Repository and Compute Environment for Sensitive Data

Research Data Management

Dipl.-Ing. Martin Weise
Technische Universität Wien, Austria
martin.weise@tuwien.ac.at

Technische Universität Wien

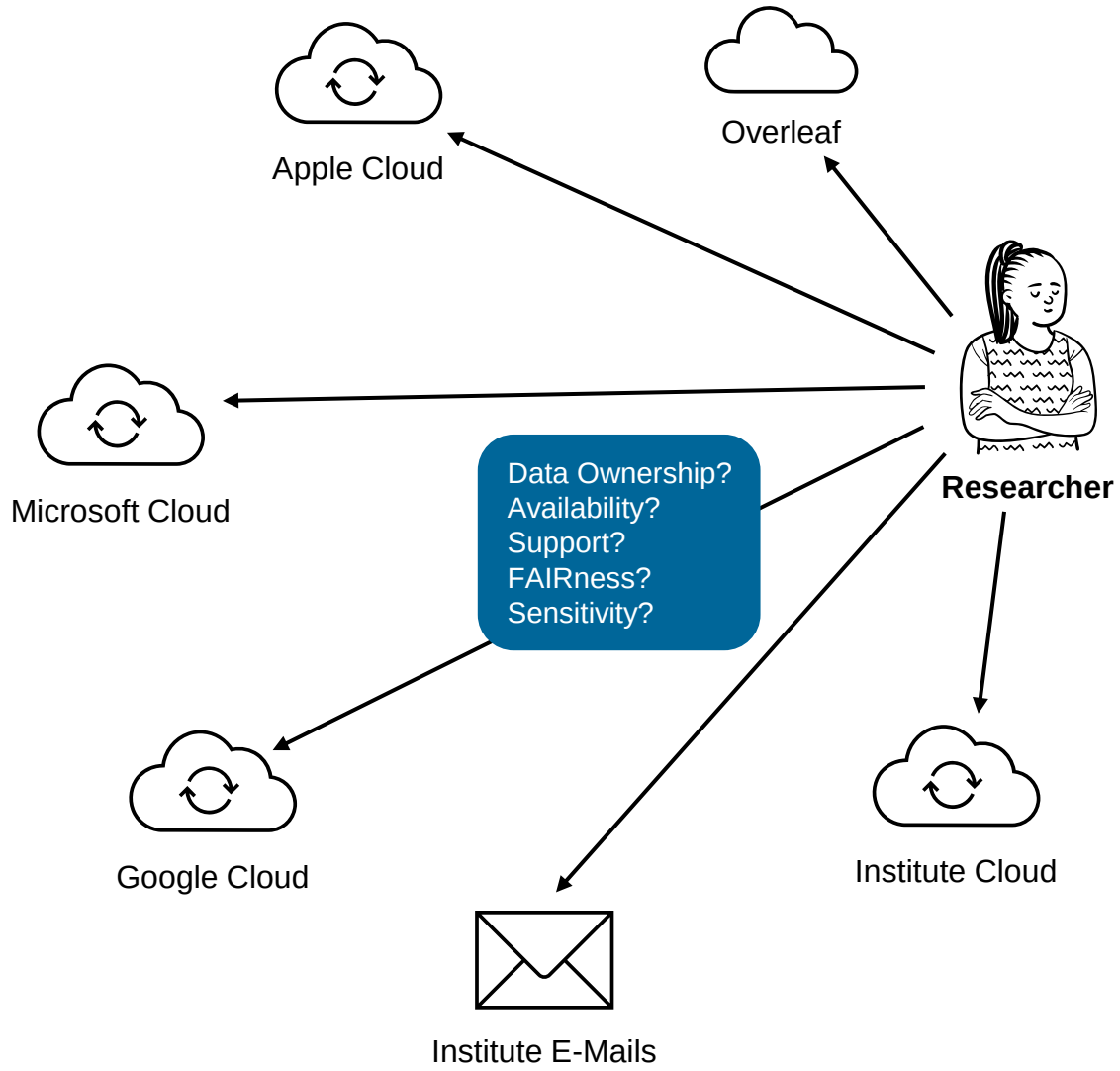
Numbers

- 5.406 employees (76% scientific)
- 26.358 students
- 8 faculties
- 51 institutes

Infrastructure

- Vienna Scientific Cluster
  VSC-5: CPU 2.31 (PFlop/s)
  GPU 1.17 (PFlop/s)
- TU.it / ADLS

Common Data Management



TECHNISCHE UNIVERSITÄT WIEN
Vienna | Austria

5. HANDLING RESEARCH DATA

Research data should from the beginning be stored and maintained in appropriate systems and made available for use in a suitable repository (see 6.1. b). Research data must be provided with persistent identifiers[5] within the repository.

It is important to preserve the integrity of research data and to comply with the FAIR principles[6]. Research data must be stored in a correct, complete, unadulterated and reliable manner. They must be findable, identifiable, accessible, traceable, interoperable and whenever possible reusable and replicable.

In compliance with intellectual property rights, and unless third-party rights, legal requirements, Rectorate decisions, other reasonable interests or property laws prohibit it, research data should be assigned an open use license.[7]

Citation norms and requirements regarding publication and future research should be followed; data sources should be explicitly traceable in order for the original sources to be acknowledged.

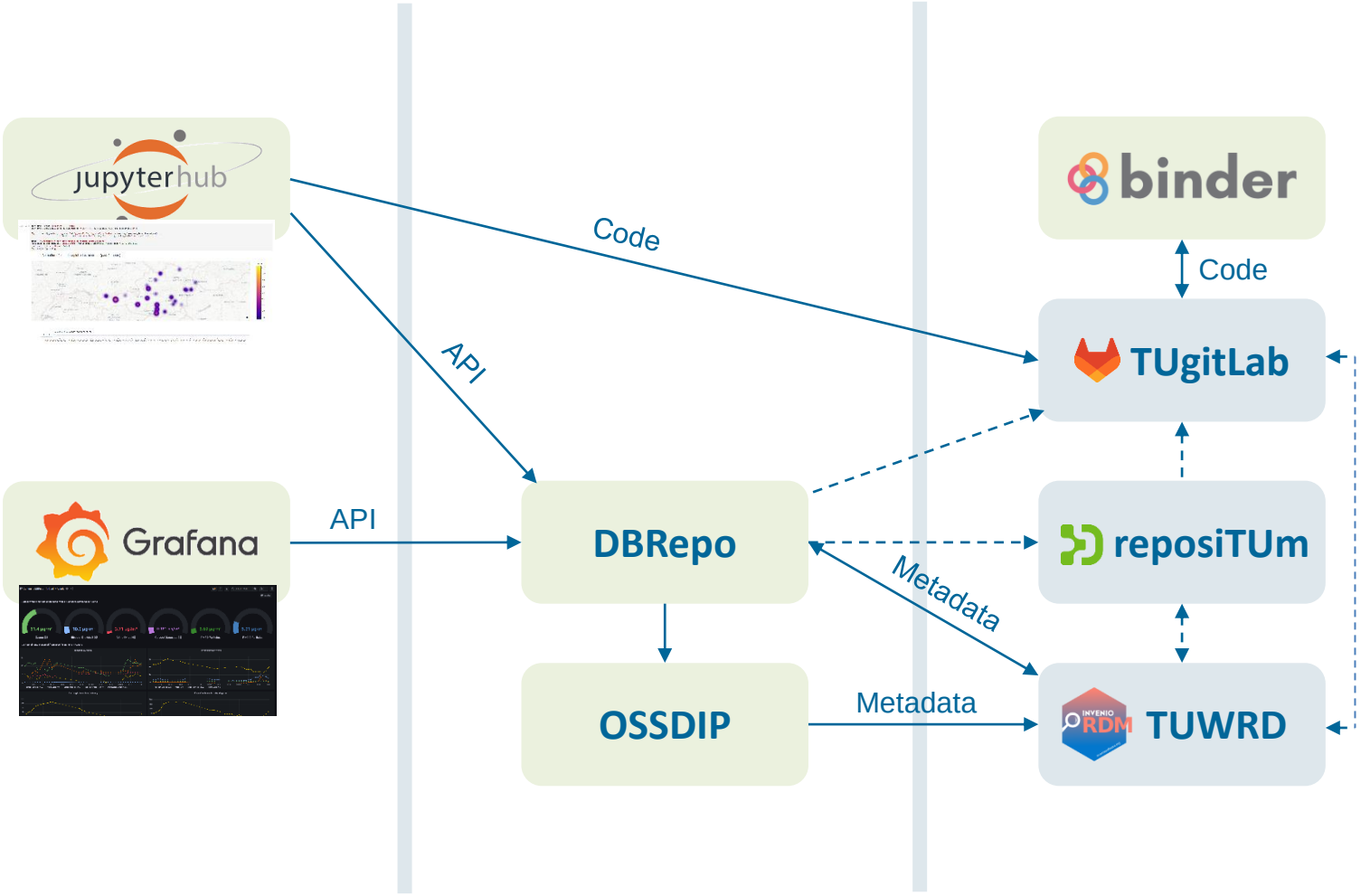
Research data and records are to be stored and made available in accordance with intellectual property laws or the requirements of third-party funders as well as applicable legal or contractual requirements (e.g. EU restrictions on where identifiable personal data may be stored). Research data that may be of future historical interest and the records accompanying them should also be archived.

The minimum retention period for research data and records is 10 years after either the assignment of a persistent identifier or the publication of a related work following research completion, whichever is later.

In the event that research data and records are to be deleted or destroyed, either after expiration of the required retention period or for legal or ethical reasons, such action is to be carried out only after consideration of all legal and ethical perspectives. The following aspects must be taken into consideration when decisions are made about the retention or destruction of research data: interests and contractual provisions of third-party funders and other stakeholders, employees and partner participants in particular, as well as confidentiality and security. Any decision taken must be documented.

Research Data Policy 2018

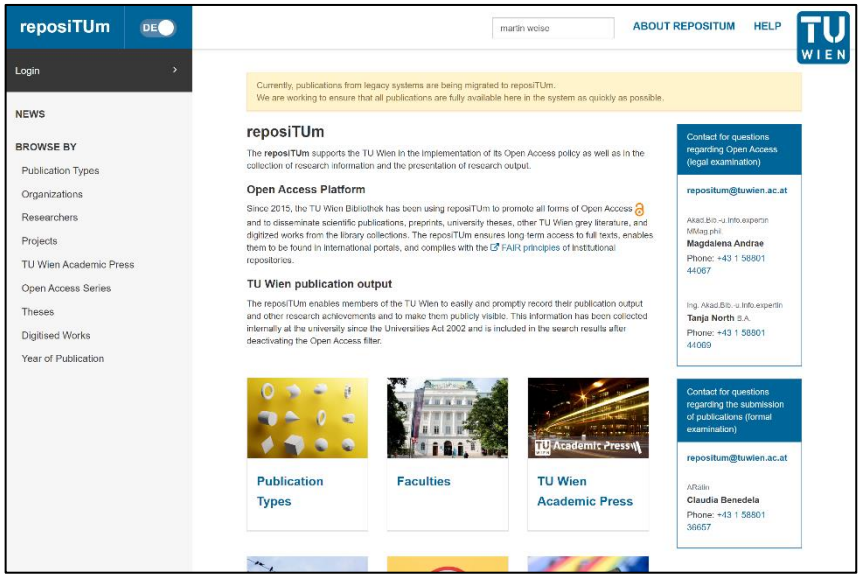
Repository & Compute Environment



reposiTUm (publications)
Overview & Goals

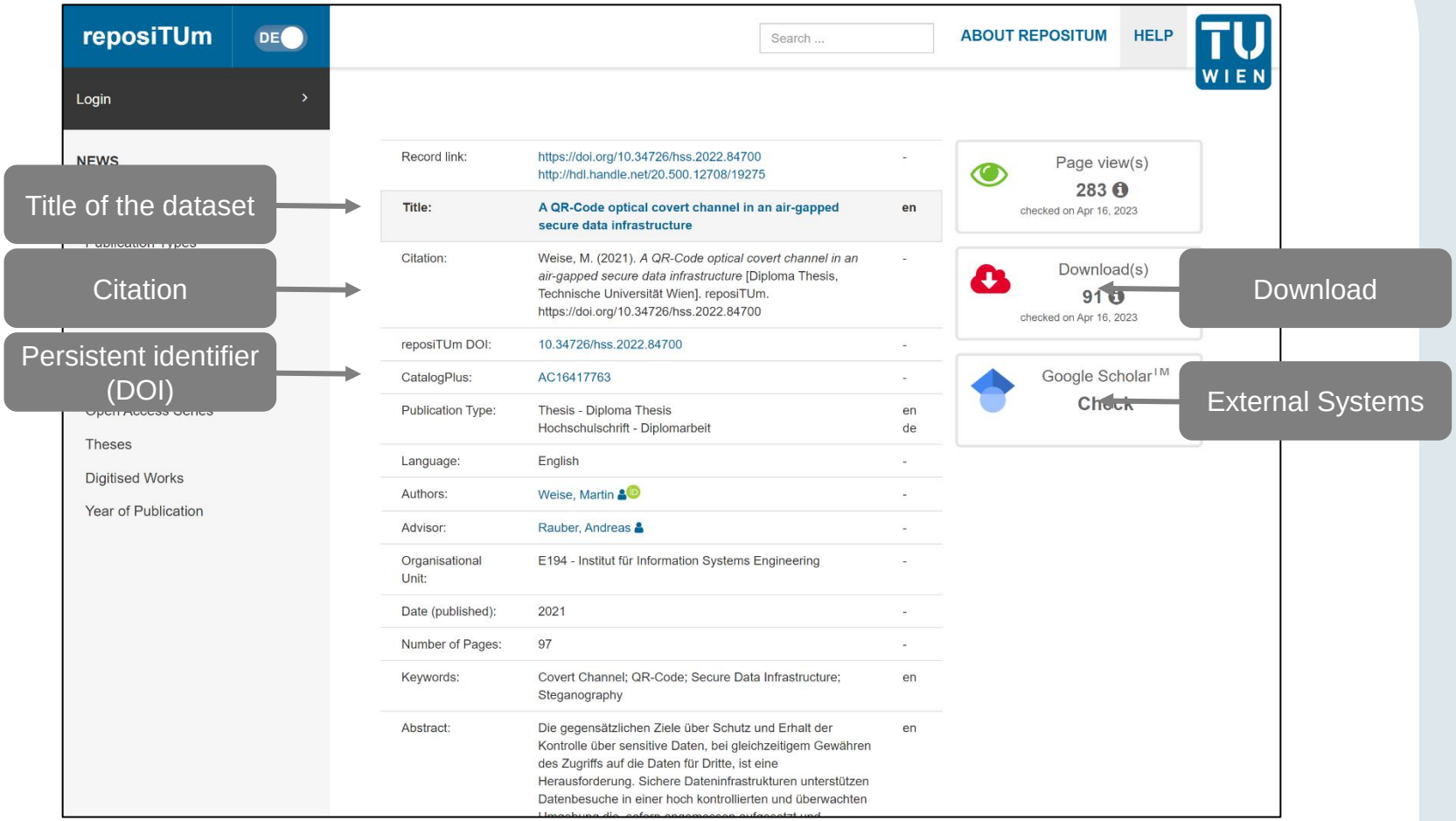
TU Wien Publication Repository

- Document-based research outputs
- Preservation
- Intellectual property
  - University ranking
  - Performance agreements
- Findability, Reusability
  - Papers
  - Presentations
  - Posters
  - Thesis’



https://repositum.tuwien.at/

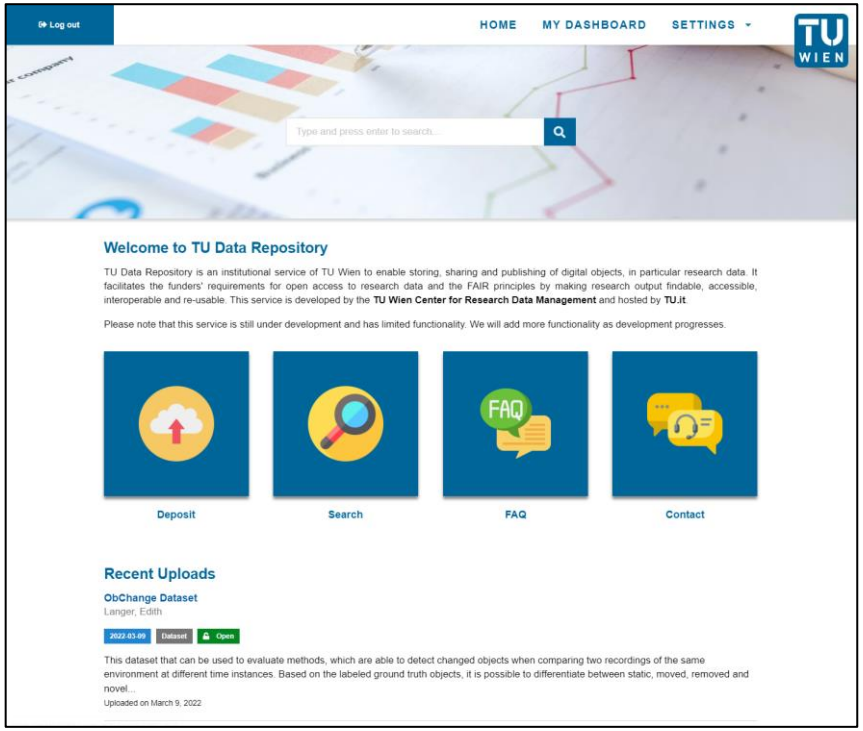
reposiTUm (publications)
Example



TUWRD (data sets)
Overview & Goals

TU Wien Research Data Repository

- File-based research data
- Individual, collections
- Extensive metadata
  - DOIs
- Not for publications
  - Other system exists
- Operational since 2022
- CEPH storage, backups
- 66 datasets
- 9 TiB



https://researchdata.tuwien.ac.at

TUWRD (data sets)

Example

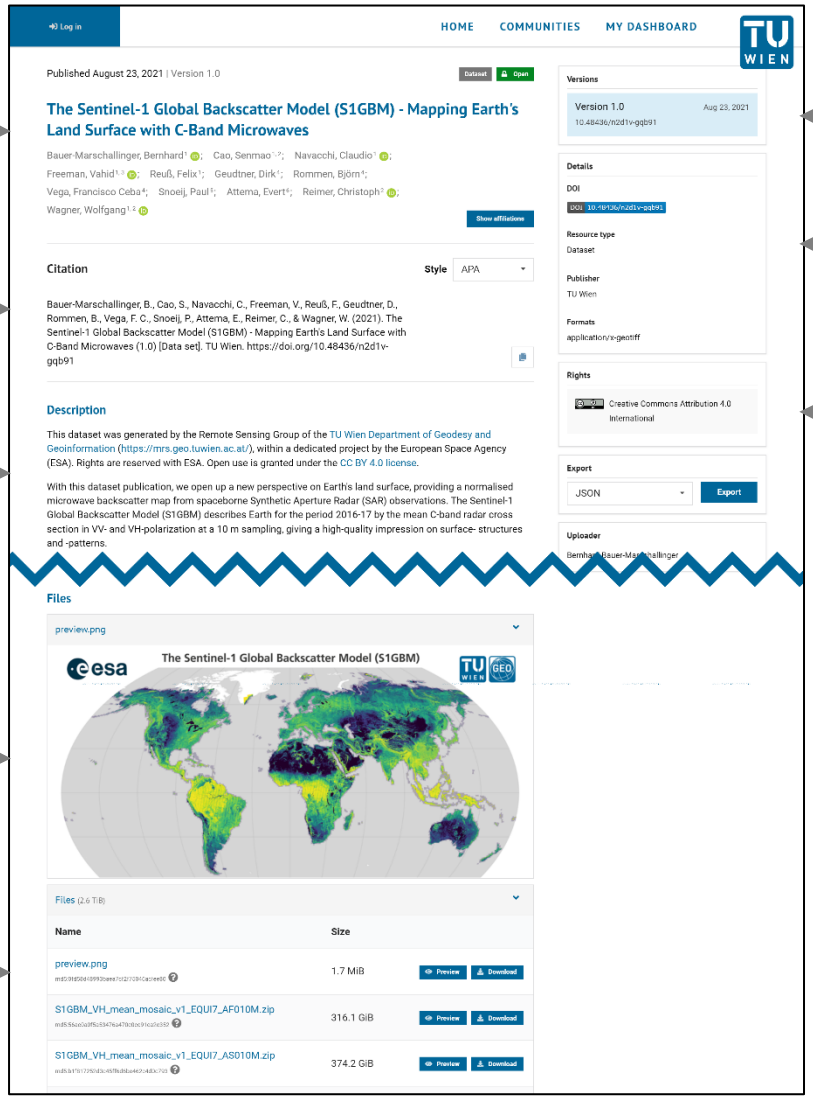
Title of the dataset

Citation

Description of the dataset

Preview file

Files for download



Version of the dataset

Persistent identifier (DOI)

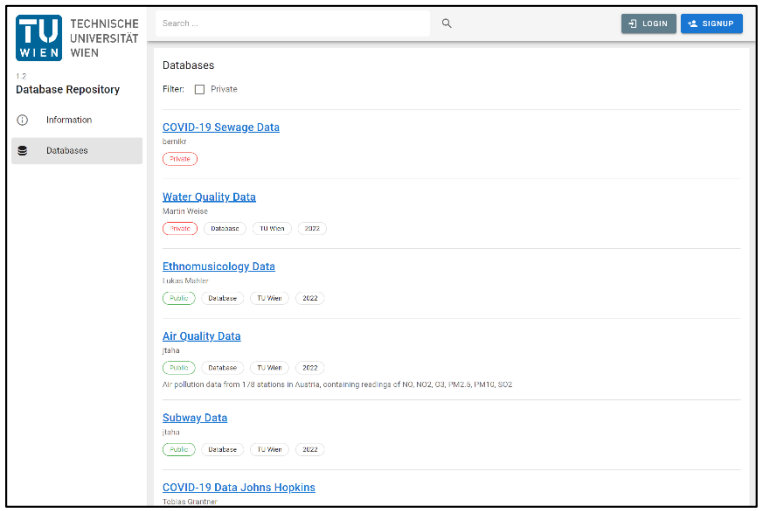
License CC-BY-NC-SA-3.0

https://doi.org/10.48436/n2d1v-gqb91

DBRepo (databases)
Overview & Goals

TU Wien Research Data Repository

- Handled ab-inition, no ex-post submission after project (no dumps)
- Handling **live data from data streams** (IoT, continuous measurements, …)
- Upload/download, **continuous feeding**, permissions, ownership
- Updates for corrections and versioning for reproducibility
- Web interface & **APIs for machine access**



https://dbrepo1.ec.tuwien.ac.at/

Supporting FAIR principles

Supporting RDA WGDC principles on data citation

DBRepo
Example

```python
class BrokerServiceClient:

    def __init__(self, exchange, routing_key, host, username, password):
        self.exchange = exchange
        self.routing_key = routing_key
        self.username = username
        self.url = f"amqp://{username}:{password}@{host}:5672/%2f"

    def send(self, payload: str):
        with rabbitpy.Connection(url=self.url) as connection:
            connection.channel() as channel:
            essage = rabbitpy.Message(channel=channel, body_value=payload, properties={"user_id": self.username})
            essage.publish(exchange=self.exchange, routing_key=self.routing_key)
            gging.debug(f"... sent tuple")
```

```python
#
# Maxx Turn 45 Power
#
mt45pa = PilotFactoryClient(hostname=hostname, port=port, topic="MaxxTurn45/Power/Active/A",
                            dbrepo_exchange="dbrepo.pilot_factory_data",
                            dbrepo_routing_key="dbrepo.pilot_factory_data.power",
                            dbrepo_hostname=__dbrepo_broker_url, dbrepo_username=__dbrepo_username,
                            dbrepo_password=__dbrepo_password)
mt45pa.get().on_message = mt45pa.on_message_maxxturn45_power
mt45pa.start()
```

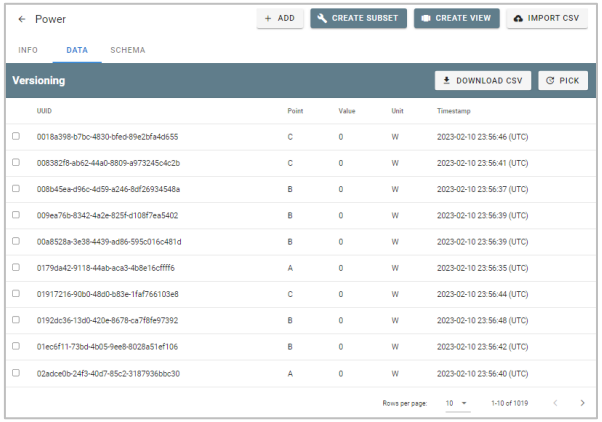
mosquitto

MQTT

rabbitpy

AMQP

RabbitMQ



DBRepo

Exchange Type

Direct (AMQP)

Exchange Name

dbrepo.pilot_factory_data

Queue Name

dbrepo.pilot_factory_data.power

Routing Key

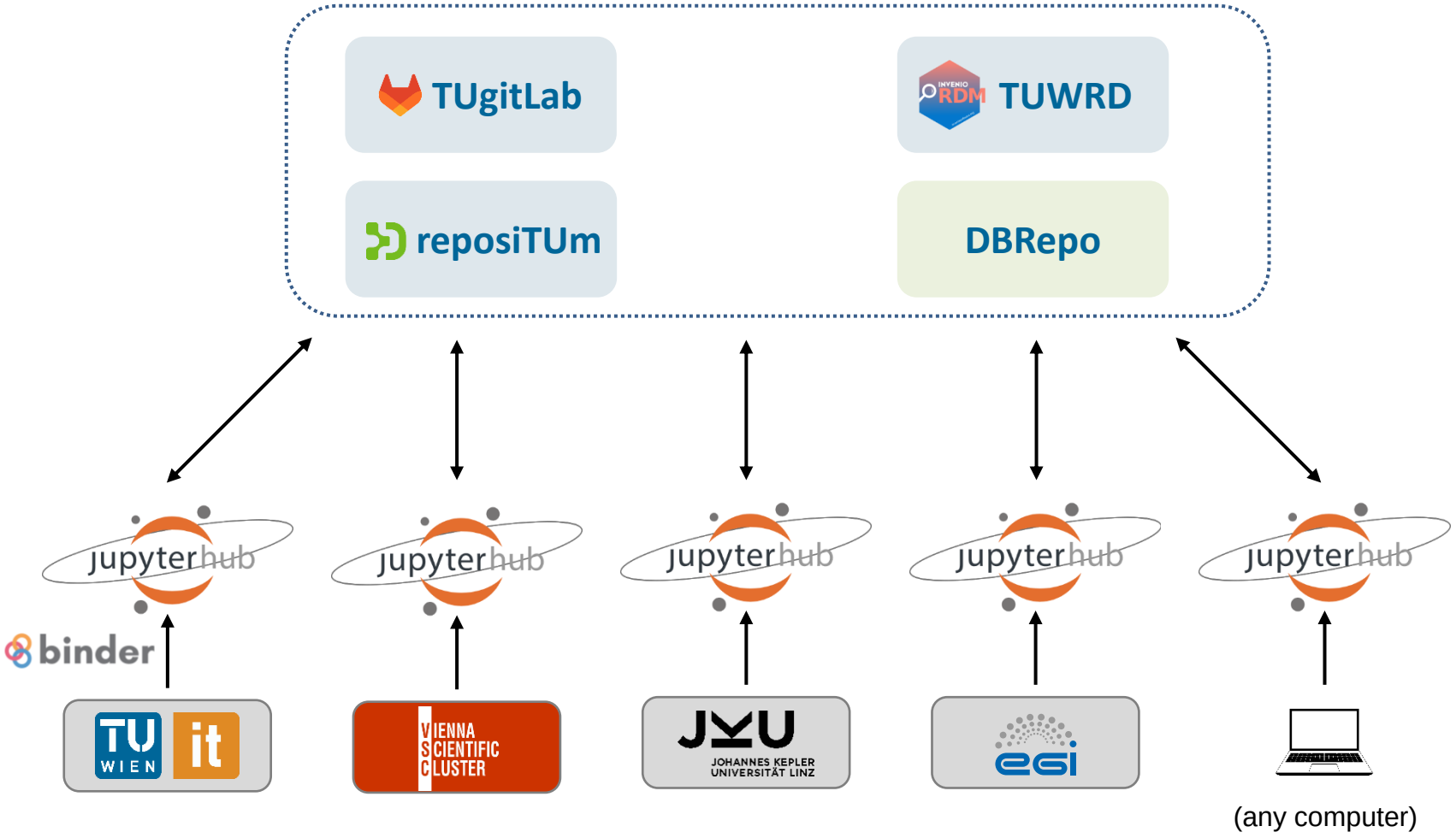
dbrepo.pilot_factory_data.power

DBRepo

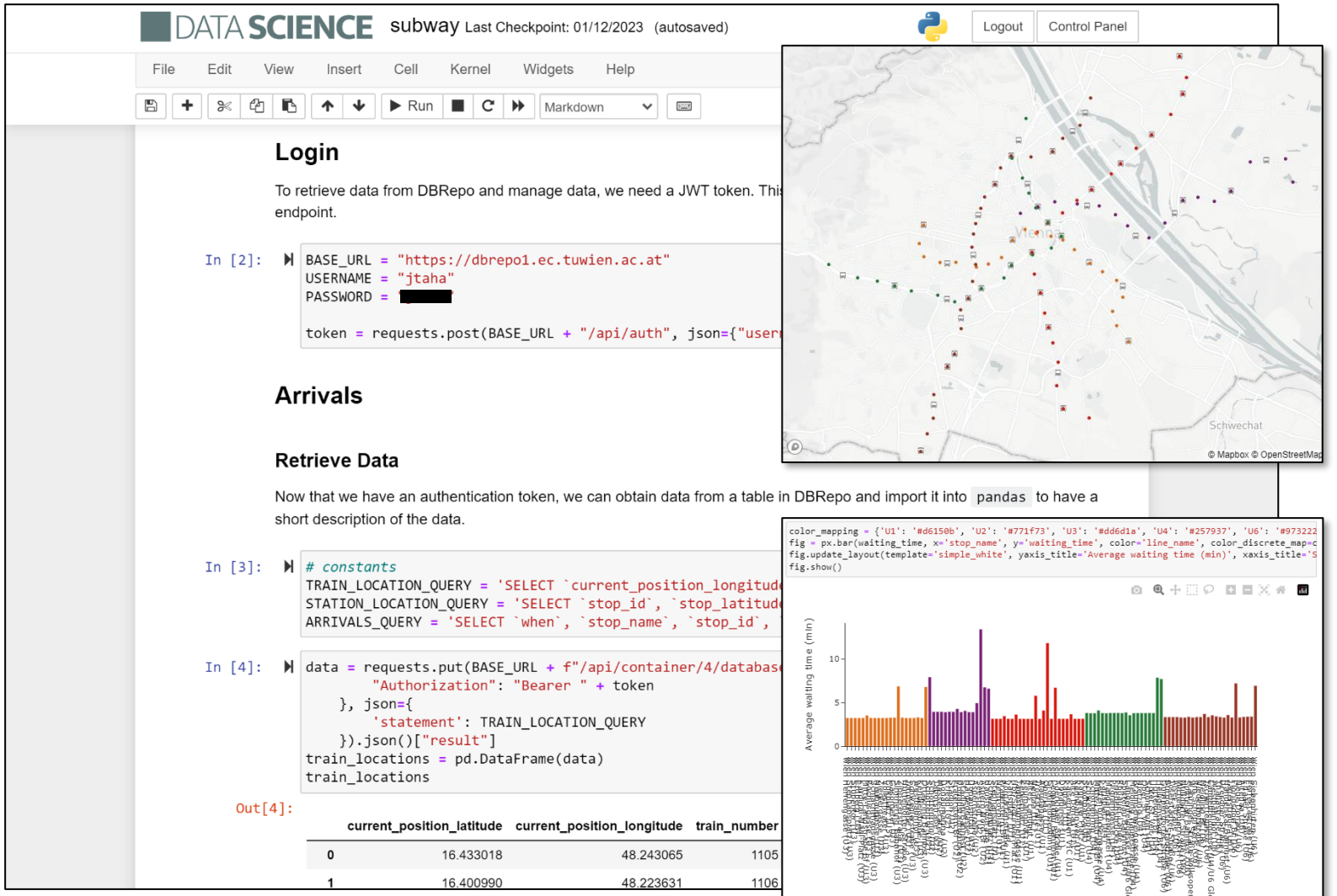
Nodes deployed worldwide

Jupyterhub

Compute environment for open data



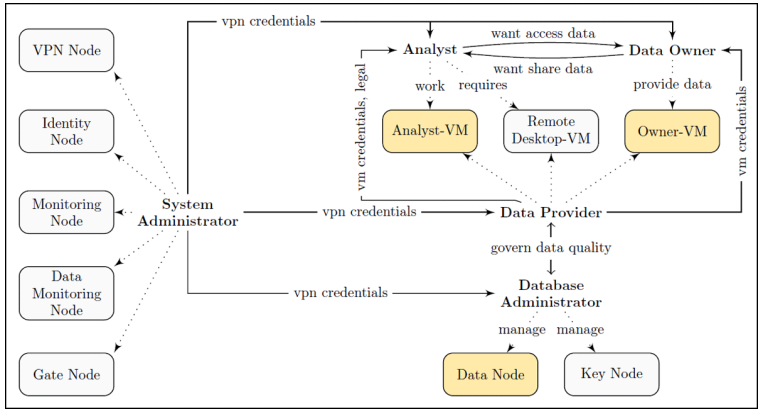
Jupyterhub
Example



OSSDIP
Overview & goals

Secure analysis environment

- TRE **model** and **reference implementation**
- Based on best-practice & open-source software
- Sensitive data (privacy issues, commercial interest), provide access for analysis, but ensure data is **not leaked** or misused
- Standard processes for involved roles



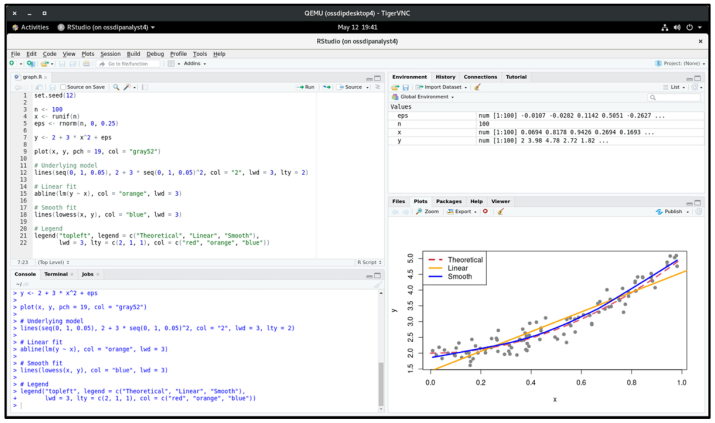
https://ossdip.at/

http://doi.org/10.5334/dsj-2022-004

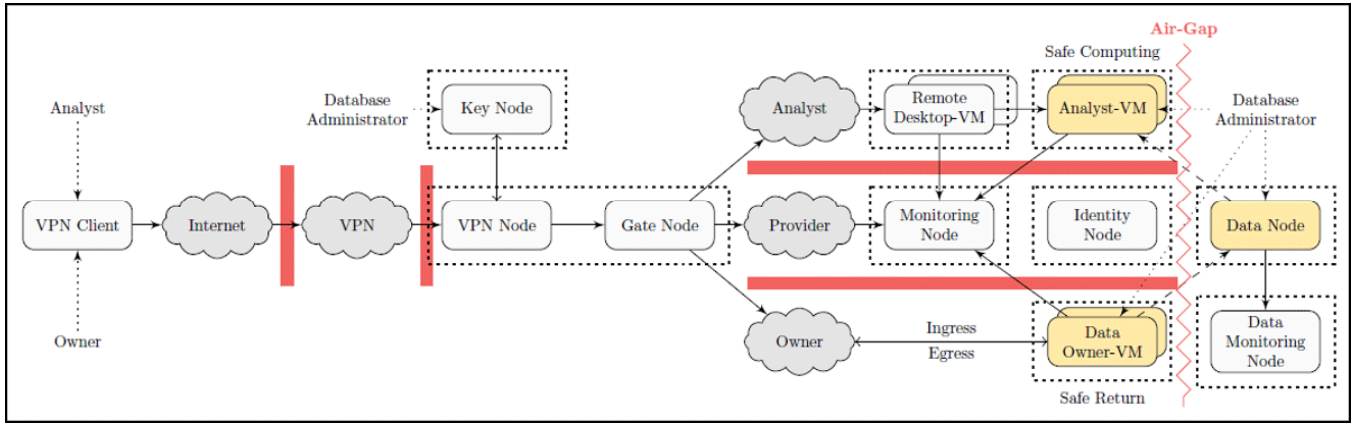
OSSDIP
Principles

Data protection

- Air-gapped Data Node
- Only **brief** connections by trusted database admin
- Copy (fingerprinted, …) subset dataset from access request
- Analysis only via **multiple secure layers** & media breaks



https://ossdip.at/



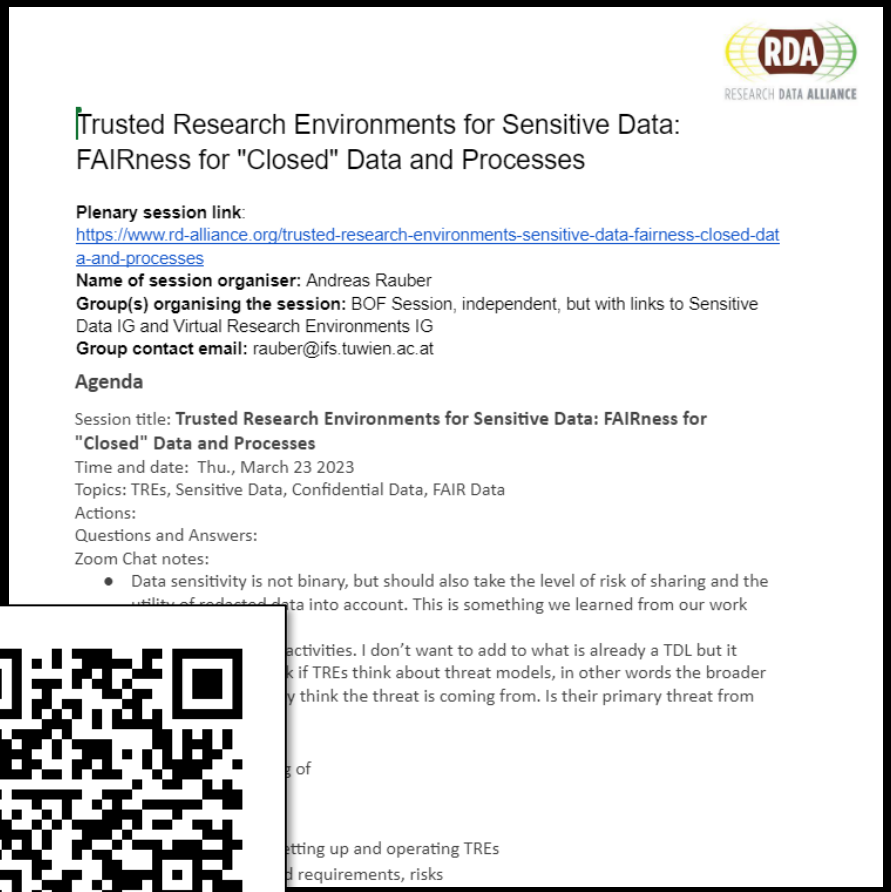
http://doi.org/10.5334/dsj-2022-004

OSSDIP

How to get involved

Research Data Alliance

- BoF Session in March (P20)
- 81/50 people in the room
- Better understanding TREs
  - Similarities
  - Differences
  - Options for setting up and operating
- Starting a Working Group



RESEARCH DATA ALLIANCE

Trusted Research Environments for Sensitive Data:
FAIRness for "Closed" Data and Processes

Plenary session link:
https://www.rd-alliance.org/trusted-research-environments-sensitive-data-fairness-closed-data-and-processes
Name of session organiser: Andreas Rauber
Group(s) organising the session: BOF Session, independent, but with links to Sensitive Data IG and Virtual Research Environments IG
Group contact email: rauber@ifs.tuwien.ac.at

Agenda

Session title: **Trusted Research Environments for Sensitive Data: FAIRness for "Closed" Data and Processes**
Time and date: Thu., March 23 2023
Topics: TREs, Sensitive Data, Confidential Data, FAIR Data
Actions:
Questions and Answers:
Zoom Chat notes:

- Data sensitivity is not binary, but should also take the level of risk of sharing and the utility of redacted data into account. This is something we learned from our work

activities. I don't want to add to what is already a TDL but it
if TREs think about threat models, in other words the broader
think the threat is coming from. Is their primary threat from

of

tting up and operating TREs
requirements, risks

Add your name if interested

https://bit.ly/rda-bof-tre-notes

Dipl.-Ing. Martin Weise
Technische Universität Wien, Austria
martin.weise@tuwien.ac.at

Moritz Staudinger, Cornelia Michlits, Kirill Stytsenko, Eva Gergely, Tobias Grantner, Josef Taha, Geoffrey Karnbach