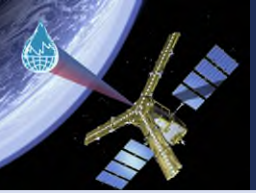


# FRM4SM Data traceability

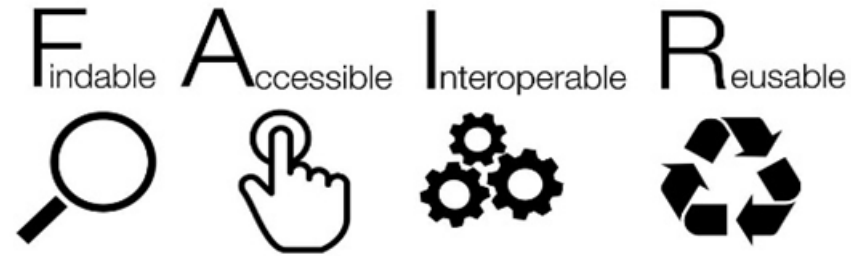
Fiducial Reference Measurement for Soil Moisture II May 2021 – May 2023

Irene Himmelbauer<sup>1</sup>, **Moritz Staudinger**<sup>1</sup>, **Tobias Hasjan**<sup>1</sup>, Daniel Aberer<sup>1</sup>, Alexander Gruber<sup>1</sup>, Wolfgang Preimesberger<sup>1</sup>, Pietro Stradiotti<sup>1</sup>, Wouter Dorigo<sup>1</sup>, Monika Tercjak<sup>2</sup>, Alexander Boresch<sup>2</sup>, Arnaud Mialon<sup>3</sup>, Francois Gibon<sup>3</sup>, Philippe Richeaume<sup>3</sup>, Yann Kerr<sup>3</sup>, Raul Diez Garcia<sup>4</sup>, Raffaele Crapolicchio<sup>4</sup>, Roberto Sabia<sup>4</sup>, Klaus Skipal<sup>4</sup>, Philippe Goryl<sup>4</sup>



# Data traceability -- FAIR principles

[1] <https://www.go-fair.org/fair-principles/>



Several levels on how to achieve traceability – gives flexibility

- Overall goal is that data changes are tracked
- Easy readable instructions on how to find data

→ *“(digital) Persistent Identifier (PI or PID)= a label which gives a unique name to an entity (a person, place, or thing) – pointing reliably to a digital entity “*

Myriad of different Persistent Identifiers (PIDs) for individual purposes/fields (e.g., BOOK → ISBN ID, researchers → ORCID ID, etc.)

- Several options can exist even within on field e.g., „research data“
  - Digital Object Identifiers (DOIs) --- costs money, well known
  - the Handle System --- free, not really known
  - Etc.

Being open source (no obstacles when fetching data – registration/login)

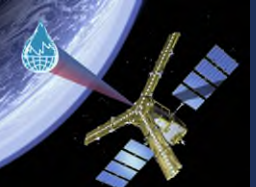
- not a must

<sup>1</sup> TU Wien: [climers.geo.tuwien.ac.at](http://climers.geo.tuwien.ac.at) & Informatics Department

<sup>2</sup> AWST: [www.awst.at](http://www.awst.at)

<sup>3</sup> CESBIO: [www.cesbio.cnrs.fr](http://www.cesbio.cnrs.fr)

<sup>4</sup> ESA: <https://earth.esa.int/eogateway>



# Several levels of FAIR

[1] <https://www.go-fair.org/fair-principles/>

## FAIR Principles

## Compliance



### Findability

Resource and its metadata are easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.

- ✓ F1. Resource is uploaded to a public repository.
- ✓ F2. Metadata are assigned a globally unique and persistent identifier.



### Accessibility

Resource and metadata are stored for the long term such that they can be easily accessed and downloaded or locally used by humans and ideally also machines using standard communication protocols.

- ✓ A1. Resource is accessible for download or manipulation by humans and is ideally also machine readable.
- ✓ A2. Publications and data repositories have contingency plans to assure that metadata remain accessible, even when the resource or the repository are no longer available.



### Interoperability

Metadata should be ready to be exchanged, interpreted and combined in a (semi)automated way with other data sets by humans as well as computer systems.

- ✓ I1. Resource is uploaded to a repository that is interoperable with other platforms.
- ✓ I2. Repository meta- data schema maps to or implements the CG Core metadata schema.
- ✓ I3. Metadata use standard vocabularies and/or ontologies.



### Reusability

Data and metadata are sufficiently well-described to allow data to be reused in future research, allowing for integration with other compatible data sources. Proper citation must be facilitated, and the conditions under which the data can be used should be clear to machines and humans.

- ✓ R1. Metadata are released with a clear and accessible usage license.
- ✓ R2. Metadata about data and datasets are richly described with a plurality of accurate and relevant attributes.

- PIDs
- Public Repositories: ZENODO, etc.
- Long term storage of metadata and resource
- Machine readability / accessible to humans
- Open source – no restrictions for fetching data
- Data & its metadata built upon (community agreed) standards
- Facility & data user metadata built on FAIR principles
- Traceable Data changes – tombstone identifiers
- Complete reproducibility of data (for a long period of time)

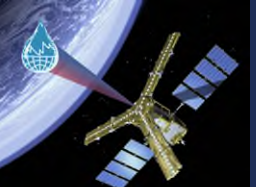
<sup>1</sup> TU Wien: [climers.geo.tuwien.ac.at](http://climers.geo.tuwien.ac.at) & Informatics Department

<sup>2</sup> AWST: [www.awst.at](http://www.awst.at)

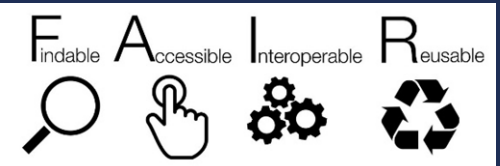
<sup>3</sup> CESBIO: [www.cesbio.cnrs.fr](http://www.cesbio.cnrs.fr)

<sup>4</sup> ESA: <https://earth.esa.int/eogateway>





# ISMN as an example: How to go FAIR?



## ISMN FAQs:

- No PIDs for data

- Data distribution = copy right – scientific use only – no onward distribution

- ✓ Free registration/login → free download
  - level of security needed
  - honest tracing of user statistics → provider reports

- ✓ ISMN built upon community agreed standards - steadily adapted

- No versioning of data changes -- data sometimes reprocessed and overwritten – news entries

## Questions:

- How to create PIDs then ?
- What are the differences here?

- How to deal with the “no onward distribution“?
- Do we even want to publish the data openly at another platform?
  - What does this mean for the statistics ?
    - Provider reports → funding base for the data provider
  - What does this mean for funding the ISMN?
  - Is that really fair?

- Can we adapt the system to track data changes better?
- How to deal with nrt updates / regular- irregular updates?

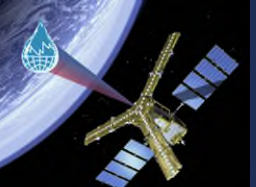
**CAN THE ISMN EVEN BE FAIR?**

<sup>1</sup> TU Wien: [climers.geo.tuwien.ac.at](http://climers.geo.tuwien.ac.at) & Informatics Department

<sup>2</sup> AWST: [www.awst.at](http://www.awst.at)

<sup>3</sup> CESBIO: [www.cesbio.cnrs.fr](http://www.cesbio.cnrs.fr)

<sup>4</sup> ESA: <https://earth.esa.int/eogateway>



# Can the ISMN even be FAIR?

**Answer: YES**

**Informatics' Department - Technical University of Vienna (TUW)**

- **Professor Andreas Rauber**
- Co chair of: Working Group on Data Citation (WGDC)



References ISMN system currently built upon:

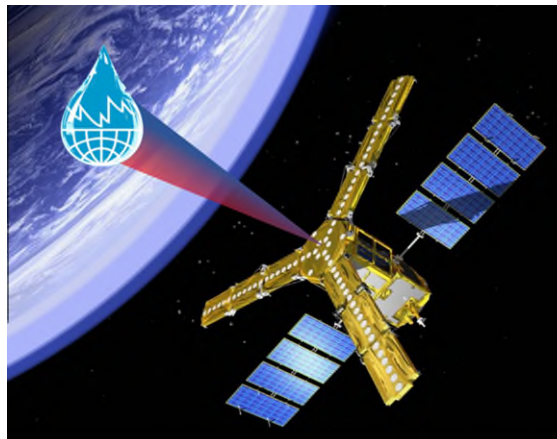
- [1] <https://www.go-fair.org/fair-principles/>
- [2] <https://zenodo.org/record/1406002/files/datacitation.pdf?download=1>
- [3] <https://hdr.mitpress.mit.edu/pub/si7wzxxa/release/1>

Going FAIR / making data traceable without not wanting to OR not being able to outsource the data

- Query based PIDs --- PIDs attached to download query not to data [3]
- Versioning of own systematic --- full reproducibility / track changes → tombstone identifiers [3]
- Recommendations for different levels on how to achieve that [2]

**ISMN is currently restructured towards:**

- Versioning of data changes [3]:
  - full reproducibility of soil moisture data
  - Metadata currently evaluated, either:
    - Full reproducibility
    - OR
    - Tombstone Identifiers: giving indication when and what has changed
- DOI system of data download requests = queries
- Data accessibility stays with ISMN



Thank you for your attention!

<https://project-frm4sm.geo.tuwien.ac.at/>

### Contact

Project lead, QA4SM: [Boresch@awst.at](mailto:Boresch@awst.at)

Science lead: [wouter.dorigo@tuwien.ac.at](mailto:wouter.dorigo@tuwien.ac.at)

SMOS expert: [arnaud.mialon@univ-tlse3.fr](mailto:arnaud.mialon@univ-tlse3.fr)

ESA main contact: [Raffaele.Crapolicchio@esa.int](mailto:Raffaele.Crapolicchio@esa.int)

ISMN coordination and EO scientist: [irene.himmelbauer@tuwien.ac.at](mailto:irene.himmelbauer@tuwien.ac.at)

Irene Himmelbauer<sup>1</sup>, Moritz Staudinger<sup>1</sup>, Tobias Hajszan<sup>1</sup>, Daniel Aberer<sup>1</sup>, Alexander Gruber<sup>1</sup>, Wolfgang Preimesberger<sup>1</sup>, Pietro Stradiotti<sup>1</sup>, Wouter Dorigo<sup>1</sup>, Monika Tercjak<sup>2</sup>, Alexander Boresch<sup>2</sup>, Arnaud Mialon<sup>3</sup>, Francois Gibon<sup>3</sup>, Philippe Richeaume<sup>3</sup>, Yann Kerr<sup>3</sup>, Raul Diez Garcia<sup>4</sup>, Raffaele Crapolicchio<sup>4</sup>, Roberto Sabia<sup>4</sup>, Klaus Skipal<sup>4</sup>, Philippe Goryl<sup>4</sup>