

The Implementation of Policies Regulating Institutional Data Management

Petra Gebauer, Ingo Kirchner, Mirjam Hirt
Free University Berlin, Institute for Meteorology, Berlin, Germany



Institutional Data Management

Initial Situation

Partially knowledge of existing data

Several types of data

Data storage on different media

Designing a questionnaire

Evaluation and needs

Solving problems together

I found some data chronicals in the old archive, a lot of documents!

Jenny documented her analysis last year and also made a list of the data she used.

city network data

climate modeling results

I'll call Tim, perhaps he knows, where we can find any data.

Questionnaire - Data Management at the Institute of Meteorology

1. How many hours a week are you working at the IfM?
2. How would you rate your activity?
3. Have you ever thought about improving your personal data management?
4. How important is data management for you?
5. How much time do you spend on the treatment and processing of data?
6. Which aspects of data management are important to your field of activity?
7. Are you responsible for data that is of institutional interest?
8. Which typical storage time have data that will be processed by?
9. How often do you move data from the workspace to an archive?
10. How long the archived data?
11. Where do you archive your data (outsourcing of data for the purpose of data request)?
12. In what kind of format is the data you work with regularly stored?
13. In what quality levels data will be stored by you?
14. Do you deliver other groups regularly data?
15. How much time is required for processing of data, which will be used by others?
16. How do you document your data?
17. Which aspects of institutional data management are important?
18. How do you evaluate the current data management at the IfM?
19. At which point would you take action to improve the institutional data management?
20. Is it a requirement for the data users subject to research projects useful?
21. Should there be a regular training for data management at the IfM?
22. Which external data portals do you know and use? Explain why?

Evaluation of Data Management activities for Meteorology

Activity	Score
memory data	1.0
usability data access	2.0
support	2.2
documentation	2.8
metadata	3.0
training	3.1

What should be done first to improve data management activities for Meteorology?

Activity	Score
memory data	1.0
usability data access	2.0
support	2.2
documentation	2.8
metadata	3.0
training	3.1

Who? When?

Where? How? What?

Policies to Regulate Data Workflow, Use and Archiving

- The **City Measuring Network Data Policy (CMNDP)** addresses issues related to the provision, exchange, availability, maintenance, use and archiving of data produced by the city measuring network of the Institute for Meteorology, Free University Berlin.
- The **purpose** of CMNDP is to set up fundamental principles in view of
 - regulating the **data workflow** (starting from measuring up to archiving the data)
 - ensuring **timely submission** of data for the use within the institute
 - easing **collaboration** among scientific community using the data
 - ensuring **preservation** of the data and the **metadata** for the future

- The CMNDP ensures
 - protecting the rights of the IfM on the data
 - providing rules for use of the data within the IfM and by third parties
 - providing easy access to the data
- Data are always being regarded as a combination of the measured data and its describing metadata.
- Metadata have to be provided by the data originators for all kinds of data. Available standards are to be considered.

- Different **data levels** depending on the degree of underlying **pre-processing and evaluation** steps are defined and with **quality flags** assigned:
 - Level 0: Original **raw data** (from the data logger, 1 min values)
 - Level 1: Automatically **quality checked data**
 - Level 2: **Reviewed and formatted data**
 - Level 3: **Gap-filled, derived, temporally aggregated data**
- The **access** to the data is restricted to certain users, depending on the different levels.
- The different **data levels** should be **available as soon as possible** within the institutional community.

- Delivery of data to external users** is handled through cooperation agreements or based on fixed charges depending on the data volume.
- Scientists, publishing data from the city measuring network or derived results, have to **cite the origin of the data**.
- All different levels of data will be kept in an **institutional repository**.
- The CMNDP regulates the **conditions of use**. The user of any data agrees:
 - not to disclose the data to **other parties**
 - not to use the data for **commercial purposes**
- All data (original measurements, different levels of processed data, metadata) will be stored in a **long term archive**.

Data Workflow from Measuring Station of the City Measuring Network up to Long Term Archive

I can use measurements from the city measuring network for my studies!

quality check

metadata/documentation

Documentation of the city measuring network at the Institute of Meteorology

- Addresses, coordinates and height of the stations in the city measuring network
- Description of the surroundings of each station
- Measured meteorological parameter at each station
- Measuring instruments (manufacturer, sensitivity, ...)
- Start and end of the measurements
- Transmission way and time (Modem, IP, ...)
- Organization of the data collecting system (MEVIS)

IfM repository

- CMN - 1 min
- Level 0 - original raw data
- Level 1 - quality checked data
- Level 2 - reviewed and formatted data
- Level 3 - gap-filled, derived, spatially and/or temporally aggregated data
- directories, data base, GUI metadata

It is easy to find information beside the data!

long term archive

city measuring network data 1 min

Cool, we learn it already during our study period!

Lectures on data management

Cooperation Partners Project EWIG

