



Demo
dbrepo.ossdip.at



Project Website
dbrepo-docs.ossdip.at

Problem Description

In many institutional settings, databases are set up **locally** at the level of labs or individual research teams, maintained by some **IT-savvy researcher** that takes care of data management tasks on top of performing actual research. This leads to several **challenges**:

- There are no repository systems for live research data held in databases
- There is no separation of concerns, i.e. researchers need to provide end-to-end support for data management themselves
- Emerging requirements towards reproducibility, FAIRness of research data, digital preservation increase the amount of non-research related activities to be addressed

Methodology

To address these problems we complement a database system with features offered by a **repository**, e.g.:

- Databases in individual Docker containers
- **Data versioning** to provide reproducible results
- Citing subsets of data via **Query Store**
- Docker deployment to provide scalability & flexibility

Database Repository for Research Data

We present DBRepo, an **open-source** database repository infrastructure for research data with machine-actionable interfaces (HTTP, AMQP, JDBC) for efficient reuse of data in services (i.e. Grafana, Jupyterhub).

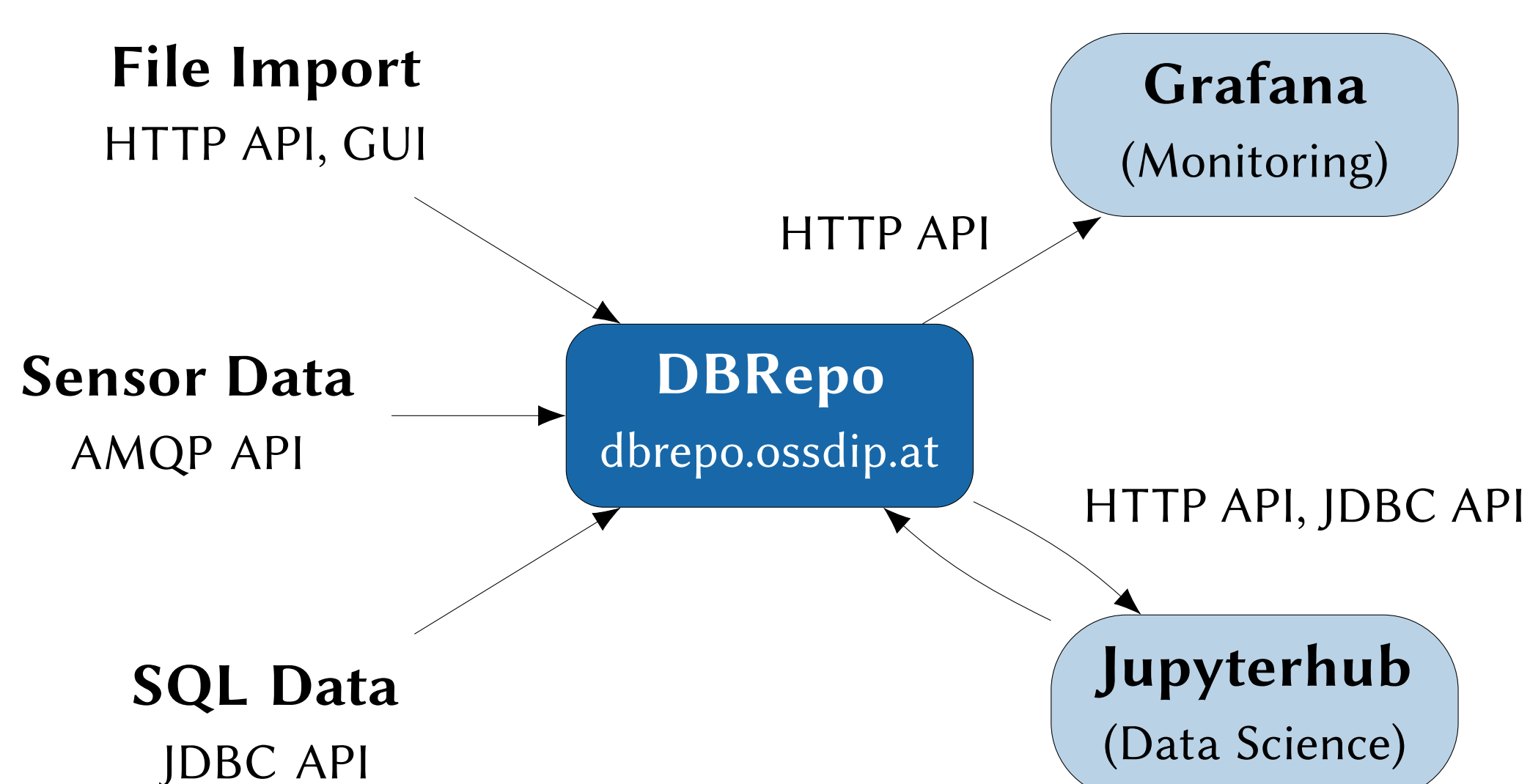


Figure 1: DBRepo offers APIs to ingest data and exposes them

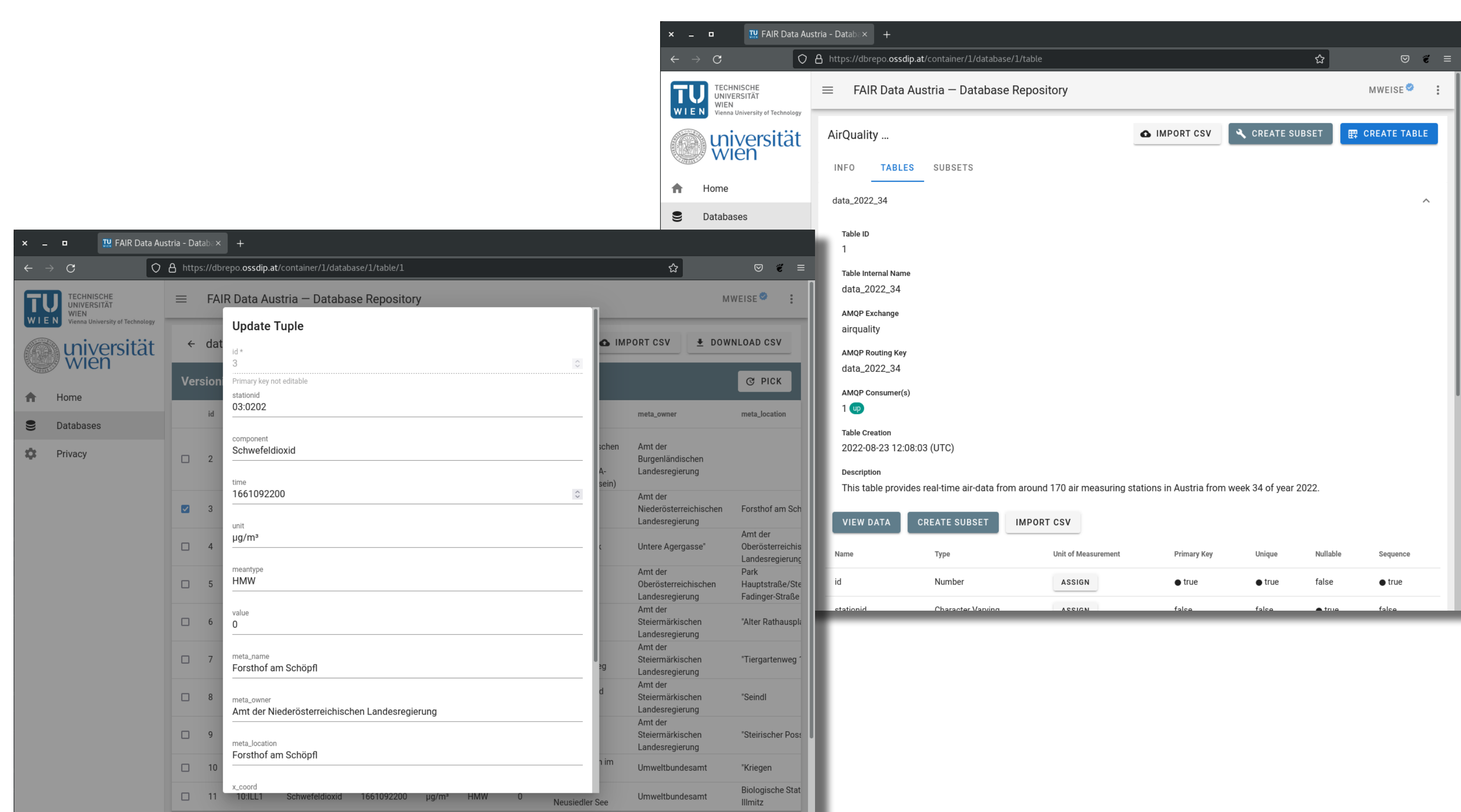


Figure 2: DBRepo's data explorer allows inline value modifications (left) and assignment of units of measurement to columns (right)

Support Research Activities

Assist in depositing the data at a managed-, central infrastructure via APIs or graphically (GUI):

- Deposit of data via APIs or GUI at **different levels** of database knowledge (novice, expert)
- **Interaction** with the data in the repository (data explorer, queries) or external analytic/monitoring tools (e.g. Grafana, Jupyterhub)
- Persistent **identification of arbitrary subsets** according to the recommendations of the RDA WG on Dynamic Data Citation

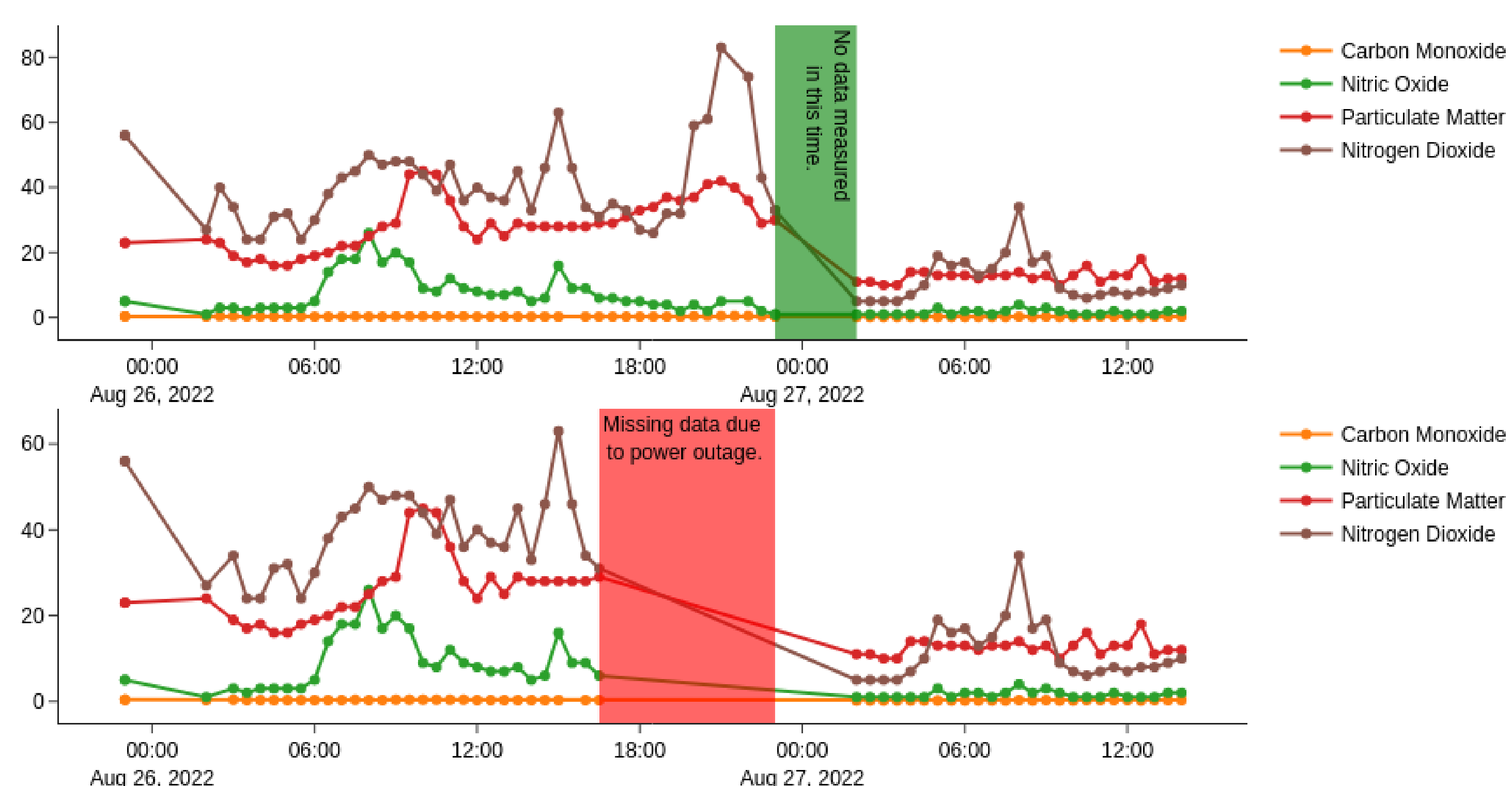


Figure 3: Air pollution measurements added via AMQP API into DBRepo, missing data can be added retrospectively

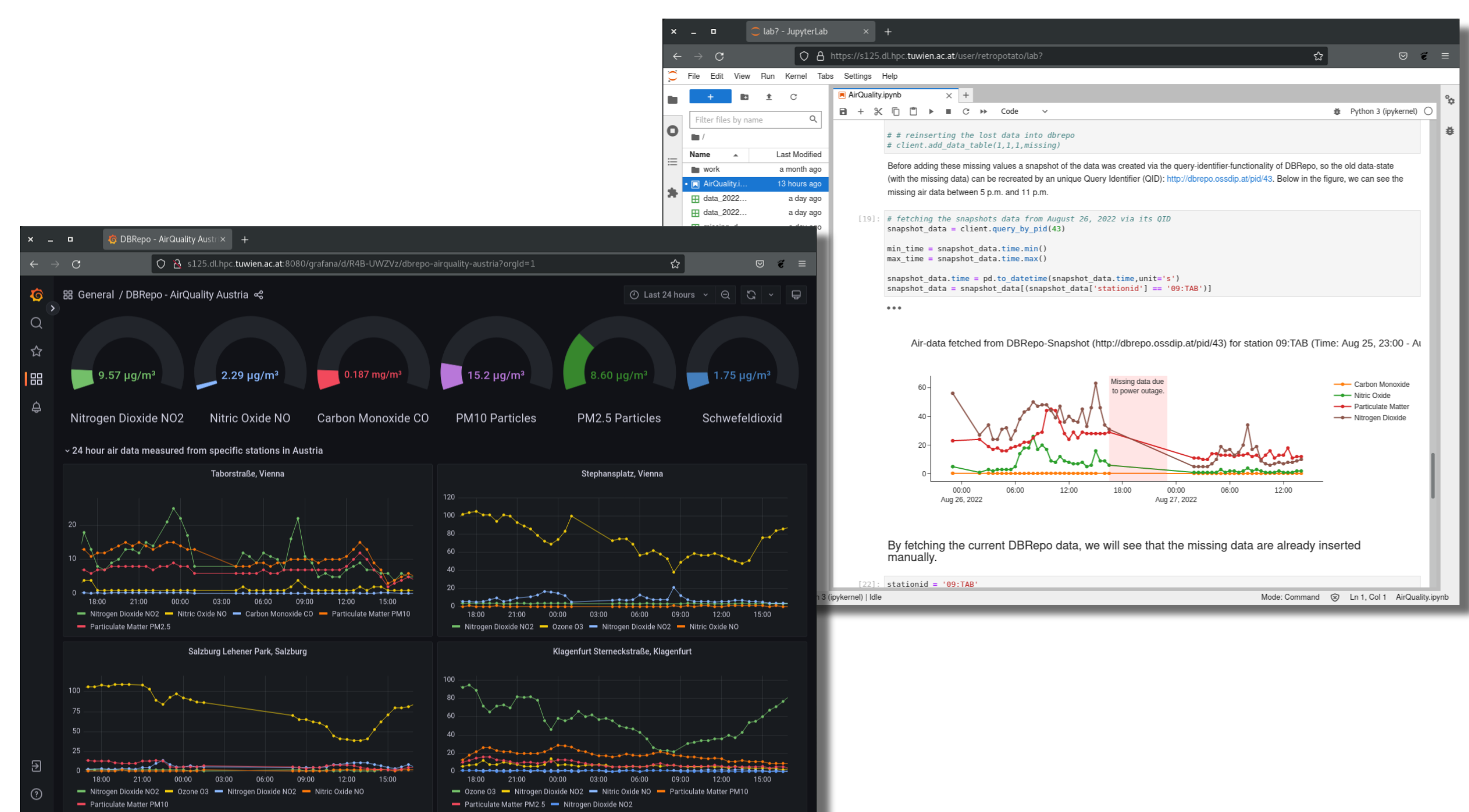


Figure 4: External analytic/monitoring tools connected to DBRepo

Conclusions

The database repository offers researchers **multiple APIs** to work with their data while **making it FAIR** and removing the burden to manage the databases.

References

- [1] Sourcecode Repository. <https://gitlab.phaidra.org/fair-data-austria-db-repository/fda-services>.
- [2] M. Weise, M. Staudinger, Michlits, E. Gergely, K. Stytsenko, R. Ganguly, and A. Rauber. DBRepo: A Semantic Digital Repository for Relational Databases. In *Proceedings of the 17th International Digital Curation Conference, 2022*. doi:10.5281/zenodo.6637333.