

An Open Source Project for Dissemination of Computational Solid Mechanics (DCSM)

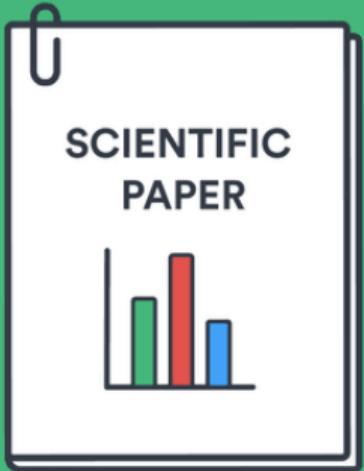
# Guillaume Anciaux

LSMS, IIC, ENAC, EPFL



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

Did it ever happen to you ?



Dataset

Can I reuse ?  
Can I re-compile ?  
Can I reproduce results ?

# Publishing datasets: Where?

## Zenodo

- Long term preservation ⇒ **for 20 more years from now**
- Generalist ⇒ stores anything, no checks
- < 50 GB limit
- Read-only datasets after publication
- Modifiable metadata

# Publishing datasets: Where?

## Zenodo

- Long term preservation ⇒ **for 20 more years from now**
- Generalist ⇒ stores anything, no checks
- < 50 GB limit
- Read-only datasets after publication
- Modifiable metadata
- **What content ? Who checks ?**
  - At the moment **no one but the owner** (Can publish anything)
  - + Curation perspective (e.g. **Horizon-Zen project**)

# Publishing datasets: Where?

## Zenodo

- Long term preservation ⇒ **for 20 more years from now**
- Generalist ⇒ stores anything, no checks
- < 50 GB limit
- Read-only datasets after publication
- Modifiable metadata
- **What content ? Who checks ?**
  - At the moment **no one but the owner** (Can publish anything)
  - + Curation perspective (e.g. **Horizon-Zen project**)

## Institutional/focused repositories ?

- Material's Cloud with **Aiida**
- Research collection at **ETHz**
- **LivMatS** with **dtool**
- **Dataverse**

many others....

# Publishing datasets: Where?

## Zenodo

- Long term preservation ⇒ **for 20 more years from now**
- Generalist ⇒ stores anything, no checks
- < 50 GB limit
- Read-only datasets after publication
- Modifiable metadata
- **What content ? Who checks ?**
  - At the moment **no one but the owner** (Can publish anything)
  - + Curation perspective (e.g. [Horizon-Zen project](#))

## Institutional/focused repositories ?

- Material's Cloud with [Aiida](#)
- Research collection at [ETHz](#)
- LivMatS with [dtool](#)
- Dataverse

many others.... **all needing curation**

# Definitions: CODATA RDM Terminology

**Data Curation** ensures usefulness for discovery and reuse.

- validation of the data (encoding, file formats)
  - description completeness (documentation, annotation)
  - discipline dependent (question of vocabulary, formats, )

# Definitions: CODATA RDM Terminology

**Data Curation** ensures usefulness for discovery and reuse.

- validation of the data (encoding, file formats)
- description completeness (documentation, annotation)
- discipline dependent (question of vocabulary, formats, )

Who does it ?

- journals
- universities (arXiv, HAL)
- specialized archivists (e.g. astronomy with NASA, ESA, neurologists and brain MRI)

# Definitions: CODATA RDM Terminology

**Data Curation** ensures usefulness for discovery and reuse.

- validation of the data (encoding, file formats)
- description completeness (documentation, annotation)
- discipline dependent (question of vocabulary, formats, )

Who does it ?

- journals
- universities (arXiv, HAL)
- specialized archivists (e.g. astronomy with NASA, ESA, neurologists and brain MRI)

**Data curation** is a central pillar in fostering scientific discoveries

# Definitions: CODATA RDM Terminology

**Data Curation** ensures usefulness for discovery and reuse.

- validation of the data (encoding, file formats)
- description completeness (documentation, annotation)
- discipline dependent (question of vocabulary, formats, )

Who does it ?

- journals
- universities (arXiv, HAL)
- specialized archivists (e.g. astronomy with NASA, ESA, neurologists and brain MRI)

**Data curation** is a central pillar in fostering scientific discoveries

Sometimes difficult to **convince** researchers to cure data

# Definitions: CODATA RDM Terminology

**Data Curation** ensures usefulness for discovery and reuse.

- validation of the data (encoding, file formats)
- description completeness (documentation, annotation)
- discipline dependent (question of vocabulary, formats, )

Who does it ?

- journals
- universities (arXiv, HAL)
- specialized archivists (e.g. astronomy with NASA, ESA, neurologists and brain MRI)

**Data curation** is a central pillar in fostering scientific discoveries

Sometimes difficult to **convince** researchers to cure data

My personal belief is that **journals** should play a role



Journal of Theoretical,  
Computational and  
Applied Mechanics

*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
- Wide spectrum: theoretical, applied, numerical, experimental



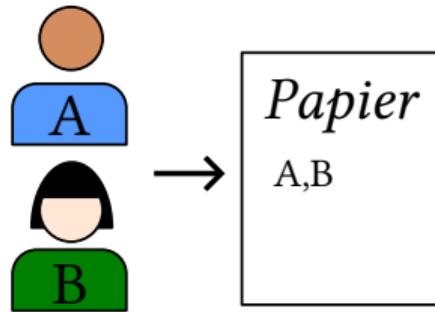
is a **Diamond open access** journal, i.e. published with **no fees** to either reader or author.

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
- Wide spectrum: theoretical, applied, numerical, experimental
- Recent attempt to include **Data curation** in the publication process

# JTCAM: publication process

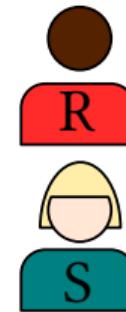
Authors



JTCAM  
Editor



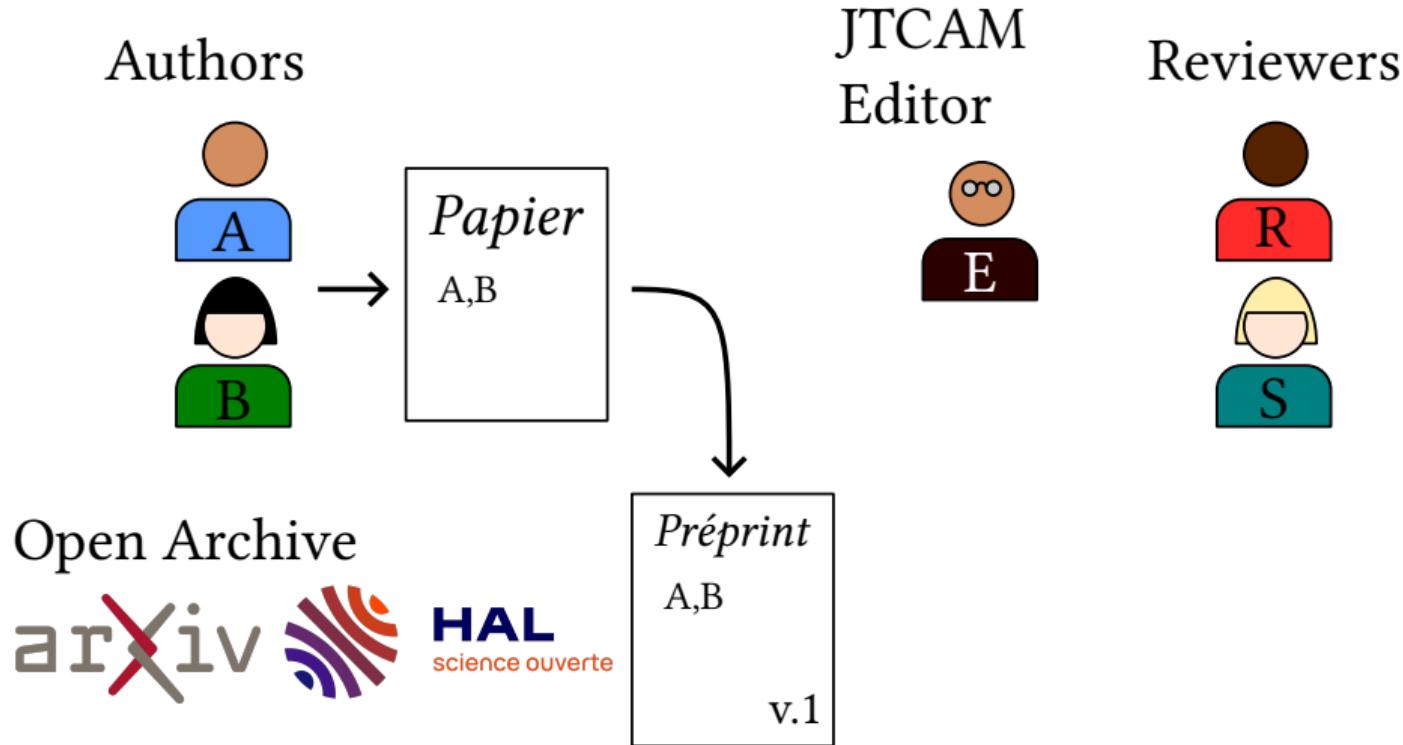
Reviewers



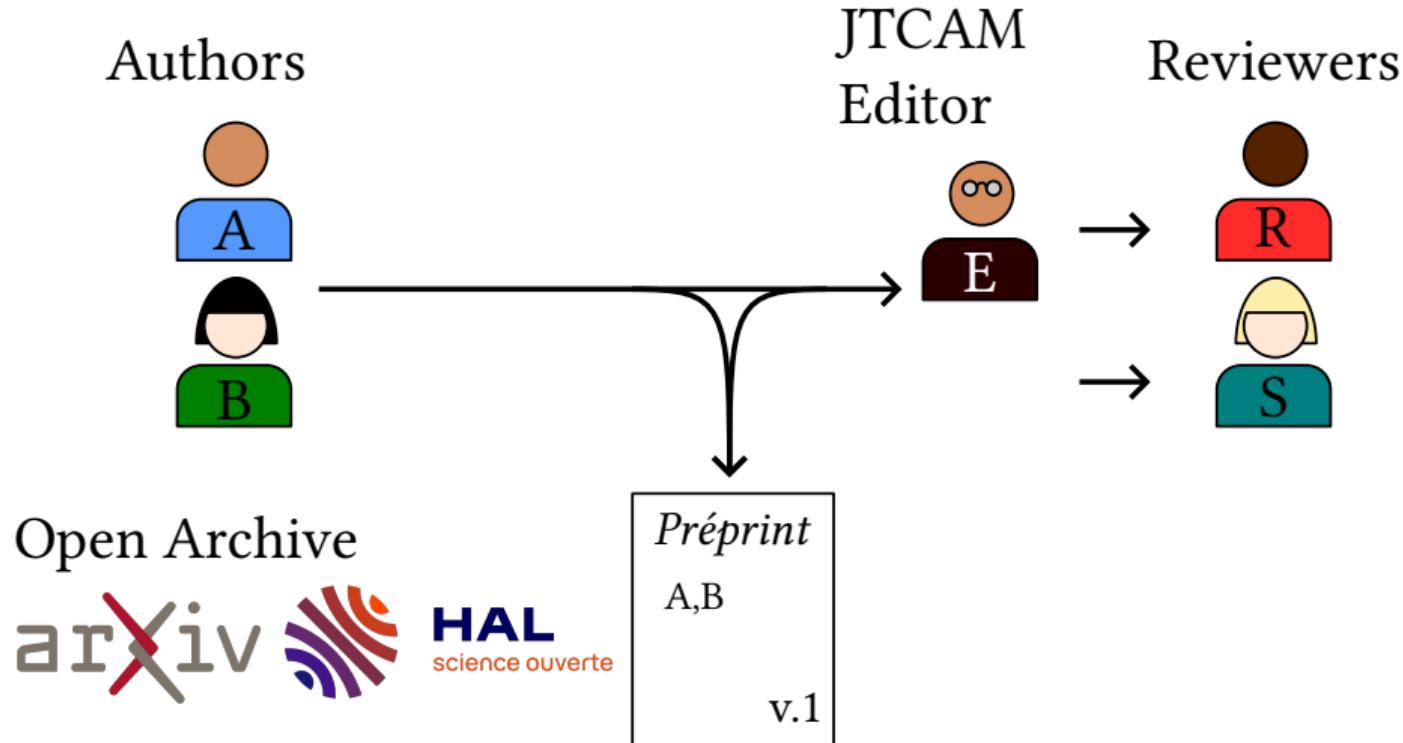
Open Archive



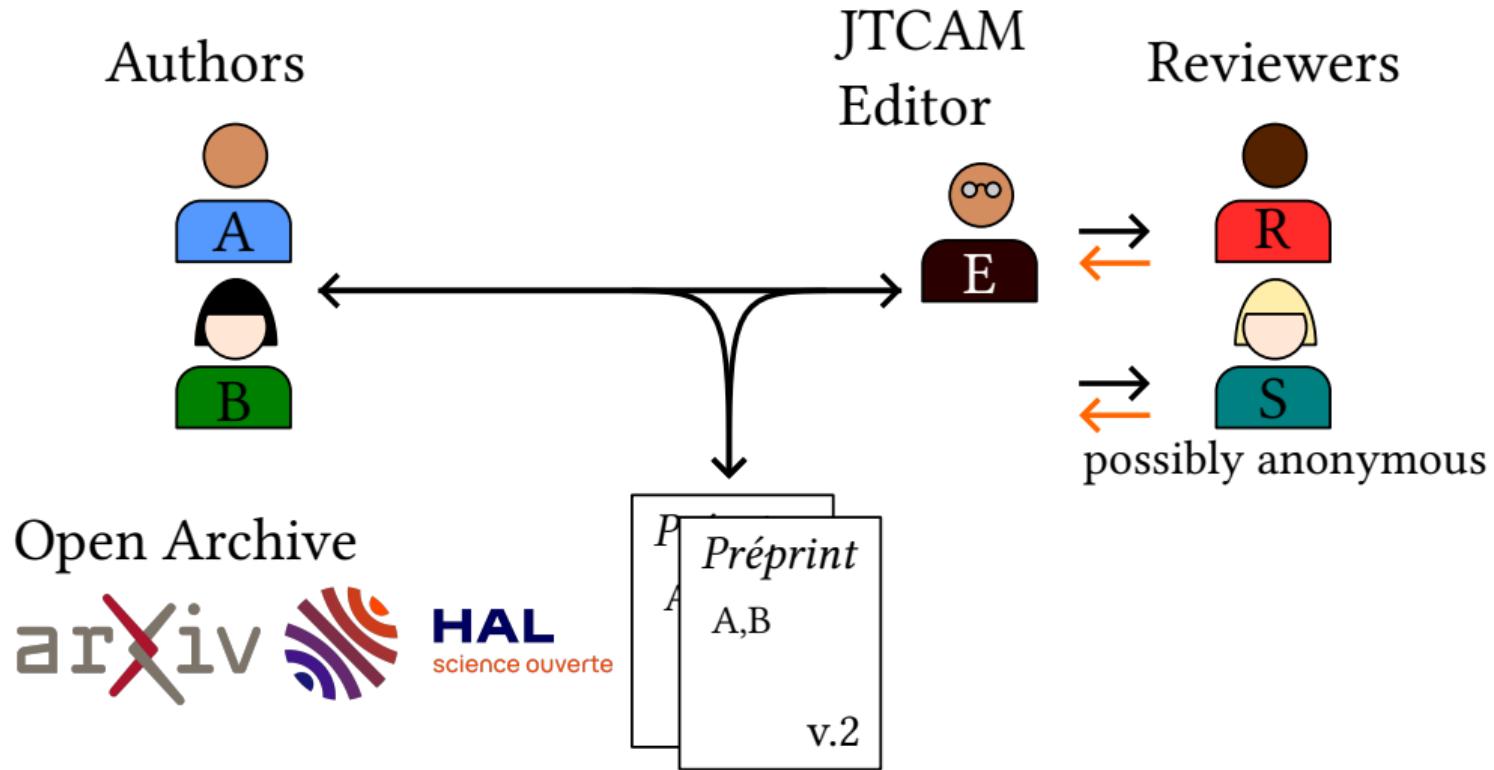
# JTCAM: publication process



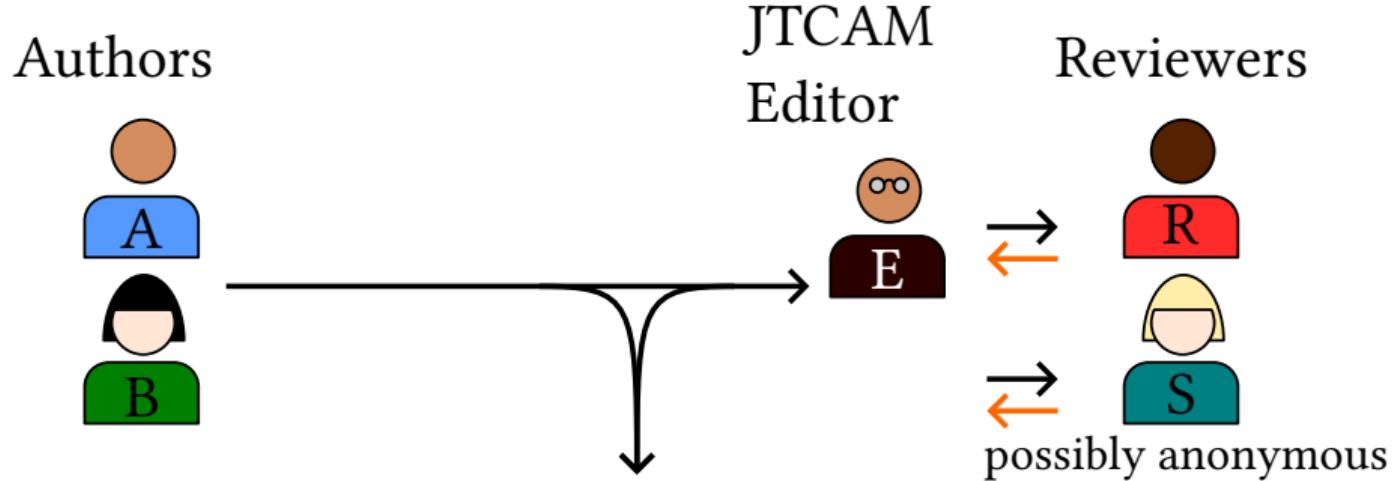
# JTCAM: publication process



# JTCAM: publication process



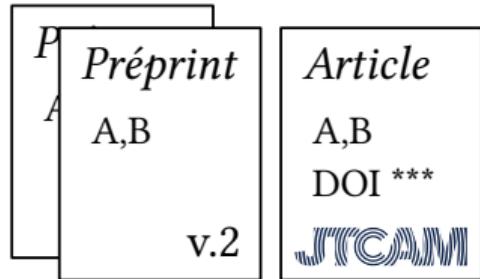
# JTCAM: publication process



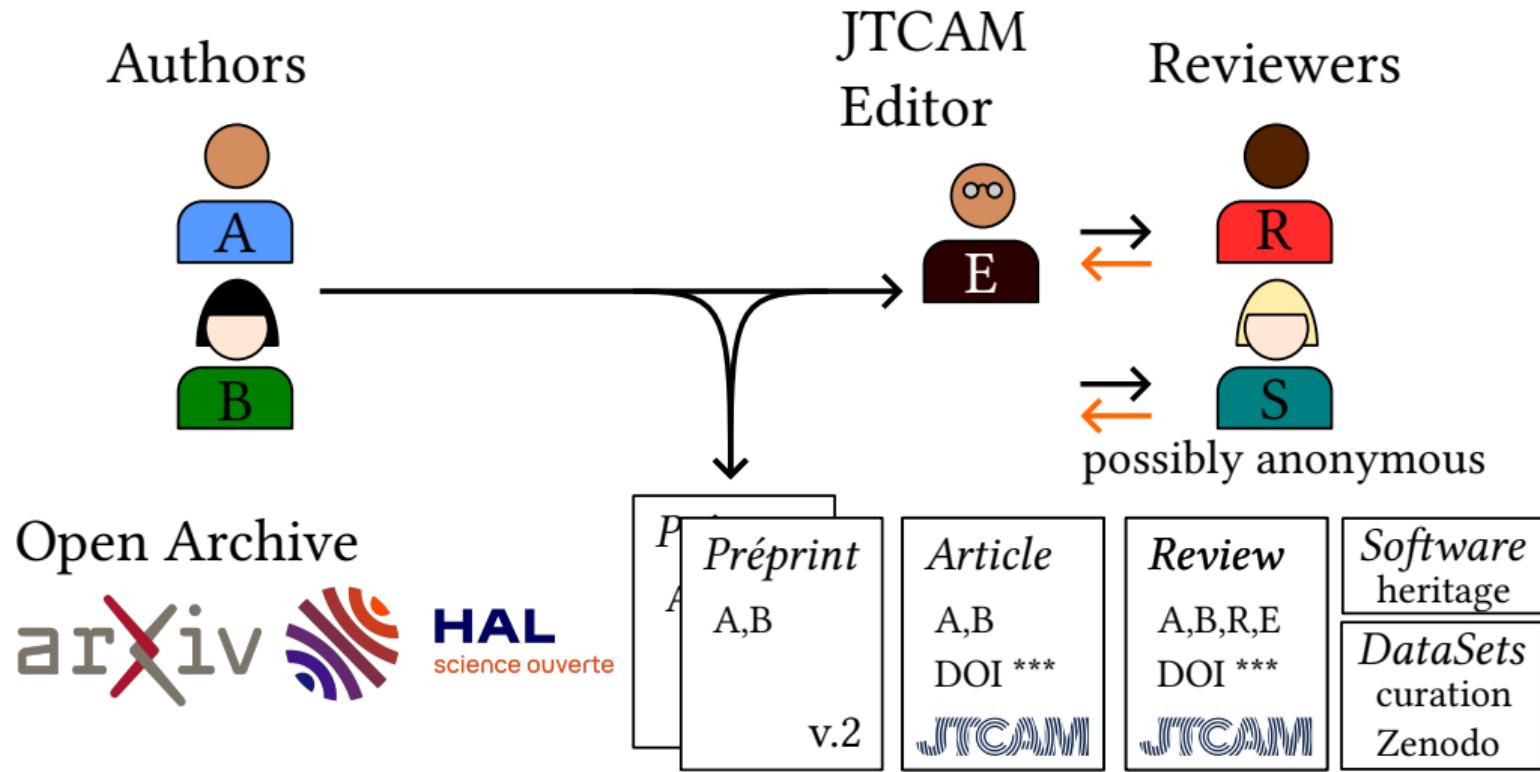
Open Archive



HAL  
science ouverte



# JTCAM: publication process



# JTCAM dataset curation policy => FAIR

The following criteria are required in order to accept a submission to the JTCAM community:

- Must be Open Access
- Ownership described in depth
- Detailed description (using standard ontologies or controlled vocabularies)
- Cross-linked reference must be added
- Software permanent links (Software Heritage)
- Acknowledged grants
- Cleaned (no unnecessary files/folders or redundancy)
- Permissive licenses are required (CC0, CC-BY-4.0)
- Files formats are open
- Workflow description

<https://zenodo.org/communities/jtcam/curation-policy>

# DCSM Project

## Project Dissemination of Computational Solid Mechanics(DCSM)

- Fund by Open Research Data (ORD)
- G. Anciaux (dev and supervision@EPFL), S. Pham-Ba (developer@EPFL)
- Young project (18 months)
- Use lots of project dependencies

## Goals

- Provide a **cloud based** repository/storage/tool for **solidmechanics** community
- Simplify the **verification, analysis and annotation** (curation) of datasets
- Stand-alone tool for researchers to manipulate data **on their personal computer**
- Web service: <https://dcsm.epfl.ch>
- Used at JTCAM for **data reviews**
- “*Overleaf*” for datasets



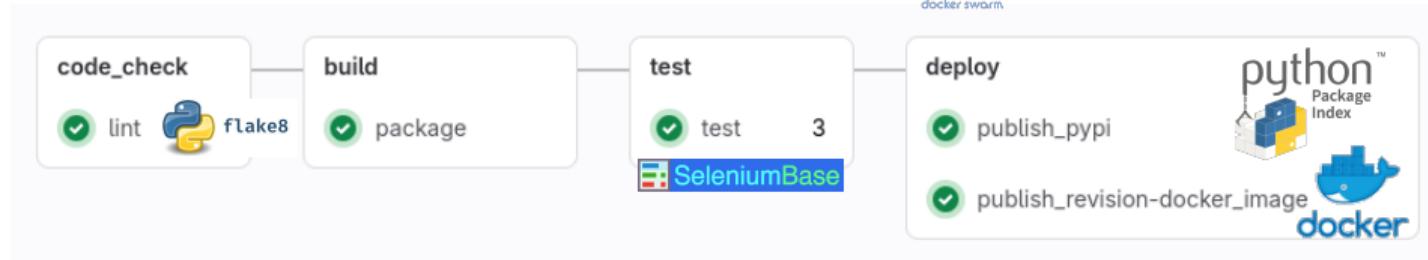
**DCSM** ●



## Dissemination of Computational Solid Mechanics



Solidiges ●  
Python package for the DCSM project



# Solidipes



***Amillaria Solidipes*** grows and spreads primarily underground, and is possibly the largest living organism on Earth by mass, area, and volume and is colloquially called the "Humongous fungus". [[Wikipedia](#)]

# Solidipes



*Amillaria Solidipes* grows and spreads primarily underground, and is possibly the largest living organism on Earth by mass, area, and volume and is colloquially called the "Humongous fungus". [[Wikipedia](#)]

... Nothing to do with **solids** ....

# Solidipes



*Amillaria Solidipes* grows and spreads primarily underground, and is possibly the largest living organism on Earth by mass, area, and volume and is colloquially called the "Humongous fungus". [[Wikipedia](#)]

... Nothing to do with **solids** ....

```
pip install solidipes
```

# Solidipes: analysis and curation tool

- 1 Access remote data-storage/repository (S3, ssh, Windows share)

# Solidipes: analysis and curation tool

- 1 Access remote data-storage/repository (S3, ssh, Windows share)
- 2 Scan files

# Solidipes: analysis and curation tool

- 1 Access remote data-storage/repository (S3, ssh, Windows share)
- 2 Scan files
- 3 For each file
  - Identify the encoding/file format
  - Extract the metadata (CSV headers, image properties, finite element field descriptions)
  - Attempt a (partial) loading of the file
  - If any perform additional validation checks

# Solidiges: analysis and curation tool

- 1 Access remote data-storage/repository (S3, ssh, Windows share)
- 2 Scan files
- 3 For each file
  - Identify the encoding/file format
  - Extract the metadata (CSV headers, image properties, finite element field descriptions)
  - Attempt a (partial) loading of the file
  - If any perform additional validation checks
- 4 Generates a validating report in either
  - text mode (terminal)
  - Jupyter notebook
  - WebApp allowing to graphical scrutiny (images, interactive 3D rendering, ...)

# Solidipes: analysis and curation tool

- 1 Access remote data-storage/repository (S3, ssh, Windows share)
- 2 Scan files
- 3 For each file
  - Identify the encoding/file format
  - Extract the metadata (CSV headers, image properties, finite element field descriptions)
  - Attempt a (partial) loading of the file
  - If any perform additional validation checks
- 4 Generates a validating report in either
  - text mode (terminal)
  - Jupyter notebook
  - WebApp allowing to graphical scrutiny (images, interactive 3D rendering, ...)
- 5 If validated: enables export to Zenodo/Renku

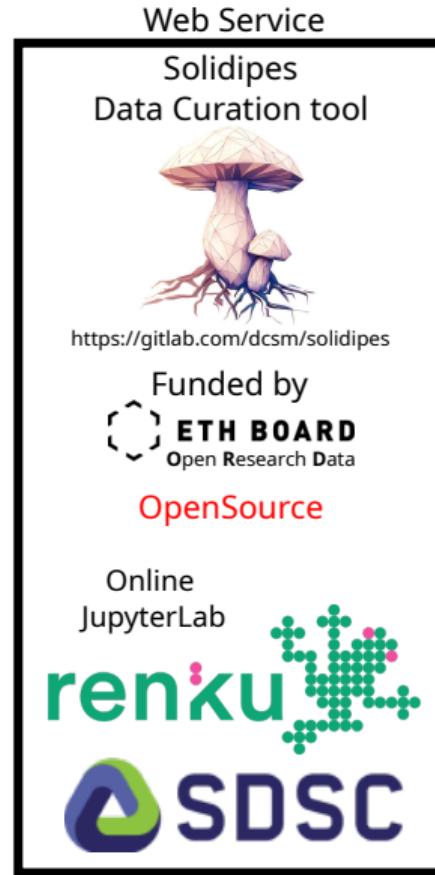
# Solidipes: analysis and curation tool

## Features

- Analysis: Jupyterlabs and context preserving
- Curation: dedicated readers&viewers (web oriented)
- Export/Import/Mount (S3, samba, nfs, Zenodo repositories)
- Operating Context saved (**Docker** containerization)

## Demo

- E. Eid, R. Seghir, & J. Réthoré. Accompanying data for the paper "Crack branching at low tip speeds: spilling the T"
- **Zenodo**
- **@Renku** (*a platform and tools for reproducible and collaborative data analysis*)
- Curation session



# Curation view

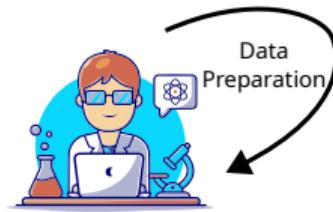
E. Eid, R. Seghir, & J. Réthoré. Accompanying data for the paper "Crack branching at low tip speeds: spilling the T"

Directory	Status	Filename	Type	Size	Open
▼  data (34)				4.84GiB	
>  Codes (6)				27.24KiB	
▼  T3DE (7)				1.25GiB	
		Distortion.mat	application/x-matlab-binary	33.87KiB	<a href="#">View File</a>
		mesh_crack_c...	application/x-matlab-binary	222.61KiB	<a href="#">View File</a>
		time.txt	text/plain	566B	<a href="#">View File</a>
>  Resu... (2)				56.01MiB	
▼  images (2)				1.2GiB	
		./data/T3DE/i...	image/tiff	1.18GiB	<a href="#">View File</a>
		ref.tiff	image/tiff	15.54MiB	<a href="#">View File</a>
>  TAF1 (7)				1.09GiB	
>  TAF2 (7)				1.24GiB	
>  THOM (7)				1.25GiB	

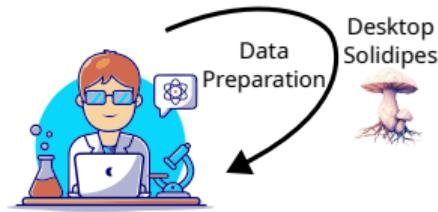
# Solidipes: analysis and curation tool



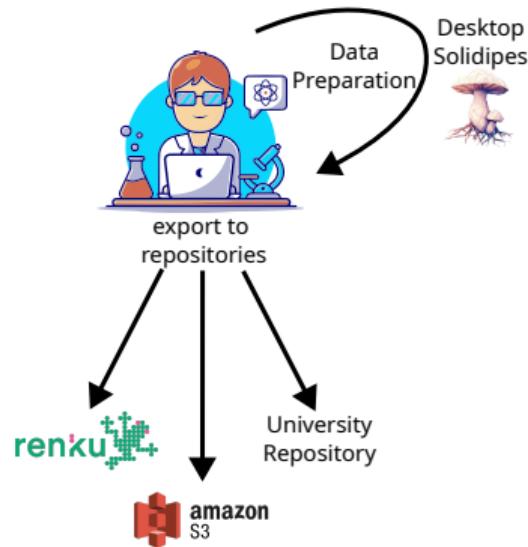
# Solidipes: analysis and curation tool



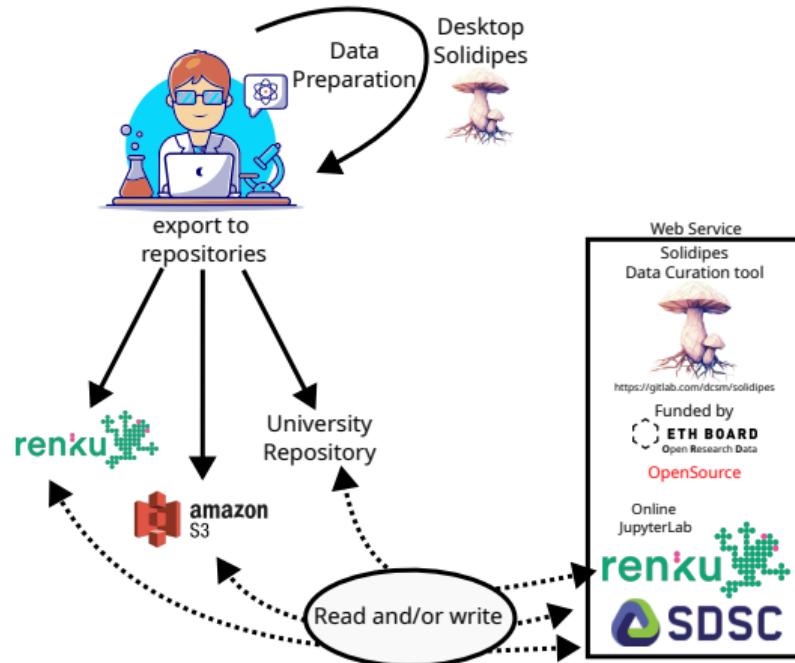
# Solidipes: analysis and curation tool



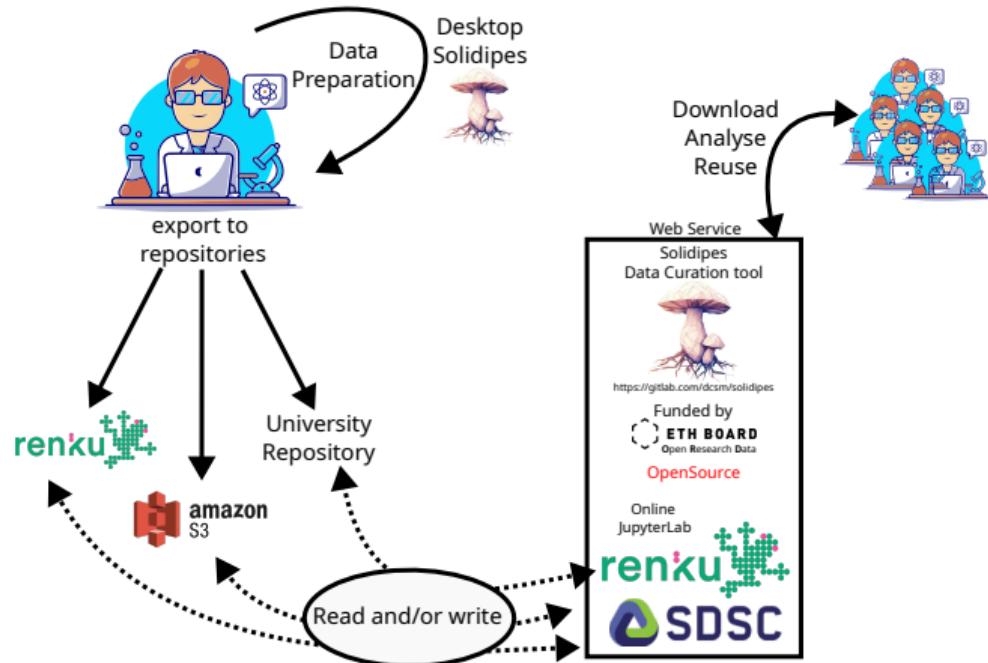
# Solidipes: analysis and curation tool



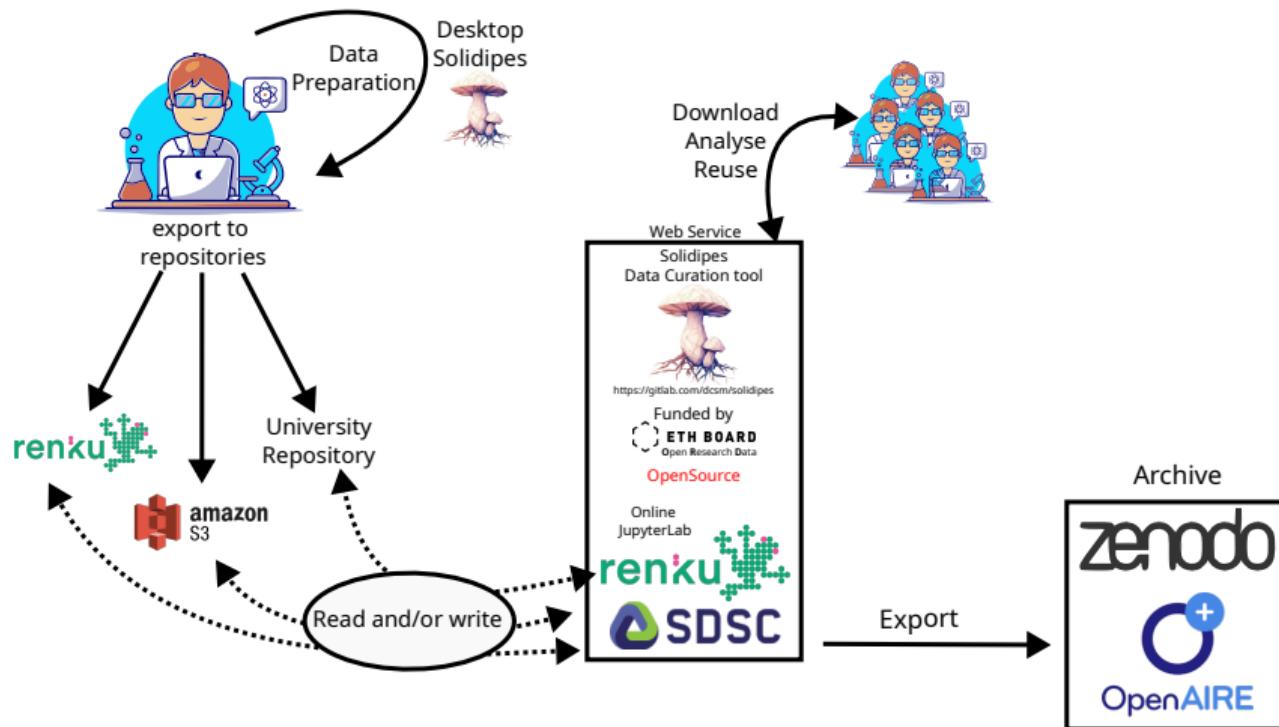
# Solidipes: analysis and curation tool



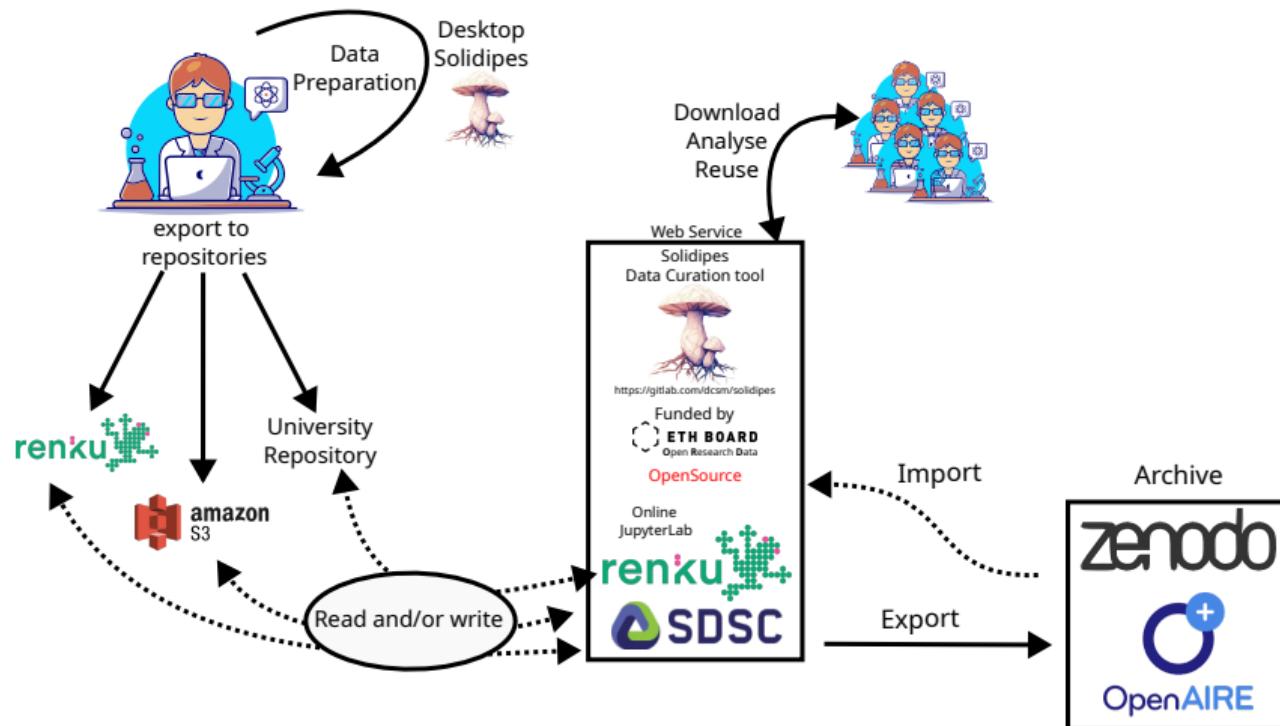
# Solidipes: analysis and curation tool



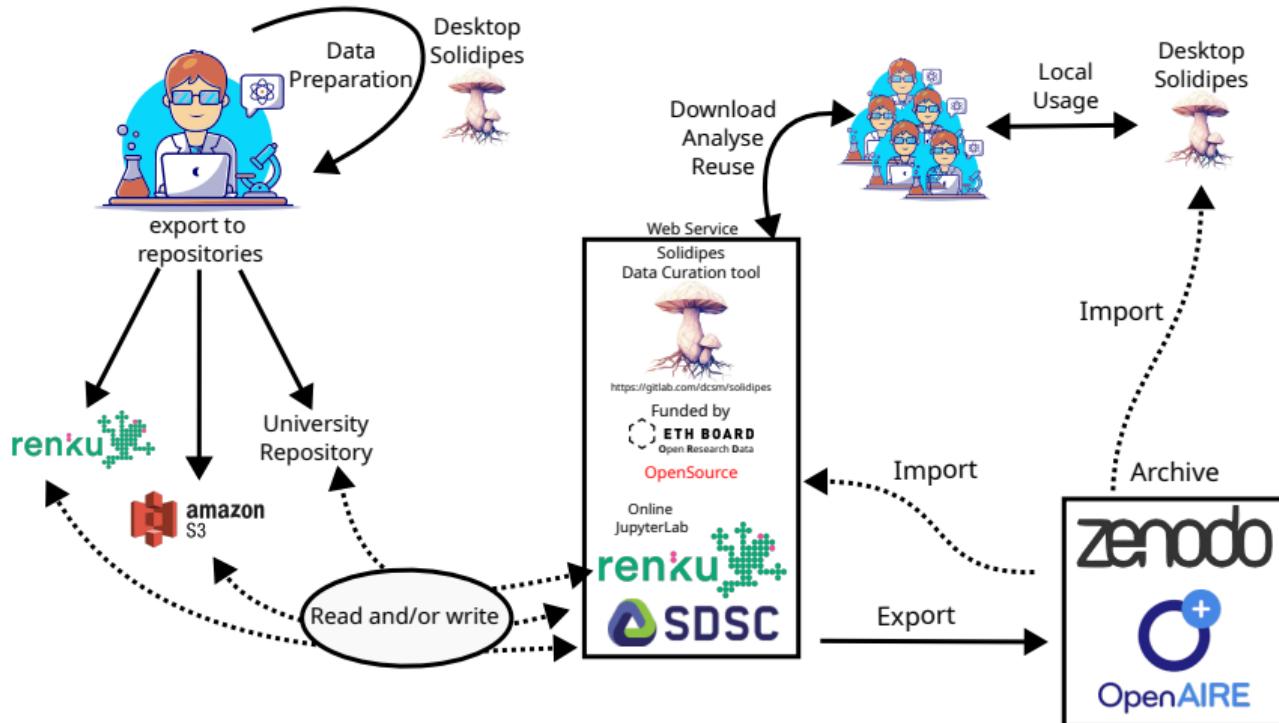
# Solidipes: analysis and curation tool



# Solidipes: analysis and curation tool



# Solidipes: analysis and curation tool

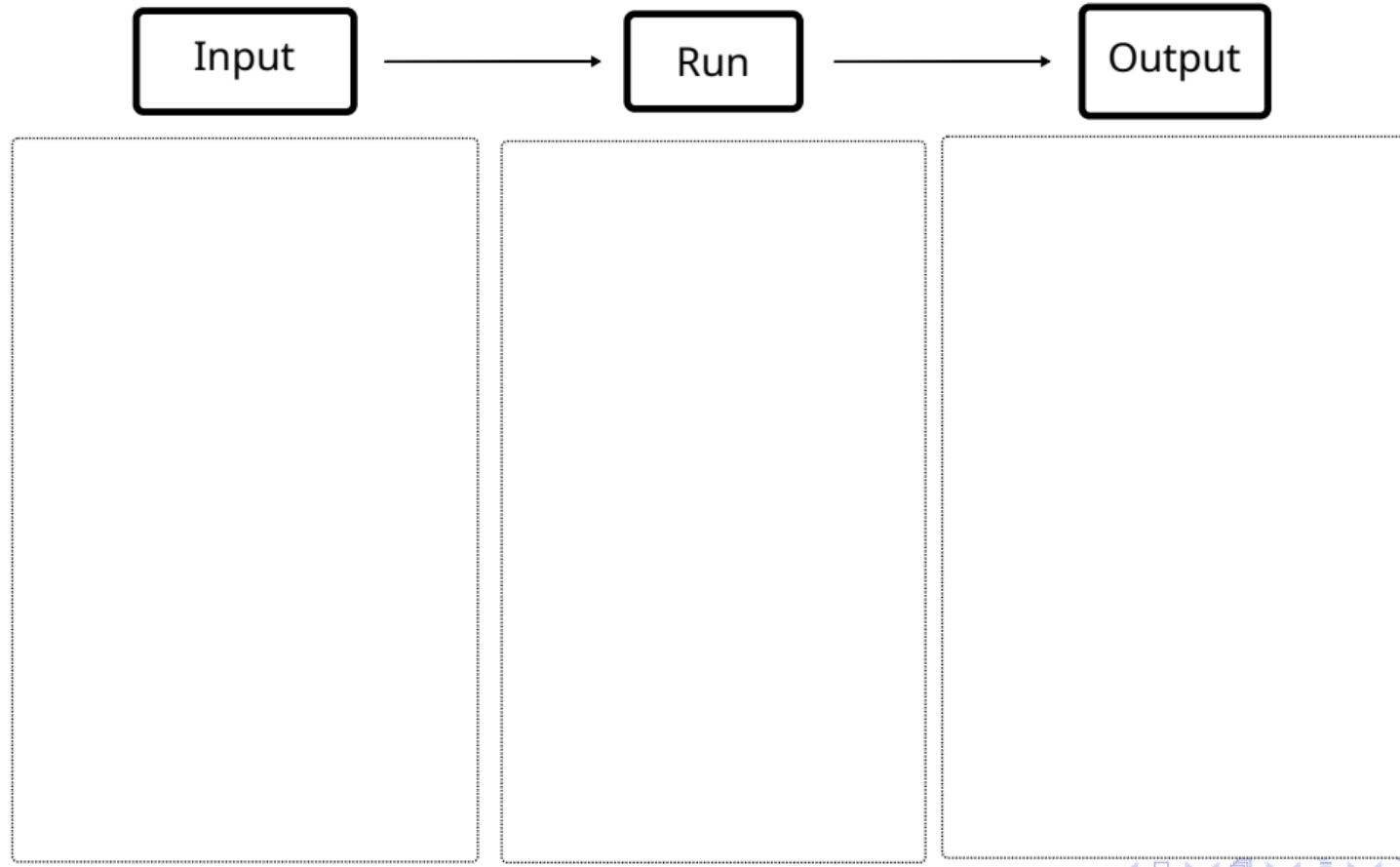


# How can open source projects help ?

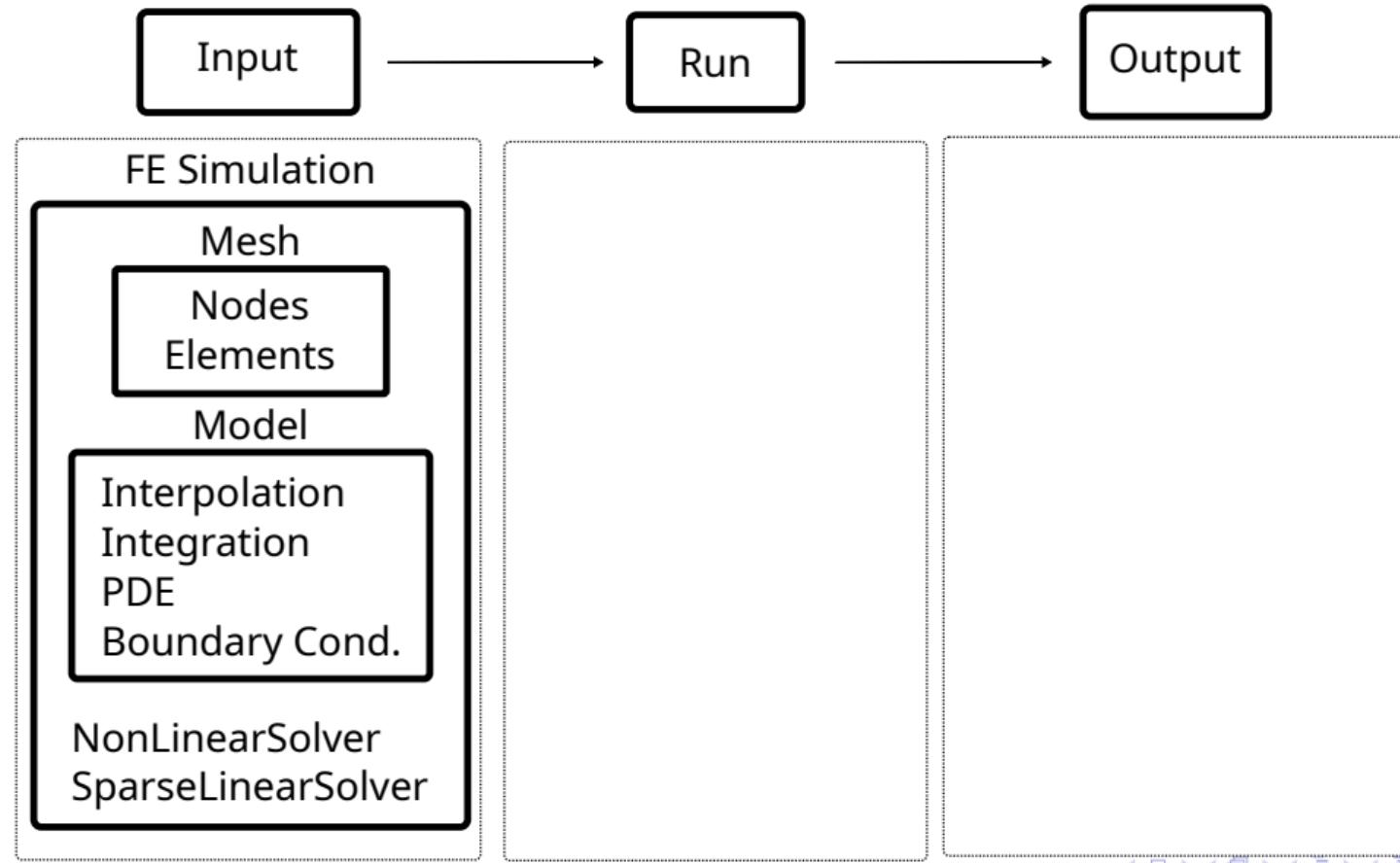
# Ideal curation

- **Versatile** data storage
- In **distinct** disciplines (experimental, theoretical, numerical, fluid mechanics, solid mechanics, ...)
- Collaborative curation (**concurrent editing**)
- Robust descriptions (**ontologies**)
- Reproducibility (**workflows**)

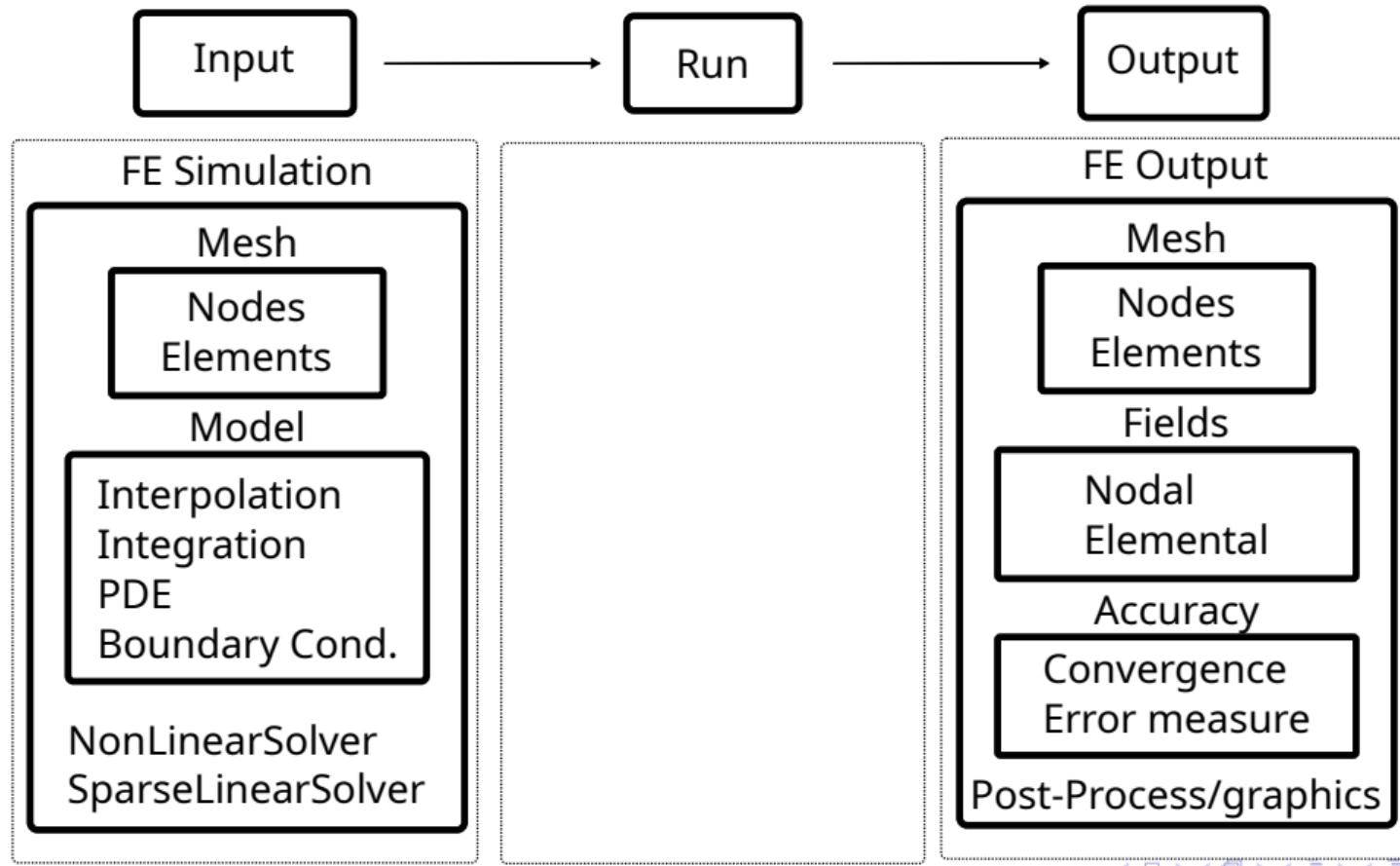
# Simulation workflow



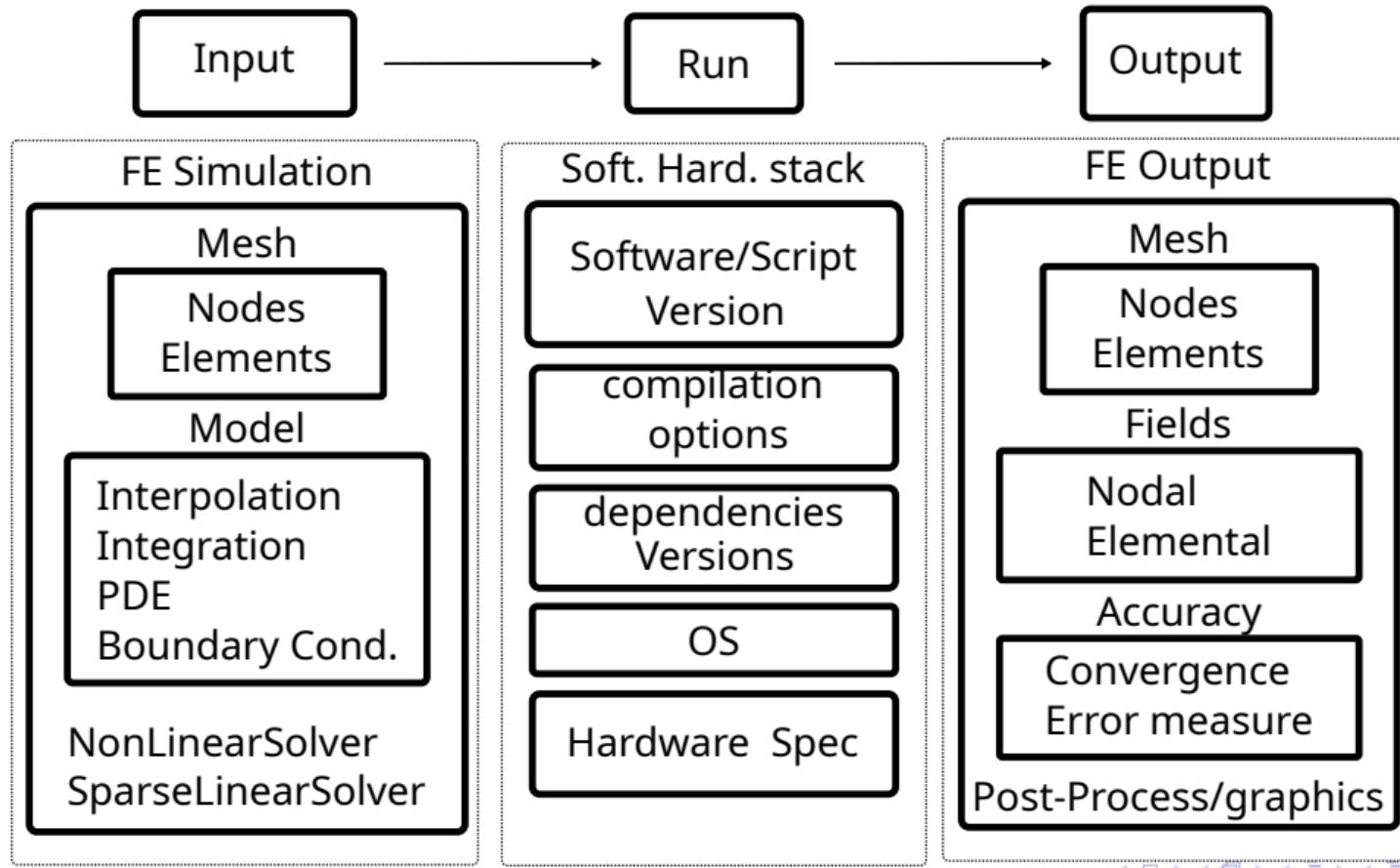
# Simulation workflow



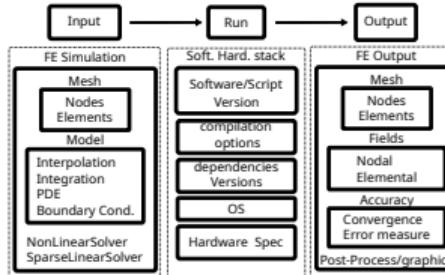
# Simulation workflow



# Simulation workflow



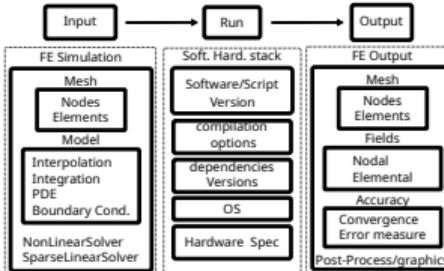
# Simulation workflow



Package

- Versions, metadata (yaml, json, ...)
- Results in standard format (i.e. readable by meshio ?)
- Limit number of files ?
- Spack/Python lock files ?
- CMakeCache.txt / build config and log files ?
- Docker file ?
- Run logs, execution times, cpu-hours ?
- How to strictly attach to output results ?

# Simulation workflow



Package

- Versions, metadata (yaml, json, ...)
- Results in standard format (i.e. readable by meshio ?)
- Limit number of files ?
- Spack/Python lock files ?
- CMakeCache.txt / build config and log files ?
- Docker file ?
- Run logs, execution times, cpu-hours ?
- How to strictly attach to output results ?

- **Docker** container: [Renku](#), [Binder](#), ...
- **Workflow management:** [AiiDA](#), [BlackDynamite](#)
- **Packagers:** [ROcrate](#), [reprozip](#)
- **Repository:** [WorkflowHub](#)

# Ontologies

## Ontology (adapted from Wikipedia)

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

# 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)*

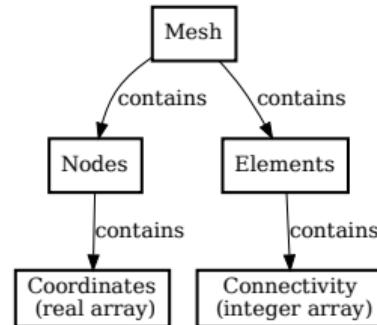
# Ontologies

## 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)*

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



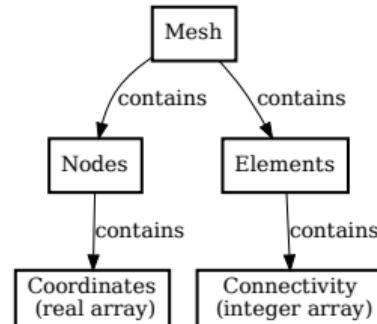
# Ontologies

## 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)*

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**, **Zébulon** input file format, which may, or may not include meshing information

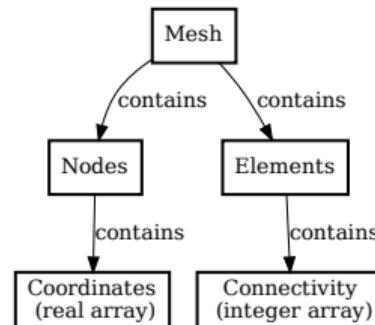
# Ontologies

## 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)*

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**, **Zébulon** input file format, which may, or may not include meshing information
- Would allow semi-automatic validation, and content check for specific purpose

# Ontologies in Astronomy

The International Virtual Observatory Alliance (IVOA), 2002

*is an organisation that debates and agrees the technical standards that are needed to make the VO possible. [...]*

THE ASTROPHYSICAL JOURNAL SUPPLEMENT SERIES, 236:24 (7pp), 2018 May

<https://doi.org/10.3847/1538-4365/aab760>

## The Unified Astronomy Thesaurus: Semantic Metadata for Astronomy and Astrophysics

Katie Frey<sup>1</sup>  and Alberto Accomazzi<sup>2</sup> 

<sup>1</sup> Wolbach Library, Harvard-Smithsonian Center for Astrophysics 60 Garden Street Cambridge, MA 02138, USA; kfrey@cfa.harvard.edu

<sup>2</sup> Astrophysics Data System, Harvard-Smithsonian Center for Astrophysics 60 Garden Street Cambridge, MA 02138, USA

Received 2017 October 1; revised 2017 December 21; accepted 2017 December 22; published 2018 May 11

### Abstract

Several controlled vocabularies have been developed and used by the astronomical community, each designed to serve a specific need and a specific group. The Unified Astronomy Thesaurus (UAT) attempts to provide a highly structured controlled vocabulary that will be relevant and useful across the entire discipline, regardless of content or platform. Because classifying articles and data will be the two major functions of the UAT, we examine the UAT in comparison with the Astronomical Subject Keywords used by major publications and the *JWST* Science Keywords used by STScI's Astronomer's Proposal Tool.

# Ontologies in mechanics

**J.-L. Hippolyte, P. Duncan, M. Bevilacqua and M. Chrubasik.** *Ontologies for Experimental Mechanics.* British Society for strain measurement, National Physical Laboratory, Hampton Rd, Teddington TW11 0LW, UK. 2022 [[link](#)]

**Marcin Skulimowski.** *An OWL Ontology for Quantum Mechanics.* Faculty of Physics and Applied Informatics, University of Lodz. Pomorska 149/153, 90-236 Lodz, Poland. 2002 [[link](#)]

**H.A. Preisig, T.F. Hagelien, J.Friis, P. Klein, N. Konchakova.** *Ontologies in Computational Engineering.* 14th World Congress on Computational Mechanics (WCCM). 2020. [[link](#)]

# Ontologies in mechanics

**J.-L. Hippolyte, P. Duncan, M. Bevilacqua and M. Chrubasik.** *Ontologies for Experimental Mechanics.* British Society for strain measurement, National Physical Laboratory, Hampton Rd, Teddington TW11 0LW, UK. 2022 [[link](#)]

**Marcin Skulimowski.** *An OWL Ontology for Quantum Mechanics.* Faculty of Physics and Applied Informatics, University of Lodz. Pomorska 149/153, 90-236 Lodz, Poland. 2002 [[link](#)]

**H.A. Preisig, T.F. Hagelien, J. Friis, P. Klein, N. Konchakova.** *Ontologies in Computational Engineering.* 14th World Congress on Computational Mechanics (WCCM). 2020. [[link](#)]

Need to integrate and mix these initiatives: Forming a committee ?

# Conclusion

## Where we are

- **Data curation** with loaders/viewers (mostly for continuum solidmechanics)

# Conclusion

## Where we are

- **Data curation** with loaders/viewers (mostly for continuum solidmechanics)
- **Flexible** (remote) data storage

# Conclusion

## Where we are

- **Data curation** with loaders/viewers (mostly for continuum solidmechanics)
- **Flexible** (remote) data storage
- Publication and archive on **Zenodo/Renku**

# Conclusion

## Where we are

- **Data curation** with loaders/viewers (mostly for continuum solidmechanics)
- **Flexible** (remote) data storage
- Publication and archive on **Zenodo/Renku**
- **JTCAM curation policy** enforced with Solidipes@DCSM already

# Conclusion

## Where we are

- **Data curation** with loaders/viewers (mostly for continuum solidmechanics)
- **Flexible** (remote) data storage
- Publication and archive on **Zenodo/Renku**
- **JTCAM curation policy** enforced with Solidipes@DCSM already  
⇒ Brings good principles to this Diamond open access initiative
- User **Documentation**

# Conclusion

## Where we are

- **Data curation** with loaders/viewers (mostly for continuum solidmechanics)
- **Flexible** (remote) data storage
- Publication and archive on **Zenodo/Renku**
- **JTCAM curation policy** enforced with Solidipes@DCSM already  
⇒ Brings good principles to this Diamond open access initiative
- User **Documentation**

## Next steps

- Extend remote storage accesses (dtool ?)
- Ontologies ⇒ **Automatic and Robust** validation&recognition (reviewer friendly)
- Complete workflow remains a **manual** task ⇒ guaranty reproducibility
- **Plugins** to become multi-disciplinary

# Conclusion

## Where we are

- **Data curation** with loaders/viewers (mostly for continuum solidmechanics)
- **Flexible** (remote) data storage
- Publication and archive on **Zenodo/Renku**
- **JTCAM curation policy** enforced with Solidipes@DCSM already  
⇒ Brings good principles to this Diamond open access initiative
- User **Documentation**

## Next steps

- Extend remote storage accesses (dtool ?)
- Ontologies ⇒ **Automatic and Robust** validation&recognition (reviewer friendly)
- Complete workflow remains a **manual** task ⇒ guaranty reproducibility
- **Plugins** to become multi-disciplinary

Community effort ?

# You are a Research Software Engineer

- You have a job in research? You write software? → **You are an RSE!**
- We often lack recognition/career progression despite contributions

→ **Join RSE communities !**

- Associations at international, national, institutional scale  
<https://researchsoftware.org/assoc.html>
- Networking and events
- Acquire skills for managing projects and building careers
- Next international event: **RSE CON 24**, September  
<https://rsecon24.society-rse.org/>