Data repositories in environmental sciences

Concepts, definitions, technical solutions and user requirements

28 Feb – 01 Mar 2011, Rauischholzhausen/Marburg

Program - Abstracts

Joint meeting of:

DFG-FOR816 Ecuador

TMF

Biodiversity Exploratories DFG-SP1374

biodiversity exploratories
functional biodiversity research

DFG senate commission on biodiversity research

DFG

Hosted by:

Philipps Universität Marburg

Introduction and aim of the meeting

Quality proofed research data build the foundation for scientific knowledge gain. Although billions of Euros are spent annually in Germany alone to acquire data only a fraction are accessible over time and are re-used in further contexts. The Alliance of German Science Organisations acknowledges this problem and pools activities for their solutions. Interdisciplinary research project repositories are at the intersection of data collection in the field, work-in-progress-storage, and long-term archiving. Against this background the workshop explores required standards of data capture, data curation and interchange with respect to technical and social issues, and addresses challenges in long-term archiving in a national and international framework.

The aim is to bring together the community of project database manager, standards developer, information scientists, librarians, and funding agencies. As results, we anticipate a better communication between the many disciplines, an overview of the current state of the art and a roadmap for future activities.

Meeting Organization

Organizing committee

Prof. Dr. Jörg Bendix	DFG-FOR816 Ecuador
Dr. Birgit Gemeinholzer	DFG senate commission on biodiversity research
Dr. Jens Nieschulze	Biodiversity Exploratories DFG-SP1374

Meeting management, secretariat

Thomas Lotz	DFG-FOR816 Ecuador
Sonja Hease	Philipps University Marburg

Conference venue

Conference Centre of the Justus-Liebig-University Gießen

Schloss Rauischholzhausen
Ferdinand-von-Stumm-Straße
35085 Ebsdorfergrund-Rauischholzhausen

A historical view on "Schloss Rauischholzhausen"

The property was first mentioned in a charter book of the monastery of Fulda between 750 and 779 and was initially a fief of the Lords of Eppstein until the Archbishop of Mainz acquired it completely in 1369. From then on the vassals called themselves Lords Rau of Holzhausen, one of the knights on the eastern bank of the Rhine.

The last member of the Rau family served as an officer in the Hessian army. When Hesse-Kassel became part of Prussia he refused to join the Prussian army and sold all his property to the ambassador's delegate Stumm.

The new owner, Ferdinand Stumm, was a member of a famous family of industrialists. He became imperial ambassador in Madrid and was ennobled by

Kaiser Friedrich in 1888. Many famous lords, earls, and dukes were his guests, among others Kaiser Friedrich and the Duke of Hesse.

He resigned as an envoy in 1890 and died in 1925, which left him 35 years to take care of the castle and its park. His eldest son, Ferdinand von Stumm, inherited the castle and sold the complete Holzhausen property in 1937. The castle was bought by the Kerkhoff Foundation in Bad Nauheim and then leased to the University of Giessen as a site for experiments in agriculture.

The forest was sold to Mr. von Waldhausen, while the castle and the adjacent park were made available to the public. A school was founded in the castle for the training of kindergarten teachers.

After having been confiscated as Nazi property by the allied forces in 1945, the castle and the park became property of the state of Hesse and were put at the disposal of the Justus Liebig University Giessen as a conference centre.

The castle of Rauischholzhausen was designed by the architect Carl Schaefer, a student of Gottlieb Ungewitter, in the style of *Klein-Potsdam*. The construction lasted from 1871 to 1878 and the castle was lavishly decorated. In 1873 the building collapsed, because the foundations had been badly layed. Carl Jonas Mylius and Alfred Friedrich Bluntschli, both students of the architect Semper, were commissioned to redesign the castle according to Schaefer's original concept. In 1875 the construction of the roof and the South East wing were completed. The main building, reminiscent of an English manor house, was finished a year later. In 1878 the half-timber wing of the building was completed.

When Rau von Holzhausen first lived on the location of today´s castle he stayed at the castle mill, which is located at the lower entrance to the park and was built in the 16th century. In today's pond there was a water castle which could be approached through the large portal which still adorns the atrium of the castle. The door frame is decorated by a lion's head with a ring in its mouth and the ionic colums may be identified as belonging to the renaissance style.

The park is designed in the English style and contains almost 300 different types of trees. Two creeks run through the park and form several ponds connected by artificial cascading waterfalls. Sculptures including a Lithuanian princess, a female slave, a virgin, and a tired rambler may be found between groups of trees.

(Source: http://www.uni-giessen.de/cms/ueber-uns/rhh/about.html 17.02.2011)

Map of venue "Schloss Rauischholzhausen"

Conference schedule

Monday, 28 Feb 2011 --- Morning

	Time	Duration	Event		by
			Arrival		
Intro	9:00	45'		Welcome Introduction to the Workshop	- Bendix - Nieschulze
Standards/Ontologies	9:45	45' oral 15' disc	***Session 1***	**Keynote** Using observational data models to enhance data interoperability for integrative biodiversity and ecological research	- Schildhauer
	10:45	15'	Coffee Break		
	11:00	20' oral 10' disc		GSC Standards, GCDML and megx.net : From minimum standard specifications through implementations to web services	- Kottmann
		20' oral 10' disc		What are data and metadata and what is the role of Media Content?	- Vogt
	12:00	1h 30'	Lunch	*Move into rooms*	

Monday, 28 Feb 2011 --- Afternoon

Concepts of Data Repositories	13:30	45' oral 15' disc	***Session 2***	**Keynote** International repository infrastructures -- connectivity between thematic and generic approaches	- Horstmann
	14:30	20' oral 10' disc		A comparative evaluation on technical solutions of long-term data repositories related to biodiversity	- Bach
	15:00	15'	Coffee Break		
	15:15	1h 30'	***Poster Session***	Guided Poster and Discussion Tour - Short Introduction (3min) on the Poster + Discussion	
	16:45	45' oral 15' disc	*cont.* *Session 2*	**Keynote** An assessment of costs and risks in the operation of long-term digital archive infrastructures	- Klump
	17:45	20' oral 10' disc		Data version management	- Reusser
	18:15			*Organisational matters*	
	18:30		Dinner		
	20:00		Social evening in the *„Schlosskeller*"		

Tuesday, 01 Mar 2011 --- Morning

	Time	Duration	Event		by
	7:30	1h 30'	Breakfast	*Move out of rooms (until 9:00)*	
User Requirements	9:00	45' oral 15' disc	***Session 3***	**Keynote** DataONE: Changing community practice and transforming the environmental sciences through access to data and tools	- Michener
	10:00	20' oral 10' disc		Morph.D.Base – Increasing transparency and reproducibility of morphological data	- Grobe
	10:30	20' oral 10' disc		Challenges and success factors in adapting IT for the management of scientific data	- Worbel
	11:00	30'	Coffee Break		
	11:30	20' oral 10' disc		The user's view on data sharing	- Enke
	12:00	1h		Open Discussion Future needs and challenges regarding user-accepted and sustainable data repositories in environmental sciences	**Plenum** Moderation - Nieschulze - Bendix - Gemeinholzer

Tuesday, 01 Mar 2011 --- Afternoon

	13:00	1h	Lunch	*Paying drinks*
	Official end of the workshop			
	14:00		individual meetings	"Members DFG Senate Commission" "Data manager of DFG funded collaborative projects"
	18:00		leaving the venue	

Notes

Abstracts of the talks

Data standards and ontologies for integrative biodiversity research

Session 1

Using observational data models to enhance data interoperability for integrative biodiversity and ecological research

M. Schildhauer

Research in the biodiversity and ecological sciences often requires discovering, interpreting, and integrating many different types of data collected at many different scales, ranging from genomic and molecular information, through human and sensor-based field observations, to remote-sensed images, etc. These data are typically collected by numerous independently operating individuals and organizations, with little or no coordination regarding the structure, syntax, and semantics used to document and store the data. In recent years, many scientific disciplines have recognized that controlled vocabularies can provide advantages for describing and interpreting heterogeneous data, especially when based on standardized web semantic approaches such as OWL-DL ontologies. What is needed in addition, however, are robust mechanisms for linking these ontological terms to data, such that powerful inferencing services can result, based on the use of standard reasoners (e.g. Pellet).

Observational data models may provide a suitable foundational construct for linking ontologies to data sets, by formalizing the interpretation of scientific data as associations of observations and measurements. A number of environmental and earth science disciplines, including the biodiversity sciences, ecology, evolutionary science, geospatial sciences, hydrology, and oceanography, have converged on the use of observation data models to help with intradisciplinary data interoperability. Thus, an opportunity now exists to leverage this commonality for facilitating cross-disciplinary data interoperability as well. In this talk I will discuss some of the main observational data models in the earth and environmental sciences, and how these are being harmonized to afford a compatible model that can assist with cross-disciplinary data interoperability. I will also describe ongoing work that demonstrates how semantically-constructed observational data models provide unique capabilities for describing and interpreting scientific data sets that are useful for integrative biodiversity and ecological research.

M. Schildhauer; National Center for Ecological Analysis and Synthesis (NCEAS), Santa Barbara CA; schild@nceas.ucsb.edu

GSC Standards, GCDML and megx.net : From minimum standard specifications through implementations to web services – Part I

R. Kottmann

Based on the rapid development of cheaper and faster next generation sequencing technologies, large scale- and mega-sequencing projects are becoming routine procedures nowadays.
Nevertheless, even billions of ACGTs provide only limited information when it comes down to a better understanding of the life style, adaptations and fitness of the organisms in their natural environments. It is now clear that the full potential of sequence analysis can only be achieved if the geographic and environmental context of the sequence data is considered, herewith referred to as contextual data (Hughes Martiny and Field, 2005). However, until recently the dedicated long-term archives of INSDC (International Nucleotide Sequence Database Collaboration) comprising GenBank, ENA (European Nucleotide Archive), and DDBJ (DNA Databank of Japan) were not ready to support the capture, curation, and interchange of contextual data.
The aim of the Genomic Standards Consortium (GSC) is to standardize contextual data. The GSC specified the "Minimum Information about a Genome Sequence" (MIGS) for all kinds of eukaryotic-, bacterial-, and archaeal genomes and the "Minimum Information about a Metagenome Sequence" (MIMS) for community sequencing approaches (Field, et al., 2008). A "Minimum Information about an ENvironmental Sequence" (MIENS) for marker genes as the ribosomal RNA (rRNA) or the Cytochrome Oxidase subunit I (COI), among others, is currently developed (Yilmaz et al., under revision). The MIGS/MIMS/MIENS family of standards requires capturing the geographic and environmental context of the sequence data as well as processing parameters about the sequencing procedure.
In order to proof the technical feasibility of the GSC's standards the GSC develops the Genomic Contextual Data Markup Language (GCDML) (Kottmann et al. 2008). GCDML is the reference implementation of MIGS/MIMS/MIENS in XML Schema and therefore makes the contextual data exchangeable between Web Services.
Several databases like INSDC, institutes as well as technology platforms (Camera, MG-Rast) have already adopted the GSC standards, see http://gensc.org/gc_wiki/index.php/Adopters. →

GSC Standards, GCDML and megx.net : From minimum standard specifications through implementations to web services – Part II

→

Among the early adopters is Megx.net (www.megx.net) a web portal for Marine Ecological GenomiX (Kottmann, et al.). Megx.net offers access to a unique combination of environmental and molecular genomic data, with manually curated habitat descriptors (Hirschman , et al., 2008) for all genomes, metagenomes, and marker genes, their respective contextual data, and additionally integrated environmental data.

Based on GSC and Open Geospatial Consortium (OGC) standards, megx.net provides digital maps of prokaryotic genomes, metagenomes, viruses and diversity datasets (rRNAs). Megx.net's Genes Mapserver service allows overlying genomic data with a rich set of environmental data such as temperature, nitrate, phosphate, etc., based on profile and remote sensing data from e.g. the World Ocean Database. All maps are highly interactive and can be flexibly combined to explore the environmental context.

The Genes Mapserver is completely based on open source software and implements existing standards. Among others the Web Map Service (WMS) standard of the Open Geospatial Consortium was implemented allowing programmatic access via web services.

R. Kottmann; Max Planck Institute Bremen, Marine Microbiology;
rkottman@mpi-bremen.de

What are Data and Metadata and what is the Role of Media Content?

L. Vogt

Due to lack of common data standards, the communicability and comparability of biological data across various levels of organization and taxonomic groups is limited. Consequently, all projects involving co-operations of scientists from different biological disciplines or from different taxonomic specializations face problems of miscommunication and data incompatibilities. The introduction of general data and metadata standards in biology could improve this situation. However, in some disciplines there is even no common consensus on what data and metadata actually are. For instance do many morphological databases use phylogenetic character matrices for recording and representing "morphological data", whereas many phylogeneticists would argue that character matrices represent sets of homology hypotheses and thus conclusions based on data which, therefore, go far beyond the factually given. I will suggest a general basic distinction between data and metadata and will also analyze the role that additional media content plays in this context. I will argue that every data and metadata standard actually comprises four different types of standards: a content, concept, nomenclatural, and format standard, each of which plays an important role in our daily communication. I will also discuss the role of data bases and their use of formalized vocabularies (i.e., ontologies) in current activities for data standardization in biology.

L. Vogt; Rheinische Friedrich-Wilhelms-University Bonn, Institue for Evolution and Ecologie; lars.m.vogt@googlemail.com

Concepts of Data Repositories

Session 2

International repository infrastructures -- connectivity between thematic and generic approaches

W. Horstmann

The question addressed is how to mediate between disciplinary data standards, for example in biodiversity research, and generic data standards for connecting repositories worldwide. Today, research publications and data from the lab or the field are offered and processed in distributed repositories around the world. Connecting these repositories allows the distributed mass of publications and data to appear as one large aggregation of research resources that supports their re-use and sharing among researchers, industry and citizens and, thereby, transcend disciplinary and geographic boundaries. But focused disciplinary research requires repositories to represent very specific aspects in the data. As an example, taxonomical information is essential in biodiversity research but not relevant to atmospheric research. However, both need geographical coordinates. Such a variety of disciplinary requirements is a challenge to the services that build aggregations of different repositories across diverse disciplinary fields. It is proposed here that generic lightweight standards for connecting repositories, i.e. how they expose and ingest data, are pivotal to create a basic international repository infrastructure that can be extended arbitrarily for disciplinary requirements.

W. Horstmann; Bielefeld University Library;
wolfram.horstmann@uni-bielefeld.de

A comparative evaluation on technical solutions of long-term data repositories related to biodiversity

K. Bach

Complex biodiversity data are severely threatened of getting lost because to date long-term data storage is neither sufficiently elaborated nor realized for both, individual and collaborative research projects. This is not only due to the lack of standardized data formats and long-term funded suitable repositories but also because of the short-term funding of research projects. On the contrary, data storage in central databases is getting increasingly important or is yet mandatory for the acceptance of peer-reviewed publications in specific fields of biodiversity research as e.g. molecular sciences. Even more important is long-term sustainable data storage for global and local scientific monitoring systems and related research activities on the impact of global environmental change on biodiversity and ecosystem services. In this study, we investigate the existing infrastructure and technical solutions for data repositories in biodiversity research. Based on comprehensive in-depth interviews with experts of existing relevant biodiversity databases we identify the core technical requirements to be addressed by a central data repository for the entire research community:

- Standards and data models for storage and exchange of data and metadata
- Dataimport/-export
- Connectivity to biodiversity networks
- Digital rights management
- Quality assurance
- Citation
- Versioning
- Workflows
- User Support

K. Bach; Philipps-University Marburg, Faculty of Geographie;
bachk@staff.uni-marburg.dw

An assessment of costs and risks in the operation of long-term digital archive infrastructures

J. Klump

The rapidly increasing amount of digital research materials raises issues relating to access, curation and preservation of these materials. A growing number of research funders are now implementing policies requiring researchers to submit data management, preservation or data sharing plans with their funding applications. New data infrastructures are needed to allow researchers to comply with these policies. On the scale of long-term data preservation most research projects are relatively short-lived and structures are needed to sustain data curation long after the project has ended. On the other hand, long-running projects, such as DFG Collaborative Research Centres, have a lifespan of 15 years. Adding ten years of data preservation after the end of the CRC results in 25 years of data curation. How are we going to fund and maintain data infrastructures over this long period of time? A number of studies have investigated the cost structures of long-term data preservation and the risks associated with different parts of the data curation life cycle. These analyses can be used to outline areference model for long-term data curation infrastructures.

J. Klump; Helmholtz Centre Potsdam – German Research Centre for Geoscience; jens.klump@gfz-potsdam.de

Data version management

D. Reusser

An important underlying principle of science is reproducibility, which is also relevant for the valuable data underlying the analyses carried out. However, in practice in many fields, documentation of data manipulation and homogenization is generally not standardized and in practice often is unstructured and heterogeneous even within the same project or working group.

An Observations Data Model (ODM) has been proposed in the field of hydrology as a standard approach to store and access time series data with a point spatial reference. While data quality levels are part of ODM, a version control management for the data is not part of the concept. We propose such a version control management based on the same mechanisms as they are used in version control systems for software development. The proposed system requires only little modifications to the ODM.

In order to assure consistent changes to the data, we suggest a standard user interface to import, access and modify the data. Differences between such a standard user interface with or without a data version management system are minimal. The use of such a system will greatly improve trackability and reproducibility of data handling.

D. Reusser; Potsdam-Institute for Climate Impact Research;
reusser@pik-potsdam.de

User Requirements

Session 3

DataONE: Changing Community Practice and Transforming the Environmental Sciences through Access to Data and Tools

W. K. Michener

Science and society are just beginning to wrestle with some of the grand environmental challenges that face humankind in the next decades—all of which will require unprecedented access to data that cross scientific domains, space (meters to the biosphere), time (seconds to centuries), and scales of biological and physical organization. Principal data challenges lie in discovering relevant data, dealing with extreme data heterogeneity, and converting data to information and knowledge. Addressing these challenges requires new approaches for managing, preserving, analyzing, and sharing data, most of which are equally challenging from technical and socio-cultural perspectives. This paper describes DataONE, a federated data network that is being built to provide universal access to data about life on earth and the environment that sustains it, and to support science by: (1) engaging the relevant science, data, and policy communities; (2) providing easy, secure, and persistent storage of data; and (3) disseminating integrated and user-friendly tools for data discovery, analysis, visualization, and decision-making. The paper provides an overview of the DataONE architecture, focusing on data submission, curation and preservation, and citation through its affiliated data centers. The paper discusses the DataONE data citation objectives and approach, linkage of publications to data, as well as other community engagement mechanisms that are employed to promote data sharing and enable new science.

W. K. Michener; Director and esa's data archive editor at LTER and esa Network Office, University of New Mexico, Albuquerque NM; wmichener@LTERnet.edu

Morph.D.Base – Increasing Transparency and Reproducibility of Morphological Data

P. Grobe

Morph·D·Base (www.morphdbase.de; online since September 2006) is an online database for storing morphological metadata and graphical material according to modern standards using controlled vocabularies like DarwinCore. You can enter data referring to individual *Specimens*, specifying their sex, sampling data, museum collection information, preparation methods, etc. Each specimen entry is linked to a *Morph·D·Base* taxonomy that is based on ITIS taxonomy and nomenclature (http://www.itis.usda.gov). It is possible to add new *Taxa* to it and submit these entries to ITIS. Furthermore, you can upload *Media* files (bitmap, z-stacks, vector graphics, audio, and movie files), specify the resolution that will be publicly available (original, 1024×768, 800×600, 640×480), classify them into general categories (e.g. cLSM, diagram, drawing, etc) and annotate their content (which part of which specimen is shown, viewing angle, etc). A *Literature* module allows uploading personal literature database files (RIS & BibTex formats) for referencing information in the abovementioned entries to its corresponding literature source. All data entries of a user can either be released to the general public or they remain editable. Users can share their editable entries by assigning 'read only' or 'edit' rights to defined groups of users and thus share content with collaborators before publication. Currently we are developing a *Matrix Editor* module. It allows generating phylogenetic character matrices, in which each cell can be linked to relevant specimen, media, and literature entries in *Morph·D·Base*. This enables detailed documentation of all relevant information, satisfying the requirement for reproducibility and transparency. The editor possesses features for managing and collaborative live editing of matrices, including dynamic update, chat, commenting, color labeling of cells/rows/columns, schedule-notification, change tracking, progress markers, import/export of NEXUS files, etc. Besides many other features that will improve character coding (e.g. automatic consistency checks), users will be able to use terms and concepts from anatomy ontologies (e.g. OBO, www.obofoundry.org), resulting in marked-up character and character state definitions/codes, which significantly increases overall semantic transparency of matrices as well as comparability between them. The combination of the different modules of *Morph·D·Base* will significantly increase transparency and reproducibility of data production, thereby mediating between data representation and phylogenetic inference, which will close the methodological gap between data production and tree evaluation.

This study was supported by the Deutsche Forschungsgemeinschaft DFG (VO 1244/3-3).

P. Grobe; University of Bonn, Institute of Evolution and Ecology;
pegrobe@googlemail.com

Challenges and success factors in adapting IT for the management of scientific data

M. Wrobel

Achieving appropriate utilization of information technology is a crucial prerequisite for efficient management of digital data. Within the context of economy and industry it can be observed that, e.g., relational database management systems have been successfully established more or less as conditio sine qua non for a multitude of data-depending applications. However, we have to constitute a remarkable lack in the adaptation of such technologies for scientific applications. For the latter, it is still not uncommon that data is managed in proprietary approaches, based on files and a variety of different formats, not unlikely to hamper, e.g., efficient identification, re-use or exchange of available data resources.

We present two applications aiming to improve the management of time series in the context of hydrology and of multi-disciplinary climate impact research. Against this background, we outline several factors that can impede the successful establishment of IT to support the management of scientific data, as well as potential success factors.

M. Worbel; Potsdam-Institute for Climate Impact Research;
wrobel@pik-potsdam.de

The user's view on data sharing

N. Enke

In the last few years scientists, funding agencies, and journals identified an increased need for the long term storage and availability of biodiversity data. This lead to a considerable effort to provide data portals and data bases for the storage of various types of biodiversity data in national, EU-wide and international initiatives. However, only few of the provided possibilities are so far accepted by the scientific communities.
An extensive survey tried to identify the reasons for the reluctant use of the already existing infrastructure for data deposition. Following questions have been raised: Is there really a need for an increased availability of biodiversity data? How diverse is biodiversity data? What are the main hindrances for data deposition? What are the minimum user requirements to ensure an active use of data bases? Scientists want repositories for biodiversity data - the demands and requirements raising the acceptance will be presented.

N. Enke; Freie Universität Berlin, Botanic Garden and Botanical Museum Berlin-Dahlem; n.enke@bgbm.org

B. Gemeinholzer; Justus-Liebig-University Giessen, Faculty of Biology; Birgit.Gemeinholzer@bot1.bio.uni-giessen.de

Notes

Poster Abstracts

Poster Session

Biodiversity – Ecosystem Functioning (BEF-China) data portal

K. Nadrowski

The BEF data portal stores and shares data of the Biodiversity – Ecosystem Functioning (BEF-China) Research Unit (FOR 891) of the German Science Foundation (DFG). 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 (1: defining raw data, 2: organizational concept, 3: international standards, 4: data provenance, 5: data dissemination, 6: meta, 7: quality control). However, since we are yet a very young research project, the focus of our application presently lies on the organizational concept of sharing and storing data (2), as well as quality control (7). Our application uses a Ruby on Rails with a PostgreSQL database as backend . We further rely on a data structure inspired by the Ecological Metadata language (EML), for which we drafted a spreadsheet to ease the import and export of data.

K. Nadrowski; Institute Special Botany and functional Biodiversity, University of Leipzig; nadrowski@uni-leipzig.de

The TR32 project database – TR32DB

C. Curdt, D. Hoffmeister, C. Jekel, G. Waldhoff, G. Bareth*

In this contribution, we introduce the centralized data management system (TR32DB) of the DFG funded research project Transregional Collaborative Research Centre 32 (TR32) "Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation". The TR32 is a joint project between the Universities of Aachen, Bonn, Cologne, and the Research Centre Jülich. The research partners are from the fields of soil and plant science, hydrology, meteorology, geophysics, geography, remote sensing, and mathematics. They work in 13 project sections on exchange processes between the soil, vegetation, and the adjacent atmospheric boundary layer (SVA). The TR32 study area is defined by the catchment of the river Rur situated in western Germany, parts of the Netherlands, and Belgium.
The TR32DB system is developed following the guidelines of the DFG (e.g. 'Good Scientific Practice', bulletin 60.06) to enable a centralized, sustainable storage and to offer exchange of the TR32 data for all project participants. Due to the various disciplines in the TR32, the TR32DB has to handle a huge amount of various research data from different research fields. Besides collected research data (including measured and modelled data), reports, presentations, publications, and pictures provided by the TR32 participants, also purchased geodata from various institutions are considered. In addition, a TR32 metadata management system was developed.
The design of the TR32DB is a combination of file management, database, and web-interface including Web-GIS functionalities. The TR32DB is physically stored at the Regional Computing Centre Cologne and accessible online (www.tr32db.uni-koeln.de). The structure is developed according to recent standards and principles (e.g. from OGC, ISO and W3C). Programming standards and languages like JavaScript, PHP, HTML, CSS, and SQL are used to develop and customize the interaction between databases, web-interface, and Web-GIS. Moreover, metadata standards like Dublin Core and INSPIRE are considered, as well as DOI guidelines.
The TR32DB offers several functionalities like management, sustainable storage and backup, visualization, exchange, and provision of TR32 data. Furthermore, the linkage of TR32 data with adequate and TR32 specific metadata is provided. All features are accessible online, e.g. representation and search of TR32 data via the TR32DB web-interface. Some functions are only available for authorized users like download of TR32 data, entry and modification of corresponding metadata, application of TR32-DOIs, temporary download of purchased and processed geodata, as well as visualization of geodata in a Web-GIS.

C. Curdt; University of Köln, Faculty of Geography; c.curdt@uni-koeln.de

Biodiversity Exploratories Information System

- **Feature Overview**
- **Technical Overview**

E. Petzold, A. Ostrowski**

The Biodiversity Exploratory Information System (BExIS) is the data repository and information exchange platform of the Biodiversity Exploratories project.
The Biodiversity Exploratories started in 2006 and is funded by the German Science Foundation (DFG) as a long-term, interdisciplinary open research platform. The goal is to advance biodiversity research in Germany and further the understanding of the relationship between biodiversity of different taxa and levels, the role of land use and management for biodiversity, and the role of biodiversity for ecosystem processes. At present 40 projects with 300 researchers work on 1000 research plots of two landuse types in three regions in Germany.
The interchange of data among projects is considered crucial for the success of the project. The diversity among the involved projects is mirrored by the diversity of file formats, -syntax and, semantics. BExIS supports file formats common among its researchers (xlsx, xls, dbf, mdb, txt) and allows data ingestion of flexible syntax, determined in collaboration with all field researchers.
At present, BExIS hosts about 470 datasets, comprising around 30 million tupel with around 400 million data items, around 7000 variables, 2000 concepts, and more than 10000 differently labelled species.
BExIS has a web-based interface to a platform for storage and management of all project data in a database. It has a fine grained user and rights management and offers upload, update, and online edit of data. BExIS supports different structured data types and unstructured data. Support of flexible data syntax is ensured by use of XML based metadata schemas and datastructure mapping.
BExIS provides further project desired functionalities, e.g. to organize and document field work, for visualising and editing common data on maps, and provides the use of statistical analysis also for online merged data sets.
BExIS is presented with two posters. One will give a "Feature Overview" with BExIS provided functionalities. The second poster shows a "Technical Overview" with focus on system design and used technology.
The problem of the semantic heterogeneity and data interoperability is dealt with the implementation of ontology and is described in the Poster "Semantic Data Access within the Biodiversity Exploratories Information System BExIS".

A. Ostrowski; Max-Planck-Institut of Biogeochemistry Jena;
aostrow@bgc-jena.mpg.de

E. Petzold; Max-Planck-Institute of Biogeochemistry Jena;
epetzold@bgc-jena.mpg.

Data management of BIOTA AFRICA, now useful for RSSC Southern Africa

G. Muche, T. Hillmann, A. Suwald*

BIOTA AFRICA (BIOdiversity Monitoring Transect Analysis in Africa) is a biodiversity monitoring network spanning Africa aiming to detect changes of biodiversity at a continental scale (Jürgens et al. 2010). Between 2001 and 2009, BIOTA AFRICA has established 71 standardised biodiversity monitoring sites (so called Biodiversity Observatories) in 8 African countries (http://www.biota-africa.org).
The Regional Science Service Centre for Adaptation to Climate Change and Sustainable Land Management in southern Africa (RSSC Southern Africa) started in 2010 and is a joint initiative of Angola, Botswana, Namibia, South Africa, Zambia, and Germany, responding to the challenges of global change. The RSSC aims at supporting the southern African countries in establishing research, capacity development and services related to climate change and adapted land management, focussing at the topics climate, water, forestry, agriculture and biodiversity (http://www.rssc-southernafrica.net).
Within BIOTA the BioMonitoring Data Facility (Muche et al. 2010) forms a joint infrastructure of all regional projects, BIOTA Maroc, BIOTA East Africa, BIOTA West Africa, and BIOTA Southern Africa of BIOTA AFRICA. In contrast, within RSSC, similar IT tools are used to map the existing and future project activities and to archive all data, present data in a contextualised format and enhance accessibility to information via internet. Also the web-facilitation of communication between all network partners is organised by the IT team.

Literature:
Jürgens, N., Haarmeyer, D. H., Luther-Mosebach, J., Dengler, J., Finckh, M., Schmiedel, U. (2010) [Eds.]: Biodiversity in southern Africa. Volume 1: Patterns at local scale – the BIOTA Observatories. – XX + 801 pp., Klaus Hess Publishers, Göttingen & Windhoek.

Muche, G., Hillmann, T., Suwald, A., Jürgens, N. (2010): Data access and availability: BIOTA Data Facility. – In: Schmiedel, U., Jürgens, N. [Eds.]: Biodiversity in southern Africa. Volume 2: Patterns and processes at regional scale: pp. 337–342, Klaus Hess Publishers, Göttingen & Windhoek.

G. Muche; University of Hamburg, Biodiversity, Evolution and Ecology of Plants; gerhard.muche@botanik.uni-hamburg.de

CRC806-Database: Implementation of a data management for an interdisciplinary research project

C. Willmes, D. Hoffmeister, O. Biessmann, C. Hütt, D. Kürner, K. Volland, A. Bolten, G. Bareth*

For the data management project of the Collaborative Research Centre 806 „Our way to Europe" (CRC806), funded by the German Research Agency (DFG), at the Universities of Cologne, Bonn and Aachen, the here presented CRC806-Database is in development. Within the 21 sub-projects of the CRC806 about 30 Post-Docs, 60 Ph.D. candidates and more than 100 graduates from the archaeological-, geo- and cultural- sciences working together and producing diverse data, to answer the questions of how modern humans (homo sapiens sapiens) migrated from Africa to Europe and to answer why and how the Neanderthals (homo neanderthalensis) disappeared. The CRC is planned for an overall period of twelve years, and is subdivided into three four-year project terms which are underlying an evaluation at the end of each term. In the sense of the DFG directive for the good scientific practice it is required for any CRC to provide a sustainable data management policy and its implementation, which guaranties the accessibility to the produced research data for at least the duration of the project duration plus ten years, which results in an up to 22 years period. The CRC806-Database web portal offers the members of the CRC806 to archive, catalogue/index and share their data and publications with colleagues and the interested public. To implement this, the application utilizes the typo3-based multilevel user access control system for fine grained access to each dataset granted by the dataset authors. To make the user access to the data as comfortable as possible, several open interfaces are implemented. For geodata access several interfaces of the OGC standards family such as WMS, WFS, WCS and CSW are implemented. For the automated access to Metadata the well known formats DublinCore and ISO19115 are used. Additionally it is taken care of avoiding proprietary data formats where possible. Another important aim of the web portal is to provide functionality and tools for data analysis, visualization and aggregation to gain additional insight into the different datasets by combining and filtering them through a WebGIS interface in its spatial and temporal context. The implementation of the web portal and the underlying data management infrastructure is built on several open source software applications, such as Typo3, Geoserver, deegree, Openlayers and GeoExt. Additionally we integrated the ESRI's ArcGIS Server technology to provide some of the OGC services of the presented data infrastructure.

More Information is available at:
http://www.sfb806.de and http://www.sfb806db.uni-koeln.de

C. Willmes; University of Köln, Faculty of Geography; c.willmes@uni-koeln.de

FOR816-Ecuador Data Warehouse: An EML-based relational project database and more

T. Lotz, M. Dobbermann*, D. Göttlicher, T. Nauss, J. Bendix*

One crucial aspect within most joint projects in ecosystem research is the way of data storage and exchange. The Data Warehouse of the DFG-Research Unit 816 (FOR816dw) provides an integrative maintenance to enable the retrieval, transfer, storage, and analysis of scientific data of various ecological disciplines, focussed on biodiversity research.

To structure the data and to make it usable by other scientists a robust metadata concept has to be implemented. The Ecological Metadata Language (EML: www.knb.ecoinformatics.org/software/eml) provides a sophisticated XML-based scheme to describe scientific ecological data sets. The relational database structure of the FOR816dw is based on the metadata logic of EML and contains at least all mandatory fields to be compliant to this specification. The metadata are transmitted by the data owner via a browser based wizard during the data upload. They can be queried for temporal, spatial, and thematically features to find data sets in the data pool. The metadata bears furthermore information on the data set creator, the intellectual rights, and the data collection/processing process of each contained attribute of a data set. The adaption of the concept to describe each attribute (variable) of a data set establishes the granulated storage and analysis of the data set values. All information are stored in the relational database and can be queried and reused. The metadata can be exported to XML (Extensible Mark-up Language) and are made user-friendly visible by an XSLT (Extensible Stylesheet Language Transformation) conversion script as a HTML web page. They are automatically attached to a data package, when a data set is downloaded.

Beside the storage and maintenance of data sets and metadata the FOR816dw provides the management of administrative data of the research group. The user management, the project structure,a news system, a mailing-lists generator, the travel reimbursement, and the station booking system is based on the same personnel and project specific data used for the metadata description of the scientific data sets. The project`s webpage (www.tropicalmountainforest.org) is the single point of contact for the project members and provides a user-friendly access to all information of the DFG-Research Unit 816 "Biodiversity and Sustainable Management of a Megadiverse Mountain Ecosystem in South Ecuador".

T. Lotz; Philipps-University Marburg, Departement of Geography;
thomas.lotz@staff.uni-marburg.de

M. Dobbermann; Philipps-University Marburg, Departement of Geography;
maik.dobbermann@staff.uni-marburg.de

Environmental Data Management with the River Basin Information System (RBIS)

S. Kralisch, F. Zander**

The pressure on environmental systems is increasing worldwide due to population growth and the consequences of climate change. Adaptable tools and methods are needed to elaborate information, develop understanding and strategies for sustainable use and management of environmental systems. Such tools should assist scientists, natural resource managers and decision makers in their work by providing them with (1) sufficient information about relevant drivers, attributes and factors from measured data, (2) tools and methods for user friendly access and integrated analyses of such data, i.e. environmental information systems (EIS), and (3) tools for estimating additional information not available as measurements, i.e. environmental simulation models. In addition, a seamless and preferably standardized exchange of data between EIS and integrated assessment tools such as simulation models is gaining increasing importance as the amount of available environmental information grows together with the complexity of current environmental challenges. The research at the Department of Geoinformatics, Hydrology and Modelling at the Friedrich Schiller University of Jena is reflecting the needs described above with the development of a number of software tools. They provide services and assistance for specific spatio-temporal related environmental problems emerging from integrated research projects. In general, all developed tools (1) use open source software wherever possible to be cost effective, (2) are provided as open source software to others and (3) are highly flexible and adaptable to ensure useability for a wide range of environmental problems. We present the River Basin Information System (RBIS) – a web-based information management system with a focus on time series and geospatial data. It provides user friendly interfaces for data input and output, powerful visualization components and an adaptable set of functions for data analysis, data management and data enrichment. Sample applications from various research projects are used to illustrate that RBIS is an appropriate platform to meet the demands for collaborative and integrated data management.

S. Kralisch; University of Jena, Faculty of Geographie;
sven.kralisch@uni-jena.de

F. Zander; University of Jena, Faculty of Geographie;
franziska.zander@uni-jena.de

The Diversity Workbench – a virtual biodiversity research environment

D. Triebel, M. Weiss, T. Schneider, G. Hagedorn, S. Jablonski, W. Reichert, B. Volz, G. Rambold*

The modularized Diversity Workbench represents a virtual research environment for multiple scientific purposes with regard to management and analysis of biological data. The framework is appropriate to store different kinds of biodiversity data and facilitates the processing of ecological, observational, collection and taxonomic data. It is capable and flexible enough to be applied as data storage unit for institutional data repositories.

The database framework Diversity Workbench consists of interacting database components. Clients of every component of the Workbench are used as stand-alone applications and provide supporting functions to clients of corresponding components. This results in a high flexibility with regard to the conceptual design, enabling sophisticated user administration and a rapid setup of project-specific and user-adapted entry forms. Further, it facilitates the dynamic integration of web services and external data resources.

Functions for field data gathering are provided by the component DiversityMobile, designed for monitoring ecological and biological data, using mobile devices. The information model of DiversityMobile is fully compatible with that of DiversityCollection, the database component for the storage of collection and observational data. DiversityCollection provides the option to export data to GBIF via the BioCase-Wrapper by using the TDWG standard schema ABCD.

D. Triebel; Botanische Staatssammlung München; triebel@bsm.mwn.de

DNA Bank Network – A shared data portal for integrative molecular and biodiversity research

G. Droege, H. Zetzsche, B. Gemeinholzer*

Access to genomic DNA becomes more and more important, as fast progress in analyses technologies continues to speed up research on the genetic diversity of all life forms. Availability of DNA samples to verify or continue molecular studies of an object is currently only possible by contacting the author or the institutions in which the studies took place. Thus, falsifiability of results and original data - the foundation of good scientific practice – is often neglected. The lack of verification of molecular data has been noted in several scientific publications.

The aim of the DNA Bank Network is to increase transparency and verification of scientific work by sharing DNA data as well as related specimen information. The DNA Bank Network, currently linking four research collections each focusing on different groups of organisms (bacteria, archaea, protists, plants, algae, fungi, and animals) is a DFG funded non-commercial scientific service based on a unique concept. While all network partners provide long-term access to information associated with their DNA samples such as DNA quality parameters, DNA sequence data, tissue and specimen information including voucher images via a central web portal partner database remain operated and curated independently. The network's data model is built upon the internationally accepted biodiversity information TDWG standards and the GBIF/BioCASE infrastructure. The Network is now open to be extended by including additional data providers. Thus, it holds the potential to serve as worldwide online platform for easy access to non-human DNA and tissue samples. The Network is fully functional; data architecture, open-source software tools, and documentation standards have been developed. In addition, contributions to DNA bank standards concerning optimized tissue and DNA handling were made.

G. Droege; Freie Universität Berlin, Botanic Garden & Botanical Museum Berlin-Dahlem; g.droege@bgbm.org.

TRY: towards a unified global database of plant functional traits

J. Kattge, S. Díaz, S. Lavorel, C. Prentice, P. Leadley, G. Bönisch, C. Wirth*

Plant functional traits – morphological, anatomical, physiological or phenological features measurable at the individual level - reflect the outcome of evolutionary processes in the context of abiotic and biotic environmental constraints and therefore provide a link between biodiversity, systematics and plant functioning. They are seen as a promising basis towards a more quantitative and predictive ecology and global change science. Here we present an ongoing initiative towards a permanent communal repository of plant trait data, named TRY (http://www.try-db.org). With a broad acceptance by the plant trait scientific community - more than 120 scientists from more than 80 research groups worldwide involved - TRY has become the largest plant trait database in the world to date. Currently the TRY database contains about 2.6 million trait entries for about 70,000 plant species, with a focus on 52 groups of traits characterising the vegetative and regeneration stages of the plant life cycle, including growth, dispersal, establishment and persistence. The TRY database is consistent with the Extensible Observation Ontology (OBOE, Madin et al. 2007, 2008; Kattge et al. 2011). The current development of a specific ontology for plant traits, headed by TraitNet (http://traitnet.ecoinformatics.org), is based on the plant traits as compiled in TRY so far.

J. Kattge; Max- Planck Institut Jena, Faculty of Biogeochemistry;
jkattge@bgc-jena.mpg.de

LifeWatch – a European e-science and observatory infrastructure supporting access and use of biodiversity and ecosystem data

M. Frenzel, S. Klotz, A. Hardisty, O. Bánki*

There are many promising earth and biodiversity-monitoring projects underway across the globe, but they often operate in information islands, unable easily to share data with others. This is not convenient: It is a barrier to scientists collaborating on complex, cross-disciplinary projects which is an essential nature of biodiversity research.

LifeWatch (www.lifewatch.eu) is an ESFRI (European Strategy Forum on Research Infrastructures) initiative which has just entered its construction phase. It is aiming at new ways of collaboration, in an open-access research environment to solve complex societal and scientific questions on biodiversity and ecosystems. It installs a range of new services and tools to help the researchers communicate, share data, create models, analyse results, manage projects and organise the community. The power of LifeWatch comes from linking all kinds of biodiversity related databases (e.g. collections, long-term monitoring data) to tools for analysis and modeling, opening entirely new avenues for research with the potential for new targeted data generation. At this level the interface with national data repositories becomes most important, as this opens the opportunity for users to gain advantage from data availability on the European level. LifeWatch will provide common methods to discover, access, and develop available and new data, analytical capabilities, and to catalogue everything, to track citation and re-use of data, to annotate, and to keep the system secure. This includes computing toolkits for researchers: for instance, an interoperable computing environment for statistical analysis, cutting-edge software to manage the workflow in scientific projects, and access to new or existing computing resources. The result: 'e-laboratories' or virtual labs, through which researchers distributed across countries, time zones and disciplines can collaborate. With emphasis on the open sharing of data and workflows (and associated provenance information) the infrastructure allows scientists to create e-laboratories across multiple organizations, controlling access where necessary.

Actually Germany is not as yet a member of LifeWatch, but efforts are underway to become a member and the German community of data providers should be prepared to jump in and consider the conceptual link to LifeWatch.

M. Frenzel; Helmholtz Centre Halle – German Research Centre for Environmental Research; mark.frenzel@ufz.de

DataONE: A virtual data center for the biological, ecological, and environmental sciences.

W. K. Michener, A. Budden, R.Koskela, D. Vieglais; DataONE Team*

Numerous grand environmental challenges face humankind in the next decades including climate change, decreased water availability, and loss of ecosystem services. The science required to understand and mitigate these challenges will require unprecedented access to data that cross science domains, distance (meters to the biosphere), time (seconds to centuries), and scales of biological and physical organization. Principal data challenges lie in discovering relevant data, dealing with extreme data heterogeneity, and converting data to information and knowledge. Addressing these challenges requires new approaches for managing, preserving, analyzing, and sharing data, which are equally challenging from technical and socio-cultural perspectives.

DataONE is designed to be the foundation of new innovative environmental research by ensuring preservation and access to multi-scale, multi-discipline, and multi-national data. DataONE is unique in that it: (1) builds on existing data centers, leveraging the global investment in scientific data preservation; (2) creates a global, federated data network by focusing on interoperability solutions, providing tools and services to enable new science and knowledge creation; and (3) supports evolving communities of practice enabled by the DataONE cyberinfrastructure (CI) and informed by best practices, exemplary data management plans, and tools that support all aspects of the data life cycle.

The CI implementation of DataONE is based on three principal components: *Member Nodes* which are existing or new data repositories that support the DataONE Member Node APIs; *Coordinating Nodes* that are responsible for cataloging content, managing replication of content, and providing search and discovery mechanisms; and an *Investigator Toolkit* which is a modular set of software and plug-ins that enables interaction with DataONE infrastructure through commonly used analysis and data management tools. There are three major development phases of the project: (1) prototype development for proof of concept; (2) release of a stable, secure, reliable core data management and discovery infrastructure; and (3) progressive addition of new features and functionality. The project is currently ending the first of these three phases. Below, we summarize the progress to date with respect to CI development, community engagement, and project management, and highlight some of the key activities planned for the next three years.

W. K. Michener; Director and esa's data archive editor at LTER and esa Network Office, University of New Mexico, Albuquerque NM; wmichener@LTERnet.edu

Climate Data for Our Future - Acquired, Analysed, Archived

A research scheme for the long-term archival storage of weather and climate data, funded by the DFG (German Research Foundation)

F. Quadt

When will lightning strike again? Will there be tropical temperatures in this area in 50 years? What were temperatures like during the last ice age? Weather and climate researchers are trying to find answers to these questions. Their findings are based on measured data – temperature, atmospheric pressure, atmospheric humidity and much more. The more comprehensive and detailed the data, the more reliable the findings. However, the researchers lack a set standard to publish and secure the valuable data for the long term.

This is the starting point of the joint research project carried out by the Bonn-Rhine-Sieg University of Applied Sciences, University of Bonn and the German Climate Computing Centre (Deutsches Klimarechenzentrum GmbH): we are developing an automated workflow system to systematically collect, examine and publish weather and climate data according to set quality standards and rules. In this way scientists have easier access to the data, which will facilitate data collection and ensure data reliability.

The problem is: if the weather data is only published on the website of a research institute, a meteorologist is unlikely to ever find it. The solution to the problem is: previous data collection results can be easily researched and compared to one's own current data using the standardised database. This enables researchers to draw on previous research work, evaluate and compare their own data with other observations and climate model data.

Once the project has been completed, the new workflow system will be made available as open source software to scientists worldwide.

F. Quadt; Hochschule Bonn-Rhein-Sieg, Faculty of Computer Science;
florian.quadt@h-brs.de

The Carbon Portal – Metadata search and data access for the Integrated Carbon Observation System

M. Köchy

The Carbon Portal will be the future search and access point for the data that will be collected in ICOS (Integrated Carbon Observation System), representing a future network of atmospheric, terrestrial, and oceanic measurement stations. In addition, data from earlier projects, documents, and supporting data will be searchable through the Carbon Portal. We suggest two search engines to deal with the heterogeneous data. The Metadata from ICOS and related greenhouse gas projects that conform to stringent metadata specifications will be included in one database. This allows the use of a more efficient search engine, i.e. the user will be able to target the search more precisely. Data sets will be searched by selecting multiple constraints (filters) from menus. The "Database" filter will include the category "ICOS" for selecting data measured in ICOS. The "Type" filter allows selecting different levels of primary data, remote sensing data, and model results. The metadata provides links to the original data stored at the ICOS Thematic Centers or World Data Centers. Non conforming metadata will be retrieved via a second search engine that has a broader search approach at the cost of precision. Data is stored independently from the Carbon Portal. Data are retrieved through the Carbon Portal via a URN. Facilities for archiving, control of access to data, and attribution of data providers are still under discussion.

M. Köchy; Johann Heinrich von Thünen Institute Braunschweig;
office@martinkoechy.de

Components and aspects of an integrated data management approach – Part I

D. Fleischer, K. Jannaschk, H. Mehrtens*, C. Schirnick, P. Springer*

The Kiel Data Management Infrastructure (KDMI) started from a cooperation of three large-scale projects (SFB574, SFB754 and Cluster of Excellence The Future Ocean) and the Leibniz Institute of Marine Sciences (IFM-GEOMAR).

KDMI key features focus on the data provenance which we consider to comprise the entire workflow from field sampling or measurements through lab work to data calculation and evaluation. Managing the data of each individual project participant in this way yields the data management for the entire project and warrants the reusability of (meta)data. Accordingly scientists provide a workflow definition of their data creation procedures resulting in their target variables.

The central idea in the development of the KDMI presented here is inspired by the object oriented programming concept which allows to have one object definition (workflow) and infinite numbers of object instances (data). Each definition is created by a graphical user interface and produces XML output stored in a database using a generic data model. On creation of a data instance the KDMI translates the definition into web forms for the scientist, the generic data model then accepts all information input following the given data provenance definition. An important aspect of the implementation phase is the possibility of a successive transition from daily measurement routines resulting in single spreadsheet files with well known points of failure and limited reusability to a central infrastructure as a single point of truth.

An interim system allows users to upload and share data files from cruises and expeditions. It relates files to metadata such as where, when, what, who etc. As a proof of concept we use a 'truncated workflow' to migrate a selection of marine chemical data files and their structured metadata into the generic data model. A web application will allow data extraction for selectable parameters, time and geocoordinates. The availability of these widely used data is expected to motivate more scientists to design their own workflows for their upcoming work and their resulting data.

This data provenance approach in terms of human workflows has several positive side effects: (1) the scientist designs the extend and timing of data and metadata prompts by workflow definitions while (2) consistency and completeness (mandatory information) of metadata in the resulting XML document can be checked by XML validation. (3) Storage of the entire data creation process (including raw data and processing steps) provides a multidimensional quality history accessible by all researchers in addition to the commonly applied one dimensional quality flag system and thus (4) improves the reuseability of the data.

→

Components and aspects of an integrated data management approach – Part II

→
(5) The KDMI concept focuses on bringing data management infrastructure into the daily measurement routines instead of the final data management hassle at the end of each project. (6) The KDMI can be extended to other scientific disciplines or new scientific procedures by simply adding new workflow definitions. The data input can start from this point while domain specific outputs with the newly added data instances will be created by the KDM-Team.

The KDMI follows scientists' requests for Web 2.0 like (net)working platforms but instead of sharing privacy or making friends it is all about sharing daily scientific work and data with project partners. For this purpose we have deployed a portal server (Liferay) where individual scientists are assigned to project communities and working groups or have their own working spaces. All these features are expected to raise the acceptance of the integrated data management applications and advance scientific collaboration.

K. Janneschk; Christian-Albrechts University, Department of Computer Sciences, Kiel

H. Mehrtens; Leibniz Institut für Meereswissenschaften (IFM-GEOMAR), Kiel; hmehrtens@ifm-geomar.de

Identifying Objects or Concepts in Ecological Data using Topic Maps

D. Seifarth

Ecological data is generally linked to identifiable objects or described concepts. However, naming conventions within one research group often do not match those of other research groups. Data sharing between groups then depends on an existing knowledge base that stores information about equivalent objects or concepts. Two objects could be merged, if they represent the same identitiy. Here Topic Maps may offer a simple solution. Topic Maps are a subject-centric method of modeling, relations between data originating from different sources may be extracted easily as Topic Maps especially addresses the identity of objects or concepts.

The Biodiversity Ecosystem Functioning China data portal stores primary research data based on a flat file structure. Data columns within and among files can be grouped in data groups to combine similar data. At the same time, nominal data values within a data group have to be unique and can be used to evolve naming conventions within the portal. Data columns, Data groups, as well as unique nominal data values can be used as entry points for semantic annotation.
Here we show how Topic Maps could be used to link species names of trees used within the data base to international identifiers for taxonomic names (IPNI identifier). Additionally we show how Topic Maps could help to map the tables of the relational database to the Darwin Core meta-data standard used by GBIF.

D. Seifarth; University of Leipzig, Topic Maps Lab, Computer Science;
redxeag@googlemail.com

Semantic Data Access within the Biodiversity Exploratories Information System BExIS

D. Heimann

In order to address the complex ecological relationships of landuse, biodiversity, and ecosystem prossess, research in ecology progressed from site-specific approaches with measurements at spatial and temporal small scales to long-term explorations on larger scales. This research paradigm is promoted by collaborations and data exchange policies between research projects, and technical advances within data management. On this account data reuse and interoperability plays an important role in information systems design of large scale and long-term ecological projects. Against this background, providing just a database system is no longer sufficient - instead a data repository should also be able to capture the semantics of its data sets to allow for the disentanglement of ecological relationships.

The poster presents our approach of enabling the Biodiversity Exploratories Information System (BExiS) for data sharing and interoperability. We have addressed the problem of semantic heterogeneity within the datasets by relating all data items to concepts of a common, upper-level ontology which reflects the relationships between individual ecological topics. The poster describes our approaches of ontology-based data integration and user-friendly annotation and shows furthermore an example of the current state of our underlying ontology.

D. Heimann; Max-Planck-Institute of Biogeochemistry Jena;
dheimann@bgc-jena.mpg.de

Notes

List of Participants

Aue, Birigt, birgit.aue@bio.uni-giessen.de, BIOLOG-Europe-Program, Justus-Liebig-University Gießen, Germany
Bach, Kerstin, bachk@staff.uni-marburg.de, Faculty of Geography, Philipps-University Marburg, Germany
Beck, Erwin, erwin.beck@uni-bayreuth.de, Department of Plant Physiology,University of Bayreuth, Germany
Bendix, Jörg, bendix@staff.uni-marburg.de, Faculty of Geography, Philipps- University Marburg, Germany
Buscot, Francois, francois.buscot@ufz.de, Helmholtz Centre Halle – German Research Centre for Environmental Research, Germany
Curdt, Constanze, c.curdt@uni-koeln.de, Faculty of Geography, University of Köln, Germany
Dobbermann, Maik, maik.dobbermann@staff.uni-marburg.de, Faculty of Geography, Philipps-University Marburg, Germany
Droege, Gabriele, g.droege@bgbm.org., Freie Universität Berlin, Botanic Garden & Botanical Museum Berlin-Dahlem, Germany
Enke, Neela, n.enke@bgbm.org, Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Germany
Fleischer, Dirk, dfleischer@ifm-geomar.de , Leibniz Institut für Meereswissenschaften, (IFM-GEOMAR), Kiel, Germany
Frenzel, Mark, mark.frenzel@ufz.de, Helmholtz Centre Halle – German Research Centre for Environmental Research, Germany
Gemeinholzer, Birgit, Birgit.Gemeinholzer@bot1.bio.uni-giessen.de, Faculty of Biology, Justus-Liebig-University Giessen, Germany
Glöckner, Frank, O., fog@mpi-bremen.de, Max Planck Institute for Marine Microbiology Bremen, Germany
Grobe, Peter, pegrobe@googlemail.com, Institute of Evolution and Ecology, University of Bonn, Germany
Güntsch, Anton, a.guentsch@bgbm.org, Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Germany
Hartig, Katja, Katja.Hartig@dfg.de, Lebenswissenschaften 2, Deutsche Forschungsgemeinschaft (DFG), Bonn, Germany
Heimann, Dennis, dheimann@bgc-jena.mpg.de, Max-Planck-Institute of Biogeochemistry Jena, Germany
Herrmann, Sylvie, sylvie.herrmann@ufz.de, Helmholtz Centre Halle – German Research Centre for Environmental Research, Germany
Herzog, Elena, elena.herzog@senckenberg.de, LOEWE - Biodiversity and Climate Research Centre BiK-F, Frankfurt am Main, Germany
Hickler, Thomas, Thomas.hickler@senckenberg.de, LOEWE - Biodiversity and Climate Research Centre BiK-F, Frankfurt am Main, Germany
Hoffmeister, Dirk, dirk.hoffmeister@uni-koeln.de, Institute of Geography, GIS/RS Group, University of Cologne, Germany
Horstmann, Wolfram, wolfram.horstmann@uni-bielefeld.de, CIO Wissenschaftliche Information University of Bielefeld, Universitätsbibliothek, Germany

Hotes, Stefan, stefan.hotes@bio.uni-giessen.de, BIOLOG-Europe-Program , Justus-Liebig-University Gießen, Germany
Kattge, Jens, jkattge@bgc-jena.mpg.de, Faculty of Biogeochemistry, Max- Planck Institut Jena, Germany
Klotz, Stefan, stefan.klotz@ufz.de, Helmholtz Centre Halle – German Research Centre for Environmental Research, Germany
Klump, Jens, jens.klump@gfz-potsdam.de, Helmholtz Centre Potsdam – German Research Centre for Geoscience; Centre for GeoInformation Technology , Germany
Köchy, Martin, office@martinkoechy.de, Johann Heinrich von Thünen Institute Braunschweig, Germany
König-Ries, Brigitta, Birgitta.Koenig-ries@uni-jena.de, Faculty of Computer Science, University of Jena, Germany
Kottmann, Renzo, rkottman@mpi-bremen.de, Microbial Genomics Group, Marine Microbiology, Max Planck Institute Bremen, Germany
Kralisch, Sven, sven.kralisch@uni-jena.de, Faculty of Geography, University of Jena, Germany
Kühn, Ingo, ingolf.kuehn@ufz.de, Helmholtz Centre Halle – German Research Centre for Environmental Research, Germany
Leimer, Sophia, sophia.leimer@giub.unibe.ch, Faculty of Geography, Universität Bern, Switzerland
Lotz, Thomas, thomas.lotz@staff.uni-marburg.de, Faculty of Geography, Philipss-University Marburg, Germany
Mertens, Hela, hmehrtens@ifm-geomar.de, Leibniz Institut für Meereswissenschaften (IFM-GEOMAR), Kiel, Germany
Michener, William K., wmichener@LTERnet.edu, Director and esa's data archive editor, LTER and esa Network Office, New Mexico EPSCoR State Program, University of New Mexico, USA
Muche, Gerhard, gerhard.muche@botanik.uni-hamburg.de, Biodiversity, Evolution and Ecology of Plants, University of Hamburg, Germany
Nadrowski, Karin, nadrowski@uni-leipzig.de, Institute Special Botany and functional Biodiversity, University of Leipzig, Germany
Nauß, Thomas, thomas.nauss@uni-bayreuth.de, Faculty of Geography, University of Bayreuth, Germany
Nieschulze, Jens, jniesch@bgc-jena.mpg.de, Database manager biodiversity exploratories, Max-Planck Institute of Biogeochemistry, Jena, Germany
Nuske, Robert, robert.nuske@nw-fva.de, Nordwestdeutsche Forstliche Versuchsanstalt, Göttingen, Germany
Ostrowski, Andreas, aostrow@bgc-jena.mpg.de, Max-Planck-Institut of Biogeochemistry Jena, Germany
Petzold, Eleonora, epetzold@bgc-jena.mpg.de, Max-Planck-Institute of Biogeochemistry Jena, Germany
Pompe, Sven, sven.pompe@uni-jena.de, Institute of Ecology, Friedrich-Schiller-University Jena, Germany
Quadt, Florian, florian.quadt@h-brs.de, Faculty of Computer Science, Hochschule Bonn-Rhein-Sieg, Germany
Reusser, Dominik, reusser@pik-potsdam.de, North-South Project - Potsdam Institue for Climate Impact Research, Research Domain 2: Climate Impacts & Vulnerabilities, Germany

Ritter, Tim, tritter@gwdg.de, Department of Geoinformatics, Biometrics and Forest Growth, Univeristy of Göttingen, Germany
Schildhauer, Mark, schild@nceas.ucsb.edu, Director of Computing, NCEAS Santa Barbara CA, USA
Schirnick, Carsten, cschirnick@ifm-geomar.de, Leibniz Institut für Meereswissenschaften, (IFM-GEOMAR), Kiel, Germany
Schönwitz, Roswitha, Roswitha.Schoenwitz@dfg.de, Lebenswissenschaften 2, Deutsche Forschungsgemeinschaft (DFG), Bonn, Germany
Seeger, Bernhard, seeger@informatik.uni-marburg.de, Department of Mathematics and Computer Science, Philipps-University Marburg, Germany
Seifarth, Daniel, redxeag@googlemail.com, Topic Maps Lab, Computer Science, University of Leipzig, Germany
Springer, Pina, pspringer@ifm-geomar.de, Leibniz Institut für Meereswissenschaften, (IFM-GEOMAR), Kiel, Germany
Thiele, Jan C., jthiele@gwdg.de, Ecoinformatics, Biometrics and Forest Growth, University of Göttingen, Germany
Triebel, Dagmar, triebel@bsm.mwn.de, Botanische Staatssammlung München, Germany
Velescu, Andre, andre.velescu@giub.unibe.ch, Faculty of Geography, University of Bern, Switzerland
Vierkant, Paul, paul.vierkant@uni-konstanz.de, University of Konstanz, Germany
Vogt, Lars, M., lars.m.vogt@googlemail.com, Institue for Evolution and Ecologie, Rheinische Friedrich-Wilhelms-University Bonn, Germany
Willmes, Christian, c.willmes@uni-koeln.de, Faculty of Geography, University of Köln, Germany
Worbel, Markus, wrobel@pik-potsdam.de, North-South Project - Potsdam-Institute for Climate Impact Research, Research Domain 2: Climate Impacts & Vulnerabilities, Germany
Wubet, Tesfaye, tesfaye.wubet@ufz.de, Helmholtz Centre Halle – German Research Centre for Environmental Research, Germany
Zander, Franziska, franziska.zander@uni-jena.de, Faculty of Geographie, University of Jena, Germany
Zhang. Jie, jie.zhang@uni-wuerzburg.de, Data Manager of the DFG Kilimanjaro Projekt, University of Würzburg, Germany

Notes

Notes