Data repositories for biodiversity and ecology research

A. Kehl

Contents

• Introduction

• IBF Project

• Institutional Data Repositories

• Conclusions
Demands of Science Policies

- DFG (1998): „Vorschläge zur Sicherung guter wissenschaftlicher Praxis“: Storage for a minimum of ten years (mostly local databases)

Traditional Dataflow

Field
Gathering of Primary Data

Project-sided Computer (local)
Data Storage and Analysis

Pen & Paper

Data Digitalization

- Post-processing of data
- Further analysis (e.g., statistics)

Data items

Journal publication
Aspired Dataflow

Field

Gathering of Primary Data

Mobile Application
with user interface and database

Data Synchronisation

Desktop Application
• Communication between the Central and Mobile Database
• Stores Metafiles in Synchronisation Database

Conflict Management System

Data Repository

Multi-User Database Management System
• Persistent Data Storage
• ER-Database Schema

Permanent Storage of Data
Aspired Dataflow

Field

- Gathering of Primary Data
- Mobile Application with user interface and database

Data Flow

- Data items
- Conflict Management System
- Desktop Application
  - Communication between the Central and Mobile Database
  - Stores Metafiles in Synchronisation Database

Data Repository

- Data items
- Multi-User Database Management System
  - Persistent Data Storage
  - ER-Database Schema

Specific Data Subsets
IBF Project

- Setting up an Information Network on Biological Research Data gained in the Field up to the Sustainable Storage in a Primary Data Repository (IBF)

www.diversymobile.net
Establishment of dataflow from and to an institutional data repository

- **Data gathering** in the field via mobile device (PDA, Smart Phone) with software (*DiversityMobile*) for
  - data entry
  - accessing taxonomy and ecology descriptor presets

- **Data transfer** to institutional data repository (via *DiversityWorkbench* database component *DiversityCollection*)
  - data storage

- **Data redistribution** to end-users by data transfer to PDAs, wrappers (ABCD schema) and diverse applications for data analysis and presentation
  - data analysis
  - data presentation
  - data re-usage
IBFProject

Field: Diversity
Mobile

Tool for Gathering Observation Data

Mobile Application
• Temporary Data Storage
• Local Windows CE Database

• Access to external data sources (e.g. standard list of taxon names)

• GPS functionality

• Collection of multimedia data (images, video, audio)

• Previously recorded data accessible in the field

• All entered data to be transferred to the data repository
IBF-Project

- Long-term data storage
- Access to data history
- Individual observations are referable (GUIDs, LSIDs)
- Selective data export for local post-processing
- Publication (traditional pathway, Internet)
IBF-Project

- Participating research groups

- Working group Fungi and Lichens
- Working group Vascular plants
- Working group Ecology
- Monitoring of fungi and lichens
- Various habitats (Regensburg surroundings)
• Field mapping of vascular plants
• Transect and various habitats (Regensburg surroundings)
• Small-scale distribution of gall-inducing sawflies (Hymenoptera, Tenthredinidae) on Salix fragilis
• Experimental plot at University Bayreuth, EBG
Tests of implemented client interface for gathering “multidimensional” data (at regular time intervals)
IBF Data Flow

Data gathering via mobile device

Data items

Local project database

Repository Project-Database

Research Projects

Institutional Data Repository

Backup Interface

Wrapper Interface (ABCD-Standard; XML)

WWW-Portal (HTTP, CF, JSP)

Database interface (JDBC, ODBC)

Data maintenance (Client-Software: Diversity Mobile)

Analysis and publication of data

External Data Interfaces (thesauri, taxonomies, maps, ...)

Archival Storage Database

Regional computation center

Data repositories

IBF Data Flow

Data gathering via mobile device
IBF Data Flow

Observatory Data, Data Flow

External Data Interfaces (thesauri, taxonomies, maps, ...)

Data set gathering (Client-Software: Diversity Mobile)

Data items

ORM: Diversity Synchronization

(Local) Project-Database

Data items

Repository Project-Database

Repository Database

(Institutional) Data Repository

Backup Interface

Wrapper Interface
(ABCD-Standard; XML)

WWW-Portal
(HTTP, CF, JSP)

Research Projects

External Data Interfaces

Database interface (JDBC, ODBC)

Data maintenance (Client-Software: Diversity Mobile)

Analysis and publication of data

(Data domain-specific) repository databases
IBF Data Flow

Observatory Data, Data Flow

Data gathering (Client-Software: Diversity Mobile)

Data items

ORM: Diversity Synchronization

Research Projects

(Local) Project-Database

Repository Project-Database

Repository Database

Backup Interface

Wrapper Interface (ABCD-Standard; XML)

External Data Interfaces (thesauri, taxonomies, maps, ...)

Database maintenance (Client-Software: Diversity Mobile)

Analysis and publication of data

Institutional Data Repository

Archival Storage Database

Regional computation center

IBF.DE

WWW-Portal (HTTP, CF, JSP)
IBF Data Flow

Observatory Data, Data Flow

External Data Interfaces (thesauri, taxonomies, maps, ...)

Data gathering (Client-Software: Diversity Mobile)

Data items

ORM: Diversity Synchronization

Local) Project-Database

Data items

Research Projects

Institutional Data Repository

Repository Project-Database

Repository Database

Repository Database

Archival Storage Database

Backup Interface

Wrapper Interface (ABCD-Standard; XML)

GBIF.DE

Data maintenance (Client-Software: Diversity Mobile)

Analysis and publication of data

Database interface (JDBC,ODBC)

WWW-Portal (HTTP, CF, JSP)

Regional computation center
IBF Data Flow

Observatory Data, Data Flow

Data gathering (Client-Software: Diversity Mobile)

ORM: Diversity Synchronization

(Local) Project-Database

Research Projects

External Data Interfaces (thesauri, taxonomies, maps, ...)

Institutional Data Repository

Repository Database

Backup Interface

Wrapper Interface (ABCD-Standard; XML)

Archival Storage Database

WWW-Portal (HTTP, CF, JSP)

Database interface (JDBC, ODBC)

Data maintenance (Client-Software: Diversity Mobile)

Analysis and publication of data
Data Repository Structure at SNSB

- Storage of structured research data
  - (Data domain-specific) structured ‘standard data’
  - structured ‘non-standard data’

- Provision/support of data standards (according to ‘Biodiversity Information Standards’, formerly TDWG)
  - Collection (and observation) data ABCD
  - Descriptive data SDD
  - Taxonomic data TCS

- Quality control and publication of primary data
  (persistent identifiers)
Data Repository Structure at SNSB

Research project

Institutional data repository

Data flow between components of scientific research projects

Data flow between components of institutional repository database

Data gathering

Custom-built Personal Database

Custom-built Project Database

Data Items

Edit, analyze and correct data with custom-built algorithms and tools

Cache for Data in Standardized Format

Secured Import Cache Database

Perform quality control

Perform corrections according to results of quality control

Data provenance allows for data quality control (who changed what when and why)

Items may be referenced within publications via a permanent identifier; the resolve service can be used for retrieving the original data item

Data provenance allows for retrieving information about data changes (who changed what when and why)

All information is subject to digital rights management - i.e. it can be restricted what will be published under which conditions

Publication of data (project specific identifiers, traditional publication)

Data Provenance Interface

Data provenance allows for retrieving information about data changes (who changed what when and why)

Data provenance allows for retrieving information about data changes (who changed what when and why)

All information is subject to digital rights management - i.e. it can be restricted what will be published under which conditions

Feedback information

Feedback Interface

Kinds of virtual hosts:
- Mirrored scientific research projects
- Provision of legacy data services
- Hosting of project environments

A virtual environment consists always of all project components (databases, interfaces, clients, ...)

Backup Interface

Data processing accesses data stored within repositories, portals and/or provision services

Reference Resolve Service (publications)

Items may be referenced within publications via a permanent identifier; the resolve service can be used for retrieving the original data item

Data provenance allows for retrieving information about data changes (who changed what when and why)

All information is subject to digital rights management - i.e. it can be restricted what will be published under which conditions

Publication of data (project specific identifiers, traditional publication)

Data Provenance Interface

Data provenance allows for retrieving information about data changes (who changed what when and why)

All information is subject to digital rights management - i.e. it can be restricted what will be published under which conditions

Feedback information

Feedback Interface

Kinds of virtual hosts:
- Mirrored scientific research projects
- Provision of legacy data services
- Hosting of project environments

A virtual environment consists always of all project components (databases, interfaces, clients, ...)

Backup Interface

Data processing accesses data stored within repositories, portals and/or provision services

Reference Resolve Service (publications)

Items may be referenced within publications via a permanent identifier; the resolve service can be used for retrieving the original data item

Data provenance allows for retrieving information about data changes (who changed what when and why)

All information is subject to digital rights management - i.e. it can be restricted what will be published under which conditions

Publication of data (project specific identifiers, traditional publication)
Data Repository Structure at SNSB

Research project

Institutional data repository

Data provision
Conclusions

Aspects of long-term storage of primary research data

• Scientific
  – Intersection and completion with standard data
  – Improvement of data quality
  – Precondition for reproducibility of data

• Economic
  – Access and subsequent use option on behalf of third parties

• Social
  – Credits to data owners and providers
Deutsche Forschungsgemeinschaft

Förderbereich LIS – Informationsmanagement

(GZ: INST 106535/1-1, INST 21946/1-1, INST 2850/1-1, INST 747/1-1)