



Data management

in the GVP

GLOWA-Strategy Meeting
Bonn
22./23.01.2007

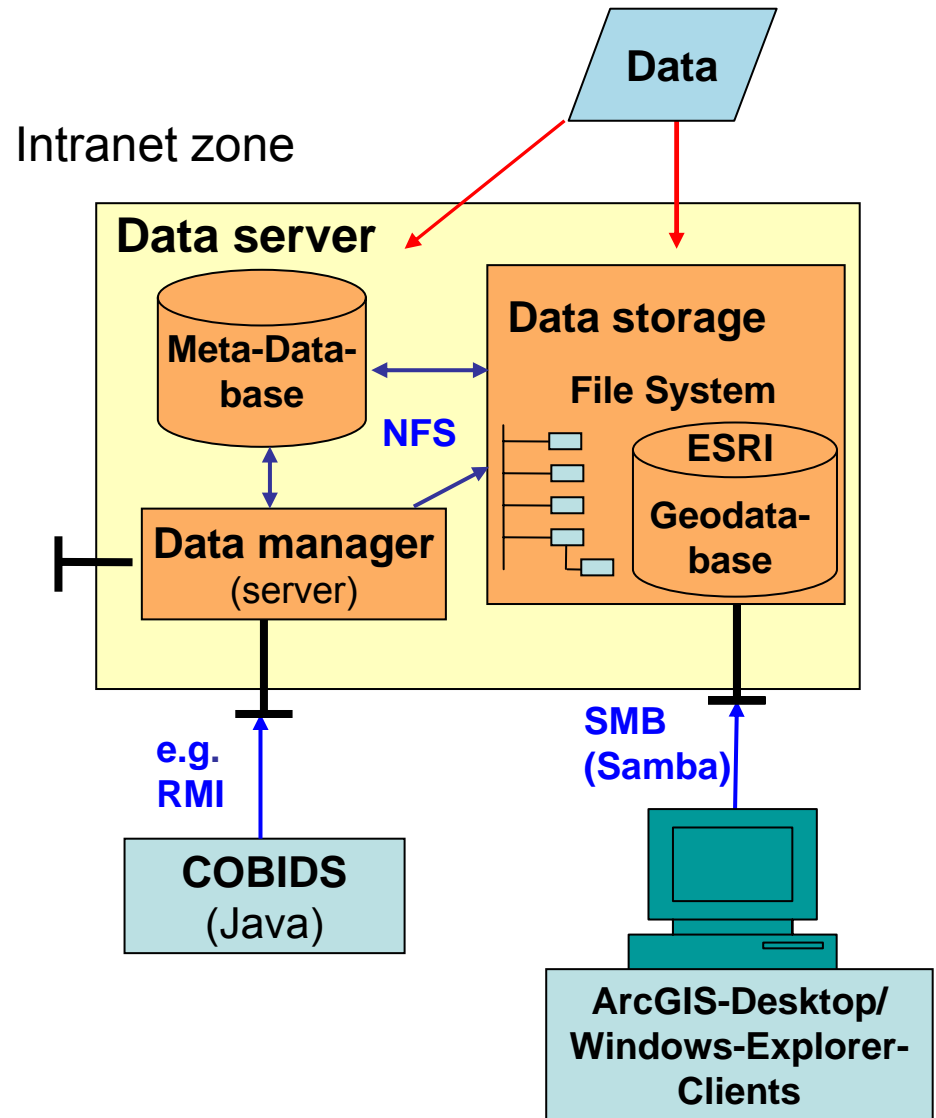
by Antonio Rogmann (ZEFc)



- Data server
 - ◆ Data storage
 - ◆ Meta-Database
 - ◆ Data manager
- Data Collection and Control

- Objectives
- Technology
- Progress
- Next Steps

→ Needs





Data Server



1. Objectives

2. Technology

3. Progress

4. Next Steps

5. Needs

- Well structured storage of data reflecting research areas and needs
Providing
 - ◆ complete overview (as possible) and description of GVP-datastock
 - ◆ information, search, update functionalities
- Direct read/write/search access
 - ◆ as plain files and also from applications (AcrGIS, JAVA)
- Graduated user concept
 - ◆ how is allowed to get/update which data for which use?



Data Server



1. Objectives

2. Technology

3. Progress

4. Next Steps

5. Needs

- Data storage
 - ◆ central storage for **all data**
 - ◆ **file server** with hierarchical directories
 - ◆ **transparent file-based access** (read/write) for Windows and ArcGIS-clients using **SAMBA**

- Filebased Geodatabase (ESRI) for GIS-data (will be available at spring 2007)
 - ◆ **proprietary ESRI** database format for vector/raster/table
 - ◆ uses **data storage** providing additional facilities for **search, management, and consistency control** of geodata
 - ◆ development of high quality data using **database schema** facilities (e.g. relationships)
 - ◆ Direct read/write/search access for **ESRI applications** (ArcGIS)



Data Server



1. Objectives

2. Technology

3. Progress

4. Next Steps

5. Needs

- Meta-database

- ◆ provides **search and cataloging facilities**
- ◆ **M3Cat** - Multistandard, Multilingual, Metadata Cataloguing Tool for the creation of geospatial metadata
- ◆ allows the **use of the Web** for metadata documentation as well as the **use of DBMS** (Access, Oracle) or XML to manage the meta-data

- Data manager

- ◆ **centralized control** of data storage and meta-database
- ◆ **interfaces** (API) for remote JAVA applications and Geoportal



Data Server



1. Objectives
2. Technology
3. Progress
4. Next Steps
5. Needs

- **Technical concept** is developed
- **Data storage**
 - ◆ **directory structure** on file server is implemented
- **ESRI Geodatabase**
 - ◆ **Personal Geodatabase** for basic geodata as test environment is created (locally on windows)
- **Meta-database**
 - ◆ **running M3Cat system** on MS Access database
- **Data manager**
 - ◆ in **conceptual development**



Data Server



1. Objectives
2. Technology
3. Progress
4. Next Steps
5. Needs

- Data manager
 - ◆ development of the **concept**
 - ◆ **implementation** of management software and Java interfaces
- ESRI Geodatabase (when published)
 - ◆ **integration** with the data storage and meta-database
 - ◆ **evaluation and testing** of functionality
 - ◆ **filling** with GVP-geodata from the data storage
 - ◆ staff member and basin-partner **training in database use**
- Meta-database:
 - ◆ change to **new database** after analyzing technical constraints
 - ▶ Native XML database (instead of MS Access) → Unix data server!
 - ◆ **upload of datasets** sorted and controlled by data management team
 - ◆ development of **graduated user** concept



Data Collection and Control



1. Objectives

2. Progress

3. Next Steps

4. Needs

● Objectives:

- ◆ collection and assessment of data
 - ▶ overview of currently given datasets
- ◆ development of the **catalogs structure** on data storage
- ◆ preparing data for data storage
 - ▶ e.g. extraction of meta-information, quality control, reformatting, etc.
- ◆ **versioning** and warranty of data quality
 - ▶ establishing specified control points within the life cycle of data

● Progress:

- ◆ **DB-team** is established: „data sorter“ in communication with data „experts“ (scientists)
- ◆ **procedure of communication** regarding data control is defined
- ◆ **data analysis and sorting** has commenced (GIS-Data)



Data Collection and Control



1. Objectives

2. Progress

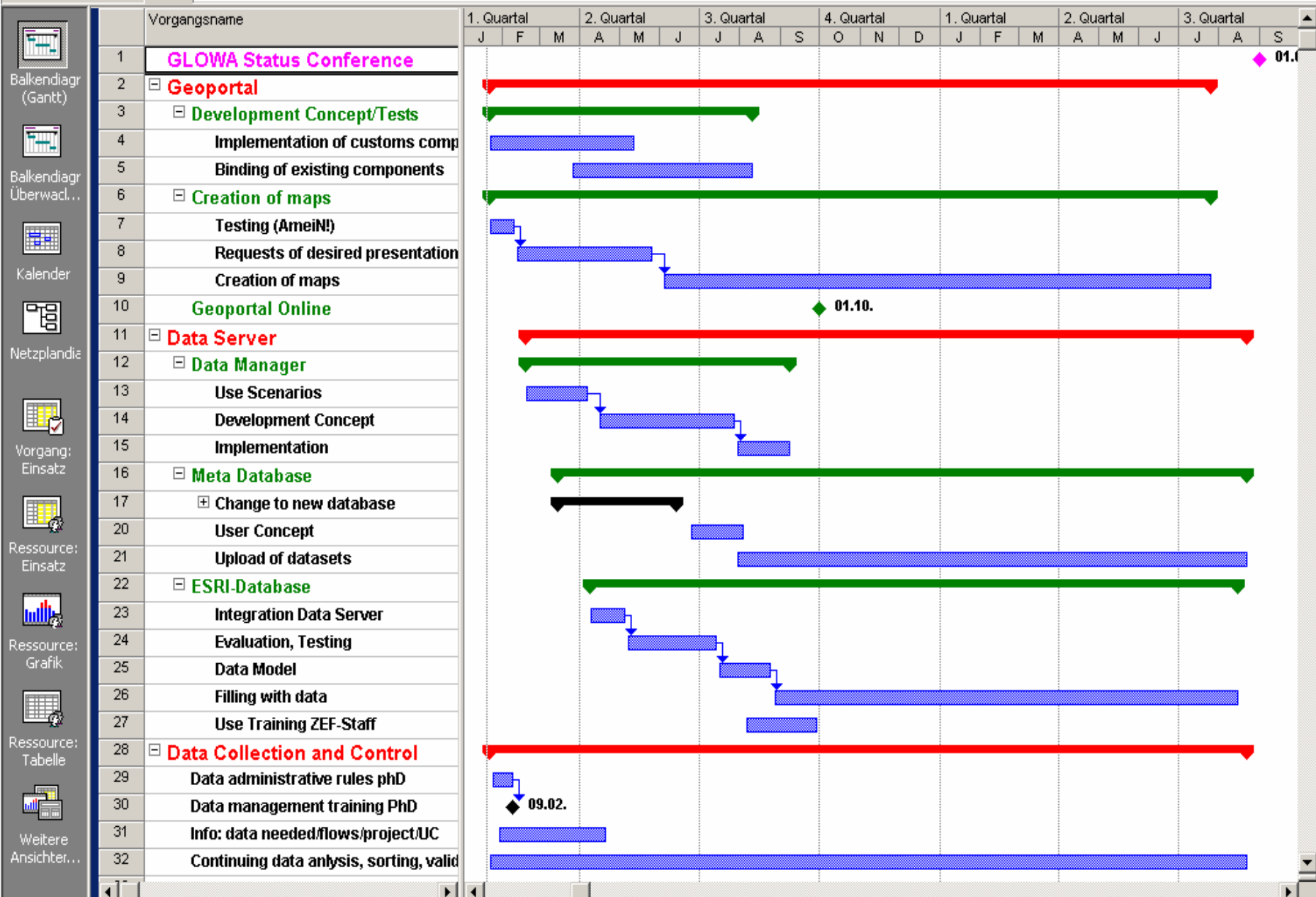
3. Next Steps

4. Needs

- Next Steps:

- ◆ continuing **data analysis, sorting, validation**
- ◆ development of data standards regarding formats, structure and meta data
- ◆ data management training for PhD students (and others)
 - ▶ will be done in february

GLOWA Status Conference





1. Objectives
2. Technology
3. Progress
4. Next Steps
5. Needs

- Data manager
 - ◆ requirements and use scenarios of the database from client applications and users
- ESRI Geodatabase
 - ◆ analyse of ESRI data model
 - ▶ arrangement of datasets reflecting their thematically framework within the GVP (projects, subprojects, use cases, models)
- Meta-database
 - ◆ description of data by GVP researchers (incl. references to projects, uses cases, models)
 - ◆ linkage with other meta data resources (partners)
 - ◆ data classification in terms of publishing and delivery



Data Server



1. Objectives
2. Technology
3. Progress
4. Next Steps
5. Needs/Solutions

- Data manager
 - ◆ requirements and use scenarios
 - ▶ *workshop/consultation with modelers*
- ESRI Geodatabase
 - ◆ data model
 - ▶ *workshop/meetings with senior staff for understanding database modeling and developing of relational data model (if needed)*
- Meta-database
 - ◆ description of data by GVP researchers (incl. references to projects, uses cases, models)
 - ▶ *direct input by dataproducers into metadatabase, in cooperation with data management. Range of necessary metadata elements has to be defined before*
 - ◆ linkage with other meta data resources (partners)
 - ▶ *consultations between GVP-management, partners and data management*
 - ◆ data classification in terms of publishing and delivery
 - ▶ *has to be done by researchers making arrangement with GVP-management*



Data Collection and Control



1. Objectives
2. Progress
3. Next Steps
4. Needs

- Dedicated Staff
 - ◆ research scientists usually have no free time or interest on this work
- Data policies
 - ◆ particularly data access and delivery rights
 - ◆ **formal administrative rules** for data delivery for PhD students, e.g. quality of data, proper description, formats, etc.
- Information about
 - ◆ inventory of existing data / data needs
 - ◆ data relation to the projects / use cases
 - ◆ data flows between research areas / models
 - ◆ processing of data by data manager in priority regarding requirements of research activities in phase III (time is running!)
 - ◆ usage of data after phase III (by whom, for what, by which applications)

Data!!



Data Collection and Control



1. Objectives

2. Progress

3. Next Steps

4. Needs/Solutions

- Dedicated Staff
 - ◆ ***one assistant responsible for getting and uploading metadata (from staff members not available)***

- Data policies
 - ◆ ***have to be discussed, developed and decided with/from the GVP-management***

- Information about
 - ◆ inventory of existing data / data needs
 - ▶ ***data investigation form (excel-sheet) and documentation of data needed***

 - ◆ data relation to the projects / use cases
 - ▶ ***data investigation form (excel-sheet) and documentation***



Data Collection and Control



1. Objectives

2. Progress

3. Next Steps

4. Needs/Solutions

- Information about

- ◆ data flows between research areas / models

- ▶ *diagramms, documentations, consultation by data management*

- ◆ processing of data by data manager in priority regarding requirements of research activities in phase III (time is running!)

- ▶ *decisions by scientists are to be made and documented, consultation by data management*

- ◆ usage of data after phase III (by whom, for what, by which applications)

- ▶ ???