The Knowledge Engineering Review, Vol. 35, e9, 1 of 19.
doi:10.1017/S0269888920000132

Matching biodiversity and ecology ontologies: challenges and evaluation results

NAOUEL KARAM[1], ABDERRAHMANE KHIAT[2], ALSAYED ALGERGAWY[3],
MELANIE SATTLER[4], CLAUS WEILAND[5] and MARCO SCHMIDT[5,6]

[1]*Fraunhofer FOKUS, Berlin, Germany*
e-mail: naouel.karam@fokus.fraunhofer.de
[2]*Fraunhofer IAIS, Sankt Augustin, Germany*
e-mail: abderrahmane.khiat@iais.fraunhofer.de
[3]*Friedrich-Schiller University of Jena, Germany*
e-mail: alsayed.algergawy@uni-jena.de
[4]*MARUM, University of Bremen, Germany*
e-mail: mbuss@marum.de
[5]*Senckenberg Biodiversity and Climate Research Center, Frankfurt, Germany*
e-mail: claus.weiland@senckenberg.de
[6]*Palmengarten der Stadt Frankfurt, Germany*
e-mail: marco.schmidt@stadt-frankfurt.de

Abstract

Biodiversity research studies the variability and diversity of organisms, including variability within and between species with particular focus on the functional diversity of traits and their relationship to environment. Managing biodiversity data implies dealing with its heterogeneous nature using semantics and tailored ontologies. These are themselves differently conceived, and combining them in semantically enabled applications necessitates an effective alignment between their concepts. This paper describes the ontology matching of biodiversity- and ecology-related ontologies. We illustrate diverse challenges introduced by this kind of ontologies to ontology matching in general. Real use cases requiring pairwise alignments between environment and trait ontologies are introduced. We describe our experience creating a new track at the Ontology Alignment Evaluation Initiative designed for this specific domain and report on the results obtained by state-of-the-art participating systems. The biodiversity and ecology use case turns out to be a strong one for ontology matching, introducing new interesting challenges. Even if most of the matching systems perform relatively well in the proposed matching tasks, there is still room for improvement. We highlight possible directions in that matter and elaborate on our plan to further progress with the track.

1 Introduction

Research practice has become increasingly data-intensive, and more and more data are being made available online by researchers all over the globe. Biodiversity research is no exception, with data ranging from observation and specimen collections to genome and environmental data, encompassing anthropogenic impacts like land use patterns (IPBES, 2018). The biodiversity community has been exploring means to discover and filter this large volume of multidisciplinary data by creating proper and rich metadata conform to the FAIR principles (Wilkinson *et al.*, 2016), foundational concepts seeking to enhance findability, accessibility, interoperability, and reusability of research data.

This highlights the need for bridging technologies to facilitate the cross-domain reuse and repurposing of biodiversity data. An interoperable exchange of such information is enabled through the use of ontologies that provide ‘meaning’ to data. Considerable effort is being put in formalizing domain knowledge in terms of ontologies and enriching data by means of annotations.

In this paper, we consider the ontology matching problem in the biodiversity and ecology domain, motivated by real use cases from two projects we are carrying out in these domains, namely GFBio (The German Federation for Biological Data)[1] and AquaDiva[2]. Both projects aim at providing semantically enriched data management solutions for data capture, annotation, indexing, and search.

Environment and trait ontologies have particular relevance within the framework of our projects and for biodiversity *big data* in general. ‘Trait’ in this context means a morphological, anatomical, or physiological feature measurable at the individual level (Kattge *et al.,* 2011). Huge datasets are provided by mass digitization efforts, which open up large amounts of trait and environmental data, link these data to digital resources like sequence databases, and foster the setup of biodiversity knowledge graphs. To enable analysis and knowledge discovery, machine learning methods based on word embeddings (Mikolov *et al.,* 2013) were developed: OPA2Vec demonstrates the usage of ontology metadata and annotations (class labels, descriptions, and synonyms) to implement new approaches for semantic similarity measures on biodiversity knowledge graphs emphasizing the need for reliable matching of the underlying ontologies (Smaili *et al.,* 2018; Weiland *et al.,* 2019).

Motivated by our use cases, we introduced a new track to the Ontology Alignment Evaluation Initiative (OAEI)[3], the first OAEI track designed to evaluate matching systems’ performance on biodiversity- and ecology-related ontologies.

In particular, we did consider two pairs of ontologies to be aligned based on their importance in the considered domain and their high degree of overlap: (i) the Environment Ontology (ENVO) (Buttigieg *et al.,* 2013a,b) and the Semantic Web for Earth and Environment Technology Ontology (SWEET) (Raskin & Pan, 2004) and (ii) the Flora Phenotype Ontology (FLOPO) (Hoehndorf *et al.,* 2016, 2018) with the Plant Trait Ontology (TO) (Walls *et al.,* 2012).

To our knowledge, the aforementioned ontologies do not contain any mappings to each other except for a mapping between ENVO and a small subset of the SWEET realm subdomain that has been manually curated. Using automated ontology matching for the generation of such mappings, with minimal manual intervention, is probably the best solution compared to the efforts needed for a manual curation process especially considering the size and dynamic evolution of those ontologies.

The main contributions of this work are the following:

- We elaborate on specific challenges in the biodiversity and ecology domain and why there is a need for automated ontology matching;
- How the considered domain ontologies introduce new and specific challenges that can lead to the improvement of ontology matching systems;
- We describe our experience developing a new track for OAEI, the challenges faced in this process and report on the results and limits of state-of-the-art ontology matching systems in coping with the track tasks.

The rest of the paper is organized as follows. In Section 2, we describe the context of this work and the need for automated ontology matching in the biodiversity domain. Preliminaries are introduced in Section 3. In Section 4, we detail the process of developing a dedicated track and present results from the OAEI 2018 edition. A detailed discussion of the challenges and future directions is presented in Section 5. Section 6 concludes the paper.

2 Context and motivation

In this section, we will elaborate on the particular challenges faced in the biodiversity and ecology domain and the importance of including automated ontology matching processes into data curation workflows.

---

[1] www.gfbio.org.

[2] www.aquadiva.uni-jena.de.

[3] oaei.ontologymatching.org/2018/biodiv/.

We introduce two real use cases where finding pairwise alignments between ecological and phenotype ontologies is needed. Finally, we present how the characteristics of the considered ontologies can raise new challenges for the ontology matching community.

Biological and ecological research has the particularity of dealing with a large number of disciplines (e.g. Botany, Microbiology, Chemistry) and addressing diverse topics ranging from marine ecosystems to species distribution (Alves *et al.,* 2018). This results in heterogeneous data stored in disparate data archives which make data acquisition more and more time consuming and a challenging task for researchers. The biodiversity community has responded to those challenges by developing a set of metadata standards as well as formalizing domain knowledge in terms of vocabularies and ontologies. This formal knowledge provides a mean to enrich data with annotations which enables efficient data discovery and curation.

The need for matching biodiversity and ecology ontologies came from real use cases related to two biodiversity projects we are involved in. The GFBio project (Diepenbroek *et al.,* 2014) is developing an infrastructure providing data management and archiving solutions for biological scientists and projects at the German national level. The projects semantic component, the GFBio Terminology Service (GFBio TS) (Karam *et al.,* 2016; Fichtmüller *et al.,* 2017), is the backbone giving meaning to the data and offering a set of semantic services enabling more efficient data capture and discovery. The GFBio TS includes ontologies and taxonomies useful for describing data in biodiversity and ecology. Those ontologies are used to annotate project-related datasets, and the integration of such datasets requires pairwise alignments between the underlying ontologies.

AquaDiva[4] is a large collaborative research center (CRC) aiming to study the effect of biodiversity on the subsurface levels in the German Hainich forest. To this end, several scientists from different domains, including ecology, biology, and geosciences, are collecting samples covering different aspects and thus generating a large number of heterogeneous datasets. In order to deal with the heterogeneity issue, the AquaDiva CRC is building a semantic layer on top of the datasets stored in a common repository (Klan *et al.,* 2017). These datasets cover a wide range of aspects, and a large number of ontologies are required to cover them. An essential step to build such a conceptual layer is to identify semantically similar entities across the considered ontologies.

In a joint effort to improve data retrieval in the search applications we are hosting within the aforementioned projects, we consolidated a set of high-level entities in the biodiversity domain that are relevant for biodiversity researchers when searching for data (Löffler *et al.,* 2017). Those entities appear across different sources and cover as much as possible of the domain spectrum both from a user and an application perspective. The identified entities served as a basis to identify and link existing ontologies useful for data annotation as illustrated by the examples in Table 1.

A strong focus of current ecological research lies on the link between the environment and functional traits (Diaz *et al.,* 2013; Bruelheide *et al.,* 2018), enabling the identification of consistent and predictable associations between plant traits and environmental conditions. In the context of this work, we focus on the traits that constitute the morphology and function of taxa corresponding to *Quality and Phenotype* and covered by FLOPO and TO, as well as the habitats and ecosystems with ENVO and SWEET, covering environmental features (*Environment*), processes (*Process*), and materials (*Material and Substance*).

The ENVO was developed in 2013 and is frequently updated and extended. As shown in Table 2, the used version has 6909 concepts. ENVO has been chosen for being a well-established ontology for describing environmental features, materials, processes as well as biomes (Buttigieg *et al.* 2013a,b). ENVO is used by many GFBio-related projects, for example, the MIxS standard (Minimum Information about any x Sequence) (Yilmaz *et al.,* 2011) and the PANGAEA database (Diepenbroek *et al.,* 2017), underlining its importance for the natural science community. In order to foster reuse and interoperability, and increase the extent of the ontology, ENVO was aligned with other Open Biological and Biomedical Ontology (OBO) Foundry ontologies, such as BFO (Basic Formal Ontology)[5], ChEBI (Chemical Entities of Biological Interest)[6], and UBERON (Uber Anatomy Ontology)[7].

---

[4] www.aquadiva.uni-jena.de.

[5] basic-formal-ontology.org.

[6] www.ebi.ac.uk/chebi.

[7] uberon.org.

Table 1 Relevant high-level entities in the biodiversity domain

Organism
quercus(`NCBITAXON:3511`), cyclothone(`NCBITAXON:48440`), globigerina bulloides(`NCBITAXON:69025`)
Environment
vegetated area(`ENVO:01001305`), ground water(`ENVO:00002041,SWEET:Groundwater`)
Quality and Phenotype
fruit color(`FLOPO:0000147,TO:0002617`), leaf area(`FLOPO:0010323,TO:0000540`), photosynthetic rate (`TO:0001015`)
Process
climate change(`ENVO:01000629,SWEET:ClimateChange`), land degradation(`ENVO:02500005,SWEET:LandDegradation`)
Material and Substance
sediment(`ENVO:00002007,SWEET:Sediment`), rock(`ENVO:00001995,SWEET:Rock`), CO2(`CHEBI:16526`)
Anatomical Entity
DNA(`PLOSTHES:2568`), leaf (`PO:0025034`), nut fruit (`PO:0030102`)
Method
sequencing assay(`OBI:0600047`), remote sensing(`PLOSTHES:3097`)

Table 2 Summary of the versions, number of classes, and source of the ontologies included in the biodiversity and ecology track

Ontology	Version	Number of classes	Download source
ENVO	2017-08-22	6909	http://purl.obolibrary.org/obo/envo.owl
SWEET	2018-03-12	4543	https://github.com/ESIPFed/sweet
FLOPO	2016-06-03	24 199	http://purl.obolibrary.org/obo/flopo.owl
TO	2017-09-11	1504	http://purl.obolibrary.org/obo/to.owl

Adding a different scope of environmental entities, the SWEET (Raskin & Pan, 2004) was selected to be matched with ENVO. With over 4500 classes as shown in Table 2, SWEET is the second-largest ontology in the environmental realm. Originally, SWEET was developed in 2003 by NASA Jet Propulsion Laboratory, but since 2017 it is hosted by Earth Science Information Partner's Semantic Technologies Committee[8]. Prospectively, SWEET will soon increase its importance for describing earth system science due to subsequent advancement.

The development of FLOPO is driven by the objective to cover the global trait spectrum of plant life with 100 000s of species. Plant data grouped by region and time period are collected in comprehensive compendia called Floras (Hoehndorf *et al.*, 2018). The creation of the FLOPO was motivated by the lack of a standardized vocabulary to systematically compare and analyze the trait information in Floras. Initially, FLOPO was completely based on an automated data-driven workflow (Vos *et al.*, 2014). The Plant Ontology (PO) (Cooper *et al.*, 2013) and the Phenotype and Trait Ontology (Gkoutos *et al.*, 2005) were used to extract entity–quality relationships from digitized taxon descriptions in Floras, and a formal ontological approach based on phenotype description patterns and automated reasoning was applied to create both the ontology FLOPO and a knowledge base containing FLOPO-annotated traits of plants (Hoehndorf *et al.*, 2016).

---

[8] github.com/ESIPFed/sweet.

The Trait Ontology (TO) was originally developed to describe rice traits and then expanded in addition to PO (which provides the vocabulary for morphological and anatomical structures as well as for stages of development (Arnaud *et al.,* 2012)) to characterize traits from all green plants (Cooper *et al.,* 2018). It is of special interest to provide mappings between FLOPO and TO, since both complement each other regarding coverage of qualitative and quantitative traits and phenotypes.

The chosen ontologies present interesting characteristics that can raise new challenges for the ontology matching community. For instance, the SWEET ontology has a particular structure, it is composed of 220 single small ontologies representing its different subdomains. Within two different subdomains, the same term can be used to express two different semantic meanings which result in a lot of homonyms in this ontology. Homonymy together with the absence of definitions may compel ontology matching developers to improve the structural matching part of their algorithms. On the other hand, the size and particular lexical content of FLOPO can boost lexical matching and particularly the use of specialized background knowledge.

3 Preliminaries

In this section, we present some preliminary notions used in the rest of the paper. We outline the definition of an ontology, a mapping, and an ontology alignment. For more details, we refer the reader to Ehrig (2006), Shvaiko & Euzenat (2013), Euzenat & Shvaiko (2013), and Kalfoglou & Schorlemmer (2005).

An *ontology* is formally represented as a 6-tuple: $O = < C, R, I, H_C, H_R, X >$ (Ehrig, 2007), where C and R are disjoint sets representing the set of concepts and the set of relations, respectively; I denotes a set of instances of C or R; H_C denotes a partial order relation on C, called hierarchy or taxonomy of concepts. It associates with each concept its super- or sub-concepts; H_R denotes a partial order relation on R, called hierarchy or taxonomy of relations. It associates with each relation its super- or sub-relations and X denotes a set of axioms. The ontologies considered in this work are represented using the OWL (Web Ontology Language) format (Grau *et al.,* 2008).

A *mapping* (also called a correspondence) (Shvaiko & Euzenat, 2013) is a quintuple $< id, e, e', n, rel >$, where id denotes a unique identifier of the mapping; e and e' are entities from two ontologies O and O', respectively; n denotes a measure of confidence, typically a value within the interval [0, 1], and rel denotes the semantic relation between e and e' (equivalence ($\equiv$), more specific ($\sqsubseteq$), more general ($\sqsupseteq$), disjunction ($\perp$)).

An *alignment* can be defined as a set of mappings, usually expressed using the RDF (Resource Description Framework) alignment format defined by the ontology matching community (David *et al.,* 2011). The alignment process receives as input two ontologies O and O' and produces as output an alignment between entities of O and O'. Other elements that may be used to complete this definition are (i) an initial alignment to be completed or refined by the process, (ii) external resources such as a thesaurus or a dictionary, and (iii) parameters, such as thresholds or weights.

Given these definitions, *ontology matching* is defined as the identification process of mappings between entities of different ontologies to ensure semantic interoperability. These mappings are the bridge that holds heterogeneous ontologies together and ensures interoperability between semantically annotated resources.

The OAEI[9] is an annual campaign for the systematic evaluation of ontology matching systems. The main objective is to evaluate ontology matching systems on the same basis and to enable the reproducibility of the results. A matching task is typically composed of one or more pairs of ontologies with their corresponding reference alignments.

Reference alignments can be categorized into gold standards, silver standards, and baselines. A *gold standard* contains mapping sets that have been manually curated by domain experts and are almost complete (Dahdul *et al.,* 2018). A *silver standard* is typically composed of automatically generated mappings that could be partially curated. Finally, *baseline mappings* are automatically generated mappings that are highly incomplete.

---

[9] oaei.ontologymatching.org.

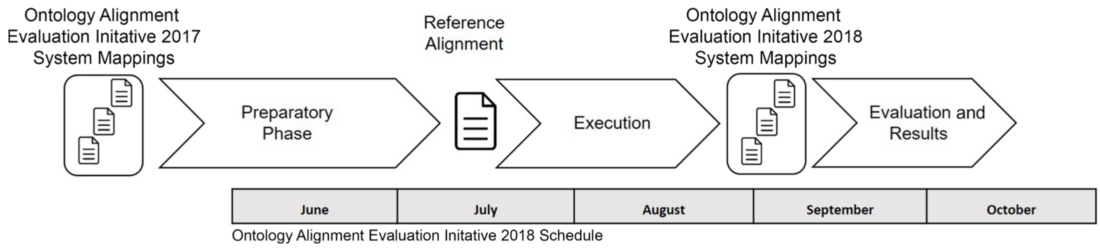

Figure 1 Phases of the Ontology Alignment Evaluation Initiative 2018

For our track, we created a silver standard reference alignment. The methodology we followed for the creation of such an alignment is presented in the following section.

In order to evaluate ontology matching systems, a set of standard evaluation measures can be used, including *precision*, *recall*, and *F-measure* (Euzenat & Shvaiko, 2013). Precision, P, determines the degree of correctness of the matching result. Recall, R, assesses the degree of completeness of the matching systems. Given a reference alignment set referred to as A_{ref} and an automatically generated alignment by a matching system denoted A_m, precision and recall are computed as follows:

$$Precision = \frac{|A_m \cap A_{ref}|}{|A_m|} \tag{1}$$

$$Recall = \frac{|A_m \cap A_{ref}|}{|A_{ref}|} \tag{2}$$

Precision evaluates the post-match effort that is needed to remove false positives, while recall evaluates the post-match effort that is needed to add false negatives from the final matching result. It is necessary to consider a trade-off between those two measures. There are several methods to handle such a trade-off, one of them is to combine both measures. The most used combined measure is the *F-measure* and it is computed as follows:

$$F\text{-}measure = 2 \times \frac{precision \times recall}{precision + recall} \tag{3}$$

4 Experience from the Ontology Alignment Evaluation Initiative

The OAEI 2018[10] edition included 12 tracks organized by different research groups and involving different matching tasks. The biodiversity and ecology track was one of the additions to the OAEI 2018 campaign. The track aims at evaluating matching systems in an important domain that has not been tackled before at OAEI. It consists of two alignment tasks that involve the alignment of ENVO with SWEET and FLOPO with TO.

We designed and developed this track based on the methodology presented in Harrow *et al.* (2017) for the creation of the disease and phenotype track which closely follows the OAEI phases, taking place between June and October as depicted in Figure 1. We performed the preparatory phase in advance using ontology matching systems from the OAEI 2017 campaign for the creation of the consensus alignments. The ontologies and reference alignments were made available online mid of July. During the execution phase, we attended to technical issues reported by matching systems' developers. Finally, we evaluated the participating systems' performance, and the results were published online mid of October.

4.1 Track preparatory phase

The main objective of this phase is to produce a set of reference mappings that will be used as a basis to evaluate ontology matching systems' performance on biodiversity- and ecology- related ontologies.

---

[10] oaei.ontologymatching.org/2018/.

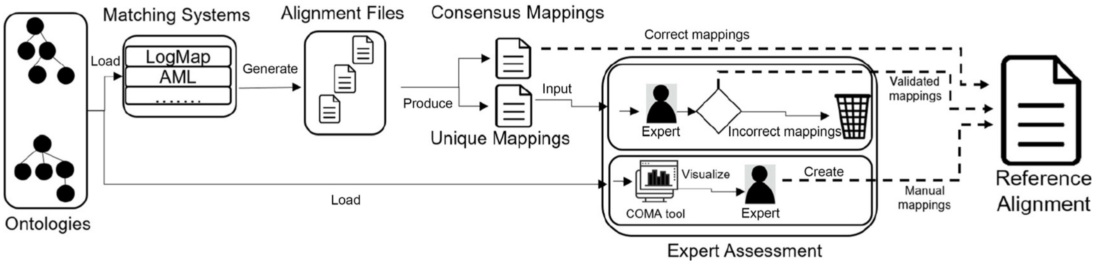

Figure 2 Biodiversity and ecology track preparatory phase

The reference alignments have been composed using (1) consensus alignments based on systems vote, (2) unique mappings validated by domain experts, and (3) manually generated mappings. Figure 2 shows the steps followed for the creation of the reference alignment for each task. In a first step, we used well-known ontology matching systems to generate alignments between the two ontologies. Those are used to generate consensus alignments based on systems vote. The unique mappings generated by only one system are then assessed by an expert, valid mappings are added to the reference as well. Finally, experts have been asked to generate manual mappings, namely ones that have not been found by any matching system. A visualization tool was used to help in this process. In the following sections, we will go through those different steps in detail and discuss our choices and the challenges we faced.

4.1.1 Creation of consensus alignments

The consensus alignments have been generated based on a voting system using the alignments produced by matching systems from the last OAEI edition. The reason for it is that we started the preparatory phase before having access to the latest versions of the participating systems. Systems were selected based on their availability and ability to cope with the matching tasks. We used AML (Agreement Maker Light) (Faria *et al.,* 2013, 2014), LogMap (Jiménez-Ruiz & Grau, 2011; Solimando *et al.,* 2017), LogMapBio, LogMapLite, LYAM (Yet Another Matcher-Light) (Duchateau & Bellahsene, 2016), POMap (Laadhar *et al.,* 2017), and YAMBio (Annane *et al.,* 2016). Additionally, we used mappings extracted from Bioportal (Whetzel *et al.,* 2011) via its RESTful API[11]. Bioportal mappings are based on the Lexical OWL Ontology Matcher tool and based on a lexical similarity algorithm (Ghazvinian *et al.,* 2009).

A vote corresponds to the number of times a mapping appeared in the sets generated by the matching systems. Consensus of vote 2, for instance, will contain mappings suggested by at least two systems. The more the votes, the smaller the size of the consensus alignment. Votes have been grouped by family, and systems with different variants like LogMap, LogMapBio, and LogMapLite have been counted only once. This has been done to reduce bias as reported in Jiménez-Ruiz *et al.* (2013) where systems' outputs have been compared and systems from the same family resulted in mapping sets that clearly overlapped. At the end, five systems families have been participating to the vote.

Table 3 (respectively, Table 4) shows the sizes of the consensus alignments for the ENVO-SWEET matching task (respectively, the FLOPO–TO matching task) as well as each systems family mappings ratio contribution. The five system families could cope with the ENVO–SWEET task and only three families could cope with the FLOPO–TO task, presumably due to the large size of the FLOPO ontology. Note that some of the matching systems could cope with this task during the evaluation phase of the track with the latest version of their system (cf. Section 5).

It is noticeable that the Bioportal mappings are quite incomplete, representing only half of the total mappings produced by at least two system families for the ENVO–SWEET matching task. They cover, on the other hand, a big part of the mappings produced by AML and LogMap families for the FLOPO–TO task which indicates that the mappings computed by the matching systems were mainly detected using lexical similarity.

Figure 3 shows the size and evolution of the average systems precision, recall, and F-measure for the different consensus alignments in the ENVO–SWEET task. The maximum precision (respectively,

---

[11] data.bioontology.org/documentation#Mapping.

Table 3 Consensus alignments for the ENVO–SWEET matching task

	Vote 2	Vote 3	Vote 4	Vote 5
Mappings	689	563	404	248
AML (%)	94.9	99.3	100.0	100.0
Bioportal (%)	50.5	61.8	79.2	100.0
LogMap (%)	81.6	89.0	93.8	100.0
POMap (%)	77.8	84.0	89.4	100.0
YAM (%)	71.6	81.7	99.0	100.0

Table 4 Consensus alignments for the FLOPO–TO matching task

	Vote 2	Vote 3
Mappings	149	140
AML (%)	100.0	100.0
Bioportal (%)	94.0	100.0
LogMap (%)	100.0	100.0

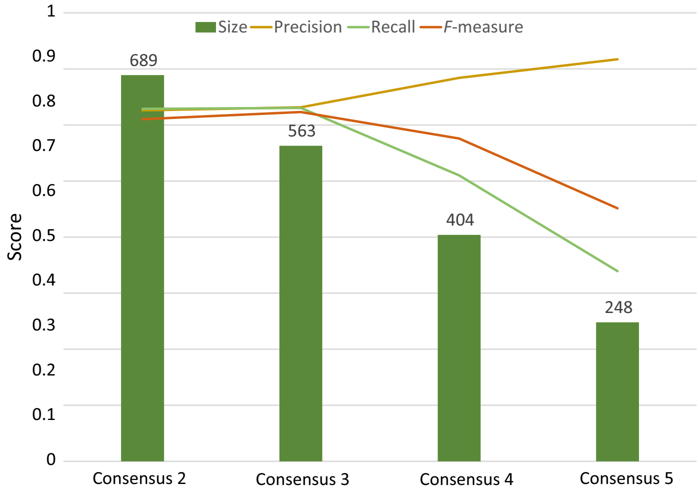

Figure 3 Evolution of precision, recall, and *F*-score for the ENVO–SWEET consensus alignments

recall) is reached with the maximum (respectively, minimum) number of votes. We have selected the consensus alignment of vote 3 (i.e. mappings suggested by at least three system families) to be included in the reference alignment as it has the best trade-off between precision and recall. We preferred not to consider consensus alignments of vote 2 as it may take only two systems to induce erroneous alignments. For the FLOPO–TO task, we considered by default consensus of vote 3 as only three system families were involved and considering the good quality of the Bioportal mappings for this particular task this was not a problem. The consensus alignments contain 563 mappings for ENVO–SWEET and 140 mappings for FLOPO–TO.

Consensus alignments are in general not sufficient to assess systems performances for many reasons. First, they allow only a comparison of the systems to each other. Second, they may contain erroneous mappings especially if the considered systems use the same background resources (e.g. the Unified Medical Language System (UMLS) Metathesaurus). And finally, valid alignments that have been found

Table 5 Assessment of unique mappings and positive and negative contributions for the task ENVO–SWEET

	Mappings	**Positive contribution (%)**	**Negative contribution (%)**
AML	153	15	37
POMap	139	11	42

Table 6 Assessment of unique mappings and positive and negative contributions for the task FLOPO–TO

	Mappings	**Positive contribution**	**Negative contribution**
AML	92	71%	29%

by only one system or none of them will be missing. For this reason, we completed this initial set with manually assessed and manually generated alignments as described in the following sections.

4.1.2 Manual assessment of unique mappings

Unique mappings are mappings generated by one and only one matching system, no other system explicitly found those mappings nor could they be derived via entailment based on the aligned ontologies. These unique mappings have been manually assessed by domain experts who are creators or curators of the ontologies and co-authors of this paper (one expert for ENVO–SWEET and two for FLOPO–TO).

Around 300 unique mappings have been manually assessed for the ENVO–SWEET matching task and around 100 mappings for the FLOPO–TO task. Two systems generated the most unique mappings for the task ENVO–SWEET, namely AML and POMap. For the FLOPO–TO task, due to the reduced number of systems which could cope with the task, only AML was able to generate a set of unique mappings.

Mappings have been validated into equivalent mappings and related mappings (subclass relationships, unidirectional synonyms). The different types of related mappings are discussed more in detail in Section 5. Only equivalent mappings have been added to the reference alignment. Seventy-five validated mappings have been added to the reference alignment for ENVO–SWEET and 66 for FLOPO–TO.

To evaluate the contribution of each system to the set of valid mappings added to the reference alignment, we used the measures of positive and negative contribution introduced in Harrow *et al.* (2017), estimating the weight of correct, respectively, incorrect unique mappings found by the systems. For a set of unique mappings U_m generated by a matching system, given that the set of all considered unique mappings is denoted as U_{all}, the positive and negative contributions (denoted PC and NC, respectively) are calculated as follows:

$$PC(U_m) = \frac{|U_m|.Precision(U_m)}{|U_{all}|} \tag{4}$$

$$NC(U_m) = \frac{|U_m|.(1 - Precision(U_m))}{|U_{all}|} \tag{5}$$

Table 5 (respectively, Table 6) shows the number of unique mappings generated by the considered systems for the ENVO–SWEET matching task (respectively, FLOPO–TO matching task) and the result of their manual assessment in terms of positive and negative contribution.

4.1.3 Manual mappings

Manually generated mappings have been curated by the same domain experts assessing the unique mappings (three authors of this paper). The goal of the manual mappings generation was to find at least 30 additional mappings, which have not been found by any of the matching systems. We chose this specific

Table 7 Examples of manually generated mappings

Entity 1	Entity 2
Divergent tectonic movement (`ENVO:01001095`)	Plate divergence (`SWEETPhenGeolTectonic:PlateDivergence`)
Tectonic movement (`ENVO:01001093`)	Continental drift (`SWEETPhenGeolTectonic:ContinenetalDrift`)
Inflorescence absent (`FLOPO:0006051`)	Inflorescenceless (`TO_0002679`)
Leaf alternate placement (`FLOPO:0001032`)	Phyllotaxy (`TO:0006014`)

number as a minimal requirement for several reasons, beside the well-known difficulty of the process especially on large ontologies, systems performed well in terms of recall making it difficult to find new unidentified mappings. The manual alignments were added to the automatically created ones in order to assess systems recall in a more reliable way.

In the following, we describe the manual creation process and report on the encountered difficulties. To begin with and in order to help the experts identify ontology areas to investigate, we proposed to visualize the ontologies and a set of automatically generated alignments. We did use COMA 3.0 (Arnold & Rahm, 2014) for the visualization with the set mappings generated by the underlying ontology matching system as well as mapping sets generated by other systems. The visualization was useful to a certain extent but the tool was not very practical to use; we faced some scalability issues when processing larger ontologies (in particular FLOPO), which left the application occasionally unresponsive. As pointed out in Dragisic *et al.* (2016), there is a need for best practices and guidelines for good user interface design in ontology alignment. Visualization tools need to be enhanced in terms of usability and user involvement support.

For the ENVO–SWEET task, an initial list of terms relevant to certain environmental domains was created. Chosen domains included geophysics (plate tectonics, volcanism, internal earth structure), glaciology (snow and ice types, glacier types), oceanography (zonation, waves and currents), geomorphology (terrestrial, fluvial, and oceanic landforms), meteorology (storm and wind types), environmental chemistry (biochemical cycling), pedology (soil types), and sedimentology (sediment types). The initial list served as a starting point to identify hot spot areas of a certain topic within the target ontologies. This step was a prerequisite due to the dispersion of topic-related terms among different ontology nodes caused by different modeling approaches by both ontologies. Topics that were barely covered by either SWEET or ENVO were excluded from further consideration. Promising areas have been chosen for further comparison. Subsequently, a tabular juxtaposition of terms from these areas was created. The table was enriched with additional domain-relevant terms and corresponding synonyms if they could be found in at least one of both ontologies, using Protégé (Musen, 2015) as a search tool. The comparison table was then used to find mappings between ENVO and SWEET. Mappings that were not already present in the consensus alignments file have been added to the reference alignment. Term equivalence was justified by literature research (dictionaries, scientific publications, and books), not only by checking the definitions within the ontology itself but also by checking the associated term relations to avoid wrong matching of homonyms. The latter was very important for the SWEET ontology due to the absence of definitions for most terms.

Additionally, a small number of subclass relation mappings derived from the juxtaposition have been saved as well. The requirement for this choice was that the subclass terms from one ontology did not have any equivalent matches to terms from the other ontology and that the superclass terms related to this subclass term did not yield any equivalent matches either so that these terms would have been missed by the matching otherwise.

For the FLOPO–TO matching task, we followed the same procedure mentioned above for ENVO–SWEET, glossaries of plant morphology (Wagenitz, 2008; Hickey, 1973; Stevens, 2001 onwards) were consulted for related and synonymous terms.

Table 6 shows examples of manually created mappings from both tasks.

4.1.4 Semantic validation of the reference alignment

The ontologies as well as the generated reference alignments have been tested for consistency. For that purpose, we used the ELK reasoner[12] on (1) the ontologies themselves and (2) their merged version together with the corresponding reference alignment.

In the first step, the ontology files have been merged using the OWL API (Application Programming Interface). Then, we created the equivalence relation between entities of both ontologies based on mappings defined in the reference alignment. Those have been added one by one to the merged ontologies, and the resulting ontology has been checked for consistency after each addition.

As a result, the consensus alignment for FLOPO–TO was consistent. However, eight mappings from the ENVO–SWEET reference alignment made the merged ontology inconsistent.

This process allowed us to detect erroneous mappings introduced by the consensus alignment, due to the presence of homonyms in the SWEET ontology. For instance, `SWEETPhenFluidDynamics:Depression` and `SWEETReprSpaceGeometry:Depression` were both matched to `ENVO:Depression`, which made the equivalence between their subclasses `ENVO:00000394` and `realmHydroBody:Strait` unsatisfiable. Due to the small number of mappings leading to inconsistency, we performed the curation and discarded invalid mappings manually. For instance, both mappings above were erroneous and had to be removed from the reference alignment. In fact, `ENVO:Depression` should have been mapped to `SWEETRealm:Depression`.

4.2 Evaluation phase and results

We have run the evaluation of the biodiversity and ecology (biodiv) track on Windows 10 (64-bit) desktop with an Intel Core i5-7500 CPU @ 3.40 GHz x 4, allocating 15.7Gb of RAM. We conducted experiments by executing each system in its standard settings, and we calculated precision, recall, F-measure as well as its run-time dealing with each task.

4.2.1 Participating systems

In the following, we are going to describe ontology matching systems that have been successfully participating to the biodiv track, namely *AML*, *Lily*, *LogMap* (with its different variants), *POMap*, and eXtended Mapping (*XMap*). We highlight the main characteristics of each system and their use of external knowledge which is needed in order to explain their performances coping with the matching tasks of the proposed track.

AML (Faria *et al.,* 2013, 2014) is a highly scalable ontology matching system, primarily designed to align biomedical ontologies. AML relies heavily on lexical matching techniques, with an emphasis on the use of background knowledge, and includes structural components for both matching and filtering. It features a logical repair algorithm ensuring that the final alignment is coherent and that it has the desired cardinality. AML uses the UBERON (Mungall *et al.,* 2012), the Human Disease Ontology (DOID) (Kibbe *et al.* 2015), and the Medical Subject Headings (MeSH) (Bodenreider, 2004) as mediators between the input ontologies.

Lily (Wang & Xu, 2009; Wang *et al.,* 2011) is a scalable matcher that implements reduction anchors to support the matching of large-scale ontologies. Lily extracts a semantic subgraph for each entity, then it uses both linguistic and structural information for similarity computation. Lily uses a structure-based matcher, which is an extension of the similarity flooding algorithm (Melnik *et al.,* 2002). Alignments are extracted and refined by mapping debugging. Lily uses the UMLS Metathesaurus (Bodenreider, 2004) as background knowledge; it relies on synonyms provided by the Metathesaurus.

LogMap (Jiménez-Ruiz & Grau, 2011; Solimando *et al.,* 2017) is a highly scalable ontology matching system with built-in reasoning and repair capabilities. LogMap implements optimized data structures for lexically and structurally indexing the input ontology. It constructs an inverted lexical index for each input ontology allowing the use of an external lexicon, such as WordNet and UMLS to find their synonyms and lexical variations. Furthermore, the matching system exploits the information in the class hierarchy

---

[12] www.cs.ox.ac.uk/isg/tools/ELK/.

Table 8. Results for the ENVO and SWEET matching task

Systems	Size	Precision	F-measure	Recall	Time (s)
AML	791	0.776	0.844	0.926	46
Lily	491	0.866	0.737	0.641	255
LogMap	583	0.839	0.785	0.738	20
LogMapBio	572	0.839	0.777	0.724	341
LogMapLite	740	0.732	0.772	0.817	18
POMap	583	0.839	0.785	0.738	210
XMap	547	0.868	0.785	0.716	21

of the input ontologies. It employs reasoning and repair techniques to minimize the number of logical errors (inconsistency) in the aligned ontology. In addition to the main matching system, *LogMap* has two more variants:

- *LogMapLite* is a lightweight variant of LogMap, which essentially only applies (efficient) string matching techniques.
- *LogMapBio* includes an extension to use BioPortal[13] as a (dynamic) provider of mediating ontologies instead of relying on a few preselected ontologies.

While LogMapBio uses Bioportal as background knowledge, LogMap uses the UMLS Lexicon[14] which is a different resource than the UMLS Metathesaurus (Bodenreider, 2004). LogMap splits and normalizes each label of each class in the input ontologies into terms and looks for its lexical variation in the UMLS lexicon.

POMap (Laadhar *et al.,* 2017) is an ontology matching system that relies on lexical and structural indexes to match scalable ontologies. The system uses the UBERON (Mungall *et al.,* 2012) (an external biomedical knowledge source) to generate the first set of semantic mappings then employs a syntactic matcher to add mappings that can be discovered at the lexical level. POMap enriches this alignment by applying a structural matcher, which exploits information from the ontologies structure.

XMap (Djeddi & Khadir, 2014; Djeddi *et al.,* 2018) implements a divide-and-conquer technique with parallel processing to enable the composition of different matchers and therefore to handle the matching of large-scale ontologies. The system relies on the use of external resources such as the UMLS Metathesaurus and WordNet[15]. It computes similarity using a terminological layer and a structural layer. The terminological layer comprises string, linguistic (based on WordNet and UMLS), and semantic matchers. The structural layer comprises matchers at the hierarchy, as well as the internal structure of concepts. These matchers are executed in parallel to tackle the scalability problem. It employs an ALCOMO technique ‘Applying Logical Constraints on Matching Ontologies’ (Meilicke, 2011) to detect and remove the inconsistent mappings.

4.2.2 Results against complete reference alignment

Table 8 shows the results for the ENVO–SWEET alignment task in terms of F-measure and the size of alignments. Regarding F-measure, most systems performed well, the top four ranked systems are AML, XMap, POMap, and LogMap. Among these, AML achieved the highest F-measure (0.84) but lower precision. XMap and Lily achieved the highest precision, followed by LogMap, LogMapBio, and POMap. In terms of number of mappings, AML and LogMapLite generated much more mappings than other systems with a lower precision which corresponds mainly to lexical-based generated mappings.

---

[13] bioportal.bioontology.org/.

[14] www.nlm.nih.gov/pubs/factsheets/umlslex.html.

[15] wordnet.princeton.edu/.

Table 9. Results for the FLOPO and TO matching task

Systems	Size	Precision	*F*-measure	Recall	Time (s)
AML	233	0.88	0.86	0.84	35
Lily	176	0.813	0.681	0.586	954
LogMap	235	0.817	0.802	0.787	24
LogMapBio	239	0.803	0.795	0.787	355
LogMapLite	151	0.987	0.755	0.611	10
POMap	261	0.663	0.685	0.709	1261
XMap	153	0.987	0.761	0.619	22

Table 10. Recall against manually assessed mappings

	AML	Lily	LogMap	LogMapBio	LogMapLite	POMap	XMap
ENVO–SWEET	0.73	0.05	0.04	0.04	0.16	0.23	0.19
FLOPO–TO	0.86	0.0	0.64	0.64	0.0	0.55	0.15

Table 9 shows the results for FLOPO–TO in terms of *F*-measure and size of alignments. Regarding *F*-measure, the top three ranked systems are AML, LogMap, and LogMapBio. Among these, AML achieved the highest *F*-measure (0.86). AML is well balanced: it achieved more than 80% recall with still a quite high precision. The systems with the highest precision but lower recall are XMap and LogMapLite, the latter being essentially based on string matching, this indicates that there are fewer issues with homonyms and lexically similar entities between the pair of considered ontologies. The lower recall, on the other hand, shows that lexical matching is not sufficient to detect all correct mappings. Furthermore, the number of mappings generated by AML, LogMap, and LogMapBio is quite the same, those systems achieved a better recall.

In terms of run-time, four out of seven systems were able to achieve the alignment task in less than 50 s for both matching tasks (see Tables 8 and 9). These are LogMapLite followed by LogMap, XMap, and AML. LogMapLite has the shortest run-time. However, the tables show that there is no correlation between the quality of the generated alignment in terms of precision and recall and the required run-time. AML is the most balanced system: it achieves more than 80% *F*-measure with a less than 50 s run-time for both tasks.

4.2.3 Results against unique and manual mappings

In the following, we show the results of the matching systems against the manually assessed mappings and the manual mappings. Table 10 shows the recall obtained by the matching systems against the manually assessed unique mappings. For the ENVO–SWEET matching task, the set of unique mappings that have been assessed were detected by AML and POMap; thus, those two systems performed the best. The highest recall has been attained by AML which already had the best positive contribution with the former version of the system (cf. Table 5). XMap was able to detect 20% of the mappings despite the fact that it was not used in the track preparatory phase.

For the FLOPO–TO matching task, AML being the only system contributing to the manually assessed mappings, it has naturally the best recall followed by LogMap, LogMapBio, and POMap which could detect a bit more than half of the mappings. In both tasks, LogMapLite performed poorly which indicates that those mappings cannot be detected with a purely lexical matcher.

We clearly did introduce bias by assessing a large number of mappings produced by some of the systems. In future editions, we will avoid this by considering the same number of manually assessed mappings per system.

Concerning manual mappings, none of the systems has been able to detect them. Detecting this kind of mappings necessitates the use of domain-specific background knowledge capturing biodiversity-related concepts.

5 Discussion

The results from the OAEI track confirmed our assumption about the need for specific consideration while dealing with biodiversity- and ecology-related ontologies. The specificities of the considered ontologies in terms of structure and vocabulary highlight well-known challenges and introduce new ones. In the following, we illustrate particular issues raised during the evaluation process.

One well-known issue in the ontology matching realm is the difference between equivalence and unidirectional synonymy. Various ontologies deal with this issue in different ways, thus having different modeling approaches.

While SWEET uses equivalence relations between terms, ENVO is modeling equivalence using the annotation relations *has_exact_synonym* and *has_related_synonym*. The annotation *has_exact_synonym* can be treated as equivalence; however, the value of *has_related_synonym* is more ambiguous. Most likely, the latter can be treated as a unidirectional synonym. An example is the class 'Archipelago' (`ENVO:00000220`). The related synonym 'Island arc' is kind of an archipelago; however, not all archipelagoes are island arcs (only the arc-shaped ones are). Those specific cases must be tackled by the matching process.

Many unnecessary and redundant mappings are the result of consideration of unidirectional synonyms. For example, ENVO 'Landslide' (`ENVO:00000520`) has been mapped to 'Slide' in SWEET (`SWEETPhenGeolGeomorphology:Slide`), although SWEET already contains Landslide. On the other hand, connections between mapped ontologies might also get lost, if these unidirectional synonyms were not considered.

In future studies, it would be interesting to investigate the influence of unidirectional synonyms on semantic precision and recall and to determine the extent of unidirectional synonym mappings among different matching systems outputs. The assessment of unidirectional synonyms could be useful for the evaluation of matching systems. The concept of synonym synergy, which has been developed and applied to calculate the coverage metric for SWEET previously (DiGiuseppe *et al.,* 2014), might be adapted for that purpose.

A different issue we have been facing is error propagation. If an ontology already contains modeling errors, especially regarding equivalence relationships, the mapping output between two ontologies will also be erroneous. For example, SWEET contains an equivalence relationship between 'Mountain', 'Orography', and 'Orographic'. This relationship does not only mix up different word types but also different lexical meanings.

Also, homonyms can lead to wrong mapping results. The same wording can have two completely different semantic meanings within two different domains. Especially for SWEET, which consists of over 220 single small ontologies corresponding to different subdomains, homonyms can be wrongly mapped to other ontologies. An example of a false positive match between SWEET and ENVO due to homonyms is ENVO 'Trough' (`ENVO:00000499`) and SWEET 'Trough' (`SWEETPhenFluidDynamics:Trough`). While the first denotes a linear depression of the earth surface, the latter denotes a linear region of low pressure.

In order to deal with such cases, ontology matching algorithms should take into account the origin of the terms which is encoded within the term URI (Uniform Resource Identifier) (e.g. http://sweetontology.net/phenFluidDynamics/Trough). This can help identify the corresponding domain. Another consideration would be to give more importance and strengthen the structural matching part of systems' algorithms.

Closely related to the homonym problematic, another origin of false positives is the wrong evaluation of differences between terms, which share a common domain in principle, but whose definitions vary to a large extent. An outstanding example is the difference between the definition of 'Ice' within SWEET and ENVO. SWEET 'Ice' (`SWEETMatrWater:Ice`) refers to solid water, while ENVO 'Ice'

(`ENVO:01001125`) refers to any crystalline, solid substance. False positive mappings could be observed for this example. The appropriate mappings should have been SWEET Ice (`SWEETMatrWater:Ice`) and ENVO 'Water ice' (`ENVO:01000277`).

FLOPO and TO agree to a minor extent by formal structure but to a greater extent by reference to a common vocabulary (provided by PO). For instance, `TO:0000906` *exocarp morphology trait* as well as `TO:0000945` *pericarp morphology trait* is a direct subclass of `TO:0000843` *portion of plant tissue morphology trait*. Compared with this, FLOPO follows the structure of PO (where *exocarp* is `part-of` *pericarp*), so `FLOPO:0006675` *exocarp morphology* is—reflecting fruit anatomy—a subclass of `FLOPO:0001339` *pericarp phenotype*.

Due to the data-driven generation of FLOPO, which is based on identifying entity–quality descriptions of traits in Floras, a class in FLOPO corresponds to at least one taxon annotation in one of the source texts (Hoehndorf *et al.,* 2016)—which implies that terminology not used in the processed Floras is not incorporated in the ontology. For instance, considering the pollen-producing organs of a flower shows that the successful matches stamen morphology (`FLOPO:0005561`) and stamen morphology trait (`TO:0000215`) are derived from conceptually different class trees: in FLOPO (`FLOPO:0000720`; androecium phenotype) from a term for the entirety of all stamina within a flower, and in TO (`TO:0000868`; microsporophyll morphology trait) from a term for spore-bearing leaves used in a broader taxonomic context (but rarely in flowering plants), which renders structural analogies difficult to identify.

6 Conclusion

In this paper, we investigated ontology matching applied to the biodiversity and ecology domain. We started by showing the specific challenges related to this particular domain which is known to be multidisciplinary and data intensive. We illustrate the specific matching needs and the ontologies to be considered based on a set of high-level entities identified in previous work to improve semantic aware data discovery applications in two of our current projects.

To investigate the actual state of automatic ontology matching systems and their ability to deal with biodiversity- and ecology-related ontologies, we proposed a new track to the OAEI. The track contains the ENVO, SWEET, FLOPO, and TO ontologies with their pairwise reference alignments. We did follow a well-established process for the generation of those alignments. The set was composed using a voting process on the basis of matching systems-generated mappings, the manual curation of mappings generated by only one system and a set of manually created alignments.

The biodiversity and ecology track was introduced in OAEI 2018 and the results obtained are very promising. Most participating systems coped well with the detection of mappings against the consensus alignments but struggled to detect mappings generated by domain experts. This deserves more attention in future developments of ontology matching systems by considering knowledge bases from the biodiversity domain.

Ontology matching systems struggled with the specific features of the considered ontologies as well. In particular, the structure of the SWEET ontology, which encodes concept meaning entirely in the ontology hierarchy, was particularly challenging for matching systems and did lead to a high number of false positives. Systems could not cope with homonyms and structural matching solely, indicating a need to strengthen the structural matching algorithms in the absence of sufficient lexical information. On the other hand, the highly specific lexical content of FLOPO and its limited structure agreement with TO puts forward the need for dynamic and personalized use of background knowledge and domain dictionaries.

This is the first work considering the ontology matching problem in biodiversity and ecology, a significant and highly dynamic domain. We confirmed with our experience at OAEI that this domain deserves a particular consideration and that the ontology matching community can greatly benefit from its specific characteristics. In future editions of the OAEI campaign, we will improve the reference alignment by involving more systems in the creation of the consensus alignment and avoiding the bias introduced by assessing a higher number of unique mappings from specific systems. In a longer run, we will expand the track with new tasks, considering additional ontologies from the biodiversity realm.

Acknowledgements

This work was partially funded by the German Research Foundation in the context of the GFBio Project (grant no. SE 553/7-1) and the CRC 1076 AquaDiva, the German Federal Ministry of Education and Research (BMBF) through the project QURATOR (grant no. 03WKDA1A), the Leitprojekt der Fraunhofer Gesellschaft in the context of the MED2ICIN project (grant no. 600628), and the German Network for Bioinformatics Infrastructure—de.NBI (grant no. 031A539B). We would like to thank Ernesto Jiménez-Ruiz and the organizers and participants of the OAEI campaign.

References

Alves, C., Castro, J. A., Ribeiro, C., Honrado, J. P. & Lomba, A. 2018. Research data management in the field of ecology: an overview. In *Proceedings of the International Conference on Dublin Core and Metadata Applications.*

Annane, A., Bellahsene, Z., Azouaou, F. & Jonquet, C. 2016. Selection and combination of heterogeneous mappings to enhance biomedical ontology matching. In *Knowledge Engineering and Knowledge Management - 20th International Conference, EKAW '16, Bologna, Italy, November* 19–23, *2016, Proceedings.*

Arnaud., E., Cooper., L., Shrestha., R., Menda., N., Nelson., R. T., Matteis., L., Skofic., M., Bastow., R., Jaiswal., P., Mueller., L. & McLaren., G. 2012. Towards a reference plant trait ontology for modeling knowledge of plant traits and phenotypes. In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* - Volume **1**: *KEOD, (IC3K '12).* INSTICC, SciTePress.

Arnold, P. & Rahm, E. 2014. Enriching ontology mappings with semantic relations. *Data & Knowledge Engineering* **93**, 1–18. Selected Papers from the 17th East–European Conference on Advances in Databases and Information Systems.

Bodenreider, O. 2004. The unified medical language system (UMLS): integrating biomedical terminology, *Nucleic Acids Research* **32**(Database-Issue), 267–270.

Bruelheide, H., Dengler, J., Purschke, O., Lenoir, J., Jiménez-Alfaro, B, Hennekens, S. M., Botta-Dukét, Z., Chytrý, M., Field, R., Jansen, F., Kattge, J., Pillar, V. D., Schrodt, F., Mahecha, M. D., Peet, R. K., Sandel, B., van Bodegom, P., Altman, J., Alvarez-Dávila, E., Arfin Khan, M. A. S., Attorre, F., Aubin, I., Baraloto, C., Barroso, J. G., Bauters, M., Bergmeier, E., Biurrun, I., Bjorkman, A. D., Blonder, B., Čarni, A., Cayuela, L., Černý, T., Cornelissen, J. H. C., Craven, D., Dainese, M., Derroire, G., De Sanctis, M., Díaz, S., Doležal, J., Farfan-Rios, W., Feldpausch, T. R., Fenton, N. J., Garnier, E., Guerin, G. R., Gutiérrez, A. G., Haider, S., Hattab, T., Henry, G., Hérault, B., Higuchi, P., Hölzel, N., Homeier, J., Jentsch, A., Jürgens, N., Kącki, Z., Karger, D. N., Kessler, M., Kleyer, M., Knollová, I., Korolyuk, A. Y., Kühn, I., Laughlin, D. C., Lens, F., Loos, J., Louault, F., Lyubenova, M. I., Malhi, Y., Marcenò, C., Mencuccini, M., Müller, J. V., Munzinger, J., Myers-Smith, I. H.,Neill, D. A., Niinemets, Ü, Orwin, K. H., Ozinga, W. A., Penuelas, J., Pérez-Haase, A., Petřík, P., Phillips, O. L., Pärtel, M., Reich, P B., Römermann, C., Rodrigues, A. V., Sabatini, F. M., Sardans, J., Schmidt, M., Seidler, G., Silva Espejo, J. E., Silveira, M., Smyth, A., Sporbert, M., Svenning, J., Tang, Z., Thomas, R., Tsiripidis, I., Vassilev, K., Violle, C., Virtanen, R., Weiher, E., Welk, E., Wesche, K., Winter, M., Wirth, C. & Jandt, U. 2018. Global trait-environment relationships of plant communities. *Nature Ecology & Evolution* **2**, 1906–1907.

Buttigieg, P. L., Morrison, N., Smith, B., Mungall, C. J. & Lewis, S. E. 2013a. The environment ontology: contextualising biological and biomedical entities. *Journal of Biomedical Semantics* **4**(1), 43:1–43:9.

Buttigieg, P. L., Pafilis, E., Lewis, S. E., Schildhauer, M. P., Walls, R. L. & Mungall, C. J. 2013b. The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *Journal of Biomedical Semantics* **7**(1), 57:1–57:12.

Cooper, L., Meier, A., Laporte, M.-A., Elser, J. L., Mungall, C., Sinn, B. T., Cavaliere, D., Carbon, S., Dunn, N. A., Smith, B., Qu, B., Preece, J., Zhang, E., Todorovic, S., Gkoutos, G., Doonan, J. H., Stevenson, D. W., Arnaud, E. & Jaiswal, P. 2018. The planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. *Nucleic Acids Research* **46**(D1), D1168–D1180.

Cooper, L., Walls, R. L., Elser, J., Gandolfo, M. A., Stevenson, D. W., Smith, B., Preece, J., Athreya, B., Mungall, C. J., Rensing, S., Hiss, M., Lang, D., Reski, R., Berardini, T. Z., Li, D., Huala, E., Schaeffer, M., Menda, N., Arnaud, E., Shrestha, R., Yamazaki, Y. & Jaiswal, P. 2013. The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant and Cell Physiology* **54**(2), e1.

Dahdul, W. M., Manda, P., Cui, H., Balhoff, J. P., Dececchi, T. A., Ibrahim, N., Lapp, H., Vision, T. J. & Mabee, P. M. 2018. Annotation of phenotypes using ontologies: a gold standard for the training and evaluation of natural language processing systems. *Database* 2018.

David, J., Euzenat, J., Scharffe, F. & dos Santos, C. T. 2011. The alignment API 4.0. *Semantic Web* 2(1), 3–10.

Diaz, S., Purvis, A., Cornelissen, J. H. C., Mace, G. M., Donoghue, M. J., Ewers, R. M., Jordano, P. & Pearse, W. D. 2013. Functional traits, the phylogeny of function, and ecosystem service vulnerability. *Ecology and Evolution* 3(9), 2958–2975.

Diepenbroek, M., Glöckner, F. O., Grobe, P., Güntsch, A., Huber, R., König-Ries, B., Kostadinov, I., Nieschulze, J., Seeger, B., Tolksdorf, R. & Triebel, D. 2014. Towards an integrated biodiversity and ecological research data management and archiving platform: the German Federation for the Curation of Biological Data (GFBio). In *44. Jahrestagung der Gesellschaft für Informatik, Informatik 2014, Big Data - KomplexitÄt meistern, September 22–26, 2014, Stuttgart, Deutschland*, pp. 1711–1721.

Diepenbroek, M., Schindler, U., Huber, R., Pesant, S., Stocker, M., Felden, J., Buss, M. & Weinrebe, M. 2017. Terminology supported archiving and publication of environmental science data in PANGAEA. *Journal of Biotechnology* **261**, 177–186.

DiGiuseppe, N., Pouchard, L. C. & Noy, N. F. 2014. SWEET ontology coverage for earth system sciences. *Earth Science Informatics* **7**(4), 249–264.

Djeddi, W. E. & Khadir, M. T. 2014. A novel approach using context-based measure for matching large scale ontologies. In *Data Warehousing and Knowledge Discovery - 16th International Conference, DaWaK 2014, Munich, Germany, September 2–4, 2014. Proceedings*.

Djeddi, W. E., Khadir, M. T. & Yahia, S. B. 2018. XMap results for OAEI 2018. In *Proceedings of the 13th International Workshop on Ontology Matching Co-located with the 17th International Semantic Web Conference (ISWC '18), USA, 2018*.

Dragisic, Z., Ivanova, V., Lambrix, P., Faria, D., Jiménez-Ruiz, E. & Pesquita, C. 2016. User validation in ontology alignment. In *International Semantic Web Conference*.

Duchateau, F. & Bellahsene, Z. 2016. YAM: a step forward for generating a dedicated schema matcher. *Trans. Large-Scale Data- and Knowledge-Centered Systems* **25**. 10.1007/978-3-662-49534-6_5.

Ehrig, M. 2006. Ontology alignment: bridging the semantic gap. In *Semantic Web and Beyond: Computing for Human Experience*. 10.1007/978-0-387-36501-5

Ehrig, M. 2007. *Ontology Alignment: Bridging the Semantic Gap*, Semantic Web and Beyond: Computing for Human Experience **4**, Springer.

Euzenat, J. & Shvaiko, P. 2013. *Ontology Matching*, 2nd Edition. Springer.

Faria, D., Pesquita, C., Santos, E., Cruz, I. F. & Couto, F. M. 2014. AgreementMakerLight 2.0: towards efficient large-scale ontology matching. In *Proceedings of the ISWC 2014 Posters & Demonstrations Track a track within the 13th International Semantic Web Conference, ISWC '14*.

Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I. F. & Couto, F. M. 2013. The AgreementMakerLight ontology matching system. In *On the Move to Meaningful Internet Systems: OTM 2013 Conferences - Confederated International Conferences: CoopIS, DOA-Trusted Cloud, and ODBASE*.

Fichtmüller, D., Gleisberg, M., Karam, N., Müller-Birn, C. & Güntsch, A. 2017. Terminologies as a neglected part of research data: making supplementary research data available through the GFBio terminology service. In *Proceedings of the 2nd International Workshop on Semantics for Biodiversity Co-located with 16th International Semantic Web Conference (ISWC '17), Vienna, Austria, October 22, 2017*.

Ghazvinian, A., Noy, N. F. & Musen, M. A. 2009. Creating mappings for ontologies in biomedicine: Simple methods work. In *AMIA '09, American Medical Informatics Association Annual Symposium, San Francisco, CA, USA, November 14–18, 2009*. AMIA.

Gkoutos, G. V., Green, E. C. J., Mallon, A.-M., Hancock, J. M. & Davidson, D. 2005. Using ontologies to describe mouse phenotypes. *Genome Biol* **6**(1), R8–R8. https://www.ncbi.nlm.nih.gov/pubmed/15642100.

Grau, B. C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P. F. & Sattler, U. 2008. OWL 2: the next step for OWL. *Journal of Web Semantics* **6**(4), 309–322.

Harrow, I., Jiménez-Ruiz, E., Splendiani, A., Romacker, M., Woollard, P., Markel, S., Alam-Faruque, Y., Koch, M., Malone, J. & Waaler, A. 2017. Matching disease and phenotype ontologies in the ontology alignment evaluation initiative. *Journal of Biomedical Semantics* **8**(1), 55:1–55:13.

Hickey, L. J. 1973. Classification of the architecture of dicotyledonous leaves. *American Journal of Botany* **60**(1), 17–33.

Hoehndorf, R., Alshahrani, M., Gkoutos, G. V., Gosline, G., Groom, Q., Hamann, T., Kattge, J., de Oliveira, S. M., Schmidt, M., Sierra, S., Smets, E., Vos, R. A. & Weiland, C. 2016. The flora phenotype ontology (FLOPO): tool for integrating morphological traits and phenotypes of vascular plants. *Journal of Biomedical Semantics* **7**(1), 65.

Hoehndorf, R., Weiland, C., Schmidt, M., Groom, Q., Gosline, G., Dressler, S. & Hamann, T. 2018. The Flora Phenotype Ontology (FLOPO) and the FLOPO knowledgebase. In *Application of Semantic Technology in Biodiversity Science*, Thessen, A. E. (ed). IOS Press, Chapter 6, pp. 107–119.

IPBES 2018. The IPBES assessment report on land degradation and restoration.

Jiménez-Ruiz, E. & Grau, B. C. 2011. LogMap: logic-based and scalable ontology matching. In *10th International Semantic Web Conference, ISWC '11*, pp. 273–288.

Jiménez-Ruiz, E., Grau, B. C. & Horrocks, I. 2013. Is my ontology matching system similar to yours? In *Proceedings of the 8th International Workshop on Ontology Matching Co-located with the 12th International Semantic Web Conference (ISWC 2013), Sydney, Australia, October 21, 2013*, Shvaiko, P., Euzenat, J., Srinivas, K., Mao, M. & Jiménez-Ruiz, E. (eds), CEUR Workshop Proceedings 1111. CEUR-WS.org, pp. 229–230.

Kalfoglou, Y. & Schorlemmer, W. M. 2005. Ontology mapping: the state of the art. In *Semantic Interoperability and Integration*.

Karam, N., Müller-Birn, C., Gleisberg, M., Fichtmüller, D., Tolksdorf, R. & Güntsch, A. 2016. A terminology service supporting semantic annotation, integration, discovery and analysis of interdisciplinary research data. *Datenbank-Spektrum* **16**(3), 195–205.

Kattge, J., Ogle, K., Bnisch, G., Daz, S., Lavorel, S., Madin, J., Nadrowski, K., Nllert, S., Sartor, K. & Wirth, C. 2011. A generic structure for plant trait databases. *Methods in Ecology and Evolution* **2**(2), 202–213.

Kibbe, W. A., Arze, C., Felix, V., Mitraka, E., Bolton, E., Fu, G., Mungall, C. J., Binder, J. X., Malone, J., Vasant, D., Parkinson, H. E. & Schriml, L. M. 2015. Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Research* **43**(Database-Issue), 1071–1078.

Klan, F., Faessler, E., Algergawy, A., König-Ries, B. & Hahn, U. 2017. Integrated semantic search on structured and unstructured data in the adonis system. In *Proceedings of the 2nd International Workshop on Semantics for Biodiversity Co-located with 16th International Semantic Web Conference (ISWC '17), Vienna, Austria, October 22, 2017.*

Laadhar, A., Ghozzi, F., Megdiche, I., Ravat, F., Teste, O. & Gargouri, F. 2017. Pomap: An effective pairwise ontology matching system. In *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - (Volume 2), Funchal, Madeira, Portugal, November 1–3, 2017.*

Löffler, F., Pfaff, C., Karam, N., Fichtmüller, D. & Klan, F. 2017. What do biodiversity scholars search for? Identifying high-level entities for biological metadata. In *Proceedings of the 2nd International Workshop on Semantics for Biodiversity Co-located with 16th International Semantic Web Conference (ISWC '17), Vienna, Austria, October 22, 2017.*

Meilicke, C. 2011. *Alignment Incoherence in Ontology Matching*. PhD thesis, University of Mannheim.

Melnik, S., Garcia-Molina, H. & Rahm, E. 2002. Similarity flooding: a versatile graph matching algorithm and its application to schema matching. In *ICDE '02.*

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS '13.* Curran Associates Inc., USA.

Mungall, C. J., Torniai, C., Gkoutos, G. V., Lewis, S. E. & Haendel, M. A. 2012. Uberon, an integrative multi-species anatomy ontology. *Genome Biology* **13**(1), R5.

Musen, M. A. 2015. The Protégé project: a look back and a look forward. *AI Matters* **1**(4), 4–12.

Raskin, R. G. & Pan, M. J. 2004. Knowledge representation in the semantic web for earth and environmental terminology (SWEET). *Computers & Geosciences* **31**(9), 1119–1125.

Shvaiko, P. & Euzenat, J. 2013. Ontology matching: State of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering* **25**(1), 158–176.

Smaili, F. Z., Gao, X. & Hoehndorf, R. 2018. OPA2Vec: combining formal and informal content of biomedical ontologies to improve similarity-based prediction. *CoRR* **abs/1804.10922**.

Solimando, A., Jiménez-Ruiz, E. & Guerrini, G. 2017. Minimizing conservativity violations in ontology alignments: algorithms and evaluation. *Knowledge and Information Systems* **51**(3), 775–819.

Stevens, P. F. Glossary of the angiosperm phylogeny website, New York Botanical Garden. sweetgum.nybg.org/science/glossary/glossary-checklist/

Vos, R. A., Biserkov, J. V., Balech, B., Beard, N., Blissett, M., Brenninkmeijer, C., van Dooren, T., Eades, D., Gosline, G., Groom, Q. J., Hamann, T. D., Hettling, H., Hoehndorf, R., Holleman, A., Hovenkamp, P., Kelbert, P., King, D., Kirkup, D., Lammers, Y., DeMeulemeester, T., Mietchen, D., Miller, J. A., Mounce, R., Nicolson, N., Page, R., Pawlik, A., Pereira, S., Penev, L., Richards, K., Sautter, G., Shorthouse, D. P., Thtinen, M., Weiland, C., Williams, A. R. & Sierra, S. 2014. Enriched biodiversity data as a resource and service. *Biodiversity Data Journal* **2**, e1125.

Wagenitz, G. 2008. *Wörterbuch der Botanik,* Nikol.

Walls, R. L., Athreya, B., Cooper, L., Elser, J., Gandolfo, M. A., Jaiswal, P., Mungall, C. J., Preece, J., Rensing, S., Smith, B. & Stevenson, D. W. 2012. Ontologies as integrative tools for plant science. *American Journal of Botany* **99**(8), 1263–1275.

Wang, P. & Xu, B. 2009. An effective similarity propagation method for matching ontologies without sufficient or regular linguistic information. In *The Semantic Web, Fourth Asian Conference, ASWC 2009, Shanghai, China, December 6–9, 2009. Proceedings*, pp. 105–119.

Wang, P., Zhou, Y. & Xu, B. 2011. Matching large ontologies based on reduction anchors. In *IJCAI '11, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16–22, 2011*, pp. 2343–2348.

Weiland, C., Kulmanov, M., Schmidt, M. & Hoehndorf, R. 2019. A machine learning based approach for similarity search on biodiversity knowledge graphs. *Biodiversity Information Science and Standards* **3**.

Whetzel, P. L., Noy, N. F., Shah, N., Alexander, P. R., Dorf, M., Fergerson, R. W., Storey, M. D., Smith, B., Chute, C. G. & Musen, M. A. 2011. Bioportal: ontologies and integrated data resources at the click of a mouse. In O. Bodenreider, M. E. Martone & A. Ruttenberg, eds, *Proceedings of the 2nd International Conference on Biomedical Ontology, Buffalo, NY, USA, July 26–30, 2011,* CEUR Workshop Proceedings **833**, CEUR-WS.org.

Wilkinson, M. D., Dumontier, M., Aalbersberg & al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018+.

Yilmaz, P., Kottmann, R., Field, D., Knight, R. & al. 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (mixs) specifications. *Nature Biotechnology* **29**(5), 415–420.