Development of a Database for Climate Research

By Reiner Krause, Nina Buchmann, Galina Churkina, Annette Freibauer and other members

of the Max-Planck-Institute for Biogeochemistry, Jena, Germany

Introduction

A database for various kinds of research data is currently being developed at the Max-Planck-Institute for Biogeochemistry in Jena, Germany. The database is particularly tailored towards CarboEurope data, e.g., continuous flux measurements, measurements of plant, animal, microbial or soil parameters, satellite images, model runs, etc.

Goals and Approaches

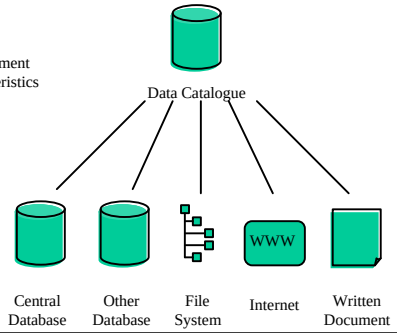
- to make data and samples available to a wide group of users
- to allow an integrated view on data from different scientific disciplines
- to archive data and samples
- to support planning and coordinating projects
- to support other forms of collaboration

- through internet based and easy to handle user interfaces and retrieval tools
- through the possibility to visualize and download free combinations of variables
- with documentation of data and links between data sets, samples and sites
- through information about projects and an easy way of data interchange
- e.g. as a platform for discussions on methods and interpretation of data

Metadata and Data

Metadata
- Name
- Abstract
- Variables and units
- Type of samples and environment
- Spatial and temporal characteristics
- Quality
- Process status
- Contributors
- Access information
- Publications

Data
- Location
- Date and time
- Sample
- Variable
- Unit
- Value



Data sets are worthless without a documentation of their meaning. This documentation - also referred to as metadata (data about data) - gives detailed information on study purpose, data origin, parameters measured, sample preparation, data quality, etc.

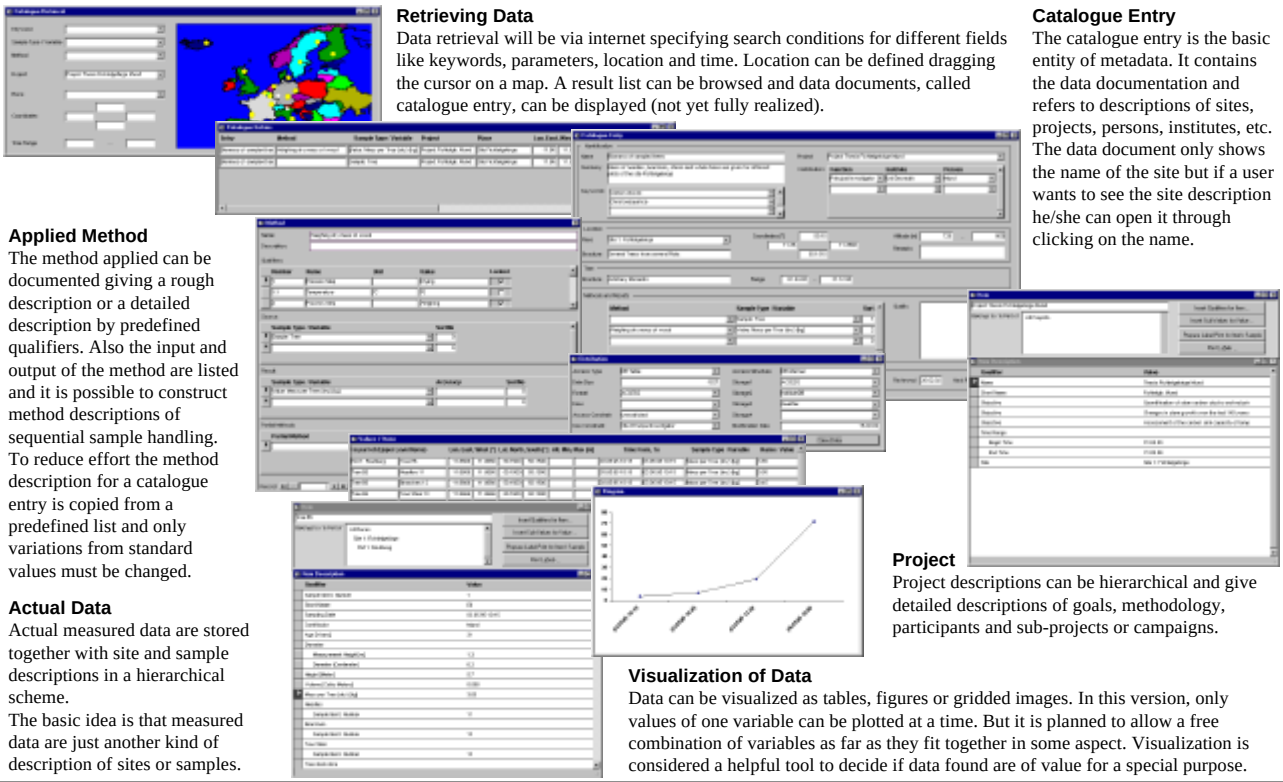
The documentation enables users to retrieve data, to understand why they were obtained, how they were created and whom to give credit.

While data sets are typically connected to their documentation, the distinction between metadata and data allows to store actual data sets in different places (e.g. different servers) whereas the data documentation can be stored in a central database. It also allows free access to metadata but restricted access to actual data, e.g. for a defined group of investigators before results are published.

Concepts to Facilitate Database Use

- Preferentially import of data (e.g. continuous measurements, data and measurement protocols in computer readable form from instruments).
- As little typing for data documentation as possible through the following approaches:
  - Descriptions of entities (e.g. sites, projects, method descriptions, institutes, persons) are typed in only once and linked to data documents.
  - Coordinates and time ranges are preset automatically by linking values of the data sets to the documentation.
- Support of producing publications through the possibility to export descriptions, figures and tables from the database.
- Avoid generating different versions of data manually through exporting data in formats suitable for other databases like FluxNet or CARBIS.

Demonstration Database



Retrieving Data

Data retrieval will be via internet specifying search conditions for different fields like keywords, parameters, location and time. Location can be defined dragging the cursor on a map. A result list can be browsed and data documents, called catalogue entry, can be displayed (not yet fully realized).

Catalogue Entry

The catalogue entry is the basic entity of metadata. It contains the data documentation and refers to descriptions of sites, projects, persons, institutes, etc. The data document only shows the name of the site but if a user wants to see the site description he/she can open it through clicking on the name.

Applied Method

The method applied can be documented giving a rough description or a detailed description by predefined qualifiers. Also the input and output of the method are listed and it is possible to construct method descriptions of sequential sample handling. To reduce effort the method description for a catalogue entry is copied from a predefined list and only variations from standard values must be changed.

Actual Data

Actual measured data are stored together with site and sample descriptions in a hierarchical scheme.
The basic idea is that measured data are just another kind of description of sites or samples.

Project

Project descriptions can be hierarchical and give detailed descriptions of goals, methodology, participants and sub-projects or campaigns.

Visualization of Data

Data can be visualized as tables, figures or gridded images. In this version, only values of one variable can be plotted at a time. But it is planned to allow a free combination of variables as far as they fit together in some aspect. Visualization is considered a helpful tool to decide if data found are of value for a special purpose.