept

numbers will rise driven by the realization that there are some problems with relying solely on Linnaean names for the identification of taxonomic concepts [5, 74, 75]. Concept taxonomy may, in fact, become even more important in the future as conservation efforts face challenges due to unresolved taxonomies [76]. Properly aligning taxonomic concepts to nomenclature across revisions [77] may be the solution.

Together with taxonomic information, the ontology allows modeling the source information in a knowledge base. This will be useful for metastudies, for the purposes of reproducible research, and other scholarly purposes. Moreover, it will be an expert system as the knowledge extracted will come from scholarly publications. We envision the system to be able to address a wide variety of taxonomic competency questions raised by researchers during pro-iBiosphere [78]. Examples include: “Is X a valid taxonomic name (in a nomenclatorial sense)?” “Which treatments use different names for the same taxon concepts?” “Which treatments are nomenclatorially linked (including homonyms!) to another treatment?”

Out immediate next efforts will be concentrated on populating the ontology with triples extracted from prospectively published Pensoft journals [79], legacy journals text-mined by Plazi [80], as well as databases such as GBIF and Bioimages [81]. Special effort will be made to link the dataset to the Linked Open Data cloud via resources such as geographic or institution names. In terms of extending the ontological model, more research needs to go into modeling the taxonomic concept circumscription—creating ontologies for morphological, genomic, or ecological traits. Also possibly refining the RCC-5 statements informed by the actual implementation. A study will be carried out to investigate the usefulness of the ontology once the LOD dataset had been created in a real-world scenario.

## Conclusions

The paper provides an informal conceptualization of the taxonomic process and a formalization in OpenBiodiv-O. It introduces classes and properties in the domains of biodiversity publishing and biological systematics and aligns them with the important domain-specific ontologies. By bridging the ontological gap between the publishing and the biodiversity domains, it will enable the creation of Open Biodiversity Knowledge Management System, consisting of (1) the ontology itself; (2) a Linked Open Dataset (LOD) of biodiversity information (biodiversity knowledge graph); and (3) user interface components

aimed at searching, browsing and discovering knowledge in big corpora of previously dispersed scholarly publications. Through the usage of taxonomic concepts, we have included mechanisms for democratization of the scholarly process and not forcing a taxonomic opinion on the users.

## Additional files

**Additional file 1:** Ontology is a plain text file containing statements in the Turtle syntax forming OpenBiodiv-O. It can be edited in a text (e.g. Sublime Text, Emacs, etc.) or in an ontology editor (e.g. Protégé). It can be loaded into a triple store (e.g. GraphDB). The prefixes that are used throughout this manuscript are defined at the beginning. This file corresponds to <http://openbiodiv.net/openbiodivo-20171103>. (TXT 22 kb)

**Additional file 2:** Vocabulary of Taxonomic Statuses is a plain text file containing statements in the Turtle syntax forming the OpenBiodiv Vocabulary of Taxonomic Statuses. Like the ontology [Additional file 1] it can be edited in a text or ontology editor or loaded in a triple store. Make sure you also load the ontology first. (TXT 7 kb)

**Additional file 3:** RCC-5 Vocabulary is a plain text file containing statements in the Turtle syntax forming the OpenBiodiv RCC-5 Vocabulary. Like the ontology [Additional file 1] it can be edited in a text or ontology editor or loaded in a triple store. Make sure you also load the ontology first. (TXT 5 kb)

## Abbreviations

LOD: Linked open data; OWL: W3C web ontology language; RCC-5: Region connection calculus 5; RDF: Resource description framework; RDFS: RDF schema; SPARQL: SPARQL protocol and RDF query language; XML: Extensible markup language

## Acknowledgements

We acknowledge É. Ó Tuama and D. Mietchen for the many helpful discussions that lead to theoretical contributions. We also acknowledge the programming team at Pensoft and in particular Georgi Zhelezov for web-development in PHP and JavaScript.

## Funding

Research financed through the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 642241.

## Availability of data and materials

A partial dataset from Pensoft’s journals has been generated with OpenBiodiv-O and can be found at the SPARQL Endpoint <http://213.145.125.72:7777/>, select repository `obkms_i6`. The endpoint is also accessible from the website, <http://openbiodiv.net/>. Demos are available as “Saved Queries” from the workbench and from the website.

## Authors’ contributions

VS: Marie-Sklodowska-Curie Ph.D. student, whose main project is OpenBiodiv. LP: principal investigator, the main academic supervisor of VS, supported each step of the way. KS and NF are co-advisors. PS consulted on the taxonomic process and on the development of the Taxonomic Status Vocabulary. KS consulted on ontological development, proof-read and improved the manuscript. NF consulted on concept taxonomy and the vision of the system, also proof-read and improved the manuscript. TC and RAM are the main authors of the Treatment Ontologies, which serve as a conceptual template for OpenBiodiv-O. They also provided proof-reading and improvements to the text. DA and GS provided many insights into biodiversity publishing. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

Not applicable

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Pensoft Publishers, Prof. Georgi Zlatarski 12, 1700 Sofia, Bulgaria. <sup>2</sup>Institute of Biodiversity and Ecosystems Research, Bulgarian Academy of Sciences, Sofia, Bulgaria. <sup>3</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria. <sup>4</sup>Arizona State University, School of Life Sciences, Tempe Campus, 4501 Tempe, AZ, USA. <sup>5</sup>Plazi, Bern, Switzerland. <sup>6</sup>University of Massachusetts at Boston, Boston, USA. <sup>7</sup>National Museum of Natural History, 1 Tsar Osvoboditel Blvd., Sofia 1000, Bulgaria.

Received: 18 August 2017 Accepted: 28 December 2017

Published online: 18 January 2018

**References**

- TDWG Past Meetings. <http://www.tdwg.org/past-meetings/>. Accessed 12 Aug 2017.
- What is GBIF. <http://www.gbif.org/what-is-gbif>. Accessed 12 Aug 2017.
- Bouchout Declaration. <http://www.bouchoutdeclaration.org>. Accessed 09 Aug 2017.
- pro-iBiosphere. <http://wiki.pro-ibiosphere.eu/>. Accessed 12 Aug 2017.
- Patterson DJ, Cooper J, Kirk PM, Pyle RL, Remsen DP. Names are key to the big new biology. *Trends Ecol Evol*. 2010;25(12):686–91. <https://doi.org/10.1016/j.tree.2010.09.004>. Accessed 11 July 2017.
- Pyle R. Towards a Global Names Architecture: The future of indexing scientific names. *ZooKeys*. 2016;550:261–81. <https://doi.org/10.3897/zookeys.550.10009>. Accessed 12 Aug 2017.
- pro-iBiosphere. Final OBKMS Brochure. Technical report 2014. [adm.pro-ibiosphere.eu/getatt.php?filename=oo\\_4749.pdf](http://adm.pro-ibiosphere.eu/getatt.php?filename=oo_4749.pdf). Accessed 12 Aug 2017.
- Senderov V, Penev L. The Open Biodiversity Knowledge Management System in Scholarly Publishing. *Res Ideas Outcomes*. 2016;2:7757. <https://doi.org/10.3897/rio.2.e7757>. Accessed 22 July 2017.
- Pensoft Plazi. OpenBiodiv knowledge system website 2017. <http://openbiodiv.net>. Accessed 12 Jan 2018.
- Momtchev V, Peychev D, Primov T, Georgiev G. Expanding the pathway and interaction knowledge in linked life data. *Proc Int Semantic Web Chall*. 2009. Accessed 22 July 2017.
- Williams AJ, Harland L, Groth P, Pettifer S, Chichester C, Willighagen EL, Evelo CT, Blomberg N, Ecker G, Goble C, Mons B. Open PHACTS: semantic interoperability for drug discovery. *Drug Discov Today*. 2012;17(21–22):1188–98. <https://doi.org/10.1016/j.drudis.2012.05.016>. Accessed 22 July 2017.
- Rebholz-Schuhmann D, Kirsch H, Couto F. Facts from text—is text mining ready to deliver? *PLoS Biol*. 2005;3(2):65. Accessed 22 July 2017.
- Kennedy JB, Kukla R, Paterson T. Scientific Names Are Ambiguous as Identifiers for Biological Taxa: Their Context and Definition Are Required for Accurate Data Integration. In: Ludäscher B, Raschid L, editors. *Data Integration in the Life Sciences: Second International Workshop, DILS 2005, San Diego, CA, USA, July 20–22, 2005. Proceedings*. Berlin, Heidelberg: Springer; 2005. p. 80–95. [https://doi.org/10.1007/11530084\\_8](https://doi.org/10.1007/11530084_8).
- Penev L, Kress WJ, Knapp S, Li DZ, Renner S. Fast, linked, and open – the future of taxonomic publishing for plants: launching the journal *PhytoKeys*. *PhytoKeys*. 2010;1(0). <https://doi.org/10.3897/phytokeys.1.642>. Accessed 22 July 2017.
- Tzitzikas Y, Allocca C, Bekiari C, Marketakis Y, Fafalios P, Doerr M, Minadakis N, Patkos T, Candela L. Integrating heterogeneous and distributed information about marine species through a top level ontology. In: *Research Conference on Metadata and Semantic Research*. Springer; 2013. p. 289–301. [https://link.springer.com/chapter/10.1007/978-3-319-03437-9\\_29](https://link.springer.com/chapter/10.1007/978-3-319-03437-9_29). Accessed 22 July 2017.
- Peroni S. The semantic publishing and referencing ontologies. In: *Semantic Web Technologies and Legal Scholarly Publishing*. 1st edn. Cham: Springer; 2014. p. 121–93.
- Peroni S, Shotton D. FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. *Web Semant Sci Serv Agents World Wide Web*. 2012;17:33–43. <https://doi.org/10.1016/j.websem.2012.08.001>. Accessed 13 Aug 2017.
- Constantin A, Peroni S, Pettifer S, Shotton D, Vitali F. The Document Components Ontology (DoCO). *Semantic Web*. 2016;7(2):167–81. <https://doi.org/10.3233/SW-150177>. Accessed 13 Aug 2017.
- Catapano T. TaxPub: An Extension of the NLM/NCBI Journal Publishing DTD for Taxonomic Descriptions. In: *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2010 [Internet]*. Bethesda: National Center for Biotechnology Information (US); 2010. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK47081/>.
- Catapano T, Morris RA. Treatment Ontologies. <https://github.com/plazi/TreatmentOntologies/blob/master/treatment.owl>. Accessed 09 Aug 2017.
- Linnæus C. In: Tomus I, editor. *Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*. Holmiæ; 1758, pp. 1–4, 1–824.
- Witteveen J. Naming and contingency: the type method of biological taxonomy. *Biol Philos*. 2015;30(4):569–86. <https://doi.org/10.1007/s10539-014-9459-6>. Accessed 13 Aug 2017.
- ICZN. The International Trust for Zoological Nomenclature, 4th edn. London: International Code of Zoological Nomenclature; 1999, p. 306. xxix.
- McNeill J, for Plant Taxonomy IA, editors. *International Code of Nomenclature for Algae, Fungi and Plants (Melbourne Code): Adopted by the Eighteenth International Botanical Congress Melbourne, Australia, July 2011. Regnum vegetabile, vol. v. 154*. Königstein, Germany: Koeltz Scientific Books; 2012. OCLC: ocn824722354.
- Dmitriev DA, Yoder M. NOMEN. <https://github.com/SpeciesFileGroup/nomen>. Accessed 22 July 2017.
- Morris PJ, Morris RA, Wang Z. Taxonomic Nomenclatural Status Terms (version 0.8). [https://github.com/pensoft/OpenBiodiv/blob/master/ontology/contrib/taxonomic\\_nomenclatural\\_status\\_terms.owl](https://github.com/pensoft/OpenBiodiv/blob/master/ontology/contrib/taxonomic_nomenclatural_status_terms.owl). Accessed 1 Dec 2018.
- Baskauf S, Webb CO. Darwin-SW: Darwin Core-based terms for expressing biodiversity data as RDF. *Semantic Web J*. 2016;7(6):629–43. <https://doi.org/10.3233/SW-150203>.
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE*. 2012;7(1):29715. <https://doi.org/10.1371/journal.pone.0029715>. Accessed 22 July 2017.
- Berendsohn WG. The concept of “potential taxa” in databases. *Taxon*. 1995;44(2):207–12.
- Franz NM, Peet RK. Perspectives: Towards a language for mapping relationships among taxonomic concepts. *Syst Biodivers*. 2009;7(1):5–20. <https://doi.org/10.1017/S147720000800282X>. Accessed 22 July 2017.
- Sternberg F, Franz NM. Taxonomy for Humans or Computers? Cognitive Pragmatics for Big Data. *Biol Theory*. 2012;7(1):99–111. <https://doi.org/10.1007/s13752-017-0259-5>. Accessed 11 July 2017.
- Taxonomic Names and Concepts interest group. Taxonomic Concept Transfer Schema (TCS), version 1.01. *Biodiversity Information Standards (TDWG)*. 2006. <http://www.tdwg.org/standards/117>. Accessed 12 Jan 2018.
- DeVries P. Taxon Concept Ontology. <http://taxonconcept.org>. Accessed 12 Jan 2018.
- Manktelow M. History of taxonomy. 2010. [http://www.atbi.eu/summerschool/files/summerschool/Manktelow\\_Syllabus.pdf](http://www.atbi.eu/summerschool/files/summerschool/Manktelow_Syllabus.pdf). Accessed 22 July 2017.
- Trontelj P, Fiser C. Cryptic species should not be trivialized. *Syst Biodivers*. 2009;7:1–23.
- Mallet J. Species, concepts of. In: *Encyclopedia of Biodiversity*, vol. 5.; 2001. p. 427–40. <http://tarjomefa.com/wp-content/uploads/2016/02/4420-english.pdf>. Accessed 11 July 2017.
- Deans AR, Yoder MJ, Balhoff JP. Time to change how we describe biodiversity. *Trends Ecol Evol*. 2012;27(2):78–84. <https://doi.org/10.1016/j.tree.2011.11.007>. Accessed 11 July 2017.
- Sokal RR. The Principles and Practice of Numerical Taxonomy. *Taxon*. 1963;12(5):190. <https://doi.org/10.2307/1217562>. Accessed 12 Aug 2017.

39. Platnick NI. From cladograms to classifications: The road to DePhylocode. Syst Assoc. 2001. <http://www.systass.org/archive/events-archive/2001/platnick.pdf>. Accessed 12 Jan 2018.
40. Page RDM. DNA barcoding and taxonomy: dark taxa and dark texts. *Philos Trans R Soc B Biol Sci*. 2016;371(1702):20150334. <https://doi.org/10.1098/rstb.2015.0334>. Accessed 07 Aug 2017.
41. Ratnasingham S, Hebert PDN. A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLoS ONE*. 2013;8(7):66213. <https://doi.org/10.1371/journal.pone.0066213>. Accessed 22 July 2017.
42. Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, Douglas B, Drenkhan T, Eberhardt U, Dueñas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lücking R, Martín MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Põldmaa K, Saag L, Saar I, Schüßler A, Scott JA, Senés C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M, Larsson KH. Towards a unified paradigm for sequence-based identification of fungi. *Mol Ecol*. 2013;22(21):5271–7. <https://doi.org/10.1111/mec.12481>. Accessed 12 Aug 2017.
43. Franz NM, Pier NM, Reeder DM, Chen M, Yu S, Kianmajid P, Bowers S, Ludäscher B. Two influential primate classifications logically aligned. *Syst Biol*. 2016;65(4):561–82. Accessed 11 July 2017.
44. Knuth DE. Literate programming. *Comput J*. 1984;27(2):97–111. Accessed 08 Aug 2017.
45. Senderov V, Franz NM, Simov K. OpenBiodiv Ontology and Guide. 2017. <http://openbiodiv.net/ontology>. Accessed 09 Aug 2017.
46. Senderov V. OpenBiodiv GitHub Repository. <https://github.com/pensoft/OpenBiodiv>. Accessed 12 Jan 2018.
47. Gruber TR. A translation approach to portable ontology specifications. *Knowl Acquis*. 1993;5(2):199–220. <https://doi.org/10.1006/knac.1993.1008>. Accessed 07 Aug 2017.
48. Obitko M. Translations between ontologies in multi-agent systems. Thesis Ph. D. Prague: Czech Technical University; 2007.
49. In: Staab S, Studer R, editors. *Handbook on Ontologies*. Berlin, Heidelberg: Springer; 2009. <https://doi.org/10.1007/978-3-540-92673-3>. <http://link.springer.com/10.1007/978-3-540-92673-3>. Accessed 07 Aug 2017.
50. Miles A, Bechofer S. SKOS Simple Knowledge Organization System RDF Schema. <https://www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html>. Accessed 08 Sept 2017.
51. Peroni S. Example of use of DoCO #2. Figshare. 2015. <http://doi.org/10.6084/m9.figshare.1513733>. [http://figshare.com/articles/Example\\_of\\_use\\_of\\_DoCO\\_2/1513733](http://figshare.com/articles/Example_of_use_of_DoCO_2/1513733). Accessed 12 Aug 2017.
52. Smith V, Georgiev T, Stoev P, Biserkov J, Miller J, Livermore L, Baker E, Mietchen D, Couvreur T, Mueller G, Dikow T, Helgen KM, Frank J, Agosti D, Roberts D, Penev L. Beyond dead trees: integrating the scientific process in the Biodiversity Data Journal. *Biodivers Data J*. 2013; 1:995. <http://doi.org/10.3897/BDJ.1.e995>. Accessed 10 Aug 2017.
53. Tillert B. A Conceptual Model for the Bibliographic Universe vol. 25: Technicalities; 2003. <http://www.loc.gov/cds/downloads/FRBR.PDF>. Accessed 08 Aug 2017.
54. Damova M, Kiryakov A, Simov K, Petrov S. Mapping the central LOD ontologies to PROTON upper-level ontology. In: *Proceedings of the 5th International Conference on Ontology Matching-Volume 689*. CEUR-WS.org; 2010. p. 61–72. <http://dl.acm.org/citation.cfm?id=2878599>. Accessed 10 Aug 2017.
55. Pyle R. Taxonomic name usage files. 2016. <http://lists.tdwg.org/pipermail/tdwg-content/2016-April/003582.html>. Accessed 13 Aug 2017.
56. Blomquist H. *The Grasses of North Carolina*. Durham: Duke University Press; 1948.
57. Poorani J, Booth R. *Harmonia manillana* (Mulsant), a new addition to Indian Coccinellidae, with changes in synonymy. *Biodivers Data J*. 2016; 4:8030. <http://doi.org/10.3897/BDJ.4.e8030>. Accessed 13 Aug 2017.
58. Koperski M, Sauer M, Braun W, Gradstein SR. *Referenzliste der Moose Deutschlands* vol. 34. Bonn: Schriftenreihe Vegetationsk; 2000, pp. 1–519.
59. Chen M, Yu S, Franz N, Bowers S, Ludäscher B. Euler/X: a toolkit for logic-based taxonomy integration. 2014;1402:1992. <http://arxiv.org/abs/1402.1992>. Accessed 11 Aug 2017.
60. DarwinCore Terms. <http://rs.tdwg.org/dwc/terms/>. Accessed 12 Jan 2018.
61. Thorpe S. *Casuarinicola australis* Taylor, 2010 (Hemiptera: Triozidae), newly recorded from New Zealand. *Biodivers Data J*. 2013;1:953. <http://doi.org/10.3897/BDJ.1.e953>. Accessed 11 Aug 2017.
62. Taylor G, D. Austin A, Jennings J, Purcell M, Wheeler G. *Casuarinicola*, a new genus of jumping plant lice (Hemiptera: Triozidae) from Casuarina (Casuarinaceae). *Zootaxa*. 2010;2601(2601):1–27.
63. Walls RL, Deck J, Guralnick R, Baskauf S, Beaman R, Blum S, Bowers S, Buttigieg PL, Davies N, Endresen D, Gandolfo MA, Hanner R, Janning A, Krishalka L, Matsunaga A, Midford P, Morrison N, Tuama EO, Schildhauer M, Smith B, Stucky BJ, Thomer A, Wieczorek J, Whitacre J, Woolley J. Semantics in Support of Biodiversity Knowledge Discovery: An Introduction to the Biological Collections Ontology and Related Ontologies. *PLoS ONE*. 2014;9(3):89606. <https://doi.org/10.1371/journal.pone.0089606>. Accessed 30 Oct 2017.
64. Huang F, Macklin JA, Cui H, Cole HA, Endara L. OTO: Ontology Term Organizer. *BMC Bioinformatics*. 2015;16(1). <https://doi.org/10.1186/s12859-015-0488-1>. Accessed 11 Aug 2017.
65. End of Project Workshop of Explorer of Taxon Concepts. Plan for the Next Step. <https://docs.google.com/document/d/1F4vai5R7ygbUD3mopJxVh8ULQfa-x3014tnTBRFVe4/edit?usp=sharing>. Accessed 12 Jan 2018.
66. Page R. Towards a biodiversity knowledge graph. *Res Ideas Outcomes*. 2016;2:8767. <https://doi.org/10.3897/rio.2.e8767>. Accessed 23 July 2017.
67. Nguyen NTH, Soto AJ, Kontonatsios G, Batista-Navarro R, Ananiadou S. Constructing a biodiversity terminological inventory. *PLoS ONE*. 2017;12(4):0175277. <https://doi.org/10.1371/journal.pone.0175277>. Accessed 12 Aug 2017.
68. GBIF Backbone Taxonomy. 2016. <http://www.gbif.org/dataset/d7ddd42cf0-4f39-9b2a-bb099caae36c>. Accessed 12 Aug 2017.
69. Catalogue of Life. <http://www.catalogueoflife.org/>. Accessed 12 Jan 2018.
70. Page RDM. The GBIF classification is broken — how do we fix it? 2012. <http://iphylo.blogspot.bg/2012/05/gbif-classification-is-broken-how-do-we.html>. Accessed 12 Aug 2017.
71. Lepage D, Vaidya G, Guralnick R. Avibase – a database system for managing and organizing taxonomic concepts. *ZooKeys*. 2014;420:117–35. <https://doi.org/10.3897/zookeys.420.7089>. Accessed 13 Aug 2017.
72. Jansen MA, Franz NM. Phylogenetic revision of *Minyomeris* Horn, 1876 sec. Jansen & Franz, 2015 (Coleoptera, Curculionidae) using taxonomic concept annotations and alignments. *ZooKeys*. 2015;528:1–133. <https://doi.org/10.3897/zookeys.528.6001>. Accessed 12 Aug 2017.
73. Franz N, Zhang G. Three new species of entimone weevils in Early Miocene amber from the Dominican Republic (Coleoptera: Curculionidae). *Biodivers Data J*. 2017;5:10469. <http://doi.org/10.3897/BDJ.5.e10469>. Accessed 12 Aug 2017.
74. Remsen D. The use and limits of scientific names in biological informatics. *ZooKeys*. 2016;550:207–23. <http://doi.org/10.3897/zookeys.550.9546>. Accessed 11 July 2017.
75. Franz NM, Chen M, Kianmajid P, Yu S, Bowers S, Weakley AS, Ludäscher B. Names are not good enough: Reasoning over taxonomic change in the *Andropogon* complex1. *Semantic Web*. 2016;7(6):645–67. Accessed 11 July 2017.
76. Garnett ST, Christidis L. Taxonomy anarchy hampers conservation. *Nature*. 2017;546:25–27. doi:10.1038/546025a.
77. Franz N, Zhang C, Lee J. A logic approach to modeling nomenclatural change. 2016. <http://doi.org/10.1101/058834>. Accessed 12 Aug 2017.
78. pro-iBiosphere. Competency Questions for RDF Treatments. 2013. [http://wiki.pro-ibiosphere.eu/wiki/Competency\\_Questions\\_for\\_RDF\\_Treatments](http://wiki.pro-ibiosphere.eu/wiki/Competency_Questions_for_RDF_Treatments). Accessed 12 Aug 2017.
79. Pensoft Journals. <https://pensoft.net/journals>. Accessed 12 Aug 2017.
80. Plazi. Treatment Bank. <http://plazi.org/resources/treatmentbank/>. Accessed 12 Aug 2017.
81. Baskauf S. Bioimages. 2017. <http://bioimages.vanderbilt.edu/>. Accessed 12 Aug 2017.