A Terminology Service for Inter-disciplinary Research Data

The German Federation for Biological Data (GFBio) is a sustainable, service oriented, national data infrastructure facilitating data sharing for biological and environmental research. The infrastructure supports scientists in all aspects of research data management covering the complete life cycle of data. The provided services include the creation of Data Management Plans (DMP), provision of tools for data acquisition, data publication in long-term archives as well as platforms for data discovery and sharing.

In this project, we are responsible for the development and maintenance of the Terminology Service (TS), a key component of the GFBio infrastructure. The TS enables a semantically enriched data management and archiving solution for GFBio and constitutes the terminological and ontological basis for data annotation, discovery and curation. It offers universal access to heterogeneous terminological resources, ranging from simple vocabularies to complex ontologies, in a uniform and transparent manner. We provide researchers with a set of tools and services for sharing and reusing terminologies useful for their research. They can enrich their own applications by accessing the public API or directly using the offered widgets to enable semantic annotation and terminology based access capabilities.

General Data Infrastructure

A core component of the GFBio infrastructure is the Data Portal that integrates existing data infrastructures into the GFBio Repository Network. The data provided are indexed and semantically enriched and thus, can be analyzed and visualized by using the GFBio VAT System. All these components are based on the assumption that the "meaning" of the data is provided by a fourth component — the GFBio Terminology Service.

Technical Architecture

The GFBio Terminology Service (TS) provides a web service’s interface that serves as a generic access point to heterogeneous terminological resources. The TS offers universal access to the various types of terminologies in a uniform and transparent manner. These terminologies are either internally hosted or accessed via their remote web services. Internal terminologies are stored in a local Semantic Web repository (Virtuoso).

The key component of the TS is the adapter that enables the mapping of both internal terminologies and external terminological resources into a common output format. The service endpoints (RESTful API) are grouped into four categories: search services, terminology services, term services and hierarchy-oriented services.

Terminology Overview

The term “terminology” refers to any terminological resource. We distinguish between five different types or formality levels of terminologies with differing levels of specifications — it ranges from simple lists of terms (controlled vocabularies) to semantically rich ontologies.