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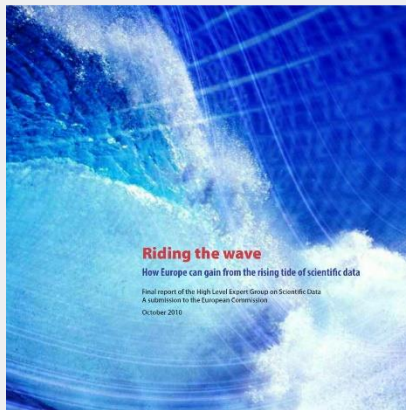
Andreas Rauber

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?

Riding the wave
How Europe can gain from the rising tide of scientific data

Final report of the High Level Expert Group on Scientific Data
A submission to the European Commission
October 2010

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/

BUILDING THE EOSC ASSOCIATION
17 December 2020 - First General Assembly

Cathrin Stöver
EOSC Executive Board Co-Chair

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

Informatics **ifs**

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)

Informatics ifs

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…

Findable Accessible Interoperable Reusable

TU WIEN Informatics ifs
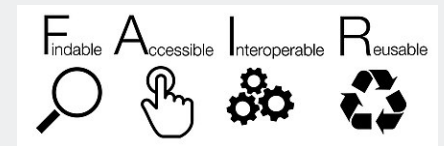
Persistent Identifiers

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

- ORCID
  - Unique user ID
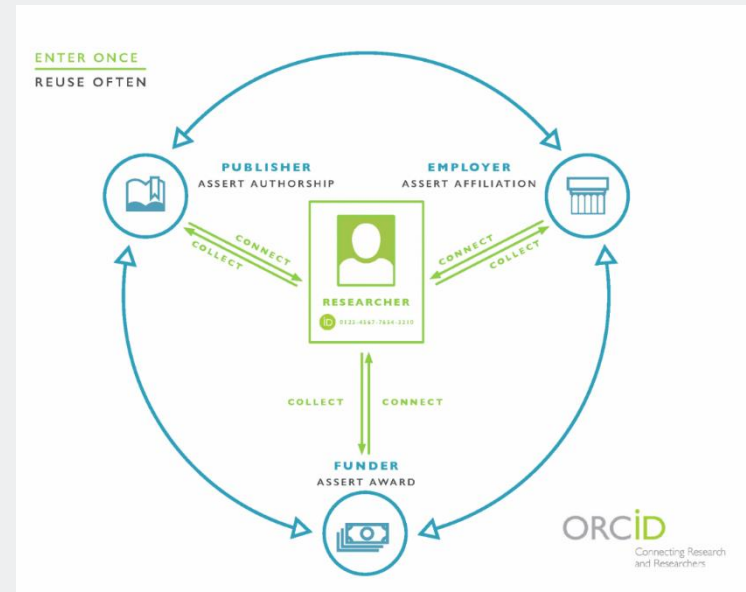
Persistent Identifiers

- DOI Service Austria, ORCID-Austria

13 Institutions in Austria

TU WIEN Informatics ifs

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**

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)
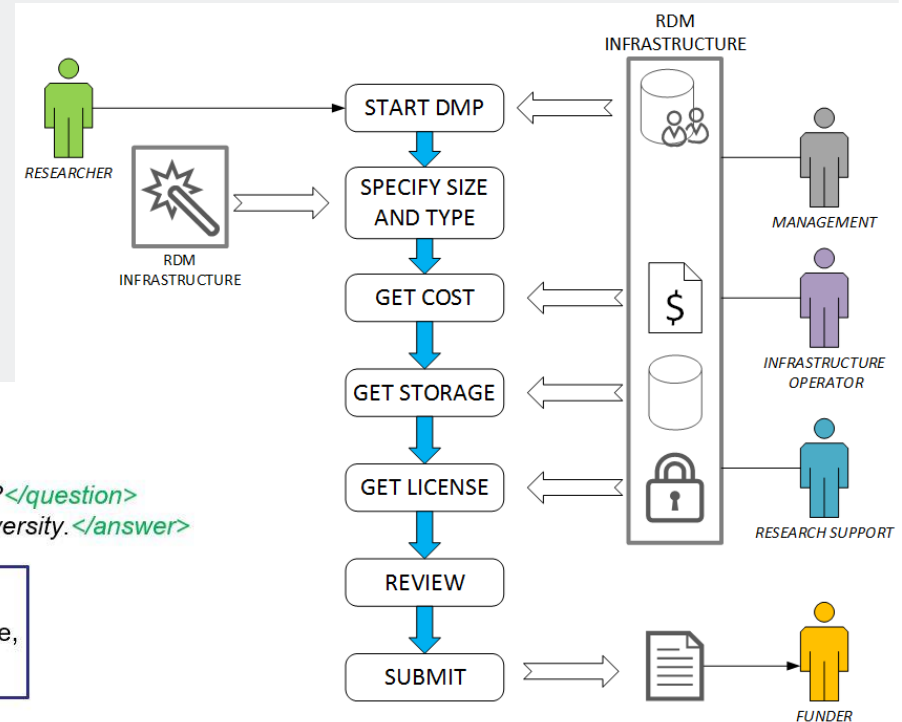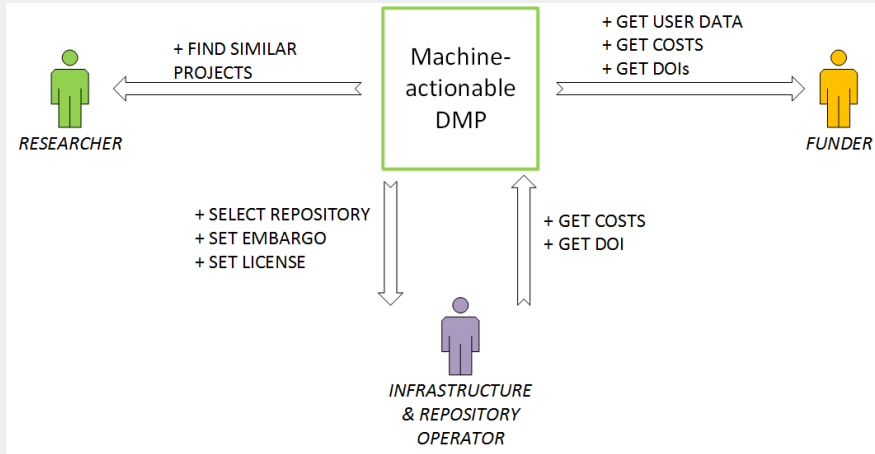
- Current DMPs – model questionnaires

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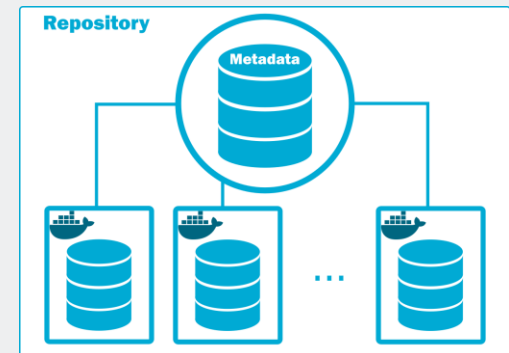
- Machine-actionable DMPs –

```
"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

FDA Repositories - Invenio

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, …
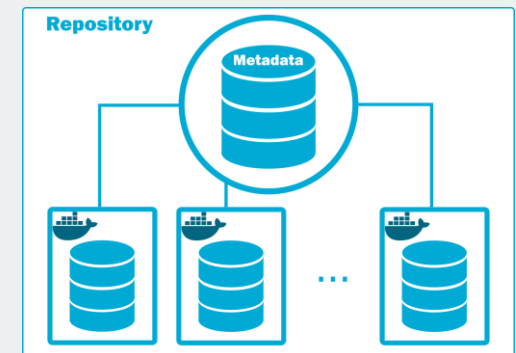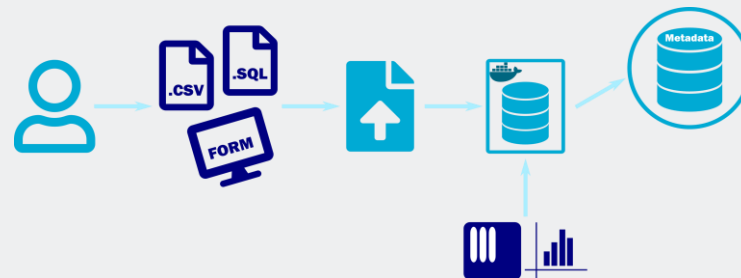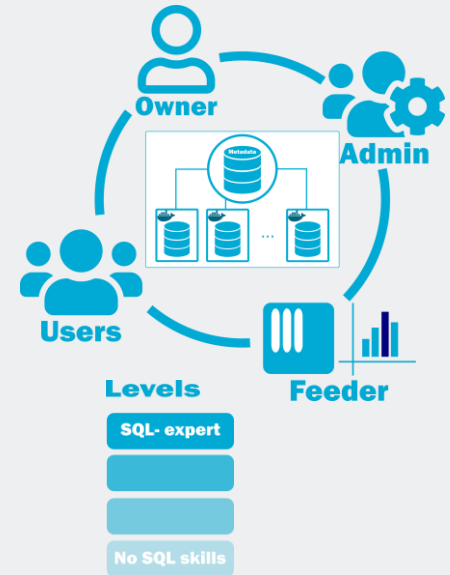
Secure Data Infrastructure (OSSDIP)

- **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
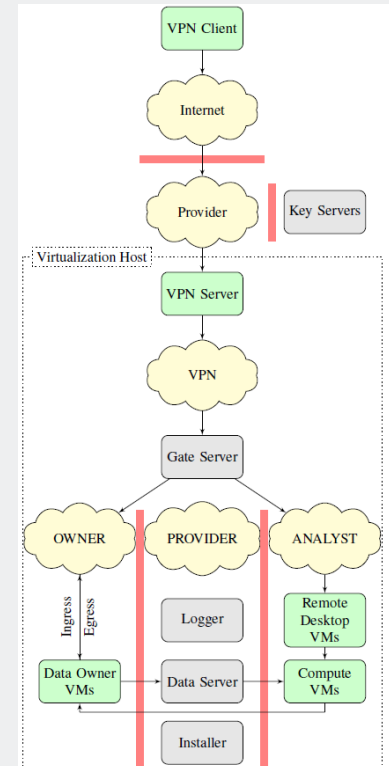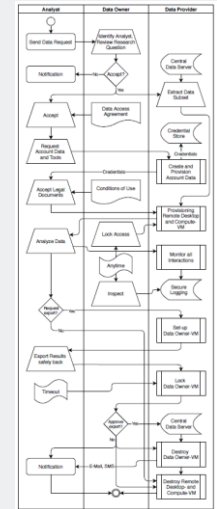
EOSC – Open Science:
Visions for Future Research Environments

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, …)

TU WIEN Informatics ifs

Visions for Future Research Environments

- 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…

TU WIEN Informatics ifs

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

- …

TU WIEN Informatics ifs

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?

...