

Interoperability/Web service experiences from Martin Schultz

- A loose collection of issues and proposals from the user and provider perspectives

- **Licenses: EUDAT B2SHARE**
- **DOI: TOAR repository at PANGAEA**
- **Data and metadata collection and harmonisation**
- **Data format**
- **Web interfaces**
- **REST services**

Licenses

<https://b2share.fz-juelich.de/> - a data publishing service



Allows specification of license under which dataset (or code) shall be published (mostly CC):

Public Domain Dedication (CC Zero)

CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

Publicly Available  

Creative Commons Attribution (CC-BY)

This is the standard creative commons license that gives others maximum freedom to do what they want with your work.

Publicly Available  

Creative Commons Attribution-ShareAlike (CC-BY-SA)

This creative commons license is very similar to the regular Attribution license, but requires you to release all derivative works under this same license.

Publicly Available  

Creative Commons Attribution-NoDerivs (CC-BY-ND)

The no derivatives creative commons license is straightforward; you can take a work released under this license and re-distribute it but you cannot change it.

Publicly Available 

Creative Commons Attribution-NonCommercial (CC-BY-NC)

A creative commons license that bans commercial use.

Publicly Available 

Creative Commons Attribution-NonCommercial-ShareAlike (CC-BY-NC-SA)

A creative commons license that bans commercial use and requires you to release any modified works under this license.

Publicly Available 

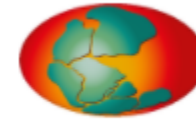
Creative Commons Attribution-NonCommercial-NoDerivs (CC-BY-NC-ND)

The most restrictive creative commons license. This only allows people to download and share your work for no commercial gain and for no other purposes.

Publicly Available 

DOIs

<https://doi.pangaea.de/10.1594/PANGAEA.876108>



PANGAEA.

Data Publisher for Earth & Environmental Science

One doi for collection (all data and software),
And one doi for each subset:

Datasets listed in this Collection

1. **Schultz, MG; Schröder, S; Lyapina, O et al. (2017):** All pre-compiled metrics data sets, link to zip archive. <https://doi.org/10.1594/PANGAEA.876110>
2. **Schultz, MG; Schröder, S; Lyapina, O et al. (2017):** Graphical products of TOAR ozone metrics, link to files in PNG format. <https://doi.org/10.1594/PANGAEA.876109>
3. **Schultz, MG; Schröder, S; Lyapina, O et al. (2017):** Pre-compiled metrics data sets, links to aggregated statistics files in CSV format. <https://doi.org/10.1594/PANGAEA.880503>
4. **Schultz, MG; Schröder, S; Lyapina, O et al. (2017):** Pre-compiled metrics data sets, links to gridded files in NetCDF format. <https://doi.org/10.1594/PANGAEA.880506>
5. **Schultz, MG; Schröder, S; Lyapina, O et al. (2017):** Pre-compiled metrics data sets, links to trend statistics files in CSV format. <https://doi.org/10.1594/PANGAEA.880504>
6. **Schultz, MG; Schröder, S; Lyapina, O et al. (2017):** Pre-compiled metrics data sets, links to yearly statistics files in CSV format. <https://doi.org/10.1594/PANGAEA.880505>
7. **Schultz, MG; Schröder, S; Lyapina, O et al. (2017):** Tropospheric Ozone Assessment Report, link to software tools. <https://doi.org/10.1594/PANGAEA.876640>

Example: gridded files are available under

<https://doi.pangaea.de/10.1594/PANGAEA.880506?format=html#download>

I am curious to see how this will be used...

Data and metadata collection and harmonisation

- Too many different data formats
- Very different level of detail in terms of metadata
- Frequently wrong or imprecise station coordinates
- Especially problematic is altitude information; verification with google maps API was very helpful, but instances where it didn't give true answer, e.g. tower measurements; generally no information on relative height of inlet available
- Sometimes a lot of metadata available, but takes time to find your way through many different web services and documents; not all online (Example: US EPA approved measurement devices)
- TOAR data submissions and downloaded files contained ~50 different spellings for „UV absorption“

Data formats

We had very good experiences with the simple TOAR data file format
(but people got even this wrong at times)

```
# metadata_key: value      (for each metadata item)
time; value; flag
2010-01-01 00:00; 34.65; 0
2010-01-01 01:00; 38.21; 0
...
```

Modellers like netCDF (but they also like gridded data)

For web services, we definitively recommend (Geo)JSON.

Web interfaces

Make it fast and simple!

It is quite challenging to make it fast.

It is even more challenging to make it simple (yet powerful).

REST services

While collecting TOAR data and metadata we often wished that there were more web services which would provide all the metadata we needed

→ try to make all controlled vocabulary items available via REST services

Ambiguities how to design REST URLs (e.g. value ranges and repetitions; separate URL or parameter in URL?):

Example: `.../search/?instrument=Dasibi` or `.../search/instrument/?name=Dasibi`

→ Within GAW this should be harmonized.

Similarly, ambiguities how to code JSON responses (dictionaries, lists, ...)

→ see <https://join.fz-juelich.de/services/rest/surfacedata/>

(technically no problem as long as documented, but multiplies effort if not standardized)