

Geoscience Data: Defining Policies and Workflow Tools for Long-term Storage of Continuously and Temporarily Collected Data

Petra Gebauer¹, Ingo Kirchner¹, Wolfgang Peters-Kottig², Jens Klump³, Roland Bertelmann³, Beate Rusch², Damian Ulbricht³

¹ Free University Berlin, Institute for Meteorology, Berlin, Germany ² Zuse Institute Berlin, Berlin, Germany

³ GFZ German Research Centre for Geosciences, Potsdam, Germany



Commonly Used Workflows

Network of Continuously Measuring Meteorological Stations

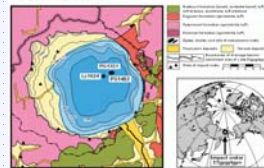
several stations in the city of Berlin



- measuring air temperature, humidity, precipitation, soil surface and ground temperature, at some stations additional wind direction/wind speed, sunshine duration, precipitation, barometric pressure and radiation quantities
- measurement interval is 1 minute

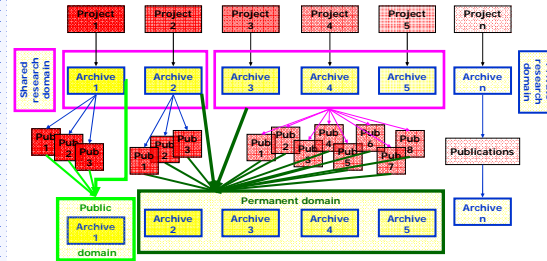
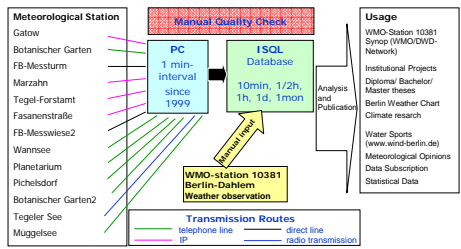
Measuring Campaigns, Temporarily Collected Data

geological information about lake El'gygytgyn

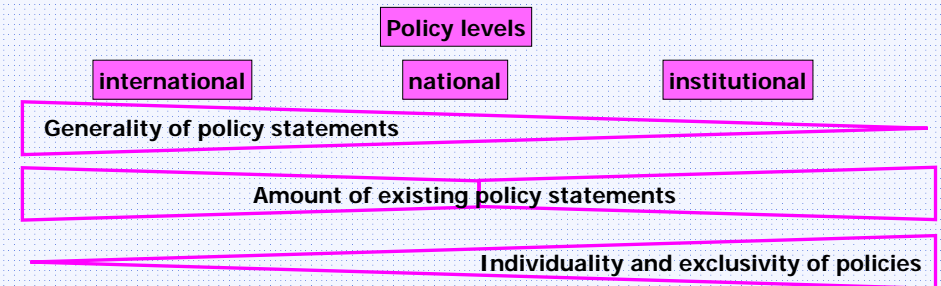


- many different projects
- measuring various geoscientific data
- measurements are variable in space and time
- large amount of data
- no homogenous data format
- international cooperating groups of scientists
- experiments may be repeatable

Data Lifecycle



Policies



Important principles, followed by policies

- good practices for methods, techniques, instruments in collection/archiving data → **Quality**
 - techniques and instruments guarantee the integrity and security of research data → **Security**
 - characteristics of different research fields, legal systems, cultures, regularity regimes
 - information on data through the Internet
 - access to research data for international research community
 - legal requirements on security, privacy, intellectual property rights
 - authorship, usage restrictions, financial arrangements, ethical rules, licensing terms, liability, sustainable archiving
 - international documentation standards
 - avoiding duplication of data collection efforts
 - evaluation of access arrangements by user
 - guarantee long term access to data
- **Flexibility**
→ **Transparency**
→ **Openness**
→ **Legal conformity**
→ **Responsibility**
→ **Interoperability**
→ **Efficiency**
→ **Accountability**
→ **Sustainability**

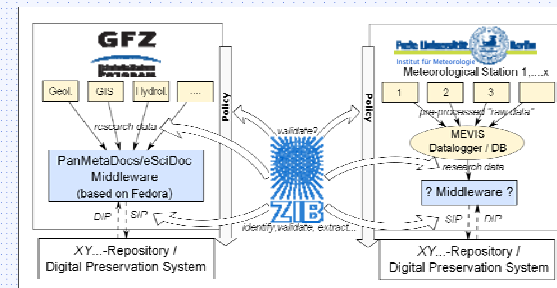
Data Management Needs

- central metadata recording
- automatic quality check programs
- final manual check
- realtime controlled data lifecycle
- defined rights of use
- long-term archive
- private research data at least while project is in progress
- easy transfer into shared research domain
- controlled storage in permanent domain, including metadata and published results
- quality checked data at last in public domain with respect to legal requirements



→ missing institutional policies for long-term storage of data

Aim of the Project



- simplify workflow up to repository
- follow digital preservation policies
- perform tests on the re-usability of produced research data packages
- design and hold an university lecture
- identify remaining gaps
- develop currently missing workflow components