

# Open Science and the EOSC: Building Blocks for FAIRness for Public and Sensitive Data

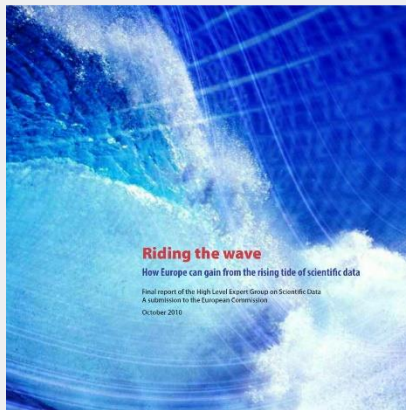
Andreas Rauber  
TU Wien

# Outline

- The EOSC is born
- Some building blocks:
  - DOIs
  - Machine-actionable Data Management Plans
  - Repositories
  - Data visiting for sensitive data
- Visions for the future

# The EOSC is born

- Many “birthdays”
  - Riding the Wave: October 2010
  - EOSC Launch, Vienna Declaration, Nov 2018
  - EOSC Association: December 2020
- What is the EOSC?



<https://www.dariah.eu/wp-content/uploads/2017/02/hlg-sdi-repordariaht.pdf>

## The Vienna Declaration on the European Open Science Cloud

Vienna, 23 November 2018

We, Ministers, delegates and other participants attending the launch event of the European Open Science Cloud (EOSC):

- 1 **Recall** the challenges of data driven research in pursuing excellent science as stated in the 'EOSC Declaration' signed in Brussels on 10 July 2017.
- 2 **Reaffirm** the potential of the European Open Science Cloud to transform the research landscape in Europe. Confirm that the vision of the European Open Science Cloud is that of a research data commons, inclusive of all disciplines and Member States, sustainable in the long-term.
- 3 **Recognise** that the implementation of the European Open Science Cloud is a process, not a project, by its nature iterative and based on constant learning and mutual alignment. Highlight the need for continuous dialogue to build trust and consensus among scientists, researchers, funders, users and service providers.
- 4 **Highlight** that Europe is well placed to take a global leadership position in the development and application of cloud services for Science. Reaffirm that the European Open Science Cloud will be both European and open to the world, reaching out over time to relevant global research partners.
- 5 **Recall** that the Council - in its conclusions of 29 May 2018 - welcomed the implementation roadmap and the federated model for the European Open Science Cloud. It invited the Commission and all Member States to set up a common governance framework that ensures participation of stakeholders from the research community based on principles of transparency, openness and inclusiveness and an effective involvement of all Member States.

<https://eosc-launch.eu/declaration/>



<https://www.eoscsecretariat.eu/news-opinion/eosc-association-first-general-assembly-17-december-2020>

# What is the EOSC?

# What is the EOSC?

- A toddler... (2 years? A few months?)

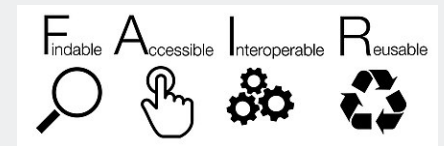
# What is the EOSC?

- **A toddler... (2 years? A few months?)**
- A toddler is a child approximately 12 to 36 months old [...]. **The toddler years are a time of great cognitive, emotional and social development.** The word is derived from "to toddle", **which means to walk unsteadily** [...]. (Wikipedia, <https://en.wikipedia.org/wiki/Toddler>)



# What is the EOSC?

- **A toddler... (2 years? A few months?)**
- A toddler is a child approximately 12 to 36 months old [...]. **The toddler years are a time of great cognitive, emotional and social development.** The word is derived from "to toddle", **which means to walk unsteadily** [...]. (Wikipedia, <https://en.wikipedia.org/wiki/Toddler>)
- The EOSC toddler should be(come) and produce things that are FAIR
- The EOSC toddler should, once grown up, provide us with joy, make our lives (as researchers) easier
- What are the things we want the EOSC to learn, to be able to perform or to provide as services?
- EOSC building blocks, or rather: services, skills, ... that are evolving to meet our needs and expectations
  - Some are familiar
  - Other less so.... and some we haven't even dreamt of yet...

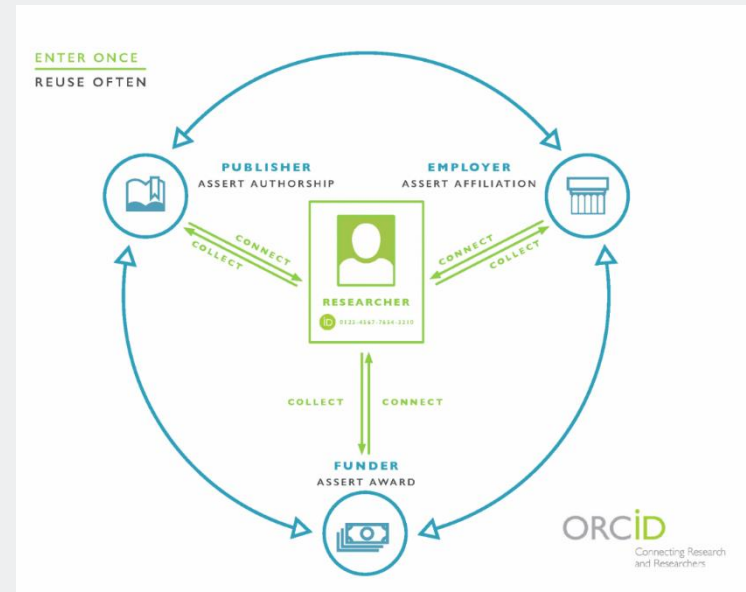


# Persistent Identifiers

- Digital Object Identifier (DOI)
  - Uniquely identify objects
  - Handle system
    - DOI assigned once
    - Physical location of data can change



- ORCID
  - Unique user ID





# Persistent Identifiers

## ■ DOI Service Austria, ORCID-Austria

### DOI Service Austria, ORCID Austria

To improve the visibility of Austrian researchers and their academic performance, TU Wien Bibliothek is leading two national initiatives: the DOI Service Austria and ORCID Austria. Not only should this raise awareness of the significance of persistent identifiers (PIDs) in academic communication, it should also create a community of practice.

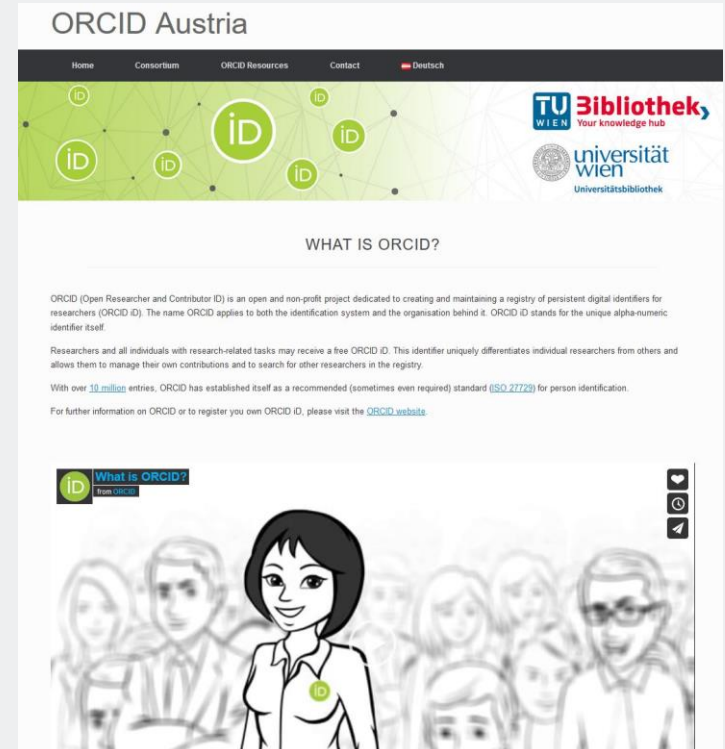
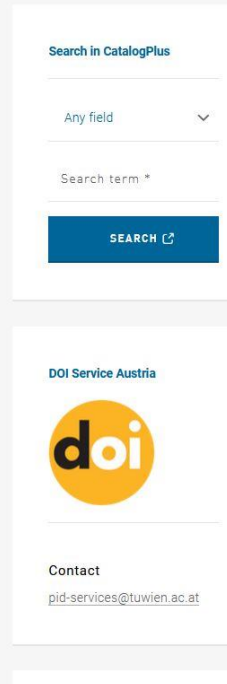
#### DOI Service Austria

Since January 2020, with the DOI Service Austria, TU Wien Bibliothek has been providing all Austrian universities, research institutions and other non-profit organisations in the research and education sector domiciled in Austria with an attractive opportunity to register and use Digital Object Identifiers (DOIs) to ensure stable retrieval of academic output via the internet.

For the first time, there is a central point of contact in Austria providing advice on the subject of DOIs, organising international developments and relaying information promptly to Austrian institutions. In order to be able to provide this service as a local authority, TU Wien Bibliothek is a member of the DataCite Association. DataCite is a DOI provider that focuses specifically on the persistent identification of objects stored in repositories and relies on uniform metadata.

Why use DOIs? DOIs are recognised and used internationally. The use of DOIs for research output published on the internet ensures reliable citations and promotes the visibility and stable findability of the document on the internet. The use of DOIs together with other persistent identifiers, such as ORCID iDs for authors and ROR for institutions, also enables improved, reliable and stable attribution of research output to particular persons, research facilities and institutions.

The DOI Service Austria enables Austrian institutions to use the [Fabrica](#) registration platform and the DataCite interfaces (MDS API, REST API); this enables both manual and automatic registration of DOIs. Customers of the DOI Service Austria receive the prefixes from us for the independent DOI assignment in the respective institutional repositories. TU Wien Bibliothek provides technical support as well as support for the quality assurance of metadata in Austrian information systems and the interoperability between IT applications. Fees for the DOI Service Austria are based on the DataCite cost model. Contact us for more details.



<https://www.tuwien.at/en/library/doi-service-austria-orcid-austria/>

<https://www.tuwien.at/kooperationen/orcid/>  
13 Institutions in Austria

# Repositories

- Many existing repositories
  - Federation?
  - Interoperability?
  - Sharing? Repository-as-a-Service?
  - A few new ones?
- **FAIR Data Austria**
  - Machine-actionable Data Management Plans (maDMPs)
  - Repositories
  - FAIR National Office, Data Stewards
  - Partners:
    - **Graz University of Technology (Lead)**
    - Technical University of Vienna
    - University of Vienna
    - University of Innsbruck
    - Medical University of Graz
    - Academy of Fine Arts Vienna
    - and **23 associated partners**

CLUSTER FORSCHUNGSDATEN

FAIR DATA AUSTRIA

NEWS ABOUT FAIR DATA AUSTRIA ORGANIZATION MATERIALS PARTNERS CONTACT US

Workshop FAIR National Office

Running Time: January 2020 to December 2022  
Project Lead: TU Graz

The FAIR Data Austria project is designed to strengthen knowledge transfer between universities, industry, and society and supports the sustainable implementation of the European Open Science Cloud (EOSC). Within the project, implementation of the FAIR principles (which mandate that research data be Findable, Accessible, Interoperable, and Reusable) plays a major role. Observation of the FAIR principles is secured through 1) integrated data management aligned with generic and discipline-specific needs of researchers, 2) development of next-generation repositories for research data, code, and other research outputs, and 3) development of training and support services for efficient research data management. FAIR Data Austria thereby offers tools to complement the Austrian Data Lab and Services as well as RIS Synergy projects.

Supporting the entire data lifecycle – from data generation all the way to data archiving – with the appropriate tools and expertise is essential to achieve efficient research data management according to the FAIR principles, a process that can only be successful when supported by all Austrian HEIs. The FAIR Data Austria project therefore supports the collaboration of Austrian universities in developing coherent services for research data, thereby securing Austria's position within the international research landscape.

NEWS

**FAIR Data Austria Online Project Meeting**  
On November 23, 2020, an online project meeting (via WebEx) was held, first, to inform members of the Cluster Forschungsdaten, associated partners and all researchers  
Read More »  
7. January 2021

**FAIRization in Austria: A preliminary concept**  
The FAIR Office Austria will help aligning tools and services for research data management with the FAIR principles. The corresponding workshop report is now available.  
Read More »  
23. November 2020

**Workshop: Tasks and Profile of a Data Steward**

**FAIR Data Austria Online Project Meeting**

TU W I E N TECHNISCHE UNIVERSITÄT WIEN

universität wien

universität innsbruck

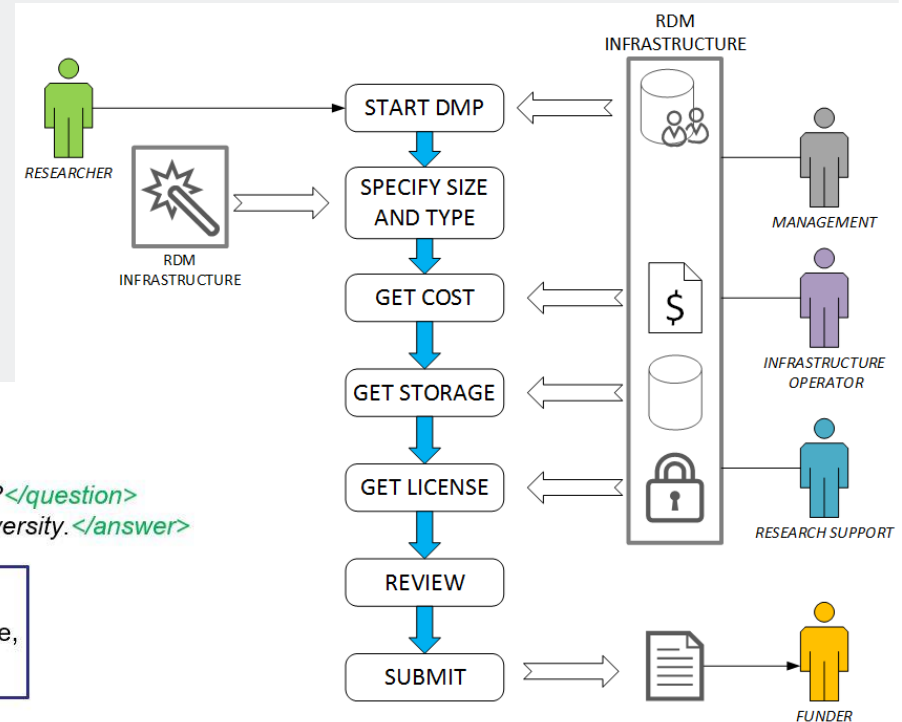
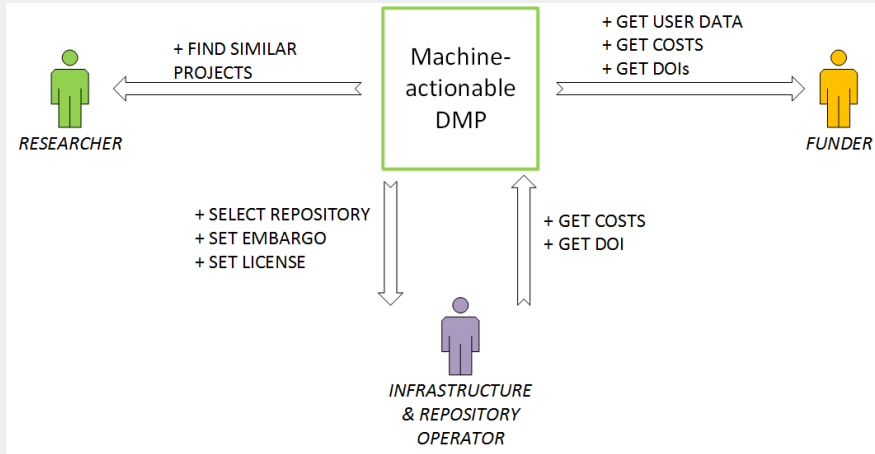
MU Medizinische Universität Graz

ja akademie der bildenden künste wien

# Machine-Actionable Data Management Plans (maDMPs)

- Writing DMPs is tedious work
  - Researchers do not like this
    - It's cumbersome
    - It's error-prone
  - Institutions / repositories don't like it
    - It's error-prone
    - It's natural language text -> no automation of processes
- Document-based DMPs are of limited usefulness
  - “Awareness-raising”
  - Hard to verify or act upon -> “promises”
- Need **automation** of DMP creation and processing
- Need **machine-actionable** DMPs (**maDMPs**)
- RDA WG DMP Common Standards  
<https://www.rd-alliance.org/groups/dmp-common-standards-wg>

# Machine-Actionable Data Management Plans (maDMPs)



- Current DMPs – model questionnaires

```
<administrative_data>
  <question>Who will be the Principle Investigator?</question>
  <answer>The PI will be John Smith from our university.</answer>
</administrative_data>
```

- Machine-actionable DMPs –

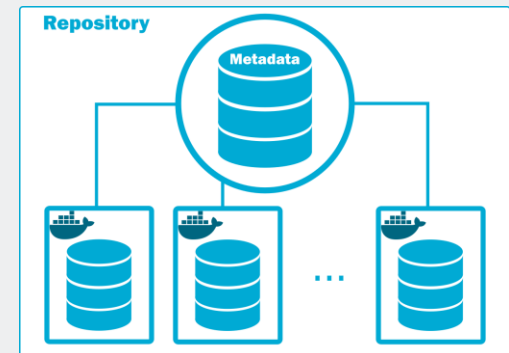
```
"dc:creator":[ {
  "foaf:name": "John Smith",
  "@id": "orcid.org/0000-1111-2222-3333",
  "foaf:mbox": "mailto:jsmith@tuwien.ac.at",
  "institution": "AT-Vienna-University-of-Technology"
}]
```

Use PIDs whenever possible, e.g. ORCID

Reuse existing standards, e.g. Dublin Core, PREMIS, etc.

# FDA Repositories

- FAIR Data Austria Repositories
  - Gitlab: source code, versioning  
(Lead: TU Wien)
  - Invenio (aka “Zenodo”)  
(Lead: TU Graz)
  - Database Repositories  
(Lead: University of Vienna)



# FDA Repositories - Invenio



The screenshot shows a web browser window with the URL <https://researchdata.dl.hpc.tuwien.ac.at>. The page has a dark blue header with the TU WIEN logo on the left and a "Log in" button on the right. Below the header, the text "TU Data" is centered, followed by a search bar with the placeholder text "Type and press enter to search". The main content area is white and features the heading "Welcome to TU Data". Below this heading, there is a paragraph of text: "TU Data is an institutional repository to enable storing, sharing and publishing of research data. It facilitates the funders' requirements for open access to research data and the FAIR principles by making research output findable, accessible, interoperable and re-usable. Please note that this service is still under development and has limited functionality. We will add more functionality as development progresses." Below the text, there is a light blue section with the heading "Brought to you by TU Wien" and a paragraph: "This repository is developed in collaboration with Austrian universities participating in the FAIR Data Austria project." At the bottom of this section is the logo for FAIR DATA AUSTRIA, which consists of four vertical orange bars of varying heights followed by the text "FAIR DATA AUSTRIA". The footer is a dark blue bar with the following text: "TU Wien", "Center for RDM", "RDM Policy", "FAIR Data Austria", and "FAIR Data Austria".



# FDA Repositories - Invenio



https://researchdata.dl.hpc.tuwien.ac.at/records/tkksf-11b75

**TU WIEN** Search Log in

January 19, 2021 | Version 1.0 Dataset Open Access

## European Sentinel-1 Forest Type and Tree Cover Density Maps

Dostalova, Alena<sup>1</sup>; Cao, Senmao<sup>1,2</sup>; Wagner, Wolfgang<sup>1,2</sup> [show affiliations](#)

**Description**  
This dataset was generated by the TU Wien Department of Geodesy and Geoinformation.

European Sentinel-1 forest type and tree cover density maps represent first continental-scale forest layers based on Sentinel-1 C-Band Synthetic Aperture Radar (SAR) backscatter data. For the year 2017 they cover the majority of European continent with 10 m and 100 m sampling for forest type and tree cover density, respectively. The maps were derived using the method described in <https://www.tandfonline.com/doi/full/10.1080/101431161.2018.1479788>.

The forest type map shows the dominant forest type class (coniferous, broadleaf). Tree cover density map shows the percentage of forest canopy cover within the 100 m pixel.

Please be referred to our peer-reviewed article at <https://doi.org/10.3390/rs13030337> for details and accuracy assessment across Europe.

### Dataset Record

The forest type and tree cover density maps are sampled at 10 m and 100 m pixel spacing respectively, georeferenced to the Equi7Grid and divided into square tiles of 100km extent ("T1"-tiles). With this setup, the forest maps consist of 728 tiles over the European continent, with data volumes of 3.12 GB and 378.3 MB.

The tiles' file-format is a LZW-compressed GeoTIFF holding 16-bit integer values, with tagged metadata on encoding and georeference. Compatibility with common geographic information systems as QGIS or ArcGIS, and geodata libraries as GDAL is given.

In this repository, we provide each forest map as tiles, whereas two zipped dataset-collections are available for download below.

### Code Availability

For the usage of the Equi7Grid we provide data and tools via the python package available on <https://github.com/TUW-Geo/Equi7Grid>. More details on the grid reference can be found in <https://www.sciencedirect.com/science/article/pii/S0098300414001629>.

### Acknowledgements

The computational results presented have been achieved using the Vienna Scientific Cluster (VSC)

https://researchdata.dl.hpc.tuwien.ac.at/records/tkksf-11b75

ForestType.zip

- ForestType
  - 2017\_FOREST\_CLASSES\_EU010M\_E029N009T1.tif 1.0 MB
  - 2017\_FOREST\_CLASSES\_EU010M\_E029N010T1.tif 670.2 kB
  - 2017\_FOREST\_CLASSES\_EU010M\_E030N009T1.tif 4.4 MB
  - 2017\_FOREST\_CLASSES\_EU010M\_E030N010T1.tif 7.5 MB
  - 2017\_FOREST\_CLASSES\_EU010M\_E030N011T1.tif 8.0 MB
  - 2017\_FOREST- 5.5 MB

Files (3.5 gb)

Name	Size	Preview	Download
ForestType.zip md5:5564390a0f93a3b209d0f1e0852d0	3.1 GB	<a href="#">Preview</a>	<a href="#">Download</a>
TreeCoverDensity.zip md5:80d815404738660b3e0d0f3e0d149	378.3 MB	<a href="#">Preview</a>	<a href="#">Download</a>

**Details**

**Licenses**

**Resource type** Dataset

**Formats** application/x-geotiff

**Related identifiers**

- isreferencedby** 10.3390/rs13030337 ( doi )
- issupplementto** 10.5281/zenodo.3515933 ( doi )  
<https://github.com/TUW-Geo/Equi7Grid> ( url )
- references** 10.1080/101431161.2018.1479788 ( doi )  
10.1016/j.cageo.2014.07.005 ( doi )

Created: January 19, 2021 | Modified: February 6, 2021

[Jump up](#)

**Invenio-madMP**  
Invenio module for integrating madMPs.

**Authors**  
Invenio module for madMP integration.  
• FAIR Data Austria <[maximilian.moser@tuwien.ac.at](mailto:maximilian.moser@tuwien.ac.at)>

**Navigation**

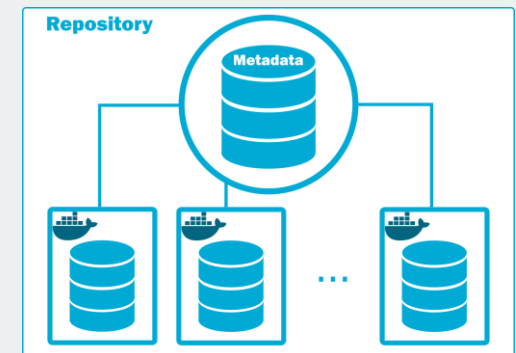
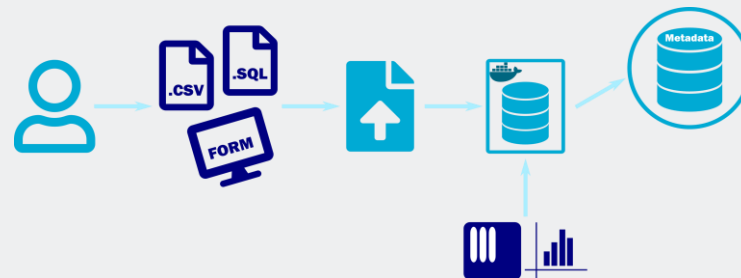
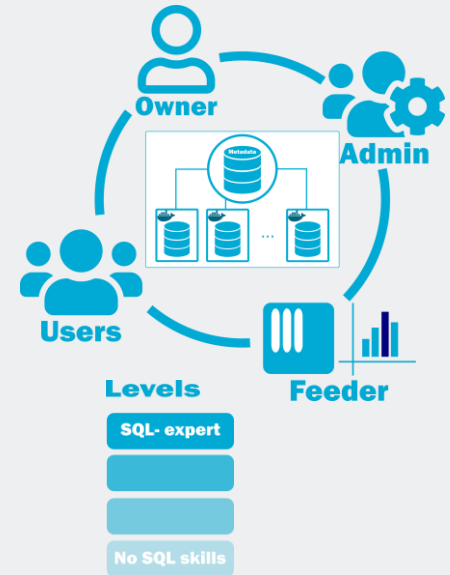
- Installation
- Configuration
- Usage
- API Docs
- Contributing
- Changes
- License
- Authors

[invenio-madmp@GitHub](mailto:invenio-madmp@GitHub)  
[invenio-madmp@PyPI](mailto:invenio-madmp@PyPI)

*For more on GitHub*

# FDA Repositories – DB-Repo: Vision

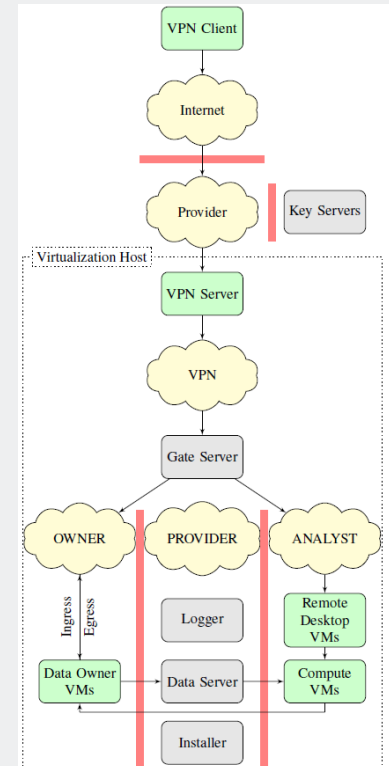
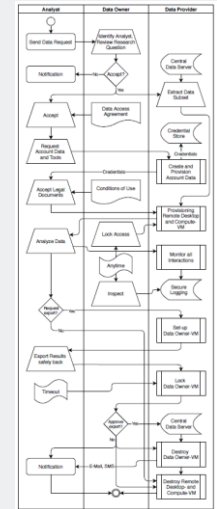
- Structured data
- Private cloud hosted relational databases
- DB is created directly in repository framework
- DB is populated and used within repository
- Metadata is generated and exposed
- Databases and data are searchable
- Data is versioned & time-stamped: reproducibility, re-use, provenance
- Data is cite-able at arbitrary levels of granularity (RDA WGDC recommendations)
- Data Management outsourced to repository infrastructure: easier for researchers, higher quality data mgt, higher security, ...





# Secure Data Infrastructure (OSSDIP)

- **FAIRness** for “closed” data!
- Sensitive Data (privacy, IPR, ...)
- **Data Visiting** instead of Data Sharing
- **Data owner maintains full control over data and use:**
  - Access by **whom**, for which period of **time**,
  - to which **subset** of data
  - for which analysis **goal** / research question
- Data infrastructure acts as data processor
- Secured IT system
  - Air-gapped virtual machines with data excerpts
  - Access solely via remote desktop
  - Complete monitoring of all interactions
- Controlled processes
- Data identification, dynamic citation, reproducibility
- Open source reference implementation



# EOSC – Open Science: Visions for Future Research Environments

**For EOSC (and other infrastructures) to be useful we need to:**

- understand the needs of stakeholders
  - What will **I as a researcher** need to do high-quality research efficiently 10-15 years from now?
  - What will **society** need to trust in and engage with research, contributing to and benefiting from it 10-15 years from now?
  - What would I like the EOSC to be?
- identify the barriers and opportunities
- **imagine possible futures!**
- have a vision for the services, tools and rules we will need
  - 2-5 years from now
  - 10-15 years from now
- Workshop and interview series with researchers  
(Nobel laureates, ERC grant holders, SciFi authors, ...)

# Visions for Future Research Environments



**Visions, needs and requirements for (future) research environments**  
An exploration series with researchers by TU Wien 

- <https://www.tuwien.at/en/research/rti-support/research-data/news/news/future-research-environments-an-exploration-series-with-researchers>
- <https://www.eoscsecretariat.eu/news-opinion/visions-needs-requirements-computer>
- <https://www.eoscsecretariat.eu/news-opinion/visions-needs-and-requirements-karl-von-wendt>
- More to come...

# EOSC – Open Science: Visions for Future Research Environments

Some services and concepts identified:

- Data capture, pre-processing
- Trust in data, Provenance
- Trust in AI, Explainability
- Automated form filling
- Automated writing
- Automated coding
- Translation beyond languages: disciplines, levels-of-expertise
- Mechanisms to save-guard against monopolies, data colonialism
- Mechanisms to balance cooperation vs. competition
- ...

# Visions for Future Research Environments

**For EOSC (and other infrastructures) to be useful:**

What will **I as a researcher** need to do high-quality research efficiently 10-15 years from now?

What will **society** need to trust in and engage with research, contributing to and benefiting from it 10-15 years from now?

What would I like the EOSC to be?  
Which building blocks do I need?