Curate and publish your study with Solidipes

Son Pham-Ba & Guillaume Anciaux
École polytechnique fédérale de Lausanne

EPFL
Funded by ETH BOARD
Open Research Data (ORD) Program

Introduction

Armillaria solidipes is the specie of the largest living organism on Earth, a fungus forming an underground network spanning 9.1 km². **Solidipes** is a Python package that aids the processes of **curating**, **publishing** and **sharing** research data, particularly tailored for the field of computational solid mechanics.

Solidipes is made with Python 3 and Streamlit.

1 Acquisition

GitLab

JupyterLab

Data can be put in **various cloud storages** and accessed seamlessly without copying.
JuiceFS, based on S3, chunks and cache data for better performance.

JuiceFS, NFS, S3, SMB, SSH

Online instances of Solidipes rely on JupyterLab (and GitLab optionaly), which can both be used to add data.

An existing dataset (and its metadata) can be **retrieved** from Zenodo.

Solidipes

On your computer

OR

On an online instance *e.g.* dcsm.epfl.ch

The curated and annotated dataset can be exported to Zenodo for **long-term storage** and creation of a DOI.

At any stage, the dataset can be exported to Renku, which would run an online instance of Solidipes, for **sharing** a live preview of the dataset with other scientists.

4 Export

Curation 2

File MIME types are checked against their extension. File headers are checked. Users and reviewers can put messages on files.

Files are loaded and visualized using Python *Loaders* and *Viewers*.

MeshIO

PyVista

The MeshIO *Loader* and the PyVista *Viewer* are tailored for solid mechanics data

3 Metadata

Metadata is mendatory:
- Authors, affiliations, and ORCID
- Title
- Keywords
- License
- Related productions

A detailed **description** must be added, with informations such as:
- Author contributions
- Data collection method and details
- Data structure
- Funding sources

The metadata and description are merged into a comprehensive formatted **README.md** file.

Solidipes is used in the dataset curation process for the **Diamond open-access** Journal of Theoretical, Computational and Applied Mechanics.

FAIR principles

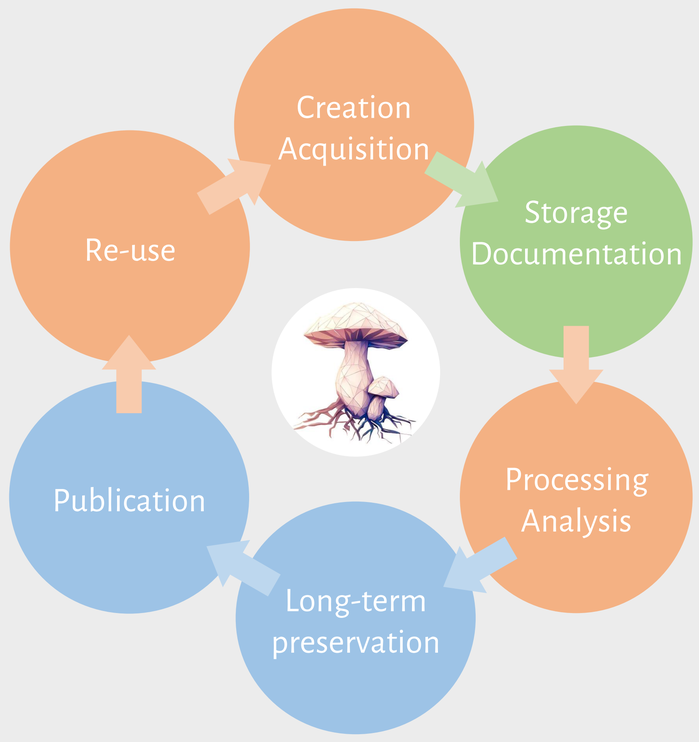
Findable
- Metadata: Authors, Keywords, cross-links, *etc.*
- Digital Object Identifier (DOI)

Accessible
- Open
- Retention time

Interoperable
- Standards: file formats, metadata, vocabularies, ontologies, *etc.*

Reusable
- Open license
- Environnent: software versions, dependencies, *etc.*

Data management plan



Next steps

- Save a *workflow*, *i.e.* the steps taken to generate the dataset
- Adapt to more software/platforms
- Provide guidance on *[store output data]* versus *[only keep scripts to regenerate data]*, to minimize CO_2 impact

https://gitlab.com/dcsm/solidipes