OSCARS
Open Science Clusters' Action for Research & Society

Funded Project

CodeMetaSoft

CodeMetaSoft

Principal Co-Investigator: **Daniel Garijo**, Universidad Politécnica de Madrid
Principal Co-Investigator: **Thomas Vuillaume**, Laboratoire d'Annecy de Physique des Particules, CNRS
Project team members: Tom Francois, Anas el Hounsri, Esteban González Guardia

Implemented by

Ontology Engineering Group

LAPP Laboratoire d'Annecy de Physique des Particules

cnrs

Funded by the European Union

CodeMetaSoft

Improving Research Software metadata good practices across OSCARS science clusters

OSCARS Funding:

€ 250000

Project Start:

01-Nov-2024

Project End:

01-Nov-2026

Field:

All clusters
Research Software
Metadata

Principal Investigators:

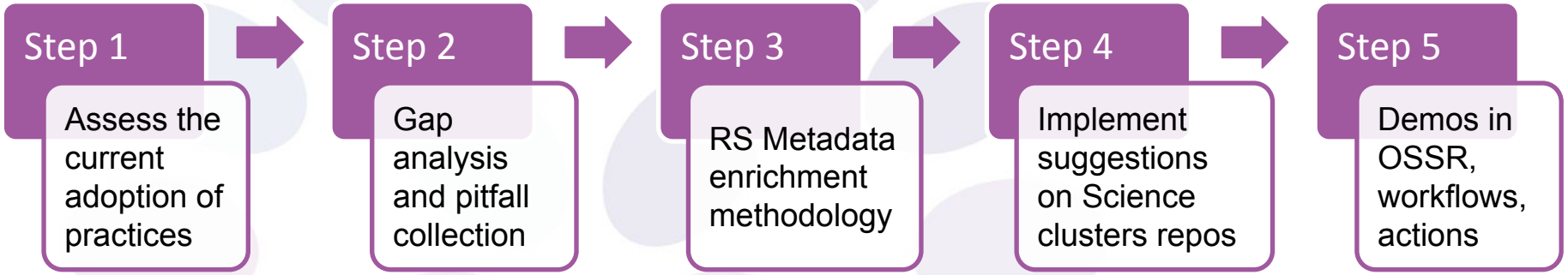
Daniel Garijo, UPM
Thomas Vuillaume, LAPP

Other Researchers involved:

Tom Francois, LAPP
Anas el Hounsri, UPM
Esteban González, UPM

Challenge addressed

Ease the adoption of Research Software metadata & good practices
Automate metadata propagation and interoperability
Propose suggestions for researchers

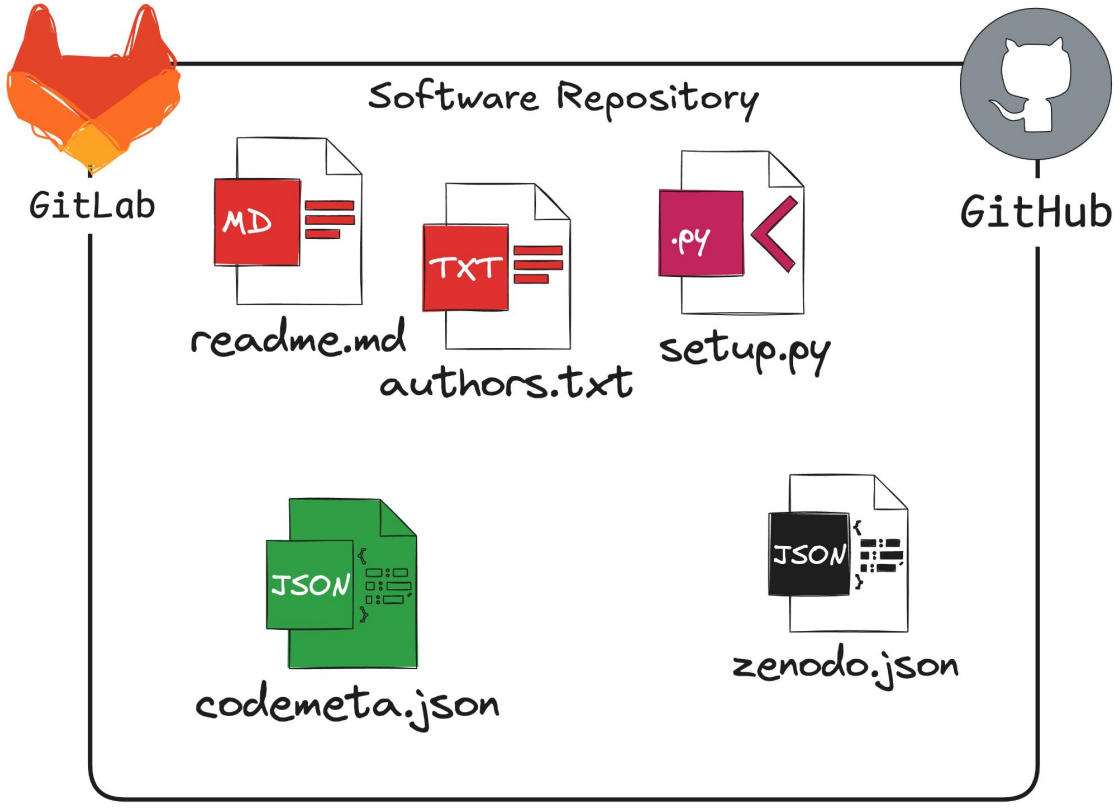


IMPACT

Improving metadata adoption and FAIR4RS principles in European Science clusters, increase the adoption of CodeMeta as a Research Software metadata standard

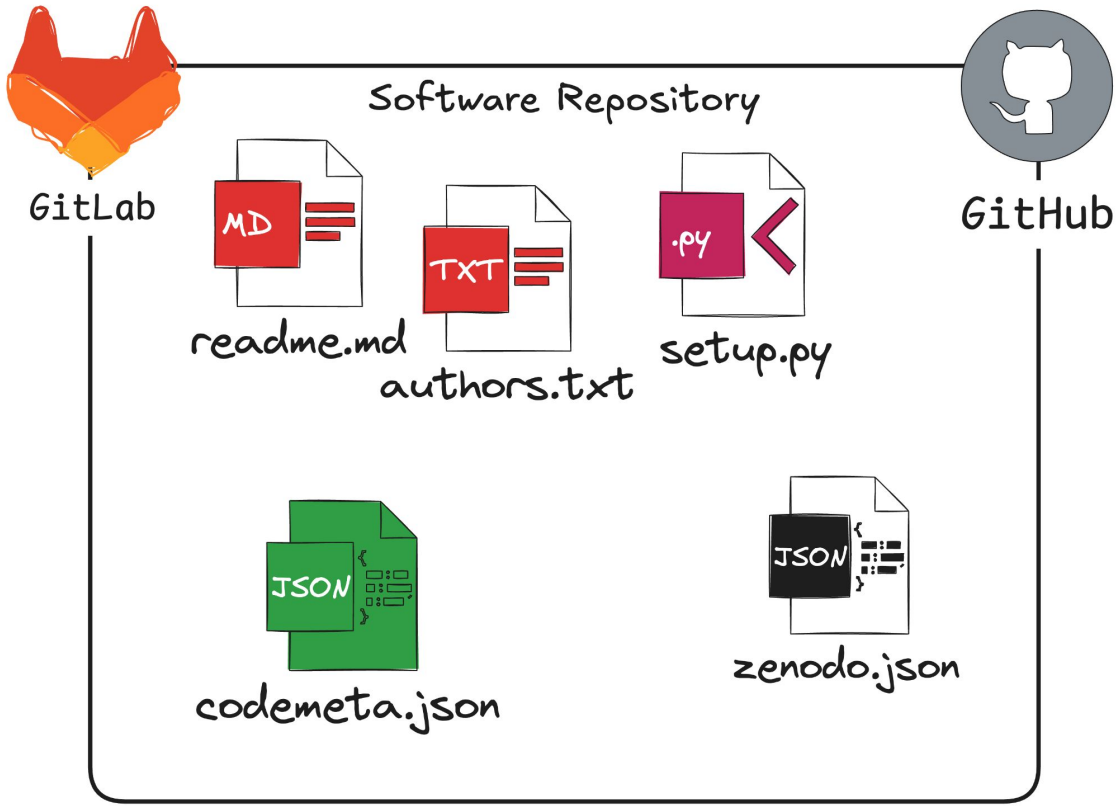
Organisations involved:

Research Software metadata are a core element of FAIRness.



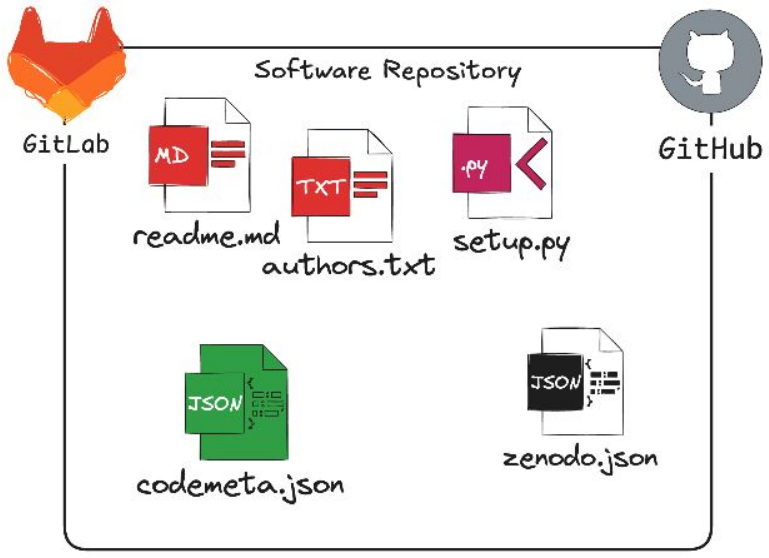
- Sources of software metadata are often project or platform specific.
  - setup.py, setup.cfg in python
  - pom.xml in Java
  - README.md
  - …
- CodeMeta is becoming the metadata standard for software metadata.

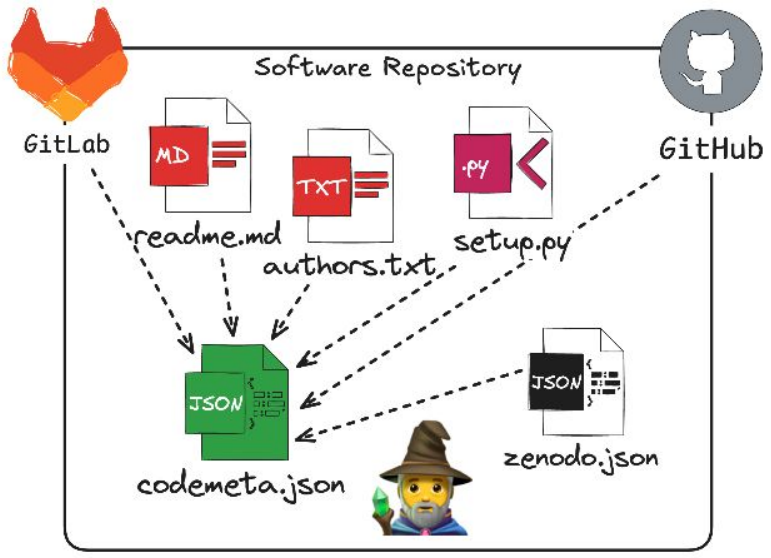
Challenge Addressed

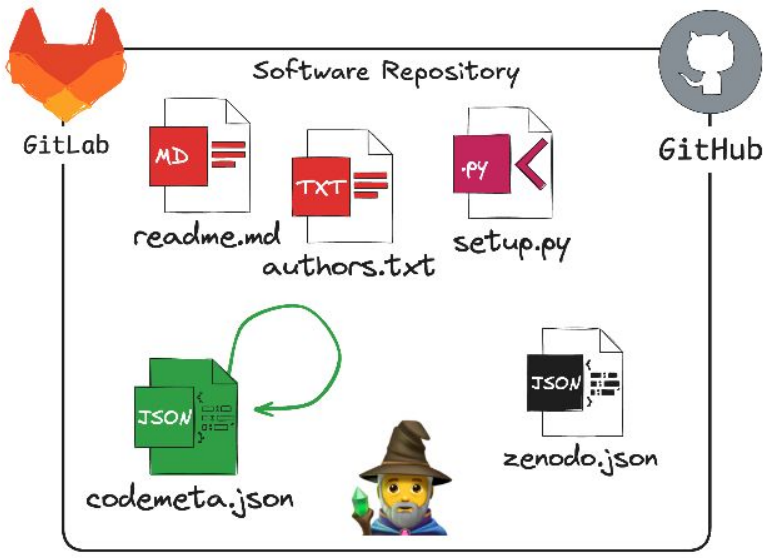


- Software metadata is currently disseminated in heterogeneous files and documentation
- Lack of automated suggestions and enrichment for improving software metadata

Proposed Solution







- Integrate and enrich Research Software (RS) metadata records
- Tools to ease metadata compliance, propagation and automated suggestions and enrichment
- Automate RS metadata maintenance workflows
- Means to measure metadata gaps and the adoption of best practices
- Methodology for RS enrichment
- Demonstrators through clusters and OSSR

Auto-codemeta wizard: https://autocodemeta.linkeddata.es/

Run-time environment

Programming Language: Python

Runtime Platform: .NET, JVM

Operating System: Android 1.6, Linux, Windows, macOS

Other software requirements

Name/Text	Version	+
astropy	>=6.1.0	-
numpy	>=1.25.0	-
packaging	>=23.2	-
parfive	ftp]>=2.1.0	-
pyerfa	>=2.0.1.1	-
requests	>=2.32.0	-
fsspec	>=2023.6.0	-
setuptools	>=62.1	

Current version of the software

Version number: v6.0.5

Release date: 2025-05-21

Download URL: https://github.com/sunpy/sunpy/releases

Release notes

```
## What's Changed
* Backport PR #7911 on branch 6.0 (Adds  support for the
timestamp %Y%m%d%H%M) by @meeseeksmachine in https://
github.com/sunpy/sunpy/pull/7915
* Backport PR #7920 on branch 6.0 (bug fix for suviclient
unit conversion) by @meeseeksmachine in https://
github.com/sunpy/sunpy/pull/7925
* Backport PR #7917 on branch 6.0 (Handling
FILLVAL_ATTRIBUTES Missing) by @meeseeksmachine in
https://github.com/sunpy/sunpy/pull/7929
* Backport PR #7933 on branch 6.0 (Updates from the
package template) by @meeseeksmachine in https://
github.com/sunpy/sunpy/pull/7934
* Backport PR #7939 on branch 6.0 (Updates from package
template) by @meeseeksmachine in https://github.com/
sunpy/sunpy/pull/7941
* Backport PR #7942 on branch 6.0 (Renable full test-
suite) by @meeseeksmachine in https://github.com/sunpy/
sunpy/pull/7944
* Backport PR #7937 on branch 6.0 (Optimizing
test_goes_suvi.py) by @meeseeksmachine in https://
github.com/sunpy/sunpy/pull/7945
* Backport PR #7954 on branch 6.0 (Updates from the
package template) by @meeseeksmachine in https://
github.com/sunpy/sunpy/pull/7956
```

Reference Publication

Reference Publication URL: https://doi.org/10.3847/1538-4357/ab4f7a

Title of publicationn: the sunpy project: open source development and status of the version 1.0 core package

DOI of publication: 10.3847/1538-4357/ab4f7a

ISSN of publication: ISSN

Date of publication: Date Published

Authors (either persons or organizations) can be added below

Authors

← Add one Remove last → Nº Authors: 35

Author #1

Type Author: Organization

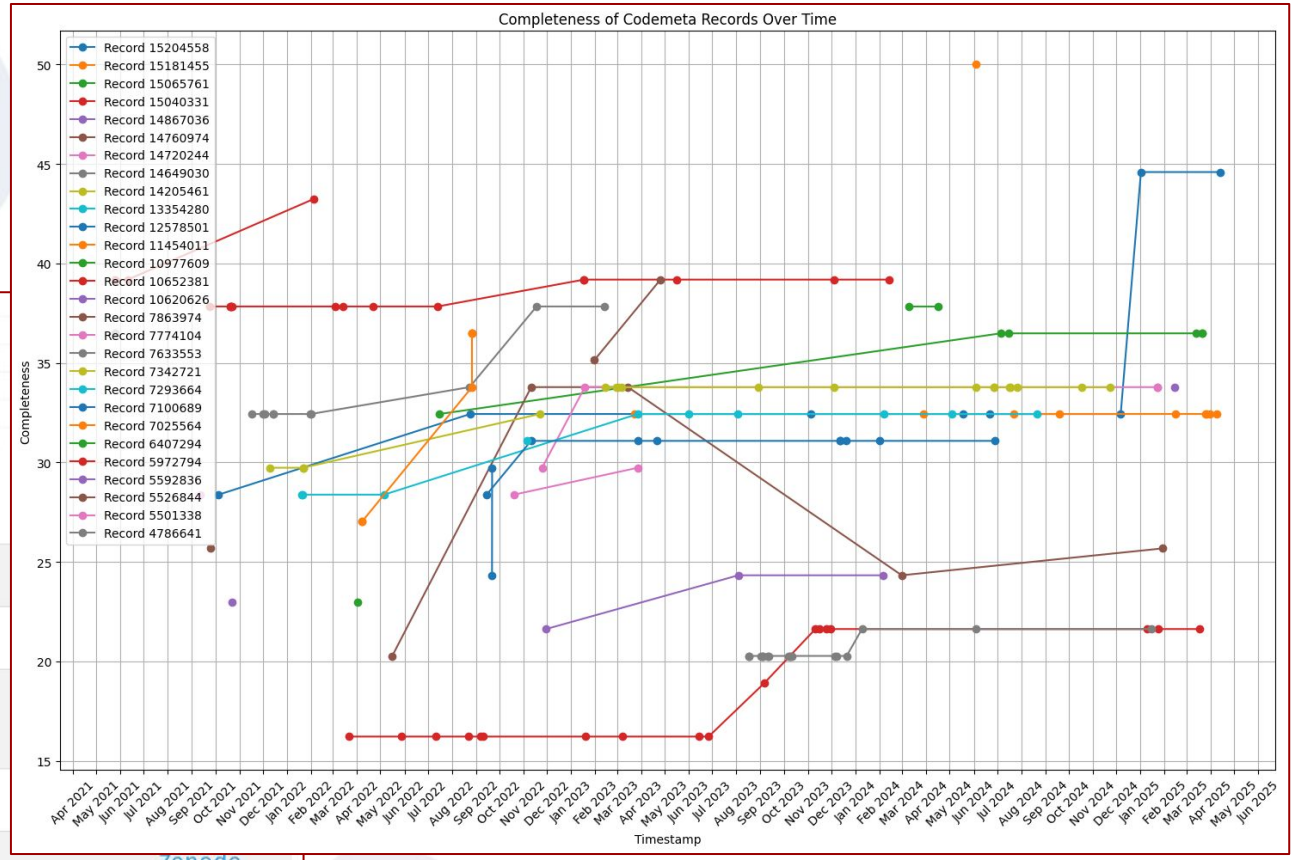
< Change priority >

Name: The SunPy Community

E-mail address: jane.doe@example.org

Metadata comparison reports:

https://rs-quality-checks-2b2333.gitlab.io/records/report/

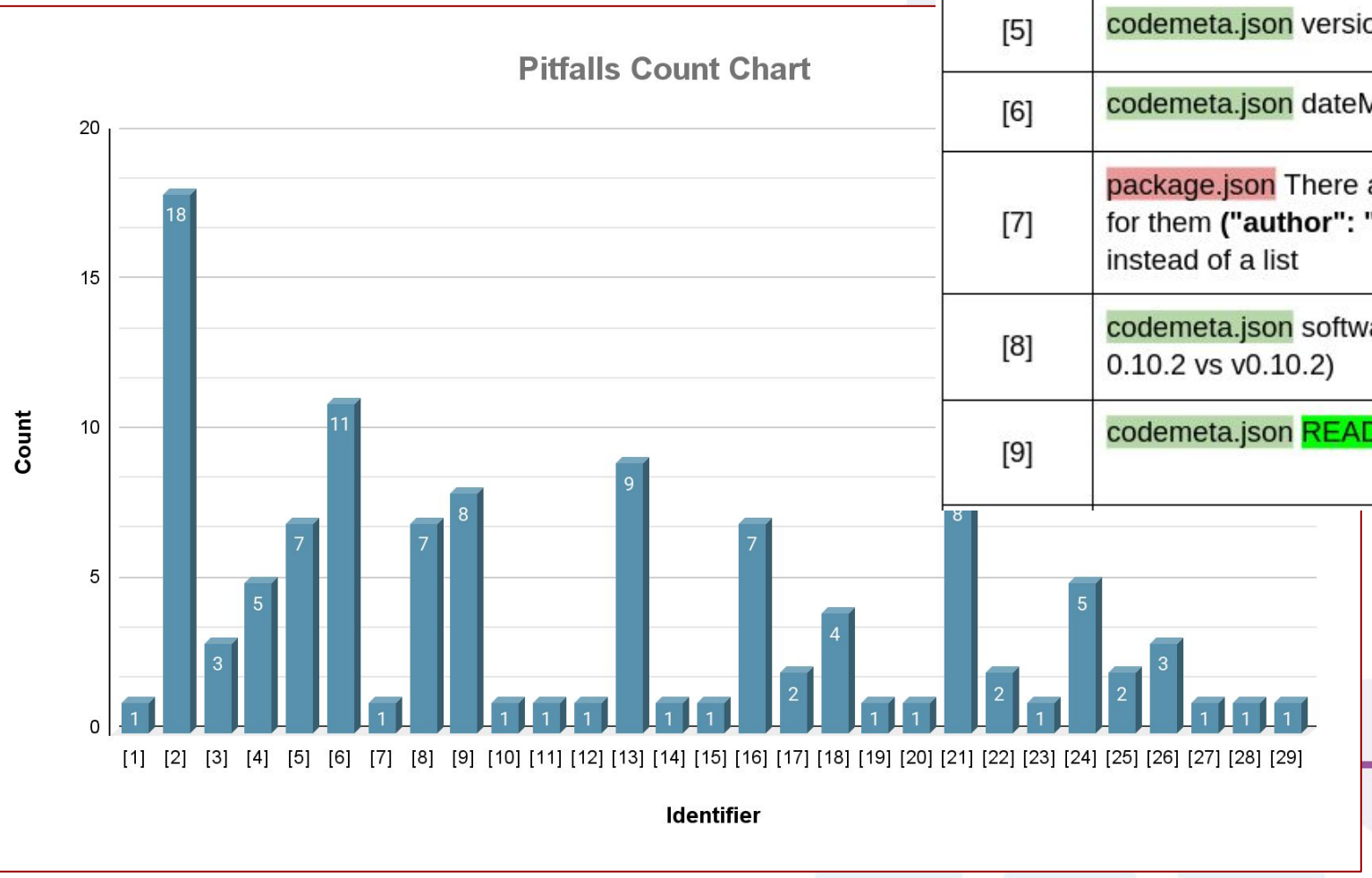


Comparison Reports

This document contains a table with individual comparison reports for each record.

Record ID	Report URL	Repository Kind	
15204558	R3BRoot	RepoType.GITHUB	
15181455	Aladin Lite	RepoType.GITHUB	
15065761	CTLearn: Deep learning for imaging atmospheric Cherenkov telescopes event reconstruction	RepoType.GITHUB	
15040331	FairMQ	RepoType.GITHUB	
14867036	Gammapy: Python toolbox for gamma-ray astronomy	RepoType.GITHUB	Zenodo
14760974	Gammapy: Python toolbox for gamma-ray astronomy	RepoType.GITHUB	Zenodo
14720244	cds-escape-tutorials	RepoType.GITHUB	Zenodo
14649030	timewise-sup: The Timewise Subtraction Pipeline v0.5.1	RepoType.GITLAB_SELF_HOSTED	Zenodo

Metadata pitfall catalog (in progress)



[4]	codemeta.json SoftwareRequirements don't have versions	5
[5]	codemeta.json version is outdated	7
[6]	codemeta.json dateModified is outdated	11
[7]	package.json There are more than two authors but only one section for them **("author": "Thomas Boch and Matthieu Baumann")** instead of a list	1
[8]	codemeta.json softwareVersion has Minor inconsistency (**example:** 0.10.2 vs v0.10.2)	7
[9]	codemeta.json README pointing to their homepage/wiki	8

Next Steps & Expected Scientific Impact

- What is going to change thanks to your project?
  - CodeMeta **maintenance in software repositories is simplified**. As a result its adoption in the Science Clusters increases, making software more FAIR globally.
  - **Gaps in metadata are identified in software catalogues**, helping Science Clusters focusing their efforts where they are most needed
- Resources that will be made available:
  - Open service(s) and actions usable by others from any community
- Sustainability:
  - Rely on **existing tooling** (e.g., CodeMeta generator) and **standards** (CodeMeta)
  - The developed solution and results will be open-source and published in Zenodo to be (re)usable by anyone.
- A first landscape analysis of good practices has been accepted at MSR'25 [1]

[1] El Hounsri, Anas and Garijo, Daniel. Good practice versus reality: A landscape analysis of Research Software metadata adoption in European Open Science Clusters. To appear in Proceedings of the Mining Software Repositories Conference, 2025. Association for Computing Machinery. MSR '25. 2025. https://dgarijo.com/papers/El_Hounsri_MSR_2025_landscape_analysis_CR.pdf


OSCARS
Thank you