
The GFBio Terminology Service: enabling research data management beyond data heterogeneity

Naouel Karam¹, Robert Harald Lorenz², Claudia Müller-Birn³

1,2,3 Institute of Computer Science, Freie Universität Berlin

Abstract. A primary goal of a research infrastructure for data management should be to enable efficient data discovery and integration of heterogeneous data. The German Federation for Biological Data (GFBio) was guided by this goal. The basic component, that enables such interoperability and serves as a backbone for such a platform, is the GFBio Terminology Service (GFBio TS). This service acts as a semantic platform for accessing, developing and reasoning about terminological resources within the biological and environmental domains. A RESTful API gives access to these terminological resources in a uniform way, regardless of their degree of complexity and whether they are internally stored or externally accessed through web services. Additionally, a set of widgets with an intrinsic API connection are made available for easy integration in applications and web interfaces. Based on the requirements of GFBios partners, we describe the added value that is provided by the GFBio Terminology Service with practical scenarios as well as the challenges we still face. We conclude by describing our current activities and future developments.

Keywords. Research data infrastructure, Interoperability, Terminology repository, Semantic Web, RESTful API, Widgets.

Introduction

Research practice has become more data-intensive over the last few decades, and this development is visible across many research disciplines. However, the sharing of research data beyond disciplinary borders is still a challenge. Thus, a research infrastructure for data management should allow for an efficient data integration and therefore, the discovery of heterogeneous research data.

The German Federation for Biological Data (GFBio) pursues this goal. GFBio aims at providing a data management platform and data archiving solutions for data capture, annotation, indexing, searching and storage in the area of biological and environmental research. The GFBio Data Portal¹ integrates existing data infrastructures such as PANGAEA² into the GFBio Repository Network.

Data generated in biodiversity and ecology research are extremely heterogeneous and pertaining to different scientific disciplines using various methods and technologies. The situation is further complicated by different understandings of employed terms within different scientific do-

1 <http://www.gfbio.org>

2 www.pangaea.de

mains. Developing interoperability and harmonizing data by using standards and terminological resources are crucial for data mobilization, integration, and discovery in the GFBio context.

The core component that enables this interoperability and serves as a backbone for the GFBio infrastructure is called the GFBio Terminology Service³ (GFBio TS) (Karam et al. 2016). The GFBio Terminology Service acts as a semantic platform for accessing, developing, and reasoning over terminological resources. The GFBio TS focuses on integrating and giving access to terminologies developed by project partners as well as external terminologies defined and maintained by related communities. These terminologies can range from simple term lists to complex ontologies. Based on the requirements of the GFBio community, the Terminology Service provides access to over 20 terminologies so far, of which GFBios partners have contributed 10 terminologies. A well-defined RESTful API gives access to all terminologies in a uniform way regardless of their degree of complexity and whether they are internally stored or externally accessed through web services. The services provided by the GFBio TS can also be integrated easily within existing web applications with the help of widgets, which are small applications with limited functionality. We developed two exemplary widget prototypes so far: a term visualization and a search widget.

We will explain the advantage of using semantic technologies for data management and highlight the utility of the Terminology Service by practical use cases of semantically enhanced components. More specifically, we will differentiate between four main usage scenarios developed so far: *Explore*, *Access*, *Download* and *Contribute*. In the *Explore* scenario, researchers can reuse ontologies that are interesting for their research. In the *Access* scenario developers can use information in ontologies programmatically to provide semantically enriched applications and web services. In the *Download* scenario, information from the ontologies can be retrieved and stored to a local information system. In the *Contribute* scenario, we consider that scientists can store their terminologies in the TS to access all provided services automatically. Finally, we discuss existing challenges in this field that are often in the social-technical context.

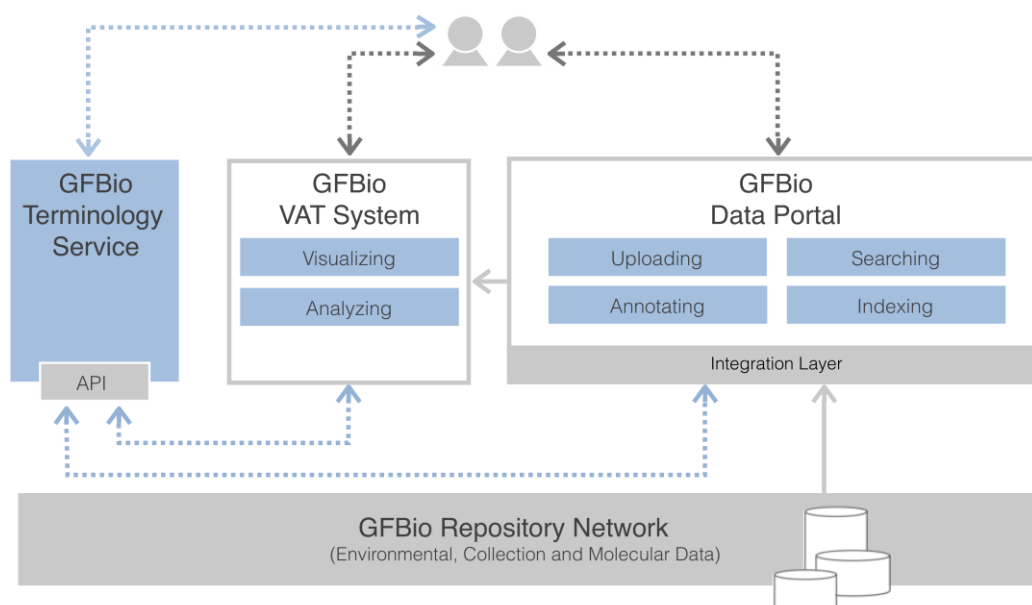


Figure 1. The GFBio components

A common infrastructure for biological data

GFBio (Diepenbroek et al. 2014) is developing an infrastructure to enable biological and environmental scientists to share and discover data more efficiently. It aims at providing data management and data archiving solutions for data capture, annotation, indexing, searching and storage. These solutions range from tailored Excel spreadsheets to virtual research environments, such as the Diversity Workbench (Triebel et al. 1999), the Bexis system (Gerlach et al. 2015) or the EDIT Platform for Cybertaxonomy (Ciardelli et al. 2009). Figure 1 presents an overview of the research infrastructure of GFBio, consisting of four main components.

The GFBio Data Portal integrates existing data infrastructures into the GFBio Repository Network (bottom in Fig. 1). The latter comprises amongst others molecular data (EMBL-EBI⁴), environmental data (PANGAEA⁵), as well as natural history and culture collection data (e.g. MfN⁶, DSMZ⁷ and SNSB⁸).

The data provided by portal users are indexed and semantically enriched, thereby providing the data with meaning. Analysis and visualization tools allow researchers to better understand the data, for example, by using the GFBio VAT System (Visualization, Analysis & Transformation system) (Authmann et al. 2015). The possibility to enrich data with semantic information is provided by a fourth component - the GFBio Terminology Service. The semantic meaning is enabled by the provision and interlinking of ontologies and taxonomies.

There are existing systems providing a comparable terminology service. These systems can be either full platforms for terminology management (Noy et al. 2009; Côté et al. 1006; Suominen et al. 2014; Hoehndorf et al. 2015; Xiang et al. 2011) or frameworks for accessing terminologies (Adamusiak et al. 2011; Viljanen et al. 2012). We defined a set of requirements related to our project needs and analyzed to what extent existing systems meet those requirements (Karam et al. 2016). One requirement was to be able to integrate well established taxonomies like the World Register of Marine Species (WORMS⁹) or the Catalogue of Life (COL¹⁰). Those taxonomies are widely used in the domain for annotating species, for example, and they are a source of valuable hierarchical information. None of the existing systems integrate this type of terminologies. Additional requirements relate to our project's philosophy, to provide tools and inference mechanisms specifically tailored to the requirements of GFBio's partners. These derived insights motivated our decision to set up our own system – the GFBio Terminology Service – that is introduced in the next section.

4 The European Bioinformatics Institute (www.ebi.ac.uk)

5 www.pangaea.de

6 Naturkundemuseum (www.naturkundemuseum.berlin)

7 Leibniz-Institut - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (www.dsmz.de)

8 Staatliche Naturwissenschaftliche Sammlungen Bayerns (www.snsb.mwn.de)

9 World Register of Marine Species (www.marinespecies.org)

10 Catalogue of Life (www.catalogueoflife.org)

The Terminology Service

We describe in this section the main building blocks of the GFBio Terminology Service. First, we introduce the basic concepts and define the meaning of *terminology* in the context of the GFBio project, then, we present the general architecture of the GFBio TS.

Basic Concepts

The term *terminology* refers to any terminological resource, this can be a formal ontology, a taxonomy, or any useful source of Semantic Web compliant collections of terms (e.g. locations available via a geographical database like Geonames¹¹). It encompasses several meanings ranging from simple lists of terms to semantically rich ontologies. Unfortunately, there are currently no commonly accepted definitions of the different terminology types (in the biological domain), which leaves room for variation causing them to be used interchangeably depending on the context.

We introduce our concept of agreed terminology formality levels, with differing levels of specifications going from the most informal to the most formal level as described in Figure 2. The different levels are illustrated by the term *water* (http://purl.obolibrary.org/obo/CHEBI_15377) that is extracted from the CHEBI ontology¹² and depicted in Figure 3.

GFBio defines five different types or formality levels in terminologies. The less formal level contains a *Controlled Vocabulary*. It is the simplest type of terminology and consists of a finite list of terms. These labels have no definitions or hierarchical ordering. Based on the example, only the label *water* is part of the terminology.

The next formality level is *Glossary*. It is a list of term labels that additionally includes an informal definition of their meaning in natural language (i.e. human readable language). Since information expressed in natural language is typically not unambiguous, these specifications are not yet adequate for further processing by computer agents. In a glossary, the definition of the term *water* is partnered by its label.

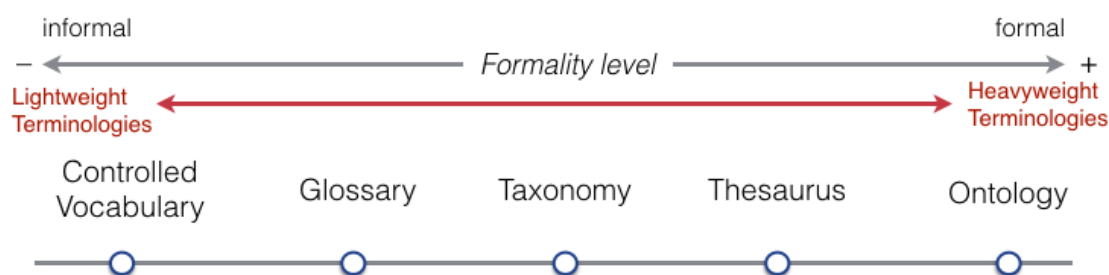


Figure 2. GFBio agreed terminology formality levels

In a *Taxonomy*, a term is a compound of a label, a definition and hierarchical information, e.g., by is-a relationships, thus providing additional semantics in the relations between the terms which can be interpreted by computer agents. The hierarchical structure depicted in Figure 3 would be part of a taxonomy describing the term *water*.

11 www.geonames.org

12 www.ebi.ac.uk/chebi/

A *Thesaurus* is a controlled vocabulary connected via relations between the terms expressing hierarchies (e.g., *narrower/broader term*), associations (e.g., *related term*), or synonym relationships. In the example, a thesaurus contains the information about the synonym *oxidane* of the term *water*.

The most formal terminology is an *Ontology*. A term consists of all the information provided at the lower levels augmented with complex relationships, allowing an unambiguous interpretation of terms and relationships according to logic-based rules. In our example, an ontology would contain the whole spectrum of relations we already considered in the other levels and additional complex or user defined relations like *has_role* and *is_conjugate_base_of*.

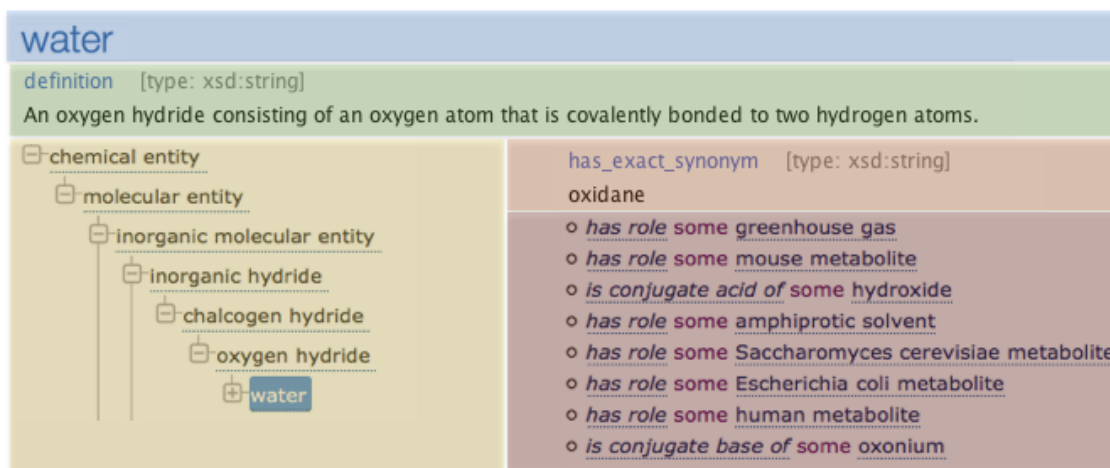


Figure 3. Excerpt of the definition of the term *water* of the CHEBI ontology

The Terminology Service Architecture

The general architecture of the Terminology Service is shown in Figure 4. In March 2017, the Terminology Service gives access to over 20 terminologies that have been requested by the GFBio partners so far. Those terminologies are either internally stored in a Semantic Web repository or remotely accessed via their web services. Internal terminologies are stored in a local RDF¹³ store in a Semantic Web compliant format such as OWL¹⁴ or SKOS¹⁵. Internal terminologies can be accessed directly via a Linked Data interface and a SPARQL¹⁶ endpoint. The included terminologies are well established ones like the CHEBI ontology¹⁷, for example, or ontologies provided by the GFBio community like the KINGDOM¹⁸ ontology, describing a GFBio agreed list of species kingdoms. The complete list and actual status of included terminologies can be found in our technical report (Karam et al. 2017).

13 www.w3.org/RDF

14 www.w3.org/OWL

15 www.w3.org/2004/02/skos/

16 <https://terminologies.gfbio.org/sparql>

17 www.ebi.ac.uk/chebi/

18 <https://terminologies.gfbio.org/describe/?url=http://terminologies.gfbio.org/terms/KINGDOM>

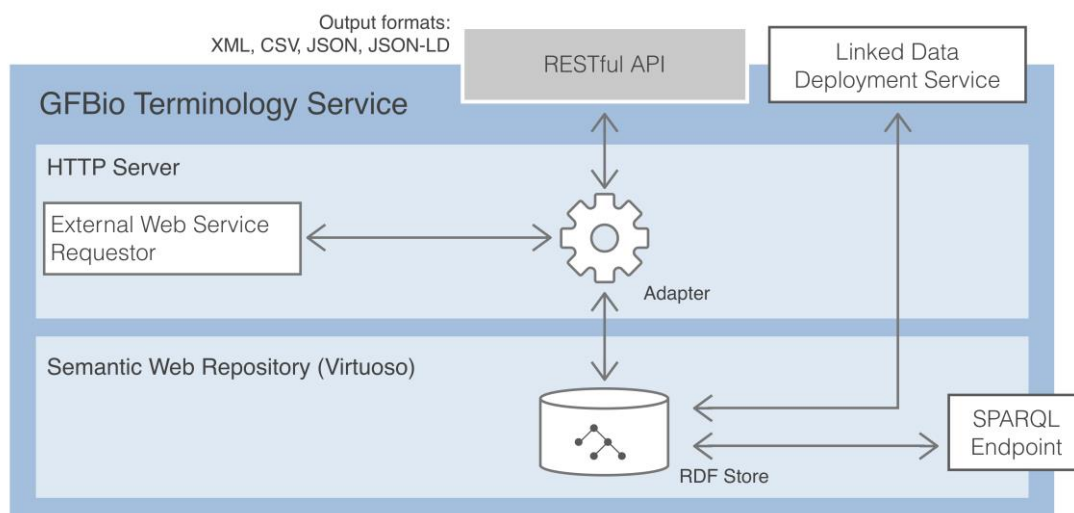


Figure 4. The GFBio Terminology Service architecture

The Terminology Service software is being developed using Java based on the Jena¹⁹ Semantic Web framework. We implemented an external web service requestor for obtaining seven external taxonomies (such as the COL - Catalogue of Life). A key component of the TS is the adapter component (cf. the gear wheel in Figure 4) that enables the schema mapping of both internal and external terminological resources into a common output format. We defined a common schema for the Terminology Service output. A mapping to this schema is required for every underlying terminology or connected external service in order to achieve a harmonized API output. For instance, the COL attribute *name* is mapped to the GFBio TS attribute *label*. Thus, all terms and terminologies can be accessed via a common interface (the RESTful API), regardless of whether they are hosted internally or externally. The service output is delivered in four formats: JSON, XML, CSV, and JSON-LD. This interface allows developers who are not familiar with semantic technologies or Linked Data to easily access the provided terminologies efficiently.

Accessing the Terminology Service

The GFBio Terminology Service can be accessed either through a common interface - the RESTful API²⁰ - or using widgets we provide; these are small web applications with limited functionality which allow for user interactions. We describe in the following both ways to access the GFBio TS.

The Terminology Service API

The RESTful API of the TS can be used programmatically by connecting the service to other web services such as the GFBio Data Portal, the VAT (cf. Figure 1) or other applications. At the mo-

¹⁹ <https://jena.apache.org/>

²⁰ Application Programming Interface

ment the API provides 14 endpoints that are organised into terminology-specific, term-specific, search, and hierarchy-oriented endpoints. Details about the call's signatures, the parameters and examples for using the service can be found in the API documentation section on our website (terminologies.gfbio.org). In the following, we describe each category briefly, a tabular description for each endpoint can be found in our technical report (Karam et al. 2017).

Terminology-specific endpoints

The four terminology-specific endpoints provide information on terminologies like the list of available terminologies and their metadata, such as the name, description and creation date.

Term-specific endpoints

Term-specific endpoints relate to particular terms from the terminologies. One can list all terms of a specific terminology, query the information about a term or get the list of its synonyms.

Search endpoints

Two search endpoints are provided, the first one returns all terms corresponding to a query string, the second is implemented for suggesting terms while users are typing.

Hierarchy-oriented endpoints

Hierarchy-oriented endpoints return information relative to the position of a term in the hierarchical structure of the terminology. Broaders and narrowers terms of a given term can be returned as well as the complete hierarchical path up to the top of the hierarchy.

The Terminology Service Widgets

The Terminology Service provides widgets – that are components, “chunks of web page” or small applications – intended to be used within web pages. The widgets deliver a restricted functionality, often for just one purpose, like displaying data or providing an interface. Typically, a widget contains a mixture of HTML, CSS and JavaScript where the complexity is ideally hidden so as to make it as easy as possible for developers to integrate the widgets in their application or website. All of our widgets use the Terminology Service API and thus, users can quickly expand their local service with all the functionalities provided by the GFBio TS API. Our goal is to provide reusable and easy to use widgets to be integrated and reused easily with none or little knowledge in web development. Furthermore, the widgets are licensed under an open source licence and will be published openly on Github soon. At the moment, we prototypically implemented two widgets: a term visualisation and a search widget. In the following, we take the latter as an example, to show the methodological approach for developing widgets.

The search widget allows users to search for terms from terminologies to determine their usefulness for their work, e.g. for annotating research data in the GFBio Data Portal. Before developing this widget, we examined 13 services which provide search functionalities in the same or related fields as ours. The majority (6) of the examined services allowing to look for classes (terms) *in* particular ontologies or vocabularies (Cropontology²¹, Finto²² (Suominen et al. 2014), Ontobee²³ (Xiang et al. 2011), Aber-Owl²⁴ (Hoehndorf et al. 2015), Bioportal²⁵ (Noy et al. 2009), OLS²⁶ (Côté et al. 2006)). The latter three are capable of searching *for* ontologies as well. Three services (Biosharing²⁷ (McQuilton et al. 2016), VEST²⁸ (Vest / AgroPortal map 2017), ANDS²⁹ (Australian National Data Service 2017)) looking for vocabularies, ontologies, policies or standards only and four (Datacite³⁰ (Datacite 2017), Dryad³¹ (Dryad Digital Repostory 2017), F1000research³² (F1000research 2017), Vertnet³³ (Vertnet 2017)) are for searching scientific papers and data resources. The appearance of the search interface differs a lot. From very simple interfaces to advanced ones with many search options and filter functionalities. We examined design criteria like the overall size of the widget, the position and layout of the submit button, the placeholder text of the search bar, the availability and presentation of advanced search functionalities and help sections. The main considerations are described in detail in our technical report (Karam et al. 2017), they resulted in the prototypical design depicted in Figure 5.

The development process included the investigation of a widget scaffold where the objective was twofold: (1) the development process for further widgets should be simplified and standardized, and (2) the process for developers to integrate the GFBio TS widgets into their websites should be supported. We then investigated three services (Google³⁴, Twitter³⁵ and ANDS (Australian National Data Service 2017)) that are providing customized widgets. With some kind of guidance users are able to click through options on the website to receive customized HTML code and references to JavaScript and style files to be embedded on their own website. As customisation is planned but not implemented yet, our goal is to deliver one JavaScript and one CSS file to be integrated in the users HTML via the corresponding HTML markups. Because our widgets will deliver a broad spectrum of functionality the scaffold consists next to the way how developers integrating it, of the module design pattern, used libraries, a shared layout file and partly shared functions.

21 www.cropontology.org

22 www.finto.fi

23 www.ontobee.org

24 www.aber-owl.net

25 <http://bioportal.bioontology.org>

26 www.ebi.ac.uk/ols

27 <http://biosharing.org>

28 <http://vest.agrisemantics.org/vocabularies>

29 <http://vocabs.ands.org.au>

30 www.datacite.org

31 www.datadryad.org

32 <http://f1000research.com>

33 <http://portal.vertnet.org>

34 <https://developers.google.com>

35 <https://dev.twitter.com>

The search service includes all labels, synonyms, common names and abbreviations when provided by terminologies.

Search

Try the following examples: *pentafluoridoarsenic, bacteria, AsF5*

Narrow search results to...

- Exact search term
- First matching terminology
- Internal terminologies only

Particular terminologies

clear selection

Figure 5. Screenshot of the GFBio TS search widget prototype

Using the Terminology Service within GFBio

Currently, the GFBio community uses the Terminology Service within four main scenarios. Each scenario has been defined and developed in cooperation with GFBios partners. Each partner provides discipline and context specific requirements to the GFBio TS. The development of these use cases is an ongoing process and further use cases will be provided in the near future.

In the *Browse* scenario, users (i.e. researchers) can peruse terminologies that are interesting for their research. For this, the visualization widget provides term details and shows a term's position within a tree structure, if the terminology is a taxonomy or in a graph structure, if the terminology is an ontology. In the GFBio Data Portal the visualization can be used in the research data submission process. When annotating the data in the submission process, the user can easily browse term details and explore existing term relations by type to identify those terms that describe their data best.

In the *Access* scenario developers can use information in terminologies programmatically to provide semantically enriched web services based on the GFBio TS. In the GFBio Data Portal, the TS allowed for developing a semantic search service for research data. Based on query expansion, the original search term is extended by related terms from different terminologies in order to provide a more comprehensive overview of existing research data.

In the *Consume* scenario, information from terminologies of the GFBio TS can be retrieved and stored to a local information system. In the GFBio context, this is needed for data management within small and medium scale projects that are carried out by virtual research environments such as BExIS (Gerlach et al. 2015) and Diversity Workbench (Triebel et al. 1999). In these contexts, the provided metadata from the terminologies of the TS can be pre-processed to support the data annotation process locally.

In the *Contribute* scenario we consider that researchers or data curators can store their individual terminologies in the GFBio TS. Instead of developing their own terminology management

system, this will allow them to access all services provided by the TS easily. For example, in the GFBio context, partners have already contributed ten terminologies. These terminologies are either internally stored like the KINGDOM ontology³⁶ or connected as external web services like the DTN Taxon Lists Services³⁷ or the Prokaryotic Nomenclature Up-to-Date³⁸ and interna. In GFBio, the mobilization of community-relevant terminologies is supported by an internal process. The terminology owner can register the terminology in the internal wiki and in collaboration with the terminology curator the needed metadata are provided. If the metadata are complete, a terminology is manually integrated into the TS.

Current activities and next steps

We introduced the GFBio TS that extends the GFBio infrastructure with semantic capabilities. This extension enables researchers to share their data despite their heterogeneous nature. After presenting the project context and the basic concepts, we described the general architecture of the Terminology Service and the way to access and integrate it using its public interface or via a set of downloadable widgets.

We described concrete use cases that support researchers at different levels in their research practice, for example, when searching for datasets or when using up-to-date terminologies in their virtual research environments.

At the moment, a high level application ontology, the GFBio ontology is being developed. It will enable interoperability between the various terminologies available by defining higher level links between them. Moreover, this ontology will serve mainly as a basis for annotations and automated faceted search.

We are working on the integration of the semantic annotation tool neonion (Müller-Birn et al. 2017) within the GFBio context. The aim is to allow scientists to annotate information in scientific texts with terminologies coming from the GFBio TS, and thus, research results and research data can be more closely connected.

The interoperability issue is due to different understandings of terms within different scientific domains or to the use of different labels to refer to the same term. This issue can be solved by annotating data with terms from the Terminology Service. Data can still be annotated using equivalent terms coming from different terminologies. In order to ensure interoperability the underlying terminologies should be interlinked. We are developing a semi-automated mapping service and interface based on a combination of matching algorithms.

The GFBio TS is continuously updated to meet partners needs. A set of tools is being developed to support terminologies selection based on query and text analysis as well as tools for transforming terminologies from text and tabular forms into a Semantic Web compliant format.

36 <https://terminologies.gfbio.org/api/terminologies/KINGDOM/>

37 http://www.diversitymobile.net/wiki/DTN_Taxon_Lists_Services

38 <https://bacdive.dsmz.de/api/>

References

- Noy, Natalya Fridman, Nigam H. Shah, Patricia L. Whetzel, Benjamin Dai, Michael Dorf, Nicholas Griffith, Clement Jonquet, et al. 2009. “BioPortal: ontologies and integrated data resources at the click of a mouse.” *Nucleic Acids Research* 37, Web-Server-Issue: 170–173.
- Côté, Richard G., Philip Jones, Rolf Apweiler, and Henning Hermjakob. 2006. “The Ontology Lookup Service: a lightweight cross-platform tool for controlled vocabulary queries.” *BMC Bioinformatics* 7 (1): 1–7.
- Suominen, Osmo, Sini Pessala, Jouni Tuominen, Mikko Lappalainen, Susanna Nykyri, Henri Ylikotila, Matias Frosterus, and Eero Hyvönen. 2014. “Deploying National Ontology Services: From ONKI to Finto.” In *Proceedings of the Industry Track at the Intl. Semantic Web Conference 2014*. Riva del Garda, Italy: CEUR Workshop Proceedings, October.
- Hoehndorf, Robert, Luke Slater, Paul N. Schofield, and Georgios V. Gkoutos. 2015. “Aber-OWL: a framework for ontology-based data access in biology.” *BMC Bioinformatics* 16 (1): 1–9.
- Xiang, Zuoshuang, Chris Mungall, Alan Ruttenberg, and Yongqun He. 2011. “Ontobee: A Linked Data Server and Browser for Ontology Terms.” In *Proceedings of the 2nd Intl. Conference on Biomedical Ontology*, Buffalo, NY, USA, July 26-30.
- Adamusiak, Tomasz, Tony Burdett, Natalja Kurbatova, K. Joeri van der Velde, Niran Abeygunawardena, Despoina Antonakaki, Misha Kapushesky, Helen Parkinson, and Morris A. Swertz. 2011. “OntoCAT – simple ontology search and integration in Java, R and REST/JavaScript.” *BMC Bioinformatics* 12 (1): 1–12.
- Viljanen, Kim, Jouni Tuominen, Eetu Mäkelä, and Eero Hyvönen. 2012. “Normalized Access to Ontology Repositories.” In *Proceedings of the Sixth International Conference on Semantic Computing (IEEE ICSC 2012)*. Palermo, Italy: IEEE Press, September.
- Karam, Naouel, Claudia Müller-Birn, Maren Gleisberg, David Fichtmüller, Robert Tolksdorf, and Anton Güntsch. 2016. “A Terminology Service Supporting Semantic Annotation, Integration, Discovery and Analysis of Interdisciplinary Research Data.” *Datenbank-Spektrum* 16 (3): 195–205.
- Diepenbroek, Michael, Frank Oliver Glöckner, Peter Grobe, Anton Güntsch, Robert Huber, Birgitta König-Ries, Ivaylo Kostadinov, 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 *44. Jahrestagung der Gesellschaft für Informatik*, Stuttgart, Germany. Ausgabe 232. LNI. GI, isbn: 978-3-88579-626-8.
- Triebel, Dagmar, Gregor Hagedorn, Stefan Jablonski, and Gerhard Rambold (eds.). 1999. “Diversity Workbench: A virtual research environment for building and accessing biodiversity and environmental data.” Online available <http://www.diversityworkbench.net>.

- Gerlach, Roman, David Blaa, Javad Chamanara, Martin Hohmuth, Nafiseh Navabpour, Sven Thiel, and Birgitta König-Ries. 2015. “BEXIS 2: A platform for managing heterogeneous biodiversity data and projects.” In *TDWG Annual Conference*.
- Ciardelli, Pepé, Patricia Kelbert, Andreas Kohlbecker, Niels Hoffmann, Anton Güntsch, and Walter G. Berendsohn. 2009. “The EDIT Cyberplatform for Taxonomy and the Taxonomic Workflow: Selected Components.” In *39. Jahrestagung der Gesellschaft für Informatik e.V. (GI)*, Lübeck, Germany, 625–638.
- Authmann, Christian, Christian Beilschmidt, Johannes Drönner, Michael Mattig, and Bernhard Seeger. 2015. “VAT: A System for Visualizing, Analyzing and Transforming Spatial Data in Science.” *Datenbank-Spektrum* 15 (3): 175–184.
- “Crop Ontology curation and annotation tool – 2011 Generation Challenge Programme, Bioversity International as project implementing agency.” Accessed: 2017-01-06.
- McQuilton, Peter, Alejandra Gonzalez-Beltran, Philippe Rocca-Serra, Milo Thurston, Allyson Lister, Eamonn Maguire, and Susanna-Assunta Sansone. 2016. *BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences*. Database 2016 (2016): baw075.
- “Australian National Data Service website.” Online available www.ands.org.au. Accessed: 2017-01-06.
- “VEST / AgroPortal - Map Of Standards website.” Online available <http://vest.agrisemantics.org/vocabularies>. Accessed: 2017-01-06.
- “Datacite website.” Online available www.datacite.org. Accessed: 2017-01-06.
- “Dryad Digital Repository website.” Online available <http://www.datadryad.org>. Accessed: 2017-01-06.
- “F1000Research website.” Online available <https://f1000research.com>. Accessed: 2017-01-06.
- “VERTNET: Distributed Databases with backbone website.” Online available <http://portal.vertnet.org>. Accessed: 2017-01-06.
- Morville, Peter, and Jeffery Callender. 2010. *Search Patterns: Design for Discovery*. I–X, 1–180. O’Reilly, isbn: 978-0-596-80227-1.
- Turbek, Steve. 2008. “Advancing advanced search.” Online available <http://boxesandarrows.com/advancing-advanced-search>. Accessed: 2017-01-13.
- Müller-Birn, Claudia, Tina Klüwer, André Breitenfeld, Alexa Schlegel, and Lukas Benedix. 2015. “neonion: Combining Human and Machine Intelligence.” In *18th ACM Conference on Com-*

puter Supported Cooperative Work & Social Computing, CSCW 2015, Vancouver, BC, Canada, March 14-18, 2015, Companion Volume, 223–226.

Karam, Naouel, Robert Harald Lorenz, and Claudia Müller-Birn. 2017. “The GFBio Terminology Service: Enabling a research data management beyond data heterogeneity”. *Technical Report Ser. B TR-B-17-01*. Freie Universität Berlin, Institut für Informatik, March.