



THE TOAR DATABASE: METADATA HARMONIZATION AND DATA QUALITY ASSURANCE ON GLOBAL AIR QUALITY DATA

EOSC Symposium 2021, Semantic interoperability session, 18 June 2021

Niklas Selke, Sabine Schröder, Mathilde Romberg, Jessica Ahring, Lukas Leufen, Sander Apweiler, Martin Schultz
Jülich Supercomputing Centre

THE TOAR DATABASE

- For the Tropospheric Ozone Assessment Report (TOAR) a database with freely accessible surface ozone data from measurement stations around the world was created

- ~13,000 stations
- ~104,000 time series
- Time series record data for ozone, ozone precursors and meteorological data (temperature, ...)
- Accessible through REST API

Contributions from:



... and many individual contributors



Metadata come in varying degrees of completeness and with varying identifiers

CONTROLLED VOCABULARY & ONTOLOGY

- By utilizing a controlled vocabulary, we can define a specific set of identifiers that is used consistently in the TOAR database
- An ontology describes the defined controlled vocabulary
- Ontology and controlled vocabulary can be accessed through the REST API

Example

Controlled Vocabulary

Station Type:

- Unknown
- Background
- Traffic
- Industrial

Ontology

StationMeta

└ StationType

└─ Unknown

└─ Background

└─ Traffic

└─ Industrial

ADDITIONAL METADATA

- Even with ontology and controlled vocabulary some data provider will have metadata fields that have no representation in our defined language
- Not to lose any data given to us, we added an additional field which saves any extra metadata
- This extra metadata will end up saved in a JSON field

Example

Given:

station_id: ...
station_name: ...

predefined keywords

inlet_tube_material: ...
flow_rate: ...

arbitrary keywords

Result:

station_id: ...
station_name: ...
additional_metadata:
{‘inlet_tube_material’: ..., ‘flow_rate’: ...}

VERSIONING

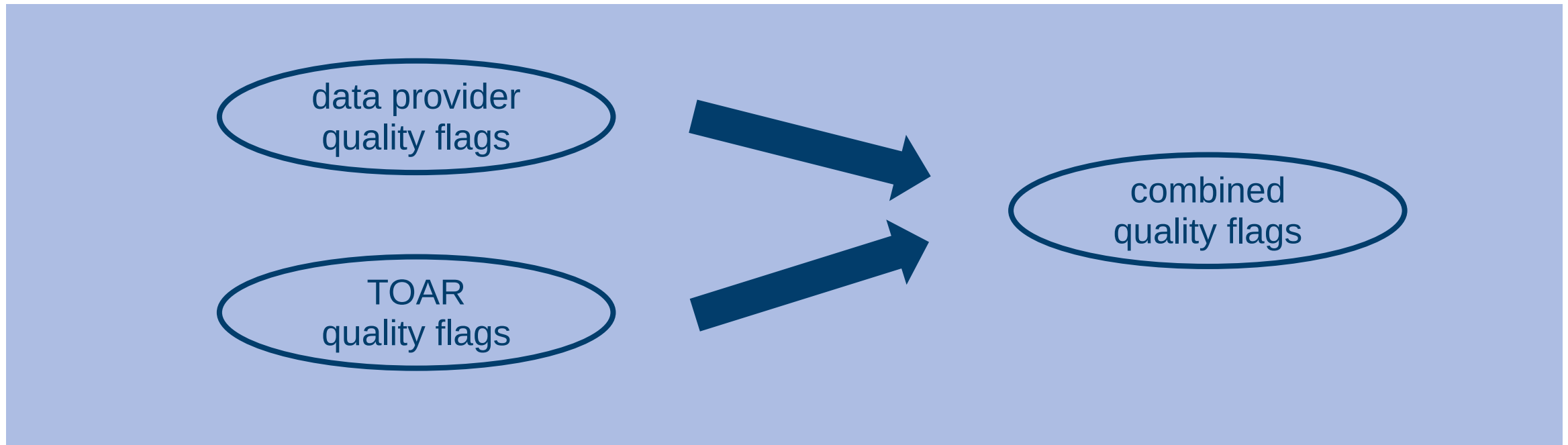
- Database is versioning the data
- When revised data is retrieved the necessary changes in the data and/or metadata are made and get a new version number
- Allows for accessing data in a version at a specific moment in time even if the data subsequently does change

Example



DATA QUALITY ASSURANCE

- Data provider will normally provide their own quality flags for the data
- For consistency in the TOAR database we run our own data quality checks
- Combining these two flags will create a new flag that is saved in the database



CONCLUSION

