UNIVERSITÄT LEIPZIG
Spezielle Botanik und Funktionelle Biodiversität

Karin Nadrowski
nadrowski@uni-leipzig.de
alfresco.uni-leipzig.de/spezbot/~nadrowski

Biodiversity – Ecosystem Functioning (BEF-China) data portal

Karin Nadrowski & Christian Wirth

The BEF data portal focuses on …

The BEF data portal stores and shares data for the Biodiversity – Ecosystem Functioning (BEF-China) Research Unit (FOR 891) of the German Science Foundation (DFG) [1]. It is built to serve two aims: preparing data for long term storage contributing to the science community as a whole and to assist consistent data exchange within the research unit itself. Here we address all of the seven recommendations put forth by the DFG (points 1-7 below) [2]. Our focus presently lies on supporting naming conventions and transparent data sharing.

Fig. 1: Browser snapshot of the BEF-China data portal displaying upload buttons, the tag list, and a list of available datasets (http://befchina.bgc-jena.mpg.de/portal).

Uploading an Excel 2003 spreadsheet in a predefined format (Fig. 2, points 2, 3) loads single raw data values into our data base (Fig. 3).

Uploading free format files allows exchange of custom information.

We are currently exploring tagging for data columns (Fig. 2, 3).

Data sets can be added to paper proposals using a mechanism similar to shopping carts (see below: transparency).

… quality control and data sharing

1: Defining data granularity Subprojects provide aggregated values at the level of individual, species or site in a generic flat file format (point 3).

4: Data provenance
Data columns as well as data sets include provenance information (Fig. 3). When exporting data, provenance information is always included.

5: Data dissemination
Our data sharing agreement states that data are free for the public after they have been published.

6: Meta data
During data upload, we provide a semi-automated user interaction to annotate data columns (see also point 3).

2: Organizational concept Convenient retrieval during active research is given by relying on a spreadsheet format for import and export of flat files (Fig. 2), since most ecologists work with spreadsheets.

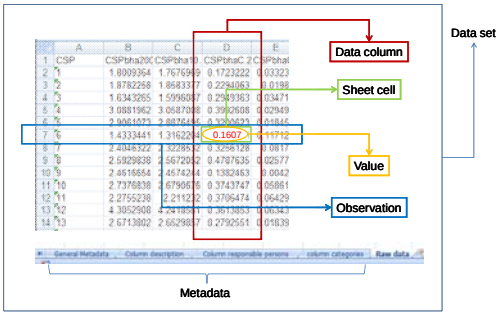

Fig. 2: Our predefined Excel 2003 spreadsheet is designed based on the Ecological Metadata Language [].

3: International data standards The structure of our flat file format (Fig. 2) as well as the structure of our data base (Fig. 3) is based on the Ecological Metadata Language [3, 4] and the Darwin Core Standard [5].

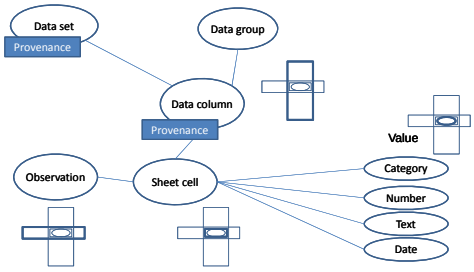

Fig. 3: The database structure allows to group data columns across datasets.

7: Quality control Our data portal stores single raw data entries and thus can handle quality control of values across data sheets. Based on our import spreadsheet (Fig. 2), we provide semi-automated user interaction that allocate data values to data groups. These data groups can be used to develop and adhere to naming conventions within portal data (Fig. 4).

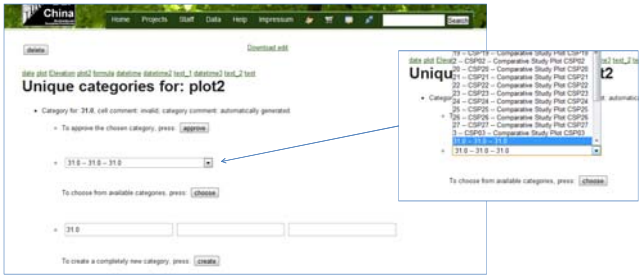

Fig. 4: During upload, unknown names (here a typo) are presented to the user. Names are unique within data groups and can be used to establish within project naming conventions. The user can accept the value from the spreadsheet, choose a name from the database or create a new name for this data group.

Transparent data sharing Our data sharing agreement states a paper proposal mechanisms involving two organizational layers (project board, data owners). We are currently developing an online registration for this process.

Summary

The BEF data portal provides a framework for sharing data within a research unit. This includes developing and adhering to custom naming convention and transparent data set sharing during active research. The same infrastructure is used to provide long term availability and open access after first publication.

Technical Details

The BEF data portal is implemented using Ruby on Rails [6], which adheres to the model – view – controller philosophy. Rails provides direct access to data base tables through it's ActiveRecord Module. We use PostgreSQL [7] as database.

[1] http://www.bef-china.de; The Chinese-European DFG Research Unit "BEF China" (FOR 891); http://www.dfg-science-tv.de/de/projekte/funktion-durch-vielfalt/
[2]Deutsche Forschungsgemeinschaft (DFG); Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme; Unterausschuss für Informationsmanagement (Hg.) (2009): Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten. (http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf)
[3] Fegraus et. al (2005): Maximizing the Value of Ecological Data with Structured Metadata: An Introduction to Ecological Metadata Language (EML) and Principles for Metadata Creation. *Bulletin of the Ecological Society of America, 86*, 158-168
[4] http://knb.ecoinformatics.org/software/eml
[5] http://www.tdwg.org/standards/450/
[6] http://rubyonrails.org
[7] http://www.postgresql.org

printed at Universitätsrechenzentrum Leipzig