BiodivPortal: Enabling Semantic Services for Biodiversity within the German National Research Data Infrastructure

Naouel Karam, Jan Fillies, Clement Jonquet, Syphax Bouazzouni, Felicitas Löffler, Franziska Zander, Birgitta König-Ries, Anton Güntsch, Michael Diepenbroek, Adrian Paschke

BiodivPortal: Enabling Semantic Services for Biodiversity within the German National Research Data Infrastructure

Naouel Karam[1*], Jan Fillies[1,2], Clement Jonquet[3], Syphax Bouazzouni[3], Felicitas Löffler[4], Franziska Zander[5], Birgitta König-Ries[5], Anton Güntsch[6], Michael Diepenbroek[7], Adrian Paschke[1,2,8]

[1*]Institute for Applied Informatics (InfAI), Leipzig, Germany.
[2]Freie Universität Berlin, Berlin, Germany.
[3]University of Montpellier & CNRS, Montpellier, France.
[4]Thuringian Ministry of Finance, Erfurt, Germany.
[5]Friedrich-Schiller-University, Jena, Germany.
[6]Botanic Garden and Botanical Museum, Berlin, Germany.
[7]GFBio - German Federation for Biological Data, Bremen, Germany.
[8]Fraunhofer FOKUS, Berlin, Germany.

*Corresponding author(s). E-mail(s): karam@infai.org;

Abstract

Research has become increasingly reliant on extensive data. The integration, sharing and reuse of research data poses a significant challenge, particularly in the context of interdisciplinary collaborative projects. An essential objective for a research infrastructure dedicated to data management is to facilitate efficient data discovery and integration of diverse data sources. This pressing need for FAIR data requires, besides persistent identifiers and data citation rules, common standards and shared vocabularies, thesauri and ontologies. These knowledge artifacts, referred to as terminologies, often exist in disconnected and distributed forms. The work presented in this paper describes our terminology repository and service, enabling a unified access, development, and maintenance of terminologies within biodiversity and environmental sciences. We characterize use cases requirements for semantically enhanced components and applications and show where the adoption of the OntoPortal technology enabled us to cover those requirements in the context of our research infrastructure.

Keywords: Research data infrastructure, Data harmonization, Data integration, Terminology repository, Terminology service, Ontologies, Taxonomies

1 Introduction

Biological and ecological research has the particularity of dealing with a substantial amount of data generated by diverse disciplines (e.g., Botany, Microbiology, Chemistry, Geo-Sciences) and addressing a multitude of topics ranging from marine ecosystems to species distribution [2]. This results in extremely heterogeneous data stored in disparate data archives, which makes data acquisition more and more time-consuming and a challenging task for researchers. The situation

is further complicated by different understandings of employed terms within different scientific disciplines. A primary goal of state of the art ecological research infrastructures for data management is to integrate those heterogeneous data into meaningful, interoperable knowledge. This enables primarily efficient data discovery with an undeniable ultimate goal of deriving new insights from merging complementary datasets.

In response to these challenges, the establishment of the German National Research Data Infrastructure (NFDI)[1] took place with the primary objective to offer comprehensive data management and archiving solutions and enhance data capture, sharing, and discovery within research institutions at the national level.

The biodiversity community has been investing a lot of effort into supporting semantic interoperability by developing a set of metadata standards as well as formalizing domain knowledge in terms of ontologies. This formal knowledge provides a means to enrich data with annotations for efficient data discovery and curation. The discipline counts a growing number of ontologies covering different aspects of the environmental spectrum, like the Environmental Ontology (ENVO) [3] for environmental features and materials, the Phenotype Quality Ontology (PATO) [9] and the Flora Phenotype Ontology (FLOPO) [11] for phenotypes. Additional valuable resources central to biodiversity management are taxonomies of biological organisms, such as the Integrated Taxonomic Information System[2] (ITIS) and the NCBI Taxonomy[3] (NCBITAXON) [8]. These are widely used since decades in biological studies, but most of them are not available in a Semantic Web compliant format. Typically, taxonomies are accessible through a set of web services and can be easily accessed programmatically.

Based on a set of initial requirements for a terminology repository and service, we developed and deployed the GFBio Terminology Service (GFBio TS) [5, 16], providing the necessary "meaning" to heterogeneous data in terms of structure, formats, and vocabulary. In the context of the NFDI4Biodiversity project [10], our objective is to continue the development and enhancement of the GFBio TS in a broader collaborative context, considering our acquired experience as well as the new requirements driven by a larger community and additional use cases. We adopted OntoPortal, a generic technology to build ontology repositories [15] for developing the next generation of the GFBio TS called BiodivPortal (https://biodivportal.gfbio.org).

The transition to the new technology mandated a thorough assessment of our use cases' requirements and the functionalities already supported by the GFBio TS. A key requirement driving the development of our custom solution was the need for federated access to the taxonomies mentioned earlier. To address this, we created transformation pipelines to convert them into Semantic Web formats and seamlessly integrate them into BiodivPortal.

In the following sections, we will describe our vision of a semantically enriched research data infrastructure. We then outline the requirements for a terminology service supporting this vision. We will delve into the transition of our solution, encompassing both technological and content aspects. Finally, we will discuss the current status, identified limitations, and outline our plans for future development.

2 A semantically enabled research data commons

The NFDI4Biodiversity consortium is developing a Research Data Commons (RDC) [7] as a scalable, cloud-based research infrastructure within the German National Research Data Infrastructure (NFDI). The primary goal of the RDC is to foster collaborative data exchange in biodiversity and related domains. It enables users to get access to heterogeneous data sources and conduct complex analyses. Adhering to the FAIR principles (findability, accessibility, interoperability, and reusability), the NFDI4Biodiversity RDC aims to offer high quality actionable data products and services through its API. Based on an incremental development strategy, the initial architecture will be refined with a growing service portfolio developed collaboratively with other NFDI consortia.

As shown in figure 1, the RDC architecture features multiple layers and data flows, with lower

[1] https://www.nfdi.de/?lang=en
[2] https://itis.gov/
[3] https://ncbi.nlm.nih.gov/taxonomy

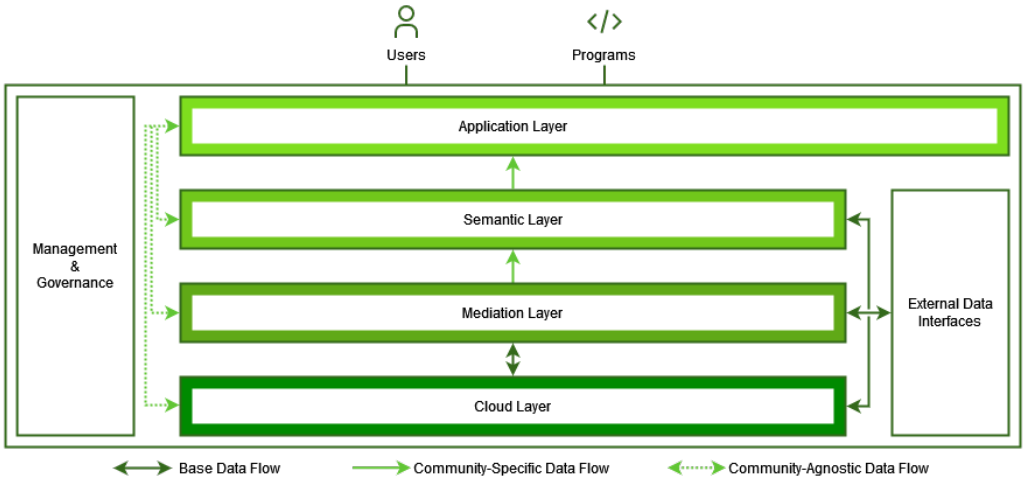

Fig. 1 NFDI4Biodiversity Research Data Commons (RDC) Architecture

layers assuring technical functionalities and upper layers tailored for end-users with domain knowledge:

- The Cloud Layer serves as the technical foundation and is based on a multi-cloud infrastructure. It offers a scalable distributed computing and cloud storage enabling users to run compute-intensive tasks and analyze large datasets. Services like Aruna Object Storage (AOS) [4] streamline data management in a unified model.
- The Mediation Layer provides a self-service toolkit of community-agnostic tools and services for creating FAIR data products. The layer provides a set of transformation services and tools to convert data and metadata and enhance data quality.
- The Semantic Layer provides community-specific data products developed and managed in the Mediation Layer, utilizing technical datasets from the Cloud Layer and external data interfaces. These FAIR-compliant, computer-actionable data products can be physically available or generated on demand.
- The Application Layer encompasses tailored applications and services for end users, which can be community-agnostic (e.g., a dataset search tool) or community-specific (e.g., a data portal for a specific NFDI4Biodiversity community).

Additionally to the horizontal layers, two crucial elements are part of the RDC architecture: the Management & Governance tools and policies for managing rules and access rights across the four layers, as well as External Data Interfaces for accessing external datasets from established data providers from the NFDI4Biodiversity partner network.

A key component of the RDC Mediation Layer is the BiodivPortal. It is a terminology repository and service for supporting the management, sharing and use of biodiversity-related terminologies. Components of the RDC can perform semantic enrichment by reusing a set of offered semantic services or by directly accessing and integrating provided terminologies.

3 Use cases requirements for a terminology service

In order to demonstrate why the new BiodivPortal is necessary within the RDC, use cases from NFDI4Biodiversity partner systems have been identified, collected, and translated into requirements. The requirements engineering process unfolded in a series of workshops with different stakeholders (developers, curators, database maintainers) from each system. Table 1 summarizes all requirements, the circles represent if a requirement is fully/partially/not covered by the terminology service. Included under [R1-R3] are the initial three requirements collected by Karam et al [16] during the first development stage of the GFBio TS. Additionally, the newly identified requirements [R4-R9] are linked to specific

Table 1 Requirements on a Terminology Repository and Service.

#	Requirement	GFBio TS	BiodivPortal
R1	A repository to store, connect, search, and browse terminologies	◐	●
R2	Provide a single access point (API) to heterogeneous terminological resources	●	●
R3	Offer access to external terminological resources (not available in a Semantic Web format)	●	●
R4	Offer efficient semantic annotation for text and tabular data	○	●
R5	Export ontologies in different formats	○	●
R6	Offer an environment for the development, curation, and publication of project terminologies	○	◐
R7	Automatically generate and store mappings between terminologies	◐	●
R8	Terminology versioning and evolution mechanisms	○	◐
R9	Recommendation mechanisms to identify relevant terminologies for an application case	○	●
R10	Community feedback mechanisms	○	●
R11	Multilingual support	◐	●

needs from the use cases. Two further nice to have requirements [R10-R11] were identified for an improved community involvement as well as a multilingual support especially for German. In the following, we describe three NFDI4Biodiversity partner systems alongside their requirements for a terminology service, namely BEXIS2, the EDIT platform and PANGAEA.

3.1 BEXIS2

BEXIS2[4] is a modular, scalable, interoperable, free, and open source system supporting large research consortia on all aspects of research data management. The software is being developed based on requirements from the biodiversity and ecology domain that mostly deal with tabular data, but it can be configured to serve other domains and data types as well.

Research questions and data in biodiversity research are quite heterogeneous. Proper data and metadata descriptions mostly depend on the scholars themselves how much effort and time they spend on describing their data. Therefore, it is almost impossible to provide only one data structure and one metadata schema that fits for all purposes. On the other hand, there is an increasing demand for data integration and data reuse. Terminology Services are essential building blocks for overcoming arbitrary data descriptions and for semantic enrichment of research data to be integrated into data management systems such as BEXIS2.

Instead of using own terms when describing a column or creating metadata, a search or suggest service in BEXIS2 to get suitable URIs [R3] would be a great benefit for data providers and curators. This is particularly helpful when working with species, data parameters, units, and chemical elements. In order to enrich poorly described metadata to support researchers with an automatic process for metadata creation or data curators for quality control, a semantic annotation service would be an advantage. For instance, if the metadata contains data parameters in the description field or if headers in tabular data give hints about whether species have been observed, a service that automatically recognizes species, parameters measured, or habitat information would speed up daily research practice [R4]. In some cases, not only one terminology entry will be found, but several URIs might match. Here, a mapping service identifying corresponding entries in other vocabularies would support semantic enrichment. For instance, if a data provider or data curator is looking for a 'butterfly' the service should not only return the information from NCBITAXON but also additional ones from the identical entry in ITIS for

[4] https://fusion.cs.uni-jena.de/bpp/

instance [R7]. For data managers and curators, it would be helpful to have a recommendation service to determine what vocabularies might be relevant for the project, as not all terminologies are suitable for all purposes [R9]. By implementing the technical base infrastructure to manage associated URIs on any level, e.g., dataset, metadata, data, and data parameters within the BEXIS2 version 3.0.0, the base has also been set to integrate provided services by NFDI4Biodiversity.

3.2 EDIT platform for cybertaxonomy

The EDIT Platform for Cybertaxonomy[5] was developed within the framework of the EU Network of Excellence EDIT (European Distributed Institute of Taxonomy, 2006-2011) with the aim of optimizing taxonomic workflows based on data standards, well-defined interfaces and services. Today, the platform offers a range of tools for e.g. taxonomic data capture and processing, the generation of maps and identification keys, data publication via portals and web services and for the generation of print publications. The Platform for Cybertaxonomy is based on a comprehensive "Common Data Model" (CDM), which was developed on the basis of existing biodiversity informatics data standards and refined over many years of practical operation.

A module was developed that allows the formal recording of descriptions of collection objects and the automated generation of species descriptions and identification keys. The use of standardized vocabularies is essential for the processing and comparability of the characteristics and characteristic values used in this process. The structures (e.g. "leaves") and properties (e.g. "length") underlying the characteristics can be provided by the GFBio Terminology Server and linked to characteristics (e.g. "leaf length") in the term editor of the EDIT platform. Missing structures and properties are then added locally, but there is no simple way to add them to the terminology service. For a further expansion of this use of a terminology service, it would be desirable that requests for additional terms could be processed. In addition to the technical requirements, this includes the development of a coordinated consensus process and a concept for handling evolving and versioned vocabularies [R8] as well as the links from infrastructures using these vocabularies [R4].

[5] http://www.cybertaxonomy.org

3.3 PANGAEA digital data repository

PANGAEA is a digital repository for Earth and Environmental scientific datasets jointly managed by the Center for Marine Environmental Sciences (MARUM), University of Bremen, and Alfred Wegener Institute for Polar and Marine Research (AWI). The repository currently holds more than 390,000 datasets. In PANGAEA, a 'parameter' of a dataset is an unstructured text, composed of primarily observed property (or a physical quantity) and other concepts such as feature-of-interest, units, aggregate functions (e.g., average and maximum), method, device, location, and time. An example of a parameter is 'Methane, daily formation rate per unit sediment mass'. Currently, there are more than 175,000 parameters of various sub-disciplines of environmental sciences defined in the repository. These parameters are unstructured due to the lack of standardized practices on naming parameters and heterogeneous data submissions.

In order to make PANGAEA parameters machine-readable and understandable, an annotation service was developed, which provides suggestions of terms relevant to a parameter [12]. Those terms are specified in external terminologies such as WoRMS[6] (World Register of Marine Species), QUDT[7] (Quantities, Units, Dimensions and Types) and PATO [9]. At present, we have developed tailor-made client applications customized to each of the terminologies [6]. The clients import terminologies from the external repositories into the PANGAEA data system and update them periodically. Building several client applications may require significant development and maintenance effort, mainly when new terminologies should be imported, or the specification of the terminologies imported changes. The OBO Foundry and BioPortal offer terminologies in several formats (e.g., OWL, RDF, OBO), however, they are limited to the biomedical domain.

[6] https://www.marinespecies.org/
[7] https://qudt.org/

Biodiversity research involves data from multiple ecological disciplines and socioeconomic data [19]. Therefore, an API that provides universal access to various biodiversity and related terminologies is essential. The API should support bulk access request, e.g., whole terminology import or a part of it (subtree). The terminologies should be accessible in serialization formats (e.g., JSON-LD) such that they can be easily integrated into the existing data system [R5]. As source terminologies may be updated over time, the API should provide access to a terminology through incremental changes [R8].

4 From the GFBio terminology service to BiodivPortal

In an effort to align with international initiatives, we joined the OntoPortal alliance,[8] a consortium of partners dedicated to the collaborative development and promotion of open ontology repositories. Members of the alliance maintain repositories for different scientific disciplines and participate in the further development of the OntoPortal open-source software [15], based originally on the BioPortal code. BiodivPortal code is publicly available at https://github.com/biodivportal. The code contains forked repositories from OntoPortal as well as our own developed tools. By adopting the OntoPortal technology, we are able to cover all the initial requirements and many of the new requirements and goals specified within the NFDI4Biodiversity project.

4.1 The OntoPortal technology

As members of the OntoPortal Alliance, we are maintaining and further developing our own instance of the OntoPortal technology[9] dedicated to biodiversity related terminologies, called BiodivPortal. BiodivPortal is a component of the Mediation Layer of the NFDI4Biodiversity RDC (c.f. Fig. 1). It is accessible to other RDC components as well as external systems.

OntoPortal is a generic technology to build terminology repositories, catalogues, and services. The system architecture of OntoPortal is structured in several layers (See figure 2):

- The *storage layer* contains a triplestore which saves each terminology in a distinct graph, as well as other data (metadata records, mappings, users, etc.). This layer also includes: (i) a storage for application caches and the Annotator dictionary datastore; (ii) a Solr search engine to index terminology content for retrieval with the Search service.
- The model layer implements all the mechanisms to parse the terminology source files using the OWL API[10] and retrieve them from the triple-store using the built-in Graph Oriented Objects library (GOO). The service layer implements the core OntoPortal services: Search, Annotator, and Recommender.
- The API layer implements a unified API for all the models (e.g., Class, Instance, Ontology, Submission, Mapping, Review, Note, User) and services supported by OntoPortal. The API returns as default a JSON-LD[11] format.
- The user interface offers a set of various views to display and use the services and components built in the API layer. The UI is customized for logged-in users and groups/organizations that display their own subset of resources. Administrators of the OntoPortal instance have access to an additional administration console to monitor, and manage the content of the portal.

4.2 Included terminologies

In our context, a terminology refers to any terminological resource. This can be a formal ontology, a taxonomy, or any useful collection of terms (e.g., locations available via a geographical database like Geonames[12]).

In a joint effort aimed at enhancing data retrieval in the search applications within the aforementioned project, we consolidated a set of high-level entities in the biodiversity domain that are relevant for biodiversity researchers when searching for data [20]. These entities span various sources and aim to encompass a broad domain spectrum both from a user and an application perspective. The identified entities formed the

[8] http://ontoportal.org/about/
[9] http://ontoportal.org
[10] https://github.com/owlcs/owlapi
[11] https://json-ld.org/
[12] http://www.geonames.org/

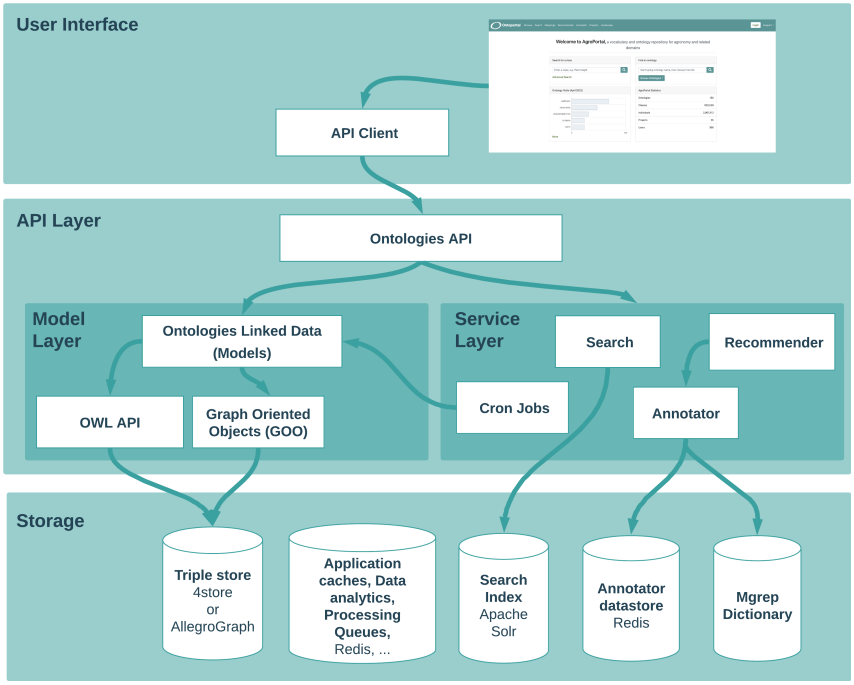

Fig. 2 The OntoPortal System Architecture

foundation for consolidating existing terminologies useful for data annotation and integration. We identified alignment needs between those terminologies in order to combine them in semantically enabled applications [1, 17].

All terminologies included in the GFBio TS are included in BiodivPortal. For terminologies that were remotely accessed in a federated manner via their web services, we developed transformation pipelines for converting structured data, e.g., CSV or relational database tables, into a semantic web format. Pipelines for ITIS and the fragment Germany of GeoNames have been implemented[13]. As for the NCBITaxonomy ontology[14], our translation methodology for taxonomies treats each taxon as a class whose instances would be individual organisms. For Geonames, the Geonames ontology is used as the basis for transforming the database dump into an OWL ontology. The transformations are triggered manually whenever a new dump is made available on the respective providers' websites.

An additional type of semantic resources especially meant for structured data integration are semantic domain specific data and metadata standards like the ABCD 3.0[15] (Access to Biological Collection Data) standard [13] or Schema.org[16].

4.3 Requirements covered by BiodivPortal

Requirement [R3] is now fulfilled through the transformation of external terminologies in a Semantic Web compliant format.

Additional requirements [R4, R5, R7, R10 and R11] are now covered by the OntoPortal technology:

- **R4 - Offer efficient semantic annotation for text and tabular data**. A key functionality of OntoPortal is the Annotator, a domain-agnostic text annotation service. It enables the identification of semantic classes within texts. The Annotator is based on a syntactic concept recognition tool leveraging concept names and synonyms, including subclass relations and

[13] https://github.com/biodivportal/rdb2rdf
[14] https://github.com/obophenotype/ncbitaxon
[15] https://abcd.tdwg.org/3.0/
[16] https://schema.org/

mappings. OntoPortal does not support tabular data annotation. We will consider tools from the SemTab challenge[17] for that purpose.

- **R5 - Export ontologies in different formats**. The technology stores all terminology versions alongside their metadata. Only the latest versions are indexed and loaded in the backend, while source files of former versions are accessible for download in different formats.
- **R7 - Automatically generate and store mappings between terminologies**. OntoPortal includes a mapping repository that stores and manages mappings between classes, supporting their identification, storage, retrieval, and deletion. The portal automatically generates syntactic mappings using the LOOM (Lexical OWL Ontology Matcher) algorithm. Users can explicitly upload mappings from external sources with corresponding provenance information. One of the main features being implemented by the alliance partners is the support for the Semantic Standard for Ontology Mapping (SSSOM) [21]. At the Ontology Alignment Evaluation Initiative (OAEI) [22], we evaluate state-of-the-art ontology matching tools that can be plugged in on demand.
- **R9 - Recommendation mechanisms** OntoPortal identifies pertinent terminologies tailored to specific application cases. It employs four criteria for recommendations: extent of the covered input data, number of views in the portal, level of detail within the covering classes, and specialization of the ontology to the input domain. Utilizing these criteria, the suitability of an ontology for the application case is assessed. Moreover, users have the flexibility to adjust the weight assigned to each criterion, influencing the final recommendation.
- **R10 - Community feedback mechanisms**. OntoPortal incorporates various community-oriented features to facilitate engagement, feedback, and project collaboration. These features include the ability for logged-in users to write reviews for ontologies, attach notes to specific artifacts or classes for discussion, and submit change requests.
- **R11 - Multilingual support**. OntoPortal presents three primary concepts within the realm of multilingualism. Firstly, it facilitates the display of content in multiple predefined languages. Secondly, it enables cross-language search functionality. Thirdly, it supports the internationalization of the user interface into multiple languages.

[17] https://www.cs.ox.ac.uk/isg/challenges/sem-tab/

5 Current state and limitations

As highlighted in the previous sections, the adoption of the OntoPortal software enabled us to cover many of the requirements we identified together with our NFDI4Biodiversity partners. Some requirements (R6 and R8) are partially covered by the new technology, as depicted in Table 1. Ongoing work in collaboration with the OntoPortal alliance partner aims to extend functionalities. For instance, covering the whole terminology development life cycle is one of the main endeavors undertaken by EcoPortal [18]. To facilitate collaboration on the content of the portal, EcoPortal incorporated a connector to the VocBench 3 system[18]. This system offers web-based, multilingual, and collaborative development capabilities for terminologies both in SKOS and OWL.

Concerning terminology versioning and evolution mechanisms, the adoption of the recommendations introduced in [14] as well as the adoption of the KGCL[19] (Knowledge Graph Change Language) standard has been agreed upon within the alliance.

Performance issues were induced by the federated framework, which was a requirement we had originally for the GFBio TS for providing access to terminologies via their APIs. Indeed, the service performance depends strongly on the capabilities of those decentralized services, the so-called external terminologies. In order to overcome this bottleneck, we decided to centrally store essential resources, which led to a drastic improvement in query performance. The corresponding requirement is fulfilled by providing local copies of the terminologies now available in BiodivPortal. The drawback of this approach is that we do not have access to the latest version of those terminologies in real time, new versions of those are loaded through the transformation pipeline when

[18] https://vocbench.uniroma2.it/
[19] https://incatools.github.io/kgcl/

new dumps are made available, for instance once a month or bi-yearly.

In the process of replacing our own developed solution with the OntoPortal technology, we lost some essential functionalities, namely Linked Data deployment (i.e. providing HTTP pages for resolving terms identifiers) as well as access to terms modeled as individuals inside a terminology (for instance QUDT units are modeled as individuals/instances not concepts/classes). Reestablishing those functionalities is essential for some applications relying on our services, and until implemented in BiodivPortal, connection to the GFBio TS will still be maintained.

6 Conclusion and perspectives

Semantic technology plays a crucial role in facilitating scientific research through various means. In this work, we illustrated how a terminology service can effectively cover specific needs of different use cases within our scientific community. Through leveraging semantic services, researchers can enhance data interoperability, and enable more efficient knowledge discovery and integration.

By joining the OntoPortal alliance, beside contributing to an international collaborative effort, we are involved in standardisation efforts and adopting international standards at an early stage, working towards more FAIR terminologies for enabling FAIR research data. Furthermore, members of the OntoPortal alliance are involved in the FAIR-IMPACT project[20] within the European Open Science Cloud (EOSC), reviewing governance models for terminologies and discussing the role terminology repositories play in this governance. Organizing and guiding the biodiversity community in the NFDI context will enhance Germany's position within broader initiatives like EOSC. NFDI4BioDiversity and its cloud-based services are designed to align with the evolving EOSC services.

The expanding use of OntoPortal based systems and the emergence of other terminology service technologies shows a clear demand for semantic services. We believe that the current technological advancements, along with the maturity of the community, are prompt to investing efforts in harmonizing those technologies. In that light, we started TS4NFDI[21], a new project within NFDI for a central terminology service encompassing and harmonizing different solutions used by its consortia. TS4NFDI will provide a single access point to existing services like among others BiodivPortal. With a national and global network, we aim for a higher level of adoption and added value for the research community.

Acknowledgements

This work was supported by the German Research Foundation DFG under the grant agreement number 442032008 (NFDI4Biodiversity). The project is part of NFDI, the National Research Data Infrastructure Programme in Germany. We would like to thank our colleagues Bernhard Seeger and Martin Zurowietz for providing material related to the NFDI4Biodiversity RDC. We are grateful to all individuals who have helped to shape and support the structures and outcomes presented in this work.

References

[1] Algergawy A, Karam N, Laadhar A, et al (2022) Too big to match: a strategy around matching tasks for large taxonomies. In: Shvaiko P, Euzenat J, Jiménez-Ruiz E, et al (eds) Proceedings of the 17th International Workshop on Ontology Matching (OM 2022) co-located with the 21th International Semantic Web Conference (ISWC 2022), Hangzhou, China, held as a virtual conference, October 23, 2022, CEUR Workshop Proceedings, vol 3324. CEUR-WS.org, pp 67–72

[2] Alves C, Castro JA, Ribeiro C, et al (2018) Research data management in the field of ecology: an overview. In: Proc. International Conference on Dublin Core and Metadata Applications

[3] Buttigieg PL, Morrison N, Smith B, et al (2013) The environment ontology: contextualising biological and biomedical entities. Journal of Biomedical Semantics 4(1)

[20] https://fair-impact.eu/

[21] https://base4nfdi.de/projects/ts4nfdi

[4] Dieckmann MA, Beyvers S, Hochmuth J, et al (2023) The aruna object storage A distributed multi cloud object storage system for scientific data management. In: Sure-Vetter Y, Goble CA (eds) 1st Conference on Research Data Infrastructure - Connecting Communities, CoRDI 2023, Karlsruhe, Germany, September 12-14, 2023. TIB Open Publishing

[5] Diepenbroek M, Glöckner FO, Grobe P, et al (2014) Towards an integrated biodiversity and ecological research data management and archiving platform: the german federation for the curation of biological data (gfbio). In: Informatik 2014. Gesellschaft für Informatik e.V., Bonn, p 1711–1721

[6] Diepenbroek M, Schindler U, Huber R, et al (2017) Terminology supported archiving and publication of environmental science data in pangaea. Journal of Biotechnology 261:177–186. Bioinformatics Solutions for Big Data Analysis in Life Sciences presented by the German Network for Bioinformatics Infrastructure

[7] Diepenbroek M, Kostadinov I, Seeger B, et al (2023) Towards a research data commons in the german national research data infrastructure NFDI: vision, governance, architecture. In: Sure-Vetter Y, Goble CA (eds) 1st Conference on Research Data Infrastructure - Connecting Communities, CoRDI 2023, Karlsruhe, Germany, September 12-14, 2023. TIB Open Publishing

[8] Federhen S (2011) The NCBI taxonomy database. Nucleic Acids Res 40(Database issue):D136–43

[9] Gkoutos GV, Green ECJ, Mallon AM, et al (2005) Using ontologies to describe mouse phenotypes. Genome Biol 6(1)

[10] Glöckner FO, Diepenbroek M, Felden J, et al (2020) NFDI4BioDiversity - A Consortium for the National Research Data Infrastructure (NFDI)

[11] Hoehndorf R, Alshahrani M, Gkoutos GV, et al (2016) The flora phenotype ontology (FLOPO): tool for integrating morphological traits and phenotypes of vascular plants. Journal of Biomedical Semantics 7(1)

[12] Huber R, D'Onofrio C, Devaraju A, et al (2021) Integrating data and analysis technologies within leading environmental research infrastructures: Challenges and approaches. Ecol Informatics 61:101245

[13] J. Holetschek AGG. Dröge, Berendsohn WG (2012) The abcd of primary biodiversity data access. Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology 146(4):771–779

[14] Jonquet C, Poveda-Villalón M (2023) About versioning ontologies or any digital objects with clear semantics. In: Castro LJ, Dessì D, Dierkes J, et al (eds) 3rd Workshop on Metadata and Research (objects) Management for Linked Open Science - DaMaLOS 2023, co-located with ESWC 2023, Hersonissos, Greece, May 29, 2023

[15] Jonquet C, Graybeal J, Bouazzouni S, et al (2023) Ontology repositories and semantic artefact catalogues with the ontoportal technology. In: Payne TR, Presutti V, Qi G, et al (eds) The Semantic Web - ISWC 2023 - 22nd International Semantic Web Conference, Athens, Greece, November 6-10, 2023, Proceedings, Part II, Lecture Notes in Computer Science, vol 14266. Springer, pp 38–58

[16] Karam N, Müller-Birn C, Gleisberg M, et al (2016) A terminology service supporting semantic annotation, integration, discovery and analysis of interdisciplinary research data. Datenbank-Spektrum 16(3):195–205

[17] Karam N, Khiat A, Algergawy A, et al (2020) Matching biodiversity and ecology ontologies: challenges and evaluation results. Knowl Eng Rev 35:e9

[18] Kechagioglou X, Vaira L, Tomassino P, et al (2021) Ecoportal: An environment for FAIR semantic resources in the ecological domain. In: Sanfilippo EM, Kutz O, Troquard N, et al (eds) Proceedings of the Joint Ontology Workshops 2021 Episode VII: The

Bolzano Summer of Knowledge co-located with the 12th International Conference on Formal Ontology in Information Systems (FOIS 2021), and the 12th International Conference on Biomedical Ontologies (ICBO 2021), Bolzano, Italy, 2021, CEUR Workshop Proceedings, vol 2969. CEUR-WS.org

[19] König C, Weigelt P, Schrader J, et al (2019) Biodiversity data integration—the significance of data resolution and domain. PLOS Biology 17(3):1–16

[20] Löffler F, Pfaff C, Karam N, et al (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 2017), Vienna, Austria, October 22nd, 2017.

[21] Matentzoglu N, Balhoff JP, Bello SM, et al (2022) A simple standard for sharing ontological mappings (sssom). Database 2022

[22] Pour MAN, Algergawy A, Buche P, et al (2023) Results of the ontology alignment evaluation initiative 2023. In: Shvaiko P, Euzenat J, Jiménez-Ruiz E, et al (eds) Proceedings of the 18th International Workshop on Ontology Matching co-located with the 22nd International Semantic Web Conference (ISWC 2023), Athens, Greece, 2023, CEUR Workshop Proceedings, vol 3591. CEUR-WS.org