

Solidipes: Open Science Through Collaborative Data Curation and Peer Review

Guillaume Anciaux

LSMS, IIC, ENAC, EPFL



ETH BOARD



Graphic from [PHD Comics](#)

FAIR Principles

What is **FAIR** (data) ?

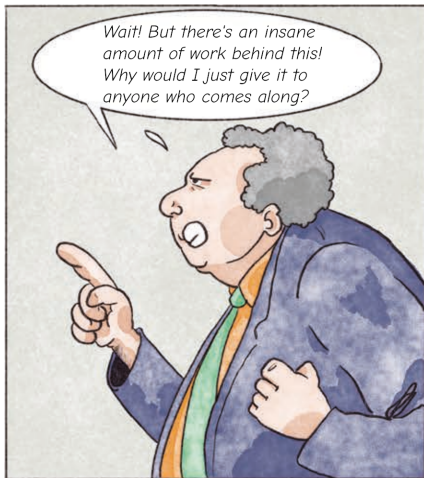
- **Findable:** metadata, permanent links, searchable databases
- **Accessible:** Open and online services, archiving (long retention)
- **Interoperable:** standards and file formats and vocabularies/ontologies
- **Reusable:** Open (source) licenses, (software) context.

Demands data curation

Definition: **Curation**

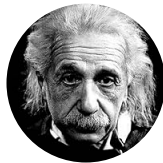
Organizing, cleaning, annotating dataset, ensuring relevance and reusability.

Why FAIR-ness ?



Translated from

<https://zenodo.org/records/10108736>



*"The right to search for the truth
implies also a duty..."*

*... one must not conceal any
part of what one has recognized
to be the truth."*

But also, making *Machine Learning Models*...

S. Leonelli. *Philosophy of open science.*

Cambridge University Press. 2023.

*[...] the UNESCO (2021) Recommendation on Open Science built on an extensive consultation [...] to emphasize the importance of **processes over products**, collaboration over competition, and inclusion over speed as necessary starting points.*

Scientific journals

Role

What is the **Role** played by **Journals** and **Publishers** ?

- **Registration:** authorship/priority claim
- **Certification:** usually peer-review
- **Dissemination:** provide (targeted) access
- **Archiving:** permanent access link (citable)

Makes sense to gather the review of **paper & datasets**

- *Journals policies with datasets*, e.g. the **Springer research data policy**, vary a lot but... are only **encouraging** at most.
- Private publishers business model $\Rightarrow$ create dataset journals (**Scientific Data**, **BMC series**, **Discover**)

How to organize FAIR-ness and data curation ?



Translated from

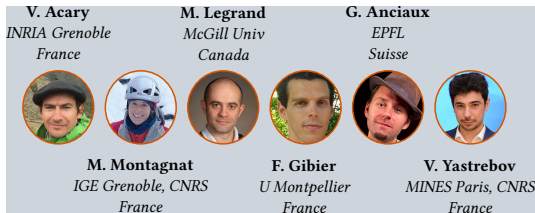
<https://zenodo.org/records/10108736>

Constraints (debatable)

- Hard work $\Rightarrow$ researchers
- Recognition
- Incentives $\Rightarrow$ journals, funding agencies, institutions
- Constantly evolving (impossible norm ?)

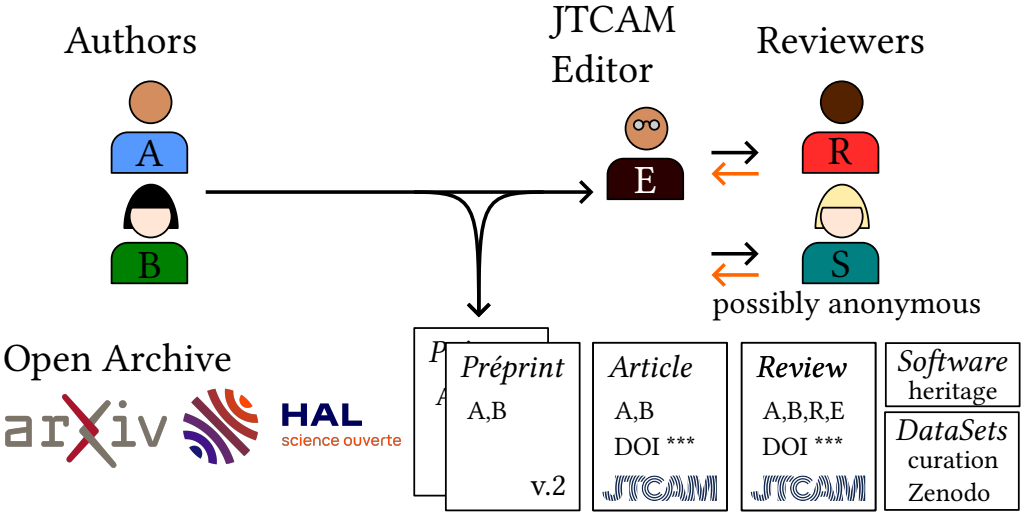
is a ***Diamond open access*** journal, i.e. published with ***no fees*** to either reader or author.

and is an ***overlay*** journal, i.e. that does not produce its own content, but selects from texts that are ***already freely available online*** (thanks to *Episciences!*).



- Editorial process entirely controlled by researchers (Reviews, Copy-editing)
- Publish Open Reviews
- Supported by *Centre for Direct Scientific Communication* (CCSD@CNRS/INRAE/Inria) and

JTCAM: publication process



JTCAM wanted to publish curated data

- Software Heritage (preserve code repositories)
- Zenodo (record < 50GB)
- Reproducibility ?

which needed a **curation tool**

DCSM: Dissemination of Computational Solid Mechanics and Solidipes: an interoperable tool for curation of research data

- Funded by **Open Research Data (ORD)** (36 months)
- G. Anciaux (LSMS), S. Pham-Ba (ENAC-IT4R), A. Borel (EPFL Library), V. Savchenko and A. Neronov (LASTRO)

Goals

- **Simplify** the verification, analysis and annotation (**curation**)
- **Usable** by researchers **on their personal computer**
- **Distribute** plugins contributed by distinct scientific communities (solid-mechanics, astro)
- **Supporting** remote data access (import/export/mounting)
- **cloud service** (<https://dcsm.epfl.ch>) $\Rightarrow$ **“Overleaf”** for datasets

Additions

- Used at **JTCAM** for **data review and curation**
- **Integrating** with new services (MMODA, Galaxy)

Solidipes: analysis and curation tool



<https://gitlab.com/solidipes/solidipes>

Software funded by  **ETH BOARD** <https://ethrat.ch/en/>

Here, you can prepare your paper's data for publication in four steps:

✓ Acquisition

Upload any files relevant to your paper, and browse them like you would in a file browser.

Curation

Automatically verify the correct formatting of your files, review their contents, and discuss potential issues.

✓ Metadata

Easily edit any metadata relevant to your paper such as authors, keywords, description, and more.

Export

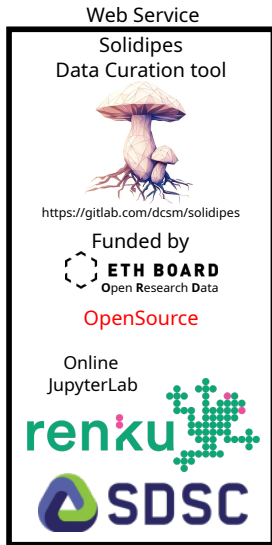
Once all previous steps are complete, review your work and export it to databases such as Zenodo.

Solidipes: analysis and curation tool

Steps to collect curate and publish authors' data



Solidipes: analysis and curation tool



<https://dscm.epfl.ch/dscm-intranet>

- 1 Scan files
 - Access remote data-storage/repository (S3, ssh, Windows share, rclone)
- 2 For each file
 - Identify the encoding/file format
 - Attempt a (partial) loading of the file
- 3 Extract metadata
 - *e.g.* CSV headers, image EXIF, finite element field descriptions
 - export **RO-crate** (json format)
- 4 Validations
 - File validation checks
 - Metadata/Ontology level validation checks
- 5 Generates a validating report in either
 - text mode (terminal)
 - WebApp allowing graphical scrutiny (images, interactive 3D rendering, ...)
- 6 Enables/facilitates export to Zenodo and others

Solidipes Demo & Tutorial

Documentation: <https://solidipes.readthedocs.io>

Curated @JTCAM

1 Zenodo

E. Eid, R. Seghir, J. Réthoré.

Accompanying data for the paper "Crack branching at low tip speeds: spilling the T"

<https://doi.org/10.5281/zenodo.8256346>

2 @Renku

data-science platform for reproducible science 



@renkulab.io

Curation Tutorial

1 Installation

```
pip install solidipes
```

2 Downloading an existing dataset

<https://sandbox.zenodo.org/records/13642>

3 Reports

```
solidipes report curation/web
```

4 Tweak with plugins



Supported Loaders/Viewers



gnuplot



image



code_snippet



yaml



table



json



xml



makefile



hdf5



zip_archive



text



matlab



pdf



notebook



video



python_pickle



rdf



FITS



abaqus



tikz



pyvista_mesh



meshio



xdmf



geo_msh



geof_mesh



binary

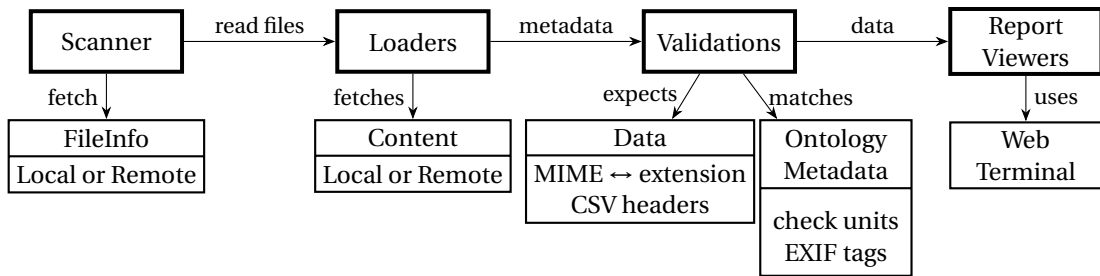


dictionary



pydantic

Plugins and Scientific Specializations



Template Plugin⇒ Adapt

- Loaders
- Viewers
- Scanners
- Validators
- Ontologies (pydantic)
- Reports

Solid Mechanics Plugin

- Loaders: meshio, abaqus, paraview
- Viewers: 3D Web-rendering of meshes
- Ontology: check Displacement fields

Astro Plugin

- Loaders: FITS
- Viewers: AladinLite, AstrovisJS
- Ontologies: TBD MModa
<https://odahub.io/ontology/>

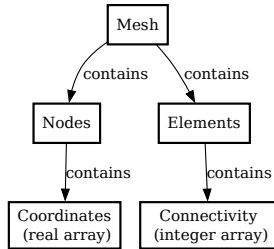
Ontologies

Ontology (adapted from Wikipedia)

*an ontology encompasses **definitions** of the **categories**, **properties**, and **relations** between the **data entities** of a (scientific) topic.*

In practice it is an **annotated graph** described with the *Resource Description Framework* (RDF) or as a set of *linked classes* (Pydantic)

e.g. a **valid** mesh file **must** contain nodes and elements.



- **XDMF** is a XML file (detected by linux) containing *nodes* and *elements* tags
- **.inp** is the **Abaqus** input file format, which may, or may not include meshing information
- Allows automatic validation/classification → Metadata based → **Fast!**
- **Solidipes** stores metadata with **ROCrates** and validates with **Pydantic**

Reproducibility with workflows

Using workflow managers → binding to **DCSM+Solidipes**

Output: Crab [2026.07.08T21:39:58] ^ x

Query parameters Log Share ⓘ API code ⓘ **Open on Solidipes ⓘ** 

Showing 1 to 2 of 2 entries

Search:

Output Name	Image
picture_output	
spectrum_astropy_table	

Show 10 ▾ entries

Previous **1** Next

Multi Messenger Online Data Analysis(MMODA)
product Example

Conclusion

Open Science through Collaborative Data Curation

Motivation

- FAIR data (Findable, Accessible, Interoperable, Reusable) requires active **curation**.

Solidipes, a **collaborative, plugin-based** curation tool for researchers.

- **Curation:** *scan* → *load* → *extract metadata* → *validate* → *report* → *publish* (Zenodo, ...)
- **Remote storage** (Rclone, ...)
- **Domain specializations** (plugins, solid mechanics, astrophysics).
- **Ontologies** (Pydantic + RO-Crate) for validation and classification.

Next steps

- **Workflows/containers** still work toward reproducibility (GUIX, Docker).
- **Reproducibility:** Author recommendations (**draft**)
- **Reproducibility:** Validations+Ontology ⇒ Quality badges

Solidipes Demo & Tutorial

Documentation: <https://solidipes.readthedocs.io>

Curated @JTCAM

1 Zenodo

E. Eid, R. Seghir, J. Réthoré.

Accompanying data for the paper "Crack branching at low tip speeds: spilling the T"

<https://doi.org/10.5281/zenodo.8256346>

2 @Renku

data-science platform for reproducible science 



@renkulab.io

Curation Tutorial

1 Installation

```
pip install solidipes
```

2 Downloading an existing dataset

<https://sandbox.zenodo.org/records/13642>

3 Reports

```
solidipes report curation/web
```

4 Tweak with plugins

