

# Data projects at IFM

nowadays and suggestion of new structure

Daniel Befort

December 17, 2012

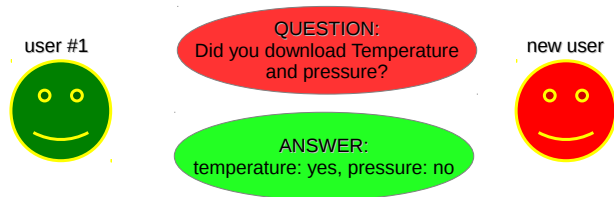
# Structure today

## Start of new project



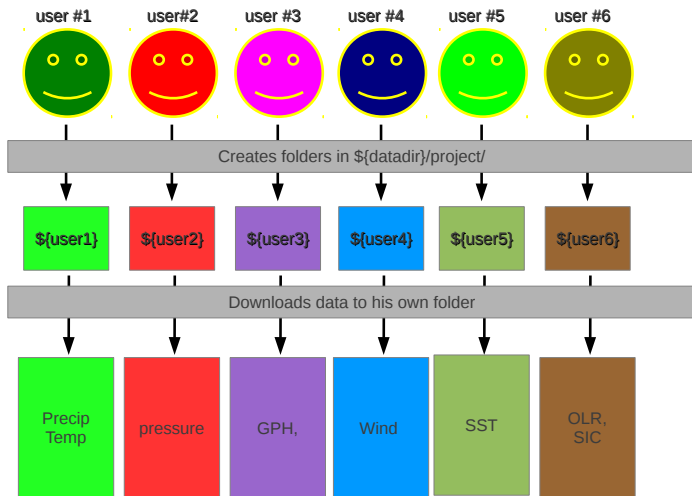
# How it should work

**new user** asks **existing users** if data is already downloaded



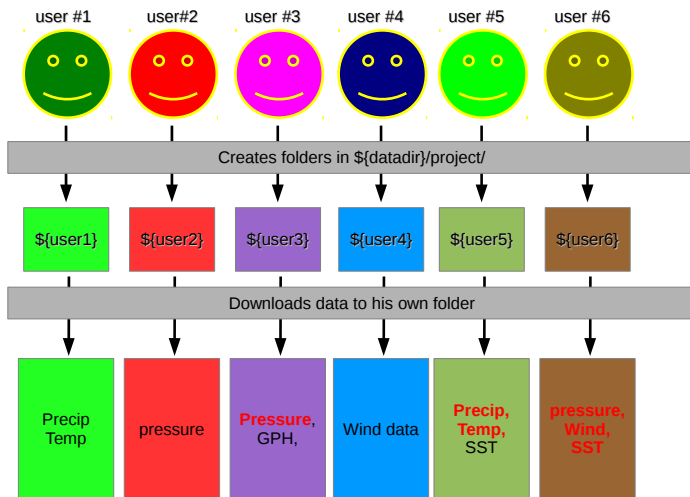
⇒ **new user** downloads only new data

# How it should be



⇒ each variable of each dataset only archived **ONCE**

# How it is



⇒ many variables are stored two, three or more times at IFM

# some problems with old structure

- 1 too many users with too many folders
- 2 new users don't know old users and don't know their structure
- 3 no README and sometimes problems with permissions of existing folders (only readable for some users)
- 4 no administrator who knows the structure of the project

This often ends up that a new user downloads data which is already available

It would be best to have a database like CERA ...  
...**BUT** this is not possible (human resources)

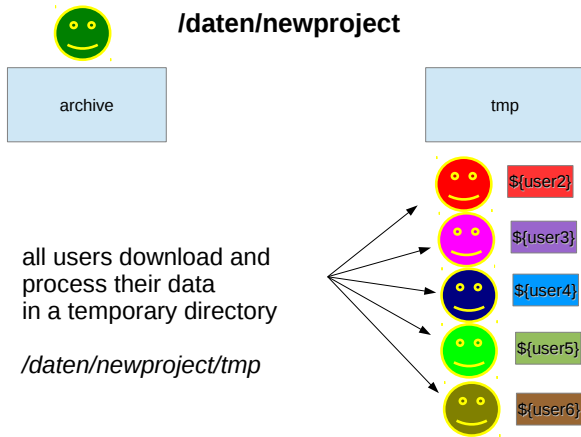
What would be desirable?

- all *global* data should be documented (README)
- new users should easily find old data
- one project administrator?

## **Problem:**

It is not possible that one persons downloads and archives all data of one dataset (e.g. ERA40) for the whole institute

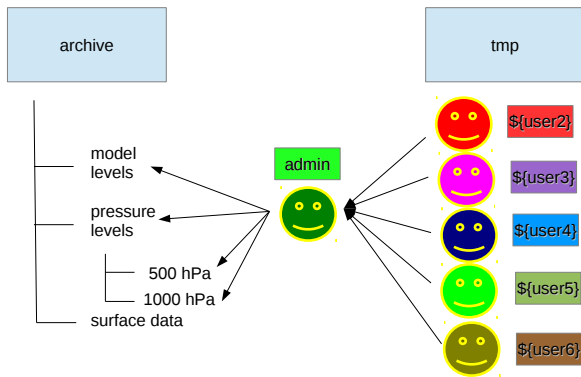
# Suggestion of new structure





# Suggestion of new structure

## `/daten/newproject`



admin gathers and archives data in  
`/daten/newproject/archive`

# Workflow

- 1 **users** download and process data in *tmp* directory
- 2 after finishing they ask **admin** to archive data
- 3 **admin** should ask **user** for README (otherwise no archiving)
- 4 **admin** adds data to *archive*
- 5 **users** delete their data from *tmp* after checking data from *archive*

# Workflow

- 1 **users** download and process data in *tmp* directory
- 2 after finishing they ask **admin** to archive data
- 3 **admin** should ask **user** for README (otherwise no archiving)
- 4 **admin** adds data to *archive*
- 5 **users** delete their data from *tmp* after checking data from *archive*

## Advantages:

- project admin knows structure of project
- archive is only writable for admin (not for users)
- new users can easily find archived data (because only one folder for all data and not for each user)
- no double entries of variables

# Workflow

- 1 **users** download and process data in *tmp* directory
- 2 after finishing they ask **admin** to archive data
- 3 **admin** should ask **user** for README (otherwise no archiving)
- 4 **admin** adds data to *archive*
- 5 **users** delete their data from *tmp* after checking data from *archive*

## Advantages:

- project admin knows structure of project
- archive is only writable for admin (not for users)
- new users can easily find archived data (because only one folder for all data and not for each user)
- no double entries of variables

## Disadvantages:

- more work for project admin