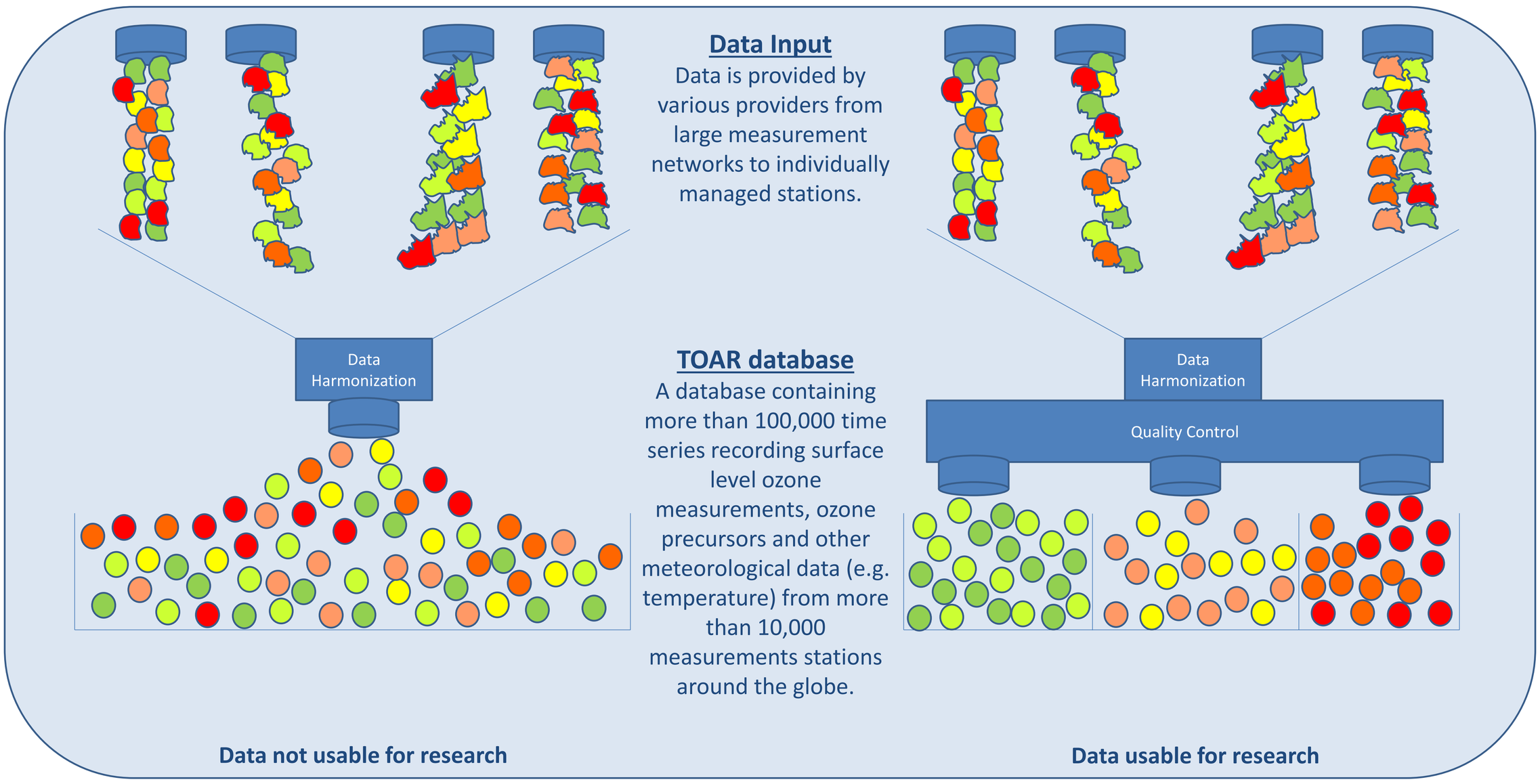
The TOAR database: data harmonization and quality assurance on global air quality data

Niklas Selke, Sabine Schröder, Martin Schultz
Contact: n.selke@fz-juelich.de Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich GmbH, Germany



Properties of the quality control tool

- Modular structure → allows for easily swapping out parts of the workflow
- Statistical tests are used with a class factory pattern → addition of new tests is possible without changing existing code

Workflow

Data is sent with quality flags from the data provider

↓

Metadata is harmonized

↓

Data is handed over to the quality control tool

↓

Statistical tests are run to determine the quality of all given data points

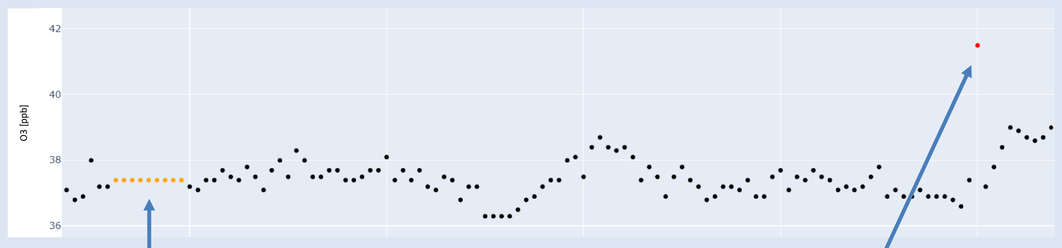
↓

Quality flags from the tool and the data provider are combined to get the final quality flag for each data point

↓

Save the quality flag alongside the data points in the TOAR database

Results



The same value was measured over a couple of hours → not directly flagged as erroneous but at least it should be investigated

This value is flagged as erroneous as it deviates from the surrounding data points

	1981	1982	1983	1984	1985	1986	1987
OK	7546 (95.90%)	7533 (98.53%)	8240 (97.60%)	8114 (95.31%)	7731 (96.52%)	8398 (97.58%)	4788 (73.26%)
Questionable	313 (3.98%)	102 (1.33%)	199 (2.36%)	311 (3.65%)	207 (2.58%)	190 (2.21%)	670 (10.25%)
Erroneous	10 (0.13%)	10 (0.13%)	4 (0.05%)	88 (1.03%)	72 (0.90%)	18 (0.21%)	1078 (16.49%)
Total	7869 (100.00%)	7645 (100.00%)	8443 (100.00%)	8513 (100.00%)	8010 (100.00%)	8606 (100.00%)	6536 (100.00%)

The quality control tool will also give an overview in form of a table to easily assess the data quality of a given measurement campaign for the different years covered.