



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna | Austria



Repository and Compute Environment for Sensitive Data

Research Data Management

Dipl.-Ing. Martin Weise

Technische Universität Wien, Austria

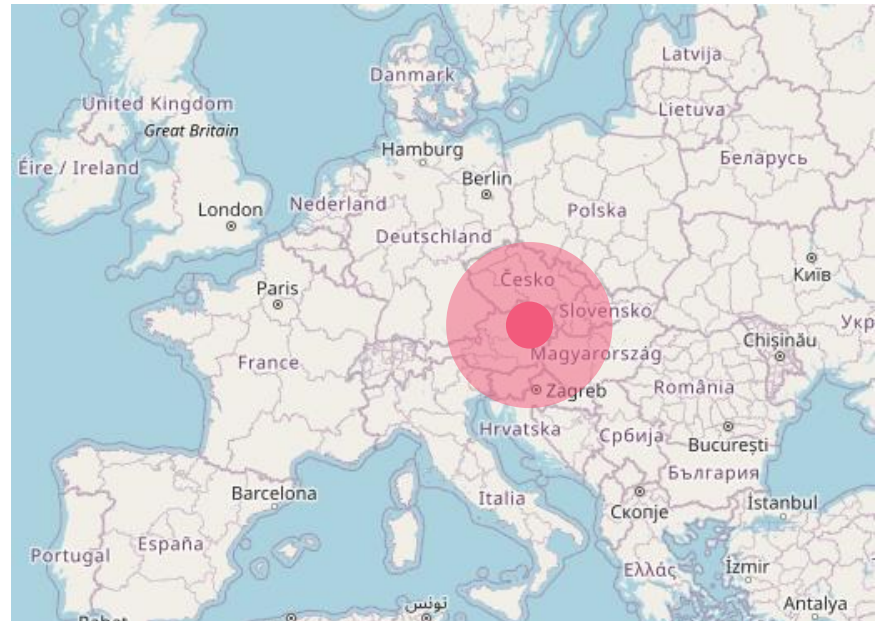
martin.weise@tuwien.ac.at

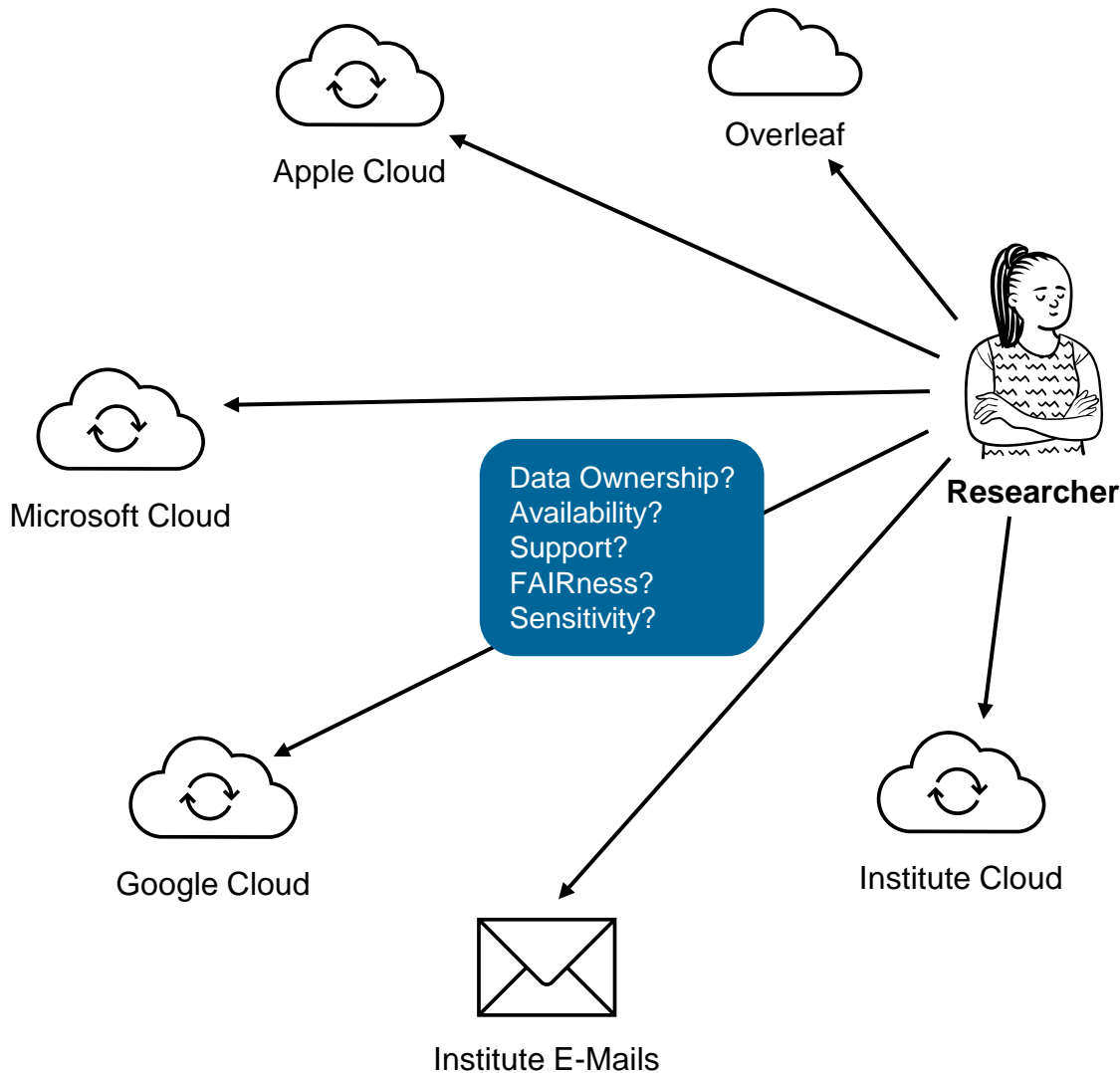
Numbers

- 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





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 identifiers⁴ within the repository.

It is important to preserve the integrity of research data and to comply with the FAIR principles⁶. 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

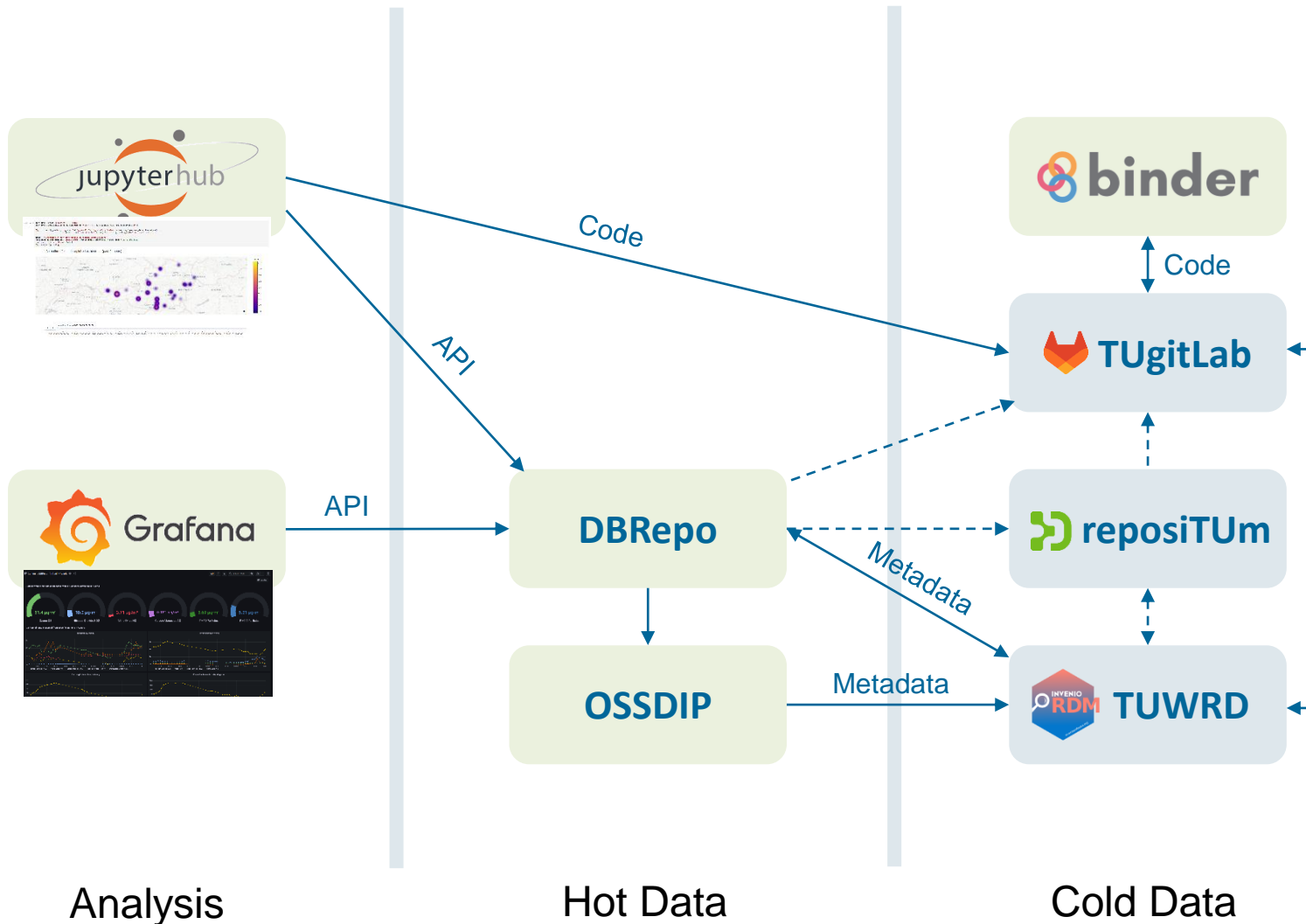
Citation norms and requirements regarding publication and future research should be followed; data sources should be explicitly traceable in order for the original sources to be acknowledged.

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.

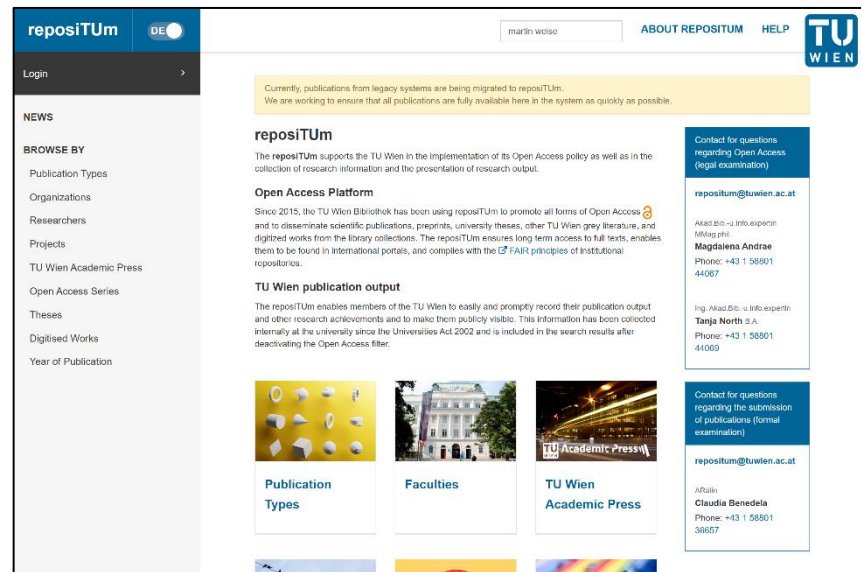
In the event that research data and records are to be deleted or destroyed, either after expiration of the required retention period or for legal or ethical reasons, such action is to be carried out only after consideration of all legal and ethical perspectives. The following aspects must be taken into 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



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/>

The screenshot displays the repositUm interface for a specific publication. The record details are as follows:

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	en
Citation:	Weise, M. (2021). <i>A QR-Code optical covert channel in an air-gapped secure data infrastructure</i> [Diploma Thesis, Technische Universität Wien]. repositUm. https://doi.org/10.34726/hss.2022.84700	-
repositUm DOI:	10.34726/hss.2022.84700	-
CatalogPlus:	AC16417763	-
Publication Type:	Thesis - Diploma Thesis Hochschulschrift - Diplomarbeit	en de
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	en
Abstract:	Die gegensätzlichen Ziele über Schutz und Erhalt der Kontrolle über sensitive 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 reformen erregene aufgesetzt und	en

Annotations on the left side of the screenshot:

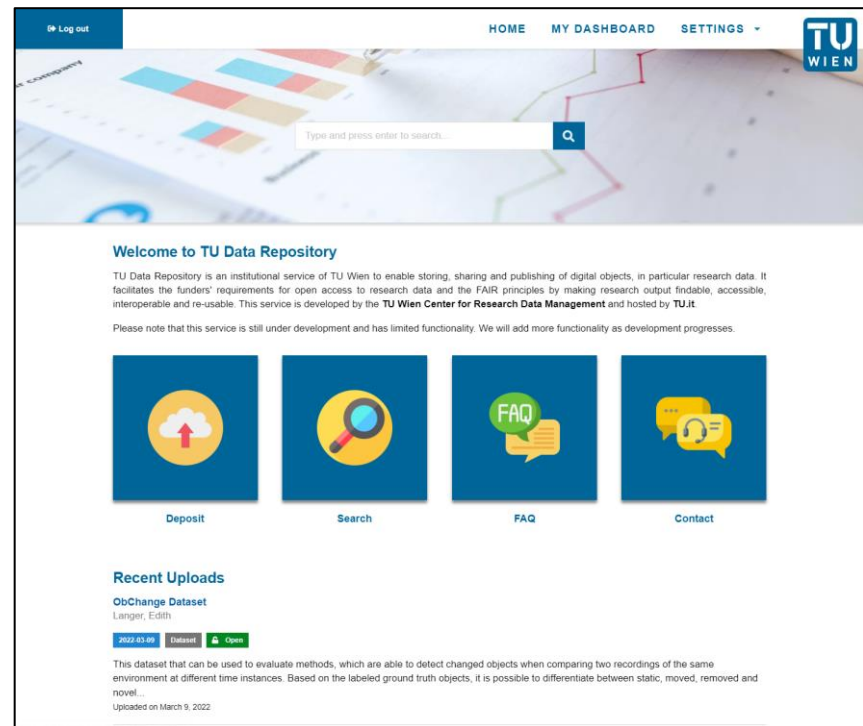
- Title of the dataset** points to the **Title** field.
- Citation** points to the **Citation** field.
- Persistent identifier (DOI)** points to the **repositUm DOI** field.

Annotations on the right side of the screenshot:

- Download** points to the **Download(s)** widget.
- External Systems** points to the **Google Scholar™ Check** widget.

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

The screenshot shows a dataset page on the TUWRD platform. The title is "The Sentinel-1 Global Backscatter Model (S1GBM) - Mapping Earth's Land Surface with C-Band Microwaves". The page includes a citation in APA style, a description of the dataset, and a list of files for download. The files include a preview image and two large zip files: "S1GBM_VH_mean_mosaic_v1_EQU17_AF010M.zip" (316.1 GiB) and "S1GBM_VH_mean_mosaic_v1_EQU17_AS010M.zip" (374.2 GiB).

Version of the dataset

Persistent identifier (DOI)

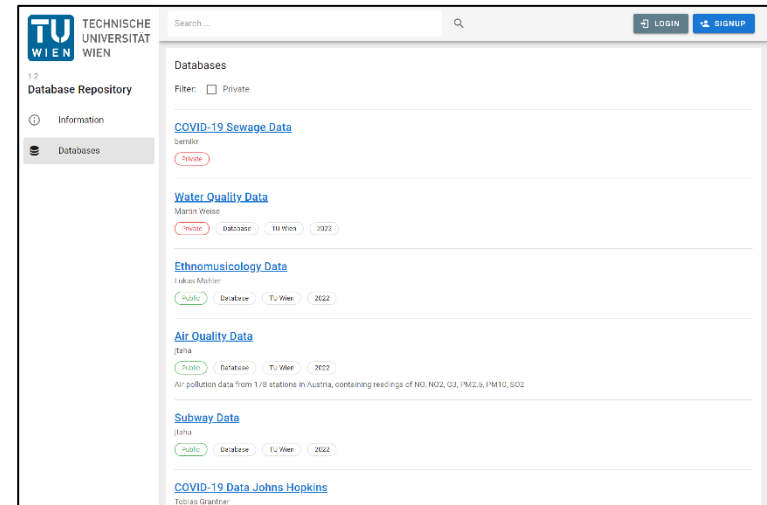
License
CC-BY-NC-SA-3.0

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



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

```
#
# 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()
```

```
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")
```



MQTT



AMQP

Exchange Type
Direct (AMQP)

Exchange Name
dbrepo.pilot_factory_data

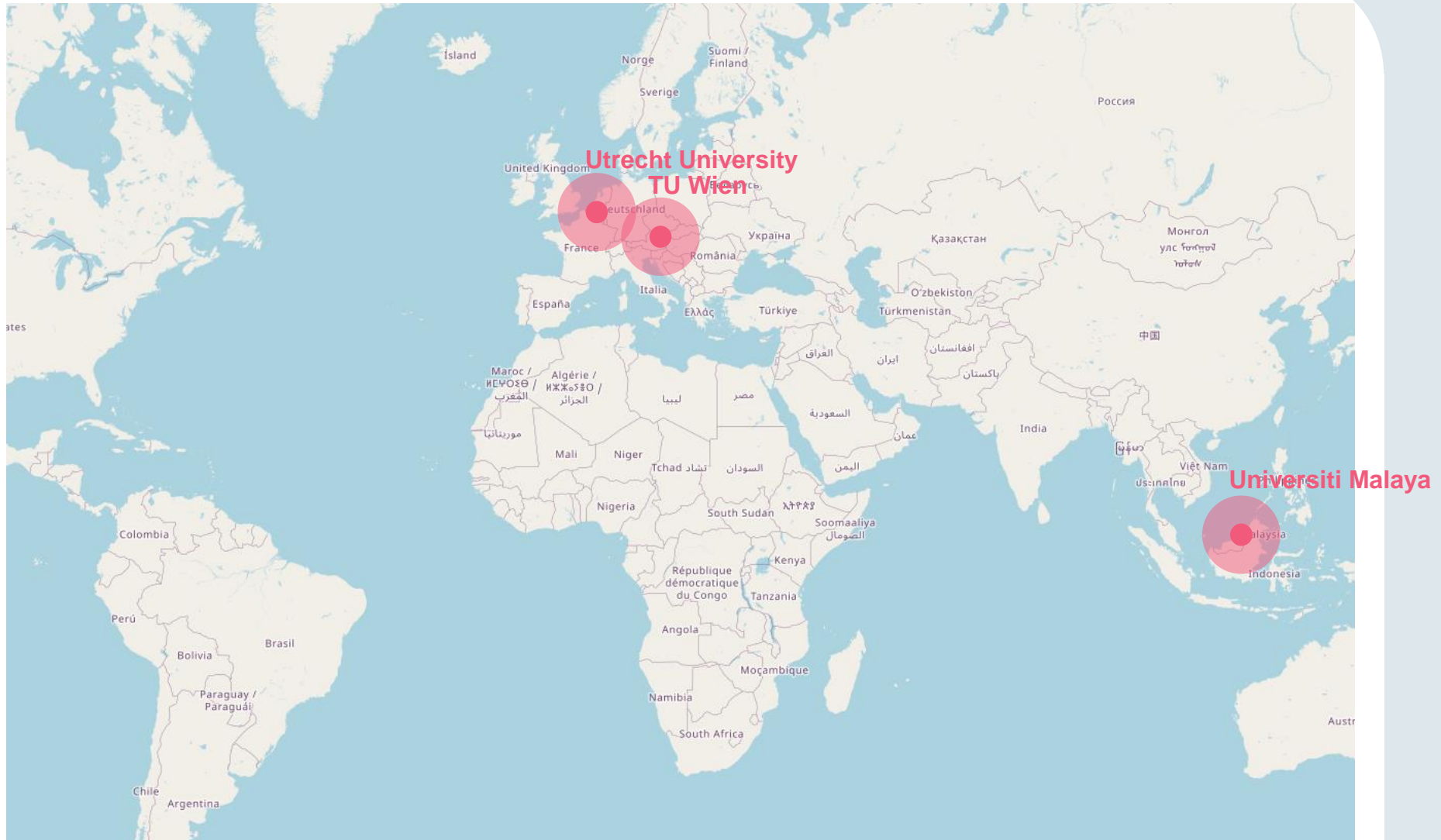
Queue Name
dbrepo.pilot_factory_data.power

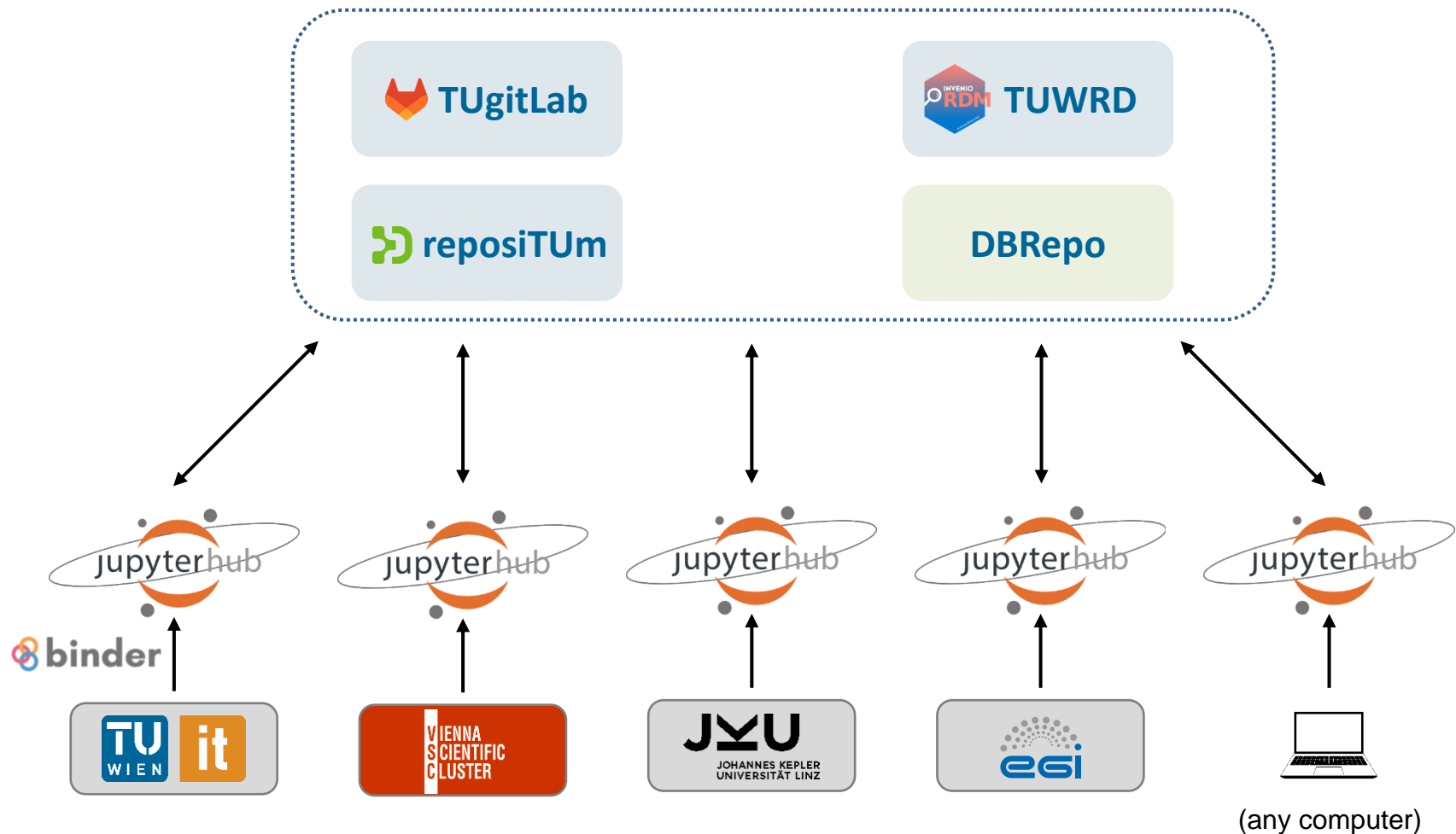
Routing Key
dbrepo.pilot_factory_data.power



PKID	Point	Value	Unit	Timestamp
<input type="checkbox"/> 0018a398-b7bc-4830-bfed-89e2f2fa4655	C	0	W	2023-02-10 23:56:46 (UTC)
<input type="checkbox"/> 008382f8-ab62-44a0-8809-49732450c2d0	C	0	W	2023-02-10 23:56:41 (UTC)
<input type="checkbox"/> 008b45ea-d96c-4d59-a246-8df26934548a	B	0	W	2023-02-10 23:56:37 (UTC)
<input type="checkbox"/> 009ea76b-8342-4a2e-825f-d10877ea5402	B	0	W	2023-02-10 23:56:39 (UTC)
<input type="checkbox"/> 00a8528a-3e38-4439-ad86-9950316c481d	B	0	W	2023-02-10 23:56:39 (UTC)
<input type="checkbox"/> 0179a42-9118-44ab-ac83-4b8e16cffff6	A	0	W	2023-02-10 23:56:35 (UTC)
<input type="checkbox"/> 01917216-90d0-48d0-b83e-1fa7766103e8	C	0	W	2023-02-10 23:56:44 (UTC)
<input type="checkbox"/> 0192dc36-13d0-420e-8678-ca78f6e77392	B	0	W	2023-02-10 23:56:48 (UTC)
<input type="checkbox"/> 01ec6f11-73bd-4b05-9ee8-8028a51ef106	B	0	W	2023-02-10 23:56:42 (UTC)
<input type="checkbox"/> 02adce0b-24f3-40d7-85c2-3187936bcc30	A	0	W	2023-02-10 23:56:40 (UTC)

DBRepo





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

Logout
Control Panel

File Edit View Insert Cell Kernel Widgets Help

Login

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

```

In [2]: > BASE_URL = "https://dbrepo1.ec.tuwien.ac.at"
        > USERNAME = "jtaha"
        > PASSWORD = "██████"

        token = requests.post(BASE_URL + "/api/auth", json={"user":

```

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]: > # constants
        > TRAIN_LOCATION_QUERY = 'SELECT `current_position_longitude`
        > STATION_LOCATION_QUERY = 'SELECT `stop_id`, `stop_latitude`
        > ARRIVALS_QUERY = 'SELECT `when`, `stop_name`, `stop_id`,

In [4]: > data = requests.put(BASE_URL + f"/api/container/4/database
        > "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

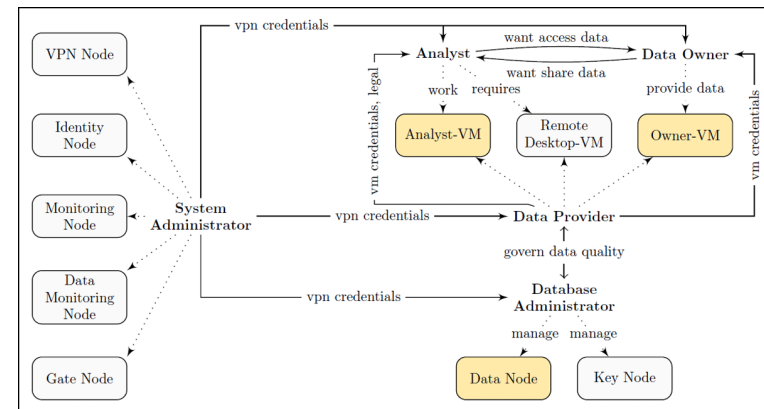
```

color_mapping = {'U1': '#d6150b', 'U2': '#771f73', 'U3': '#dd6d1a', 'U4': '#2579b7', 'U6': '#973222'}
fig = px.bar(waiting_time, x='stop_name', y='waiting_time', color='line name', color_discrete_map=
fig.update_layout(template='simple_white', yaxis_title='Average waiting time (min)', xaxis_title='S
fig.show()

```

Secure analysis environment

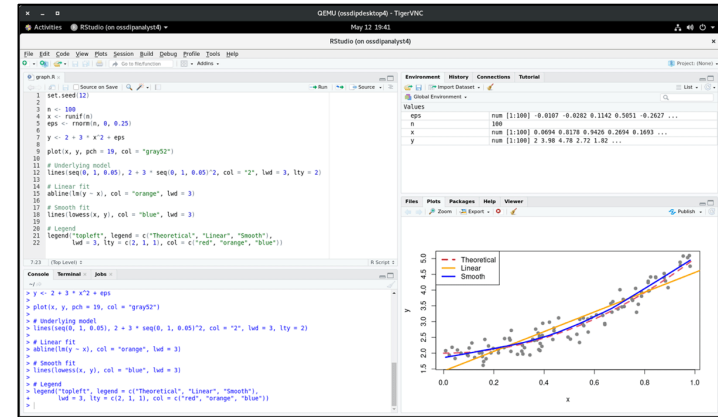
- TRE **model** and **reference implementation**
- Based on best-practice & open-source software
- Sensitive data (privacy issues, commercial interest), provide **access for analysis**, but ensure data is **not leaked** or misused
- Standard processes for involved roles



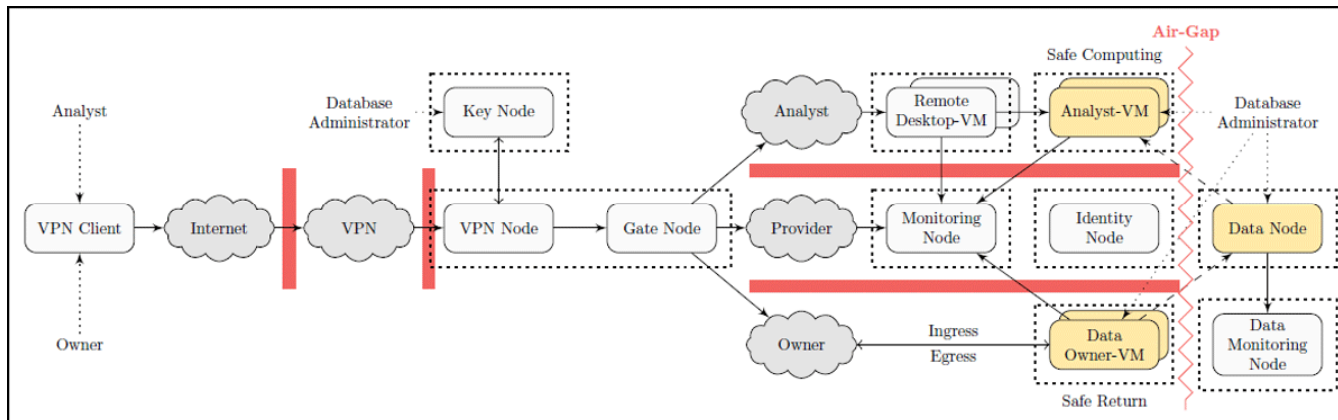
<https://ossdip.at/>

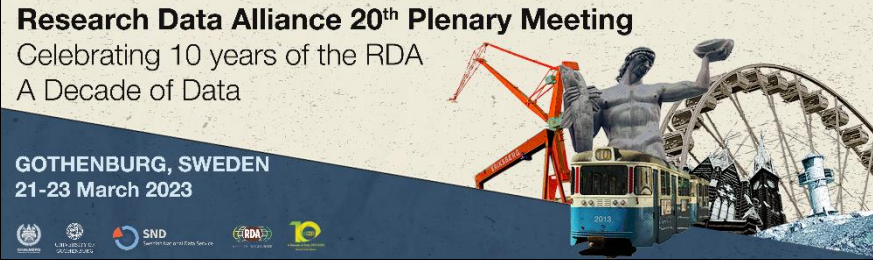
Data protection

- 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://ossdip.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>

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
 k if TREs think about threat models, in other words the broader
 y think the threat is coming from. Is their primary threat from

g of

etting up and operating TREs
 d requirements, risks



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna | Austria



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