A Model to Represent Nomenclatural and Taxonomic Information as Linked Data. Application to the French Taxonomic Register, TAXREF

Franck Michel, Olivier Gargominy, Sandrine Tercerie, Catherine Faron Zucker

A Model to Represent Nomenclatural and Taxonomic Information as Linked Data. Application to the French Taxonomic Register, TAXREF

Franck Michel[1][0000-0001-9064-0463], Olivier Gargominy[2], Sandrine Tercerie[2] and Catherine Faron-Zucker[1][0000-0001-5959-5561]

[1] Université Côte d'Azur, Inria, CNRS, I3S, Sophia Antipolis, France
[2] Muséum national d'Histoire naturelle, Paris, France

Abstract. Taxonomic registers are key tools to help us comprehend the diversity of nature. Publishing such registers in the Web of Data, following the standards and best practices of Linked Open Data (LOD), is a way of integrating multiple data sources into a world-scale, biological knowledge base. In this paper, we present an on-going work aimed at the publication of TAXREF, the French national taxonomic register, on the Web of Data. Far beyond the mere translation of the TAXREF database into LOD standards, we show that the key point of this endeavor is the design of a model capable of capturing the two coexisting yet distinct realities underlying taxonomic registers, namely the nomenclature (the rules for naming biological entities) and the taxonomy (the description and characterization of these biological entities). We first analyze different modelling choices made to represent some international taxonomic registers as LOD, and we underline the issues that arise from these differences. Then, we propose a model aimed to tackle these issues. This model separates nomenclature from taxonomy, it is flexible enough to accommodate the ever-changing scientific consensus on taxonomy, and it adheres to the philosophy underpinning the Semantic Web standards. Finally, using the example of TAXREF, we show that the model enables interlinking with third-party LOD data sets, may they represent nomenclatural or taxonomic information.

Keywords: Linked Data, Taxonomy, Nomenclature, Data Integration.

1 Introduction

Started in the early 2000's, the Web of Data has now become a reality [6]. It keeps on growing through the relentless publication and interlinking of data sets spanning various domains of knowledge. Building upon the Linked Data paradigm [5,14] to connect related pieces of data, this new layer of the Web enables the integration of distributed and heterogeneous data sets, spawning an unprecedented, distributed knowledge graph.

A wealth of existing data sources exists out there, that would valuably populate the Web of Data. For instance, taxonomic registers are key tools to comprehend the diversity of nature and develop natural heritage conservation strategies, *e.g.* by crossing the

myriad records of occurrence data and biological traits. Taxonomic registers are commonly used as the backbone of thematic databases and applications, such as the Global Biodiversity Information Facility[1] that aggregates 54 taxonomic data sources. They may adopt a certain perspective and purpose. For instance, Agrovoc [8] is a controlled vocabulary covering all areas of interest of the Food and Agriculture Organization. In this respect, it lists the names of species related to agriculture, fishery and forestry. The NCBI Organismal Classification [12] is another vocabulary covering the organisms specifically referenced in the NCBI nucleotide and protein sequences database. Hence, there does not exist one central register of the taxonomic knowledge. Instead, multiple taxonomic registers cover complementary and often overlapping regions, epochs or domains. Consequently, publishing them as RDF data sets while drawing links between related resources is a way of integrating multiple data sources into a world-scale, biological knowledge graph.

Two coexisting yet distinct realities underlie taxonomic registers, namely the taxonomy (the description and characterization of biological entities called biological taxa, taxon concepts or simply taxa), and the nomenclature (the rules defining how to assign scientific names, or nominal taxa, to these biological entities). The nomenclatural rules are compiled in several Codes. In particular, the Codes for animals [15], plants and fungi [19] and bacteria [17] are used in the TAXREF taxonomic register. The nomenclature yields a controlled thesaurus of scientific names. Each of these scientific names consists of a Latinized name, an authority and a taxonomic rank, along with the original publication and the type specimen bearing that name. Taxonomic registers distinguish each biological taxon from all nominal taxa by retaining a unique reference name for it. For example, taxonomists decided that "*Delphinus capensis* Gray, 1828" and "*Delphinus delphis* Linnaeus, 1758" are the same biological entity, based on morphological or molecular data [10]. In addition to this, the Code of zoological nomenclature rules that this species must be called "*Delphinus delphis* Linnaeus, 1758" as per the principle of priority.

In this paper, we present an on-going work related to TAXREF [13], the French national taxonomic register for fauna, flora and fungus. Our goal is to publish TAXREF on the Web of Data while adhering to standards and best practices for the publication of Linked Open Data (LOD) [11]. First, we analyze how some international taxonomic registers have been published as Linked Data so far. We describe the different modelling choices made to represent the information using the Semantic Web technologies, and the issues that stem from these choices. Then, far beyond the mere translation of the TAXREF database into LOD standards, we show that the key point of this endeavor is the design of a model capable of capturing nomenclatural and taxonomic information. The model we propose has several key advantages: (i) it separates nomenclatural from taxonomic information; (ii) it is flexible enough to accommodate the ever-changing scientific consensus on taxonomy; (iii) it adheres to the philosophy underpinning the Semantic Web standards and it enables drawing links with third-party data sets published as Linked Data, may they represent nomenclatural or taxonomic information.

[1] Global Biodiversity Information Facility: https://www.gbif.org/

The rest of this paper is organized as follows. Section 2 analyzes the Linked Data modelling choices of several taxonomic registers. Section 3 describes the model we propose to distinguish between nomenclature and taxonomy. In section 4, we report on more technical aspects of this work, notably the publication of TAXREF according to this model and the production of rich metadata in line with LOD guidelines. Finally, section 5 draws a few conclusions and envisions future actions to be conducted with the biodiversity community.

2 Representing Taxonomic Registers as Linked Data

Several international taxonomic registers have already been published as Linked Data. They adopt somewhat different approaches to model nomenclatural and/or taxonomic information using the Semantic Web stack of technologies. To figure this out, we looked into the following ones: NCBI Organismal Classification [12], Vertebrate Taxonomy Ontology (VTO) [21], Agrovoc Multilingual agricultural thesaurus [8], Encyclopedia of Life (EOL) [7], GeoSpecies Knowledge Base[2] and TaxonConcept Knowledge Base[3]. We also considered the models of two well-adopted generic data sets: DBpedia [18] and BBC Wildlife Ontology[4]. Fig. 1 illustrates the different modelling choices taking the example of the *Delphinus delphis* species and the *Delphinus* genus. Properties with no namespace (*rank* and *genus*) are generic names conveying the idea of such properties; they may be implemented using properties from different ontologies.

- A first option, adopted by NCBI and VTO, is to represent a taxon as an RDFS or OWL class[5] (Fig. 1(a)). The taxonomic ranks are represented by separate classes (*Genus* and *Species* in this example), and a taxon is related to its rank with an appropriate *rank* property. The relationship between a taxon and its parent taxon is modelled by the *rdfs:subClassOf* property.
- Closer to the nomenclature mindset, the model in Fig. 1(b), adopted by Agrovoc, utilizes the SKOS vocabulary[6] to build a thesaurus. Yet, although it could seem that each SKOS concept (an instance of the *skos:Concept* class) solely depicts a scientific name, the model embeds synonymy relationships that are typical of taxonomic information. The child-to-parent relationship between two scientific names is represented by the *skos:broader* property.
- The EOL database is queried by means of an API[7] that returns results in the RDF JSON-LD syntax. A response makes use of the Darwin Core standard for biodiversity data exchange [23]: each taxon is rendered as an instance of the *dwc:Taxon* class, as depicted in Fig. 1(c), that is meant to denote taxonomic information (*dwc:Taxon*

[2] https://bioportal.bioontology.org/ontologies/GEOSPECIES
[3] http://lod.taxonconcept.org/
[4] BBC Wild Life Ontology : http://www.bbc.co.uk/ontologies/wo
[5] OWL2: https://www.w3.org/TR/2012/REC-owl2-rdf-based-semantics-20121211/
[6] SKOS: https://www.w3.org/2009/08/skos-reference/skos.html
[7] EOL API : http://eol.org/info/api_overview

is equivalent to *Taxon* and *TaxonConcept* in the TDWG Ontology[8]). Nomenclatural information is hardly separated from taxa.

- The model in Fig. 1(d) defines specific classes for each taxonomic rank, such as *Species* and *Genus*. Unlike models (a) to (c), the taxonomic rank is not denoted by a specific property but by the belonging to a class, *e.g. Delphinus delphis* is an instance of the *Species* class. The child-to-parent relationship is represented by a per-rank property, *genus* in this case. This model has been adopted by GeoSpecies, DBpedia and the BBC Wildlife Ontology.
- Lastly, TaxonConcept's model (Fig. 1(e)) is very similar to model (d), with the difference that only the species rank is represented as a class. Higher ranks are simply mentioned by means of a per-rank property whose object is a literal (property *genus* and literal "*Delphinus*" in the example).

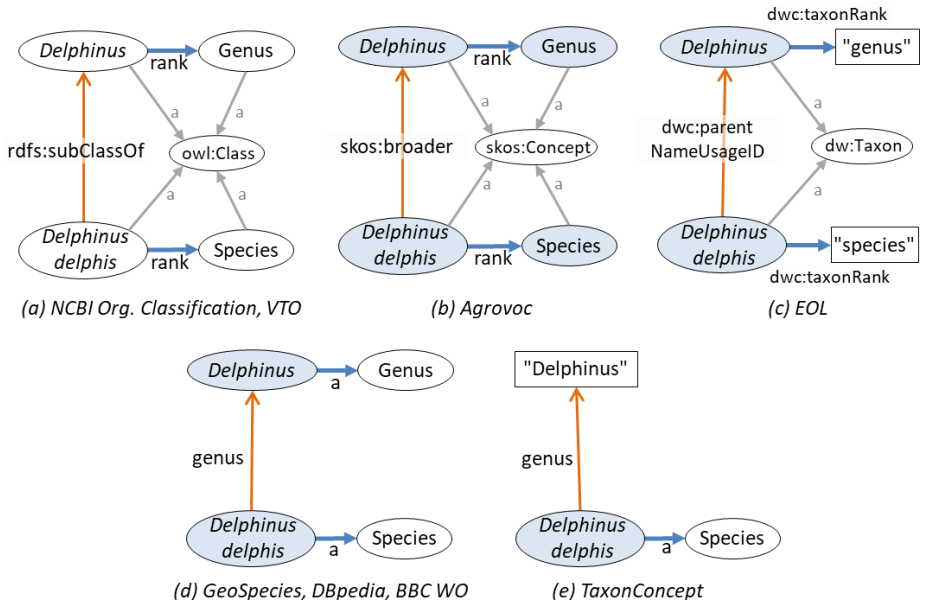

Fig. 1. Various models to represent taxa and/or scientific names using OWL classes (a), SKOS concepts (b) or instances of other classes (c, d and e). Boxes depict literals. White bubbles are OWL classes whereas blue bubbles are class instances. Orange arrows depict the child-to-parent relationship between the *Delphinus delphis* species and the *Delphinus* genus. Blue arrows relate a taxon with a taxonomic rank.

In spite of these differences, all those models seem to depict the same reality. Nevertheless, a careful look suggests that they convey somewhat varying mindsets. In the Semantic Web ethos, OWL classes are defined by extension as a set of instances (or individuals). Intuitively, the *Delphinus delphis* class in (a) comprises the individuals of

[8] https://github.com/darwin-sw/dsw/wiki/ClassTaxon#equivalence-of-taxon-and-taxoncon-cept-in-the-tdwg-ontology-and-the-darwin-core-standard

that species. This is in line with the models of NCBI and VTO that mostly provide a biological description, *i.e.* taxonomic information. By contrast, SKOS is commonly used to describe a nomenclatural system as a thesaurus, *i.e.* a hierarchy of concepts connected by semantic relationships. Yet, the generic term "nomenclatural system" must not be confused with the nomenclature in its biological sense. Indeed, Agrovoc (b) models a hierarchy of concepts that not only represent nomenclatural information (scientific names) but also taxonomic information (how names are assigned to taxa) intertwined with each other. Similarly, EOL (c) chooses to model a taxon as an instance of the *dwc:Taxon* class. Using OWL classes on the one hand, or SKOS concepts or *dwc:Taxon* instances on the other hand, are equally valid solutions. Only, they indicate different perspectives of the same reality: an instance (in particular of the *skos:Concept* class) characterizes a taxon as one concept within a thesaurus of taxon concepts, while an OWL class characterizes a taxon as the set of individuals of that biological entity. The use of instances to represent species in GeoSpecies (d) and TaxonConcept (e) makes them close to the SKOS mindset. Both describe scientific names along with occurrence data, thus, again, interweaving taxonomic and nomenclatural information.

Hence, to some varying extent, it occurs that all these approaches intertwine taxonomic information and nomenclatural information. When we consider a broader picture, these discrepancies entail several impediments:

- Firstly, the scientific consensus about taxonomy constantly evolves. For instance, Linné described most snails as species belonging to genus *Helix* in 1758, but many of them now belong to another family, *e.g.* "*Helix glauca* Linnaeus, 1758" is a synonym of "*Pomacea glauca* (Linnaeus, 1758)" which is the valid name. Similarly, "*Delphinus capensis* Gray, 1828" became a synonym of "*Delphinus delphis* Linnaeus, 1758" in 2015 in light of new scientific evidences [10]. When nomenclatural and taxonomic information is intertwined, the model pictures a snapshot of the use of scientific names at a certain time, that can hardly accommodate changes. A workaround to this issue consists in versioning the whole data set but this entails setting up a mechanism to track the changes from one version of the data set to the next. Editorial notes can be used to document such changes but these are mainly meant for humans and are hardly machine-processable. For a model to accommodate such changes in a flexible manner, it is necessary to distinguish explicitly between the nomenclatural and taxonomic levels. This distinction may allow not only to follow up on taxonomical changes, but also to track and characterize them as proposed by Chawuthai et al [9].
- Secondly, the power of Linked Data spawns from the number and quality of links. Interlinking two data sets requires that they model the same kind of information. If it is unclear whether the focus of a data set is about nomenclature (scientific names) or biology (taxa), then drawing *owl:sameAs* links with resources of other data sets may be erroneous: a species name is not the same thing as the group of individuals of that species. Furthermore, a more technical limitation can occur when interlinking data sets: good practices generally discourage the alignment of class instances with classes since reasoners for Description Logics rely on the distinction between terminological and assertional knowledge [1]. Interestingly enough, this issue is strikingly

evidenced by the data sets that we analyzed: NCBI and VTO, both based on OWL classes, are linked with each other using the *owl:equivalentClass* property, but they have no link whatsoever with the data sets based on instances[9] (models b, c, d and e of Fig. 1). This absence of links does not result from a conceptual mismatch; it results from a sheer technical issue, although conceptually, it would make perfect sense to link NCBI and VTO with these other data sets.

In the next section, we propose a model intended to tackle these issues in the context of the TAXREF taxonomic register.

3 A Generic Model to represent Nomenclatural and Taxonomic Information as Linked Data

TAXREF [13] is the French national taxonomic register for fauna, flora and fungus, maintained and distributed by the National Museum of Natural History of Paris (France). It is a manually curated register of all the species inventoried in metropolitan France and overseas territories, organized as a hierarchy of over 500.000 scientific names that mark a national and international consensus. From the temporal perspective, all living beings are considered as well as those of the close natural history, from the Paleolithic until now. Available through a Web site[10], a Web service[11] or a downloadable text file, TAXREF enables the interoperability between biological databases (mainly occurrence databases), thus supporting biodiversity studies and natural heritage conservation strategies. A new version of TAXREF is published every year, that acknowledges synonymy or hierarchy changes.

Our goal is to design a model to represent TAXREF as Linked Data, that works out the issues and limitations discussed in section 2. More specifically, we seek to achieve three objectives:

1. the model must be relevant to biologists by reflecting the distinction between nomenclature and taxonomy, as well as to computer scientists by adhering to the philosophy that underpins the Semantic Web standards;
2. the model must be flexible enough to accommodate taxonomic changes from one version of TAXREF to the next;
3. the model must enable the alignment with third-party data sets published as Linked Data, may they represent nomenclatural or taxonomic information.

Fig. 2 sketches the model we propose to publish TAXREF as Linked Data, that we denote TAXREF-LD. It is the outcome of a thorough reflection during which we confronted concepts from the biology (taxonomy, systematics) with Semantic Web modelling practices and LOD publication pragmatic concerns.

---

[9] Here we refer to proper LOD links using HTTP URIs. NCBI and VTO embed cross references to third-party database identifiers (using *e.g.* property *obo:hasDbXref*), but these do not comply with LOD principles.

[10] https://inpn.mnhn.fr/programme/referentiel-taxonomique-taxref?lg=en

[11] https://taxref.mnhn.fr/taxref-ws

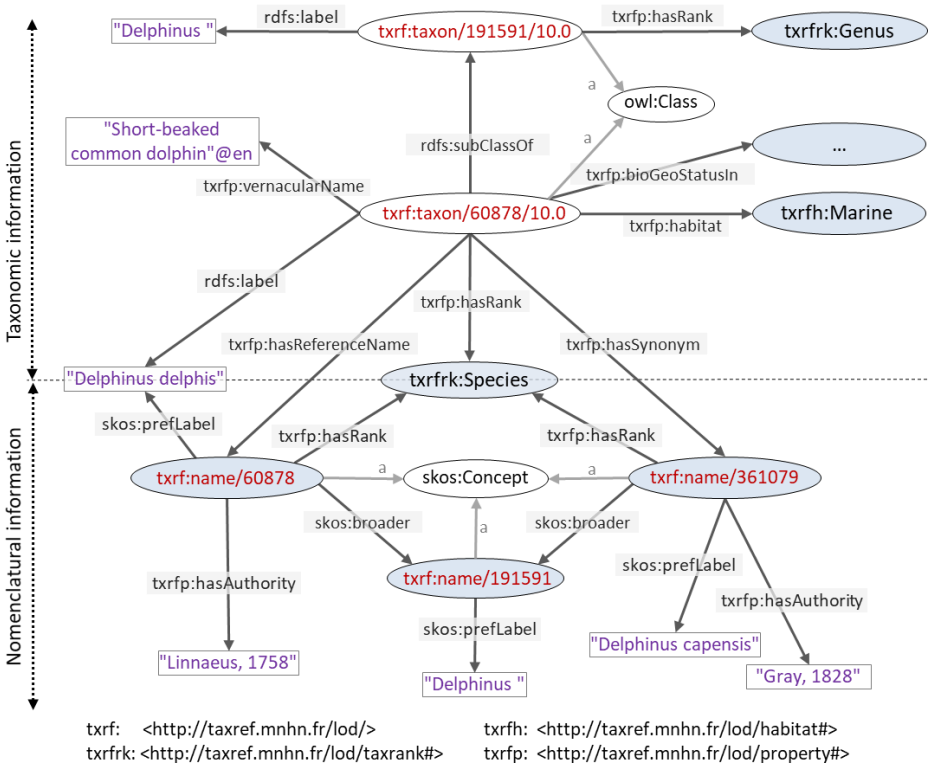

Fig. 2. TAXREF-LD: a Linked Data model for TAXREF, separating nomenclatural from taxonomic information. White bubbles are OWL classes whereas blue bubbles are class instances.

To achieve objective 1, two distinct levels are modelled. At the nomenclatural level (lower part of Fig. 2), each scientific name is modelled as a SKOS concept along with a label, an authority and a taxonomic rank. The child-to-parent relationship between scientific names is expressed with the *skos:broader* property. At the taxonomic level (upper part), a biological taxon is modelled as an OWL class. As a mirror of the nomenclatural level, the child-to-parent relationship between taxa is expressed with the *rdfs:subClassOf* property. Vernacular names are not governed by nomenclatural rules, but account for a property of a group of individuals. Hence, they are attached to the taxon's OWL class. Likewise, biological traits (currently habitat and biogeographical status) are attached to the OWL class. Both levels are connected by the links between a taxon and its reference name (property *txrfp:hasReferenceName*), and between the taxon and its synonyms (property *txrfp:hasSynonym*). A taxon gets the label of its reference name, hence the *rdfs:label* property with the same value as the reference scientific name's *skos:prefLabel* property. It also takes the taxonomic rank of its reference name, hence the *txrfp:hasRank* property at both levels.

In this model, instances of a taxon's OWL class are not depicted. They would typically be the biological individuals of that taxon. In particular, an instance may be created to represent and characterize the type specimen attached to a scientific name.

Note that, for the sake of clarity, details of biogeographical statuses are not depicted in Fig. 2. Also, taxonomic ranks and types of habitats are instances of the *skos:Concept* class but this is not depicted.

OWL class vs. Darwin Core Taxon. Arguably, an alternative model could represent taxa as instances of the *dwc:Taxon* class, rather than OWL classes. The Darwin Core terms were initially designed as a means to exchange taxonomic data using flat text files. As of today, the journey towards a proper ontological representation in RDF is still on-going, as pointed out by Baskauf et al [4]. Despite efforts of the Darwin-SW project to define object properties relating organisms, identifications, taxa, occurrences and locations [2], some issues have not been addressed yet, as underlined in [3]: "the object properties necessary to relate *dwc:Taxon* instances to name entities, references, parent taxa, and child taxa do not exist and the exact relationship between taxonomic entities such as taxon concepts, protonyms, taxon name uses, etc. has not been established using RDF". Accordingly, it occurred to us that the RDF representation of Darwin Core terms is not mature enough yet to fulfill the distinction we wish to model between the nomenclatural and taxonomic information levels.

URI naming scheme. The nomenclatural level is stable in time: new scientific names may be coined but the information associated with a name shall not change, as ruled by the Codes of nomenclature. Consequently, URIs of SKOS concepts are fixed once for all versions of TAXREF. For instance, *Delphinus capensis* is associated a SKOS concept whose URI is *http://taxref.mnhn.fr/lod/361079/name*. Conversely, the taxonomic level must be able to accommodate changes (objective 2). Our point is not to characterize and keep track of the changes that may occur through time (in contrast to *e.g.* [9]), but simply to allow changes in the use of scientific names by taxon concepts, between two versions of TAXREF. Toward this end, we append TAXREF's version number to the URIs of OWL classes. For instance, *Delphinus capensis* was a reference name in version 9.0, thus it was associated an OWL class (*http://taxref.mnhn.fr/lod/taxon/361079/9.0)* and a SKOS concept (given above). Since version 10.0, it has become a synonym of *Delphinus delphis*, hence it has no corresponding OWL class in version 10.0, only the SKOS concept remains.

Interlinking. The separate modelling of the nomenclatural and taxonomic levels provides greater flexibility for the interlinking with third-party data sets (objective 3). For instance, NCBI's classes model biological taxa that are linked with TAXREF-LD's taxonomic level using the *owl:equivalentClass* property (section 4 discusses further the choice of relevant linking properties). The distinction between nomenclatural and taxonomic levels may also be useful to avoid linking biological entities that bear the same scientific name although they denote different entities throughout data sets. For example, the IUCN Red List of Endangered Species[12] still considers *Delphinus delphis* and *Delphinus capensis* as separate species, although *Delphinus capensis* is now considered as a synonym for *Delphinus delphis*. Consequently, 'their' *Delphinus delphis* taxon does not denote the same biological entity as the one in TAXREF, thence a link at the taxonomical level would be erroneous. Yet, a link at the nomenclatural level (names) makes sense since it does not depend on synonymy relationships.

[12] http://www.iucnredlist.org

4 Publishing TAXREF-LD as High Quality Linked Data

To perform the translation of the TAXREF database into the model presented in section 3, we used the Morph-xR2RML software[13], an implementation of the xR2RML generic mapping language [20] designed to address the translation of heterogeneous data sources into RDF. This produced a graph of approximately 8.5M RDF triples, accounting for 509.148 scientific names (SKOS concepts) and 236.507 taxa (OWL classes).

Access methods. An on-going work intends to set up a server enabling the sustainable dereferencing of TAXREF-LD URIs. Until then, a temporary server hosts the RDF graph for test purposes. It provides a dereferencing method[14] as well as a public SPARQL endpoint[15].

Metadata. In order to ensure discoverability, understandability and exploitability of TAXREF-LD, we have taken great care of providing rich and informative metadata while adhering to best practices for the publication of data on the Web [11]. Using the DCAT vocabulary[16], we defined a catalog (*http://taxref.mnhn.fr/lod/TaxrefCatalog)* wherein the different versions of TAXREF are represented by separate DCAT data sets. Each data set comes with three distributions: a Web service, a downloadable text file and a Linked Data distribution *i.e.* TAXREF-LD (*http://taxref.mnhn.fr/lod/Taxref-ld/10.0* in TAXREF version 10.0). Additional annotations are provided with respect to the number of triples, vocabularies used, links with other data sets, provenance, etc., using notably the VoID vocabulary[17]. The TAXREF-LD resource is also the SKOS thesaurus (of type *skos:ConceptScheme*) that registers all the SKOS concepts representing scientific names. *Biota* (*http://taxref.mnhn.fr/lod/name/349525*) is its top concept.

Links with other taxonomic registers. To achieve significant interlinking, we first manually aligned the TAXREF-LD classes and properties (related to taxonomical ranks, habitats, authority, etc.) with their counterparts from other ontologies. Then, we developed a plugin for the Silk Framework [22], that ports a matching algorithm previously developed by TAXREF experts. We leveraged the distinction between the nomenclatural and taxonomic levels to link TAXREF-LD with datasets based on the multiple models presented in Fig. 2. NCBI Organismal Classification and VTO both define classes that we aligned with the taxonomic level of TAXREF-LD, as illustrated in the upper part of Fig. 3. With a model based on SKOS concepts, Agrovoc's SKOS concepts are more likely linked with TAXREF-LD's nomenclatural level using the *skos:exactMatch*. Yet, this equivalence is controversial since taxonomic information is interweaved in Agrovoc's model. An alternative may be to use the weaker *skos:closeMatch* property, or to assume that Agrovoc's concepts represent taxa and declare TAXREF-LD's SKOS concepts as reference or synonymous names of these taxa. Likewise, with

---

[13] Morph-xR2RML: https://github.com/frmichel/morph-xr2rml/

[14] Any TAXREF-LD URI can be dereferenced by pointing to http://erebe-vm2.i3s.unice.fr:8890/describe/?url=<URI>. For instance, this tiny URL leads to the description of taxon *Delphinus delphis*: https://frama.link/RJd-_xq8

[15] http://erebe-vm2.i3s.unice.fr:8890/sparql

[16] DCAT: https://www.w3.org/TR/2014/REC-vocab-dcat-20140116/

[17] VoID: https://www.w3.org/TR/2011/NOTE-void-20110303/

an instance-based modelling of taxa intertwined with some nomenclatural information, TaxonConcept and GeoSpecies are controversial cases. As discussed in section 2, good practices recommend not to align these instances with OWL classes of TAXREF-LD's taxonomic level, unless utilizing a semantically-poor property such as *rdfs:seeAlso*. Thus, we opted for an alignment at the nomenclatural level of TAXREF-LD, yet using the weaker *skos:relatedMatch* or *skos:closeMatch* SKOS properties, depending on how close they are to our model. This is illustrated in the lower part of Fig. 3. The linking with EOL is still on-going at the time of writing. Overall, we created 267.155 links towards resources of these taxonomic registers. Additionally, TAXREF maintains references to Web pages of on-line scientific databases. We used these references to produce 992.722 *foaf:page* links from TAXREF-LD classes and concepts towards related Web pages (not depicted in Fig. 3).

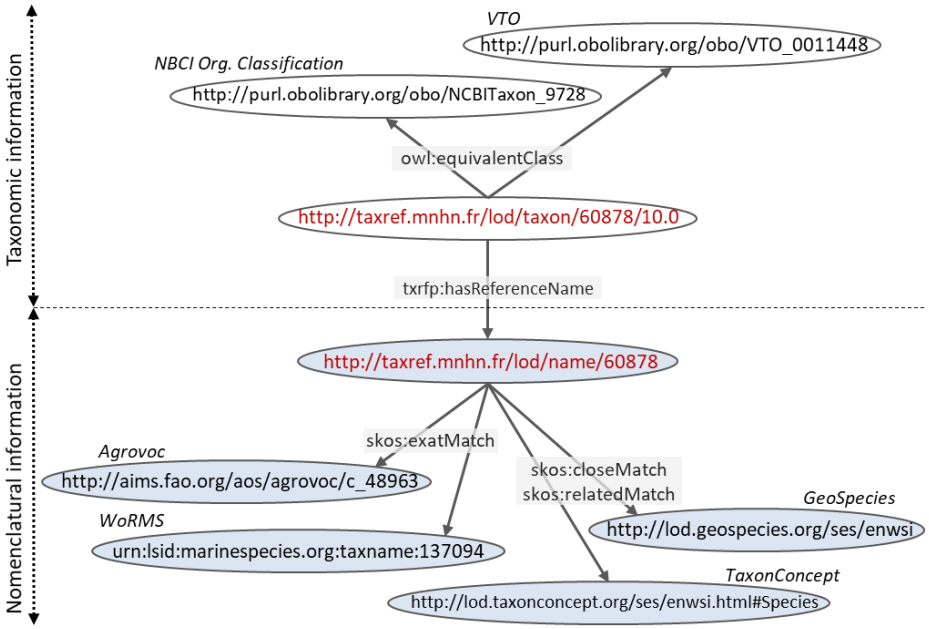

Fig. 3. Interlinking of the *Delphinus delphis* species with six other LD taxonomic registers

5 Conclusion and Perspectives

Taxonomic registers are key tools for the integration of biological databases. As such, they stand out as promising candidates to populate the Web of Data. In this paper, we reported on the publication of the French taxonomic register (TAXREF) in the Web of Data, by adhering to Linked Open Data best practices.

We first analyzed the varying modelling choices made in the past years to represent some international taxonomic registers as Linked Data. We pointed out that these models convey different mindsets that can make their interlinking difficult. Furthermore,

these models do not easily accommodate the ever-changing scientific consensus about taxonomy.

Consequently, we proposed a model tackling these issues and capable of capturing two distinct levels of information: nomenclatural information (scientific names assigned to biological entities) is represented as a SKOS thesaurus, and taxonomic information (the description and characterization of these biological entities) is represented by OWL classes. We argue that this model is relevant to biologists as well as Semantic Web experts, it is flexible enough to accommodate taxonomy changes and it enables interlinking with third-party data sets published as Linked Data, whatever the model they adopted. We applied this model to the case of TAXREF, that is now publicly accessible through a SPARQL endpoint and a Linked Data server, and we seek to achieve proper dereferencing of the URIs in the near future. To increase its visibility, we are in the process of registering TAXREF-LD on the DataHub.io portal, and we are considering its publication on the AgroPortal ontology repository for agronomy [16] .

Furthermore, our goal with this paper is to engage in a discussion with the stakeholders of the biodiversity community, may they be data consumers or producers of sibling taxonomic registers covering complementary regions, epochs or domains. Our point it to delineate some scientific questions and the underlying data integration scenarios, and engage in actions to pursue these objectives.

More generally, the publication of taxonomic registers as Linked Data is a way to contribute to a large, distributed, biological knowledge base. This knowledge base may be beneficial in many ways. For instance, taxonomists may leverage it to compare and discuss their conceptions of biological entities throughout the world. Navigating through interlinked data sets related to domains as diverse as the biology, genetics, medicine, resources management, sociology, etc., could pave the way to inferring new knowledge on organisms and spur new research areas.

Acknowledgement. We thank the Université Côte d'Azur for its financial support to this work (IADB project).

References

1. F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider: The Description Logic Handbook: Theory, Implementation, and Applications, Cambridge University Press, New York, NY, USA (2003).
2. S.J. Baskauf, C.O. Webb: Darwin-SW: Darwin Core-based terms for expressing biodiversity data as RDF, Semantic Web. 7 (2016) 617–627.
3. S.J. Baskauf, J. Wieczorek, J. Deck, C. Webb, P.J. Morris, M. Schildhauer: Darwin Core RDF Guide, Biodiversity Information Standards. (2015).
4. S.J. Baskauf, J. Wieczorek, J. Deck, C.O. Webb: Lessons learned from adapting the Darwin Core vocabulary standard for use in RDF, Semantic Web. 7 (2016) 617–627.
5. T. Berners-Lee: Linked Data, in Design Issues of the WWW, (2006).
6. C. Bizer: The Emerging Web of Linked Data, IEEE Intelligent Systems. 24 (2009) 87–92.
7. R. Blaustein: The Encyclopedia of Life: Describing Species, Unifying Biology, BioScience. 59 (2009) 551–556.

8. C. Caracciolo, A. Stellato, A. Morshed, G. Johannsen, S. Rajbhandari, Y. Jaques, et al.: The AGROVOC linked dataset, Semantic Web. 4 (2013) 341–348.
9. R. Chawuthai, H. Takeda, V. Wuwongse, U. Jinbo: Presenting and Preserving the Change in Taxonomic Knowledge for Linked Data, Semantic Web. 7 (2016) 589–616.
10. H.A. Cunha, R.L. de Castro, E.R. Secchi, E.A. Crespo, J. Lailson-Brito, A.F. Azevedo, et al.: Molecular and Morphological Differentiation of Common Dolphins (Delphinus sp.) in the Southwestern Atlantic: Testing the Two Species Hypothesis in Sympatry, PLOS ONE. (2015).
11. B. Farias Lóscio, C. Burle, N. Calegari: Data on the Web Best Practices, W3C Recommandation. (2017).
12. S. Federhen: The NCBI Taxonomy database, Nucleic Acids Research. 40 (2012) D136–D143.
13. O. Gargominy, S. Tercerie, C. Régnier, T. Ramage, C. Schoelink, P. Dupont, et al.: TAXREF v10. 0, référentiel taxonomique pour la France: méthodologie, mise en oeuvre et diffusion, Muséum National d'Histoire Naturelle, Paris. (2016).
14. T. Heath, C. Bizer: Linked Data: Evolving the Web into a Global Data Space, 1st ed., Morgan & Claypool, (2011).
15. International Commission on Zoological Nomenclature: International Code of Zoological Nomenclature, Fourth Edition, International Trust for Zoological Nomenclature, (1999).
16. C. Jonquet, A. Toulet, E. Arnaud, S. Aubin, E. Dzalé-Yeumo, V. Emonet, et al.: Reusing the NCBO BioPortal technology for agronomy to build AgroPortal. In Proceedings of the 7th International Conference on Biomedical Ontologies, ICBO'16, Demo Session, Corvallis, Oregon, USA (2016).
17. S.P. Lapage, P.H. Sneath, E.F. Lessel, V.B.D. Skerman, W.A. Clark, H.P.R. Seeliger: International Code of Nomenclature of Bacteria: Bacteriological Code - 1990 Revision, ASM Press, (1992).
18. J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P.N. Mendes, et al.: DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia, Semantic Web. 6 (2014) 167–195.
19. J. McNeill, F.R. Barrie, W.R. Buck, V. Demoulin, W. Greuter, D.L. Hawksworth, et al.: International Code of Nomenclature for algae, fungi, and plants (Melbourne Code). Regnum Vegetabile 154, Koeltz Scientific Books, (2012).
20. F. Michel, C. Faron-Zucker, J. Montagnat: Translation of Heterogeneous Databases into RDF, and Application to the Construction of a SKOS Taxonomical Reference. In Revised Selected Papers of the 11th International Conference on Web Information Systems and Technologies (WebIST), Springer, (2016): pp. 275–296.
21. P.E. Midford, T.A. Dececchi, J.P. Balhoff, W.M. Dahdul, N. Ibrahim, H. Lapp, et al.: The Vertebrate Taxonomy Ontology: a framework for reasoning across model organism and species phenotypes, Journal of Biomedical Semantics. 4 (2013) 34.
22. J. Volz, C. Bizer, M. Gaedke, G. Kobilarov: Silk - A Link Discovery Framework for the Web of Data. In 2nd Workshop about Linked Data on the Web, Madrid, Spain (2009).
23. J. Wieczorek, D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, et al.: Darwin Core: An Evolving Community-Developed Biodiversity Data Standard, PLOS ONE. 7 (2012).