



TECHNISCHE  
UNIVERSITÄT  
WIEN  
Vienna | Austria



# Repository and Compute Environment for Sensitive Data

*Research Data Management*

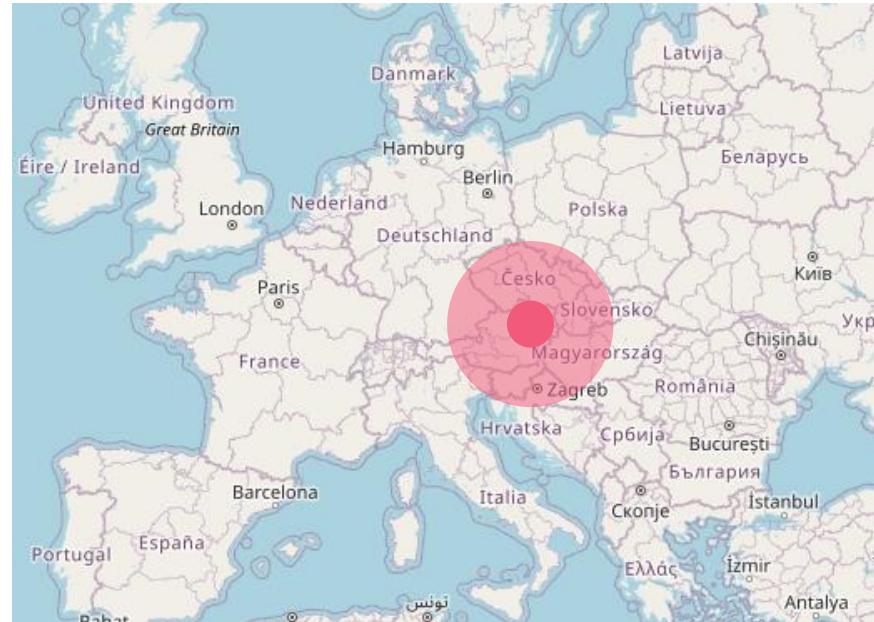
**Dipl.-Ing. Martin Weise**  
Technische Universität Wien, Austria  
[martin.weise@tuwien.ac.at](mailto:martin.weise@tuwien.ac.at)

## Numbers

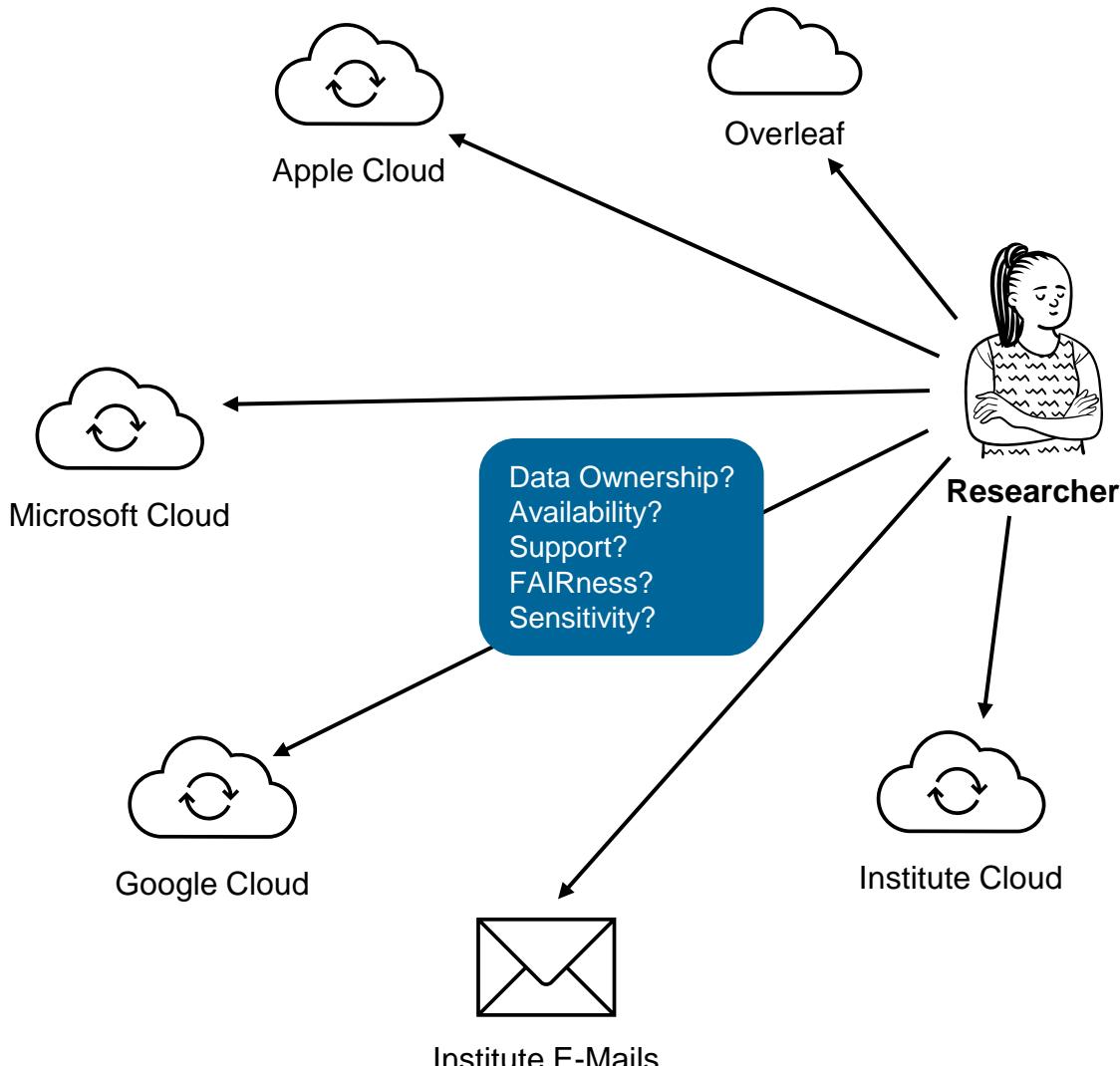
- 5.406 employees (76% scientific)
- 26.358 students
- 8 faculties
- 51 institutes

## Infrastructure

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



# Common Data Management



**TU WIEN 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<sup>6</sup> within the repository.

It is important to preserve the integrity of research data and to comply with the FAIR principles<sup>6</sup>. 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.<sup>7</sup>

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.

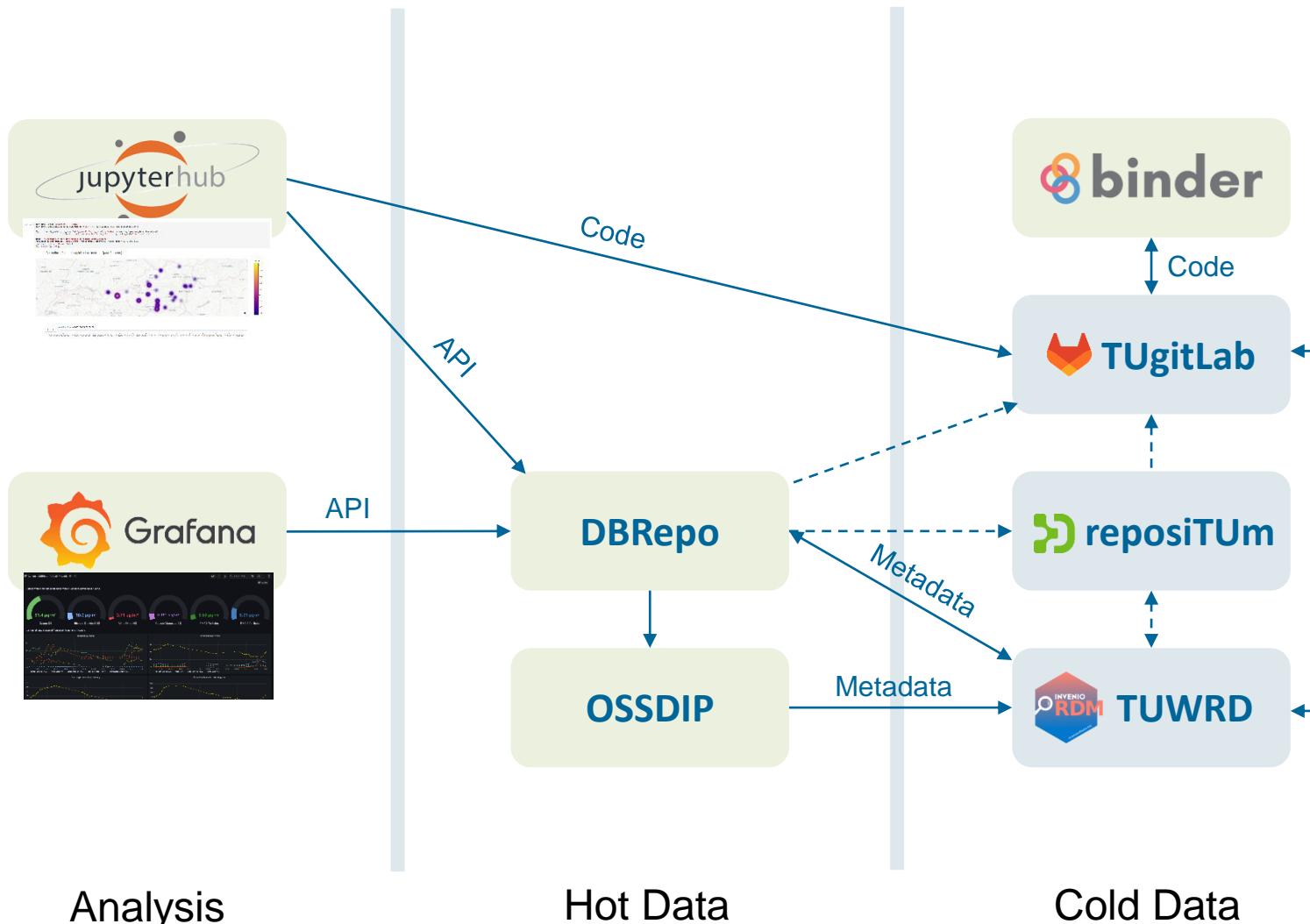
Research data and records are to be stored and made available in accordance with intellectual property laws or the requirements of third-party funders as well as applicable legal or contractual requirements (e.g. EU restrictions on where identifiable personal data may be stored). Research data that may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



Analysis

Hot Data

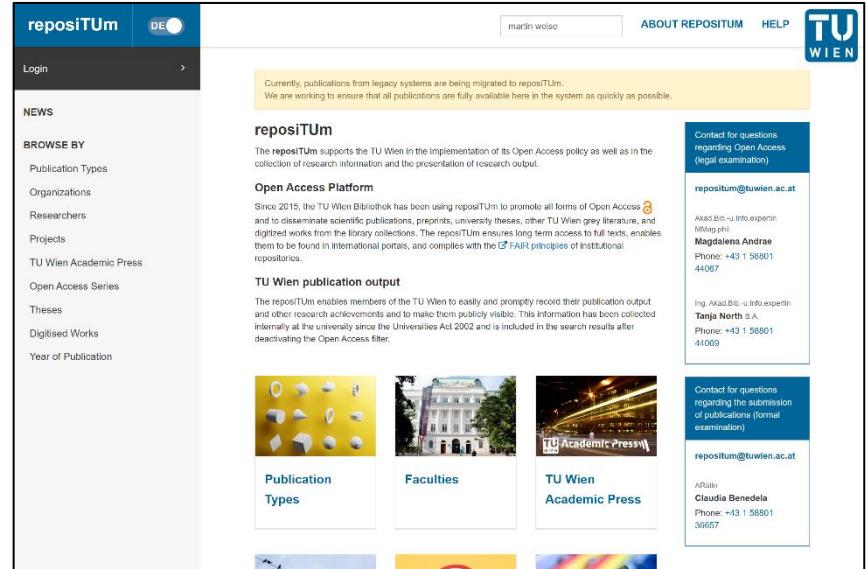
Cold Data

# 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'

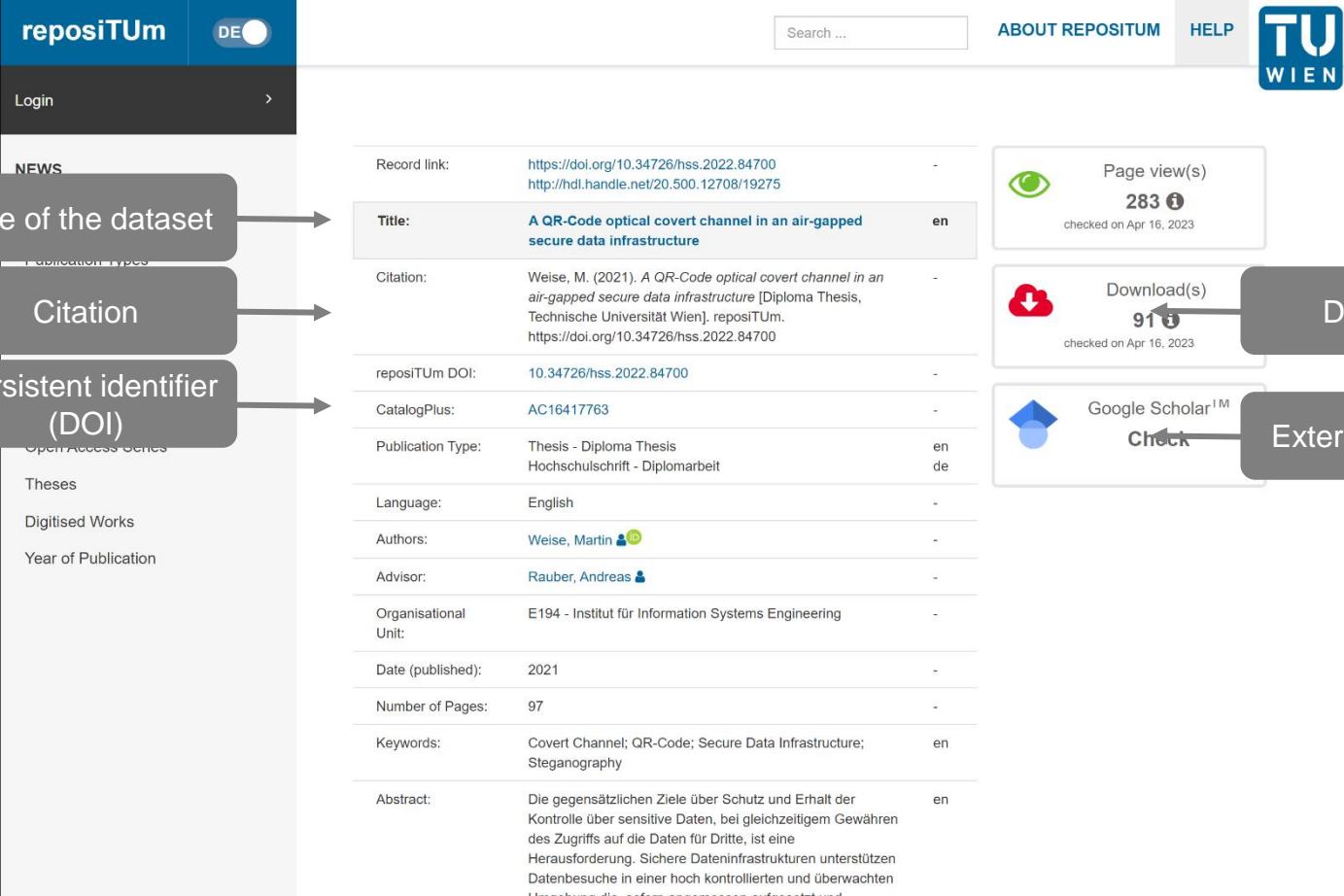


The screenshot shows the homepage of the repositUM (publications) website for TU Wien. The header includes the TU Wien logo and navigation links for 'ABOUT REPOSITORY' and 'HELP'. A search bar shows the query 'martin weise'. The main content area features a message about migration to repositUM, followed by sections for 'repoitUM', 'Open Access Platform', and 'TU Wien publication output'. There are also sections for 'Publication Types', 'Faculties', and 'TU Wien Academic Press'. Contact information for various staff members is provided in boxes on the right.

<https://repositum.tuwien.ac.at/>

# repositUM (publications)

## Example



The screenshot shows a detailed view of a publication record on the repositUM platform. The record is for a Thesis - Diploma Thesis titled "A QR-Code optical covert channel in an air-gapped secure data infrastructure". The record includes fields such as Record link, Title, Citation, repositUM DOI, CatalogPlus, Publication Type, Language, Authors, Advisor, Organisational Unit, Date (published), Number of Pages, Keywords, and Abstract. To the right of the record details, there are metrics: 283 page views and 91 download(s). Below these metrics are buttons for "Download" and "External Systems" (Google Scholar Check). On the left side of the record, three specific fields are highlighted with arrows pointing to them: "Title of the dataset", "Citation", and "Persistent identifier (DOI)".

**Title of the dataset**

**Citation**

**Persistent identifier (DOI)**

Record link: <https://doi.org/10.34726/hss.2022.84700>  
<http://hdl.handle.net/20.500.12708/19275>

**Title:** A QR-Code optical covert channel in an air-gapped secure data infrastructure

**Citation:** Weise, M. (2021). A QR-Code optical covert channel in an air-gapped secure data infrastructure [Diploma Thesis, Technische Universität Wien]. repositUM.  
<https://doi.org/10.34726/hss.2022.84700>

**repositUM DOI:** [10.34726/hss.2022.84700](https://doi.org/10.34726/hss.2022.84700)

**CatalogPlus:** AC16417763

**Publication Type:** Thesis - Diploma Thesis  
Hochschulschrift - Diplomarbeit

**Language:** English

**Authors:** Weise, Martin 

**Advisor:** Rauber, Andreas 

**Organisational Unit:** E194 - Institut für Information Systems Engineering

**Date (published):** 2021

**Number of Pages:** 97

**Keywords:** Covert Channel; QR-Code; Secure Data Infrastructure; Steganography

**Abstract:** Die gegensätzlichen Ziele über Schutz und Erhalt der Kontrolle über sensible Daten, bei gleichzeitigem Gewähren des Zugriffs auf die Daten für Dritte, ist eine Herausforderung. Sichere Dateninfrastrukturen unterstützen Datenbesuche in einer hoch kontrollierten und überwachten Umgebung, die sofern nötig den aufgetretenen

**Page view(s):** 283  checked on Apr 16, 2023

**Download(s):** 91  checked on Apr 16, 2023

**Download**

**External Systems**

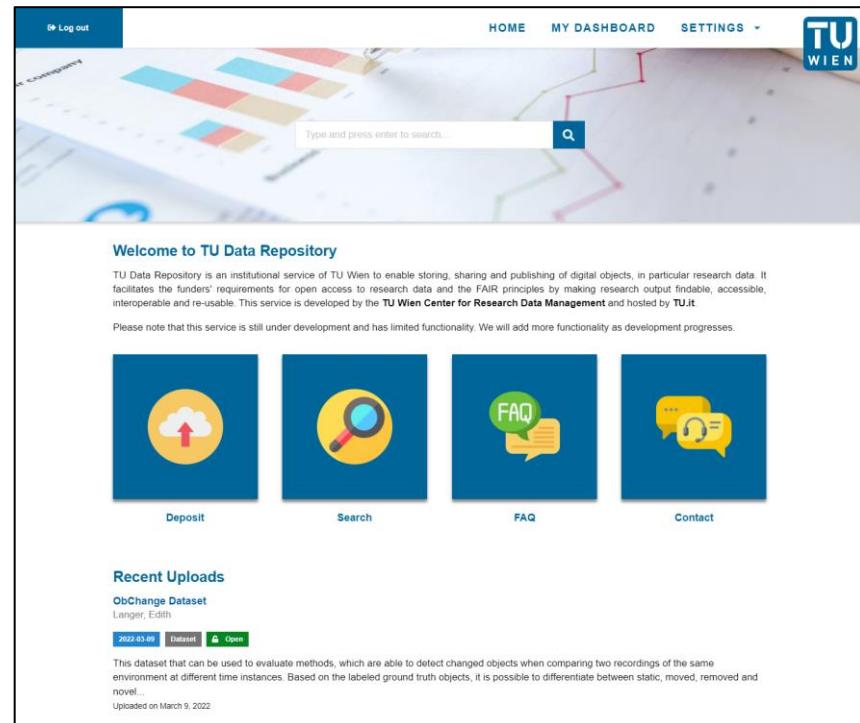
**Check**

# 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

Diagram illustrating the structure of a dataset page on TUWRD, showing various components and their corresponding labels.

The dataset page URL is: <https://doi.org/10.48436/n2d1v-gqb91>

**Labels pointing to specific sections:**

- Title of the dataset**: Points to the main title "The Sentinel-1 Global Backscatter Model (S1GBM) - Mapping Earth's Land Surface with C-Band Microwaves".
- Citation**: Points to the citation section, which includes authors (Bauer-Marschallinger, Bernhard; Cao, Senmiao; Navacchi, Claudio; Freeman, Vahid; Reul, Felix; Geudtner, Dirk; Rommen, Björn; Vega, Francisco Cebal; Snoeijs, Paul; Attema, Evert; Reimer, Christoph; Wagner, Wolfgang) and a reference (Bauer-Marschallinger, B., Cao, S., Navacchi, C., Freeman, V., Reul, F., Geudtner, D., Rommen, B., Vega, F. C., Snoeijs, P., Attema, E., Reimer, C., & Wagner, W. (2021). The Sentinel-1 Global Backscatter Model (S1GBM) - Mapping Earth's Land Surface with C-Band Microwaves (1.0) [Data set]. TU Wien. <https://doi.org/10.48436/n2d1v-gqb91>)
- Description of the dataset**: Points to the descriptive text below the citation, stating: "This dataset was generated by the Remote Sensing Group of the TU Wien Department of Geodesy and Geoinformation (<https://rsg.geo.tuwien.ac.at/>), within a dedicated project by the European Space Agency (ESA). Rights are reserved with ESA. Open use is granted under the CC BY 4.0 license."
- Preview file**: Points to the preview image titled "preview.png" showing a world map with green and yellow colors representing land surface backscatter.
- Files for download**: Points to the "Files" section at the bottom left, listing three files:
 

| Name                                     | Size      |
|--|-----------|
| preview.png                              | 1.7 MiB   |
| S1GBM_VH_mean_mosaic_v1_EQUI7_AF010M.zip | 316.1 GiB |
| S1GBM_VH_mean_mosaic_v1_EQUI7_AS010M.zip | 374.2 GiB |

**Labels pointing to specific dataset details on the right side:**

- Version of the dataset**: Points to the "Versions" section showing "Version 1.0" and its date "Aug 23, 2021".
- Persistent identifier (DOI)**: Points to the "DOI" section showing "DOI: [10.48436/n2d1v-gqb91](https://doi.org/10.48436/n2d1v-gqb91)".
- License CC-BY-NC-SA-3.0**: Points to the "Rights" section showing "Creative Commons Attribution 4.0 International".

# DBRepo (databases)

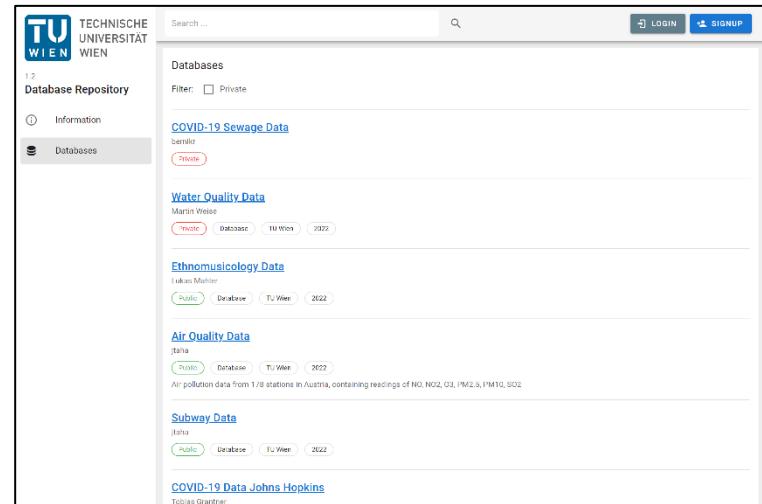
## Overview & Goals

### TU Wien Research Data Repository

- Handled ab-initio, 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**

Supporting FAIR principles

Supporting RDA WGDC principles on data citation



The screenshot shows the TU Wien Database Repository web interface. The top navigation bar includes a search bar, a login button, and a sign-up button. The main content area is titled "Databases" and shows a list of datasets. Each dataset entry includes a thumbnail, the name, the author, the database type, the university, and the year. For example, the "COVID-19 Sewage Data" entry is by Martin Weise, a Database at TU Wien from 2022.

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

# DBRepo Example

```
# 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()
```



|               |                                 |
|---------------|---------------------------------|
| Exchange Type | Direct (AMQP)                   |
| Exchange Name | dbrepo.pilot_factory_data       |
| Queue Name    | dbrepo.pilot_factory_data.power |
| Routing Key   | dbrepo.pilot_factory_data.power |

```
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:
                message = rabbitpy.Message(channel=channel, body_value=payload, properties={"user_id": self.username})
                message.publish(exchange=self.exchange, routing_key=self.routing_key)
                logging.debug(f"... sent tuple")
```

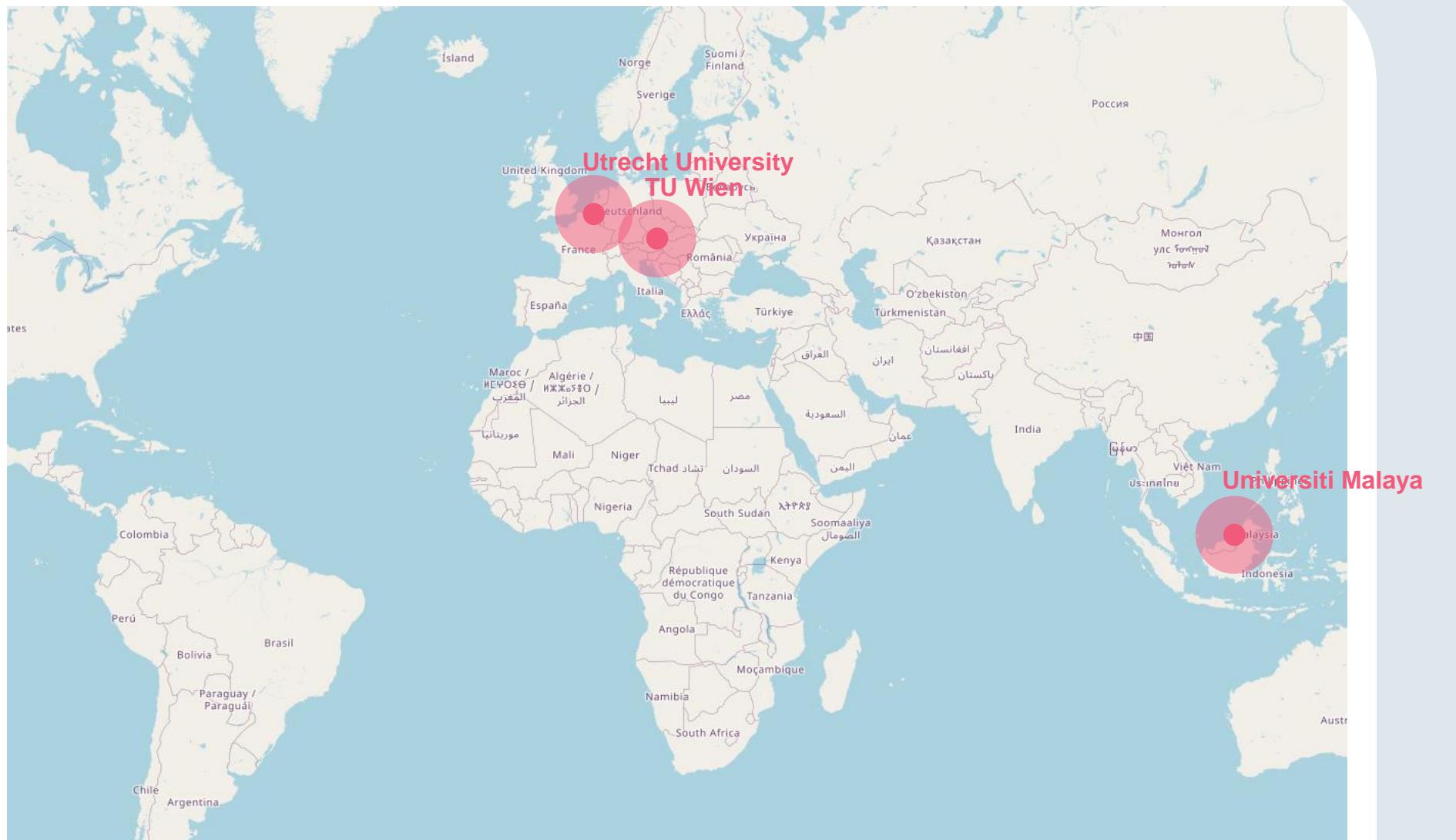
RabbitMQ

| Power                                |       |        |      |                           |
|--------------------------------------|-------|--------|------|---------------------------|
| INFO                                 | DATA  | SCHEMA |      |                           |
| <b>Versioning</b>                    |       |        |      |                           |
| UUID                                 | Point | Value  | Unit | Timestamp                 |
| 0018a398-b7bc-4830-bfed-89e2f9a4d655 | C     | 0      | W    | 2023-02-10 23:56:46 (UTC) |
| 008382fb-ab62-44a0-8809-9773245c4c2d | C     | 0      | W    | 2023-02-10 23:56:41 (UTC) |
| 008b45ea-d96c-4d59-a246-8df26934548a | B     | 0      | W    | 2023-02-10 23:56:37 (UTC) |
| 009fe7b-9342-4x2e-825f-d1087e5402    | B     | 0      | W    | 2023-02-10 23:56:39 (UTC) |
| 00a8528a-3e38-4439-ad86-595016c481d  | B     | 0      | W    | 2023-02-10 23:56:39 (UTC) |
| 0179db42-9118-44ab-ac35-408e16cffff6 | A     | 0      | W    | 2023-02-10 23:56:35 (UTC) |
| 01917216-9000-48d0-838e-1fe766103e8  | C     | 0      | W    | 2023-02-10 23:56:44 (UTC) |
| 0192d36-13d0-420e-8678-ca7f6e97392   | B     | 0      | W    | 2023-02-10 23:56:42 (UTC) |
| 01ec6f11-73bd-4b05-9ee8-8028a51ef106 | B     | 0      | W    | 2023-02-10 23:56:42 (UTC) |
| 02adce0b-24f3-407-85c2-31797936bbc30 | A     | 0      | W    | 2023-02-10 23:56:40 (UTC) |

DBRepo

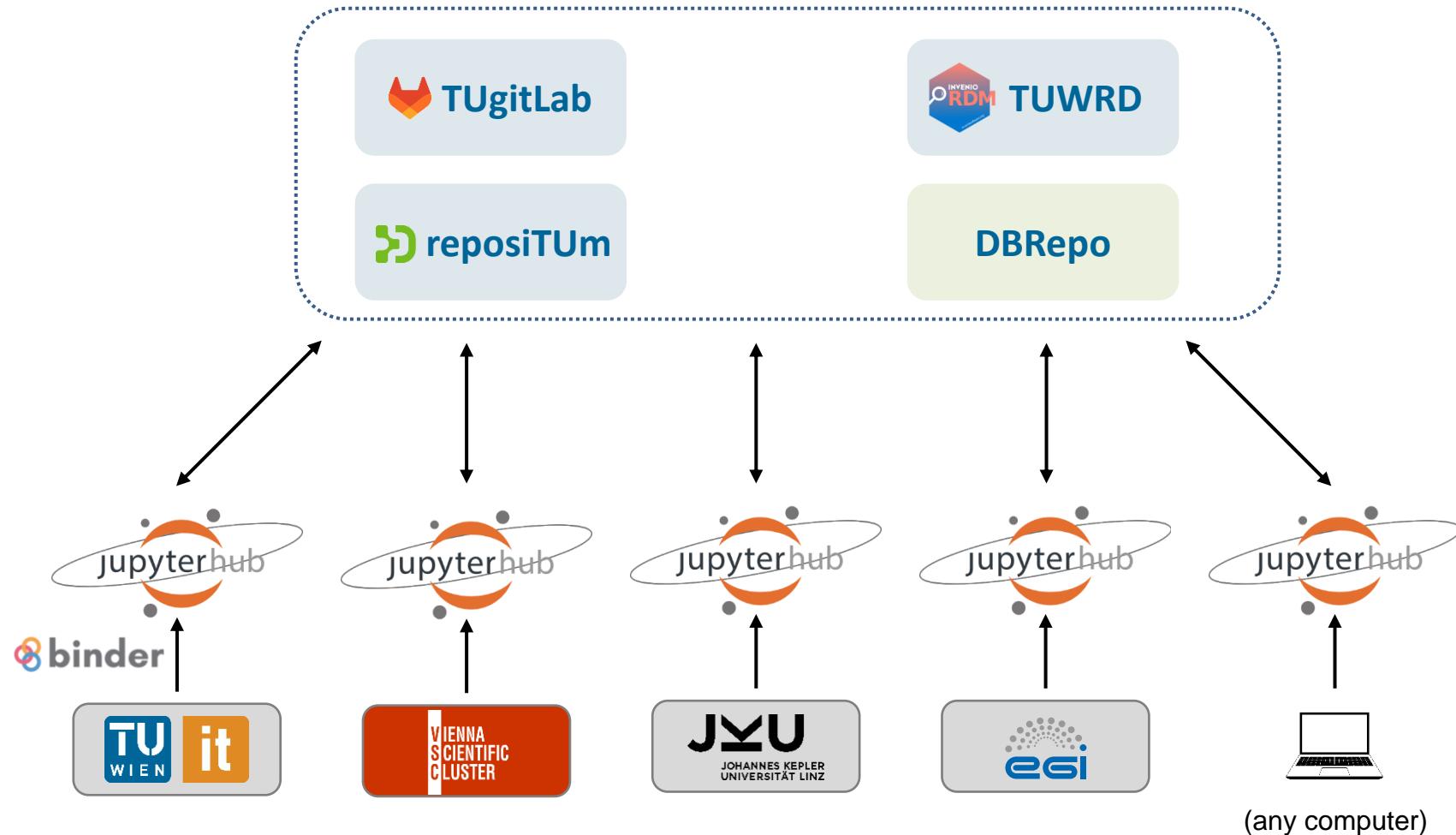
# DBRepo

## Nodes deployed worldwide



# Jupyterhub

## Compute environment for open data



# Jupyterhub Example

**DATA SCIENCE** subway Last Checkpoint: 01/12/2023 (autosaved)

Logout | Control Panel

File Edit View Insert Cell Kernel Widgets Help

Run | Kernel | Help | Markdown | Cell | Kernel | Widgets | Help

**Login**

To retrieve data from DBRepo and manage data, we need a JWT token. This endpoint.

```
In [2]: #!/usr/bin/env python
# coding: utf-8
from requests import post
import json
import base64
import requests
from IPython.display import display, JSON
```

token = requests.post(BASE\_URL + "/api/auth", json={"username": USERNAME, "password": PASSWORD}).json()["token"]

**Arrivals**

**Retrieve Data**

Now that we have an authentication token, we can obtain data from a table in DBRepo and import it into pandas to have a short description of the data.

```
In [3]: #!/usr/bin/env python
# coding: utf-8
from IPython.display import display, JSON
import requests
import pandas as pd
from IPython.core.display import HTML
```

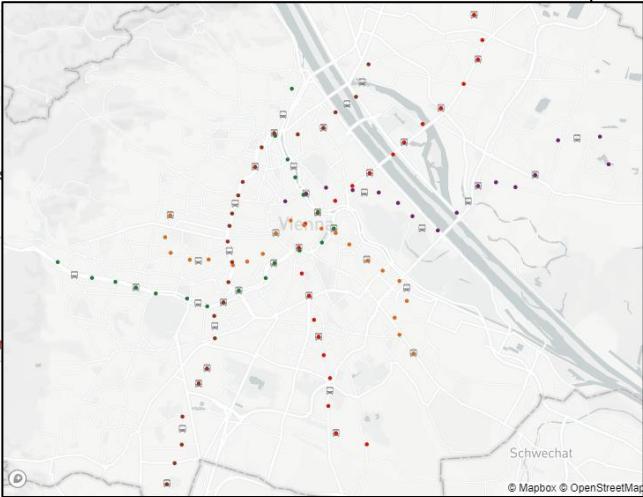
TRAIN\_LOCATION\_QUERY = 'SELECT `current\_position\_longitude`, `current\_position\_latitude` FROM `train\_locations` WHERE `train\_number` = 1106;  
STATION\_LOCATION\_QUERY = 'SELECT `stop\_id`, `stop\_latitude`, `stop\_longitude` FROM `station\_locations`;  
ARRIVALS\_QUERY = 'SELECT `when`, `stop\_name`, `stop\_id`, `train\_number` FROM `arrivals` WHERE `train\_number` = 1106;  
color\_mapping = {'U1': '#d6150b', 'U2': '#771f73', 'U3': '#dd6d1a', 'U4': '#257937', 'U6': '#973222'}

```
In [4]: #!/usr/bin/env python
# coding: utf-8
from IPython.display import display, JSON
import requests
import pandas as pd
from IPython.core.display import HTML
```

data = requests.put(BASE\_URL + f"/api/container/4/database",  
headers={"Authorization": "Bearer " + token},  
json={  
"statement": TRAIN\_LOCATION\_QUERY  
}).json()["result"]  
train\_locations = pd.DataFrame(data)  
train\_locations

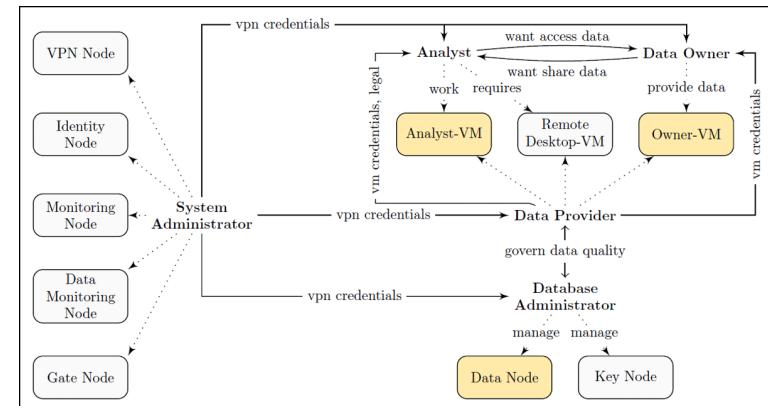
**Out[4]:**

|   | current_position_latitude | current_position_longitude | train_number |
|---|---------------------------|----------------------------|--------------|
| 0 | 16.433018                 | 48.243065                  | 1105         |
| 1 | 16.400990                 | 48.223631                  | 1106         |




## Secure analysis environment

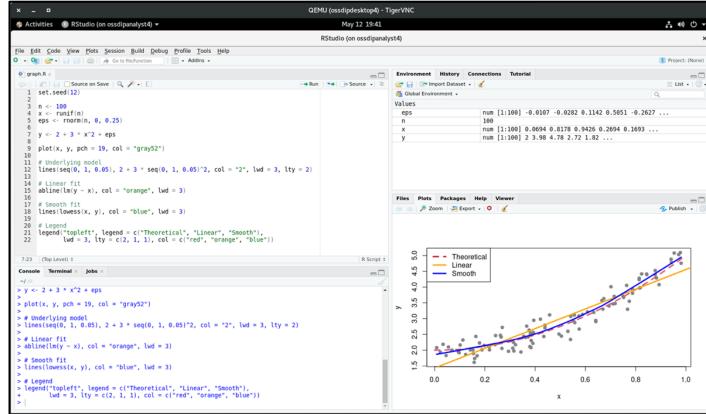
- TRE model and reference implementation
- Based on best-practice & open-source software
- Sensitive data (privacy issues, commercial interest), provide access for analysis, but ensure data is **not leaked** or misused
- Standard processes for involved roles



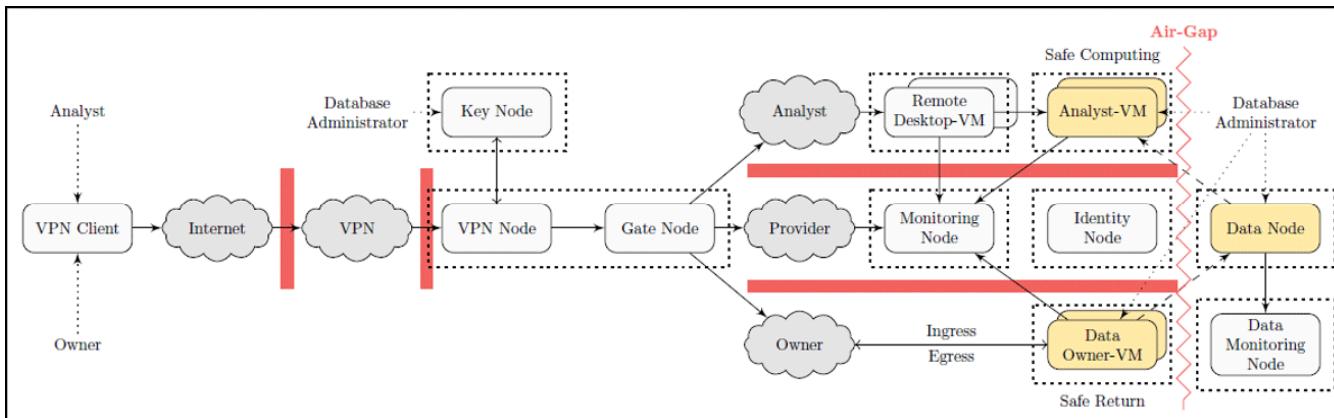
<https://ossdip.at/>

### 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://osstdip.at/>



## 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

Add your name if interested

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

### Research Data Alliance 20<sup>th</sup> Plenary Meeting

Celebrating 10 years of the RDA  
A Decade of Data

GOTHENBURG, SWEDEN  
21-23 March 2023



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 sensitivity of conducted 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 makes sense to think if TREs think about threat models, in other words the broader context of the threat. Who is the primary threat? Who is the secondary threat? Who is the tertiary threat? Is their primary threat from



Setting up and operating TREs  
Requirements, risks



TECHNISCHE  
UNIVERSITÄT  
WIEN  
Vienna | Austria



**Dipl.-Ing. Martin Weise**

Technische Universität Wien, Austria

[martin.weise@tuwien.ac.at](mailto:martin.weise@tuwien.ac.at)

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