GFBio - building a sustainable and service-oriented biodiversity data infrastructure for Germany

Marco Schmidt[*1], Katja Schiffers[†1], Aidin Niamir[†1], Claus Weiland[1], Eva-Maria Gerstner[1], Thomas Hickler[†1,2], and Katrin Böhning-Gaese[†1,2]

[1]Senckenberg Biodiversity and Climate Research Centre (BiK-F) – Senckenberganlage 25 60325 Frankfurt (Main), Germany
[2]Goethe-University Frankfurt am Main – Grüneburgplatz 1 60323 Frankfurt am Main, Germany

Abstract

In the context of biological research, data is being produced in larger quantities each year. The need for such data also increases, as research approaches are increasingly data-driven and relying on 'big data'. However, data are often scattered, hard to integrate, and their long-term preservation is not secured. To overcome this mismatch, the German Federation for Biological data (GFBio) is building a sustainable and service-oriented national data infrastructure to facilitate data sharing and stimulate data intensive science. GFBio has been in an intensive design and development phase since 2014, resulting in deployment of a stable portal framework, considerable progress towards mobilizing and harmonizing data sources, and prototype implementations of advanced applications and services. GFBio includes representatives from key institutions, including major data providers in the biological and environmental sciences community. Established and well-proven data archiving infrastructures like Pangaea for environmental data and Germany's major natural history collection data repositories will be improved and integrated in a new infrastructure with a common technological and organizational outline.

The approach of GFBio is to support sustainable research data management throughout the 'data life cycle' on an integrated platform: from data management plans in research proposals, via data collection, quality assurance, data deposition and publication to secure long-term archiving and semantic integration. GFBio has developed a series of web-based tools for retrieval, aggregation and visualization of organismic and environmental data (VAT), supporting scientists addressing fundamental questions of biodiversity research and making large-scale computing resources for data-intensive processing available.

Senckenberg Biodiversity and Climate Research Centre established an office to align GFBio internationally, fostering information and data exchange, networking and cooperations, and to promote GFBio beyond national frameworks. Here, we present the conceptual framework of GFBio, along with some of VAT tools and functionalities, with particular focus on needed data infrastructures, data products, and data support for the development of biodiversity and ecosystem services, and to promote data management as funded part of science projects.

Keywords: Biodiversity, Biological, Ecosystem Services, Informaiton, Knowledge, Data Management, Germany

[*]Corresponding author: marco.schmidt@senckenberg.de
[†]Speaker