

Creating a common data management at MPIC for your data



David Walter & Hartwig Harder (Architecture of the project folder structure)
Bärbel Sarbas (MPIC data curator)
Thomas Disper (IT infrastructure)

Contact: david.walter@mpic.de
All: Max Planck Institute for Chemistry

MAX PLANCK INSTITUTE
FOR CHEMISTRY

Further involved: M. Pöhlker, T. Könnemann, T. Klimach, F. Ditas, M. Peißker, C. Pöhlker, U. Pöschl

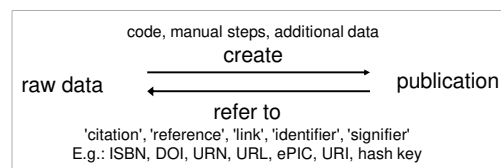
MOTIVATION

Demand by publishers

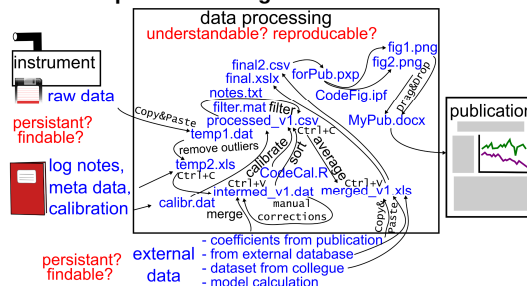
- "Data underpinning any research finding should be **findable, accessible, interoperable, and reusable** (FAIR) – not only for humans, but also for machines." [from: Copernicus Publications, 2018-11-05, "https://www.copernicus.org/news_and_press/2018-11-05_enabling-fair-data-commitment-statement.html", accessed on 2018-11-19]
- "Therefore, Copernicus Publications requests depositing data that correspond to journal articles in reliable (public) data repositories, assigning digital object identifiers, and properly citing data sets as individual contributions." [from ACP Data policy, "https://www.atmospheric-chemistry-and-physics.net/about/data_policy.html", accessed on 2018-11-19]

Claim

It should be possible to reconstruct published results based on the raw data, even if the author of the publication can not be asked any more.



How data are processed - e.g. like that?



Strategies

- Make data management plan (ideally before you start the measurements)
- Uniform, easy document structure
- Proper documentation
 - code
 - intermediate steps
 - additional (external) data
- Make data citeable and persistent

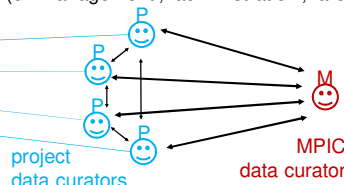
MPIC DATA

Places where you find data at MPIC

- Logbooks on paper
- MPIC file server
 - + \\mpic.de\file
 - + ftp
 - + groups
 - + AG_Muster
 - + homes
 - + f.mustermann
 - + projects
 - + AQABA
 - + ATTO
 - + CAFE
 - + OMO
 - + Template
- Laptops, hard drives etc.

Data curation at MPIC - project folders

Q: What happens with the data ...
- while project ongoing (regular input of new data)?
- after project is finished?
- after PI leaves the institute?
→ Concept of 'data curation' (or 'management', 'administration', 'archiving')



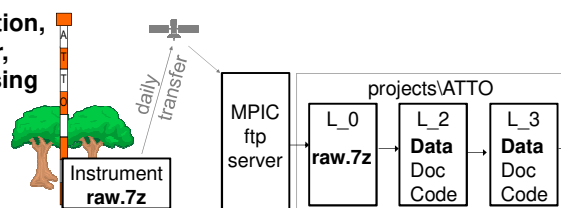
External places

- Papers and supplements
- Open access repositories
 - EDMOND (https://edmond.mpdl.mpg.de)
- Closed access storage
 - KEEPER (https://edmond.mpdl.mpg.de)
- Databases
 - ATTO data portal

ATTO

Data

Acquisition, Transfer, Processing



ATTO Data Portal

Goals: Data exchange within ATTO community, publication of ATTO data
Type: Relational database, hosted at GWDG, maintained by MPI-BGC, software: BEXIS2
First step: Upload of L_3-data, i.e. 30-minute averages



ATTO Folder structure

```
\\mpic.de\file\projects\ATTO\
+ FinalDataShortcuts
+ GeneralOverview
+ INST
+ _F_Definition
+ _Labbook
+ _Photos
+ CCN
+ CeiloCHM15kx
+ CRM
+ ExampleInst
+ MAX_DOAS
+ MicroClim
+ Pic21
+ Pic89
+ Meetings
+ Organisation
+ SuppData
+ Topics
```

data not obtained at ATTO, but used for analysis and interpretation

Scientific topics based on cross-instrument analysis

Instrument folders

```
\\mpic.de\file\projects\ATTO\INST\ExampleInst\
+ Doc
+ L_0
+ CalData
+ Data
+ L_2
+ F_0
+ Code
+ Data
+ Doc
+ L_3
+ db30
+ Code
+ Data
+ Doc
+ Labbook
+ Preliminary_0_2
+ SharedCode
+ Tools
```

raw data

processed on native time grid

processed on equidistant time grid

General Overview

Data for calibration

Measurement data
(In case of ATTO, a folder named "dtr" is used for daily transferred files.)

Name of data product type, "F_0" for unfiltered standard product *

Code used for creating files in L_2\Data

Human readable description: how content of L_2\Data has been created

Name of data product type, "db30" for 30 minute average standard product *

Preliminary data, recoverable intermediate steps for data processing from L_0 to L_2

colors:
red: fix folder names
blue: variable folder names

* For filtered products: E.g. "F_NoContamination", with additional subfolder named "Filter"

Further info: \\mpic.de\file\projects\Template\ProjectsQuickStart.pdf

And now: Your data

If not already done: Please upload your data (including documentation and code) to the project folder(s) of your project(s)!
How? Which project? → Please contact us!