

Why Data Management ?

Urs Beyerle - IAC/C2SM (ETH Zürich)

urs.beyerle@env.ethz.ch

Just because ...



Data Management helps you to ...

- keep your data organized
 - easy to find
 - easy to reproduce
 - easy to publish
 - easy to archive
- have a disaster recovery plan
 - backup
 - rolling back to previous versions

Your Data Management helps us to ...

- save storage costs
- (re-)use your
 - data
 - methods (scripts, programs, etc.)
 - results
- simplify long term storage (publishing / archiving)

Your data is very valuable, but without **metadata** useless !

What is Metadata?

- Metadata is, literally, «data about data» (e.g. README.txt)
 - Who created it?
 - What is it?
 - When was it created?
 - How was it generated?
 - Which raw data was used?
 - How may it be used?
 - Are there restrictions on it?

Take-Home Message

- Design a logical folder structure and use meaningful filenames
- Document each step from raw data to your final results
- Metadata is the key – it provides important information about the content of the data and how it was generated
- Add metadata as early as possible
- Prevent data loss
- Use a version control system

What's next

- Best Practices for your Thesis – DM in the Dynamic Meteorology Group
(Michael Sprenger)
- Data Workflow: From Raw Data to Archiving
(Sylvaine Ferrachat)
- Version Control System: Introduction to GIT
(Colombe Siegenthaler)