How to organize, process and archive collection and occurrence data

using GFBio services

provided by Germany’s major natural history and culture collection data repositories

Dagmar Triebel, Peter Grobe, Anton Güntsch, Gregor Hagedorn, Joachim Holstein, Carola Söhngen, Claus Weiland, Tanja Weibulat

Funded by
Challenges for biologists: How to organise data?

Biologists traditionally produce data in spreadsheets or simple-structured databases designed by themselves.

This concerns all disciplines and main data domains, e.g.:

- collection, observation
- ecological (time series)
- molecular data
- taxon-related data
However: Organisation means...

- keep complex data coherent
- manage heterogeneous data sets and data objects
- exchange data with researchers from other working groups
- make data ready to be curated by other persons by offering appropriate metadata
- make data ready to be archived and published
Use powerful management tools ....

- which allow flexible high-level data quality control for complex data
- allow flexible export of complex data for analysis and visualisation tools
- allow data exchange following standards
- which provide access to (web) services
The result is….

Your data will be

- interoperable with international platforms
- could be published via services of publishers
- could be used for analysis and visualisation pipelines
The result is ...

Your biodiversity, ecology and molecular data will be compliant to GFBio
Objective of the DFG-funded initiative
Sustainable, service oriented, national data infrastructure facilitating data sharing for biological and environmental research
GFBio consortium: 19 partners
What is new about GFBio?

- Common long-term data archiving policies & workflows based on know-how of trusted data centers
- Harmonization of structures and workflows across archives
- Interconnection of data centers in Germany for common data provision and publication mechanisms
- Combination of environmental-, collection- and genome-data
Architecture of the services

Data producer and user

Visualisation/Analysis

Environment/Collection/Genome Data Archives

Domain Databases

Project Databases

External Datasources

Search

Data Submission

Help Desk

Terminology Services

Curation

Interfaces

Storage

Portal

Search

Data Submission

Help Desk

Terminology Services

Curation

Interfaces

Storage

Visualisation/Analysis

Environment/Collection/Genome Data Archives

Domain Databases

Project Databases

External Datasources

Search

Data Submission

Help Desk

Terminology Services

Curation

Interfaces

Storage
Direct benefit for data producers?

- Training courses and education modules
- Support to fulfill published DFG recommendations on data handling and long-term archiving
- **Support** with user help desk for workbenches (BExIS and DWB)
- **Support** with user help desk for archiving and publishing biodiversity data
Many GFBio services rely on data archiving and provision guided by the **federation** of the seven collection data centers:

- They build (together with PANGAEA) the archive backbone.
- They guarantee the long-term access on structured biological research data.
- They offer - as a federation - IT services for the common GFBio system architecture.
Services provided by collection data repositories

- Managing, archiving and providing collection-related data of various data domains
- Training of staff at the data centers „data curators“ to ensure long-term management and archiving of research data
- Documentation and publication of existing IT infrastructure and services
- Development of best practices for data sharing and archiving
Services follow ISO standard for OAIS archives

Functional Entities of GFBio Collection Data Centers/Archives with assignment of WP5 tasks
(The functional entities and major terms are those of the reference model for an Open Archival Information System, OAIS environment)
Services follow GFBio agreed schemes and standards

ABC ... DwC

various GFBio agreed/published schemes

EML ... SDD

Functional Entities of GFBio Collection Data Centers/Archives with assignment of WP5 tasks
(The functional entities and major terms are those of the reference model for an Open Archival Information System, OAIS environment)

SIP (Submission Information Package)
AIP (Archival Information Package)
DIP (Dissemination Information Package)
Back to data producers: How to archive and publish your data?

again: **organise your data** already during the research project

and **contact** GFBio data repositories/ data centers
Thank you for your attention!

http://www.gfbio.org/