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 Engineer ingGroup

UNIVERSIDAD POLITÉCNICA MADRID

LAPP
Laboratoire d'Annecy de Physique des Particules

cnrs

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

Step 1
Assess the current adoption of practices

Step 2
Gap analysis and pitfall collection

Step 3
RS Metadata enrichment methodology

Step 4
Implement suggestions on Science clusters repos
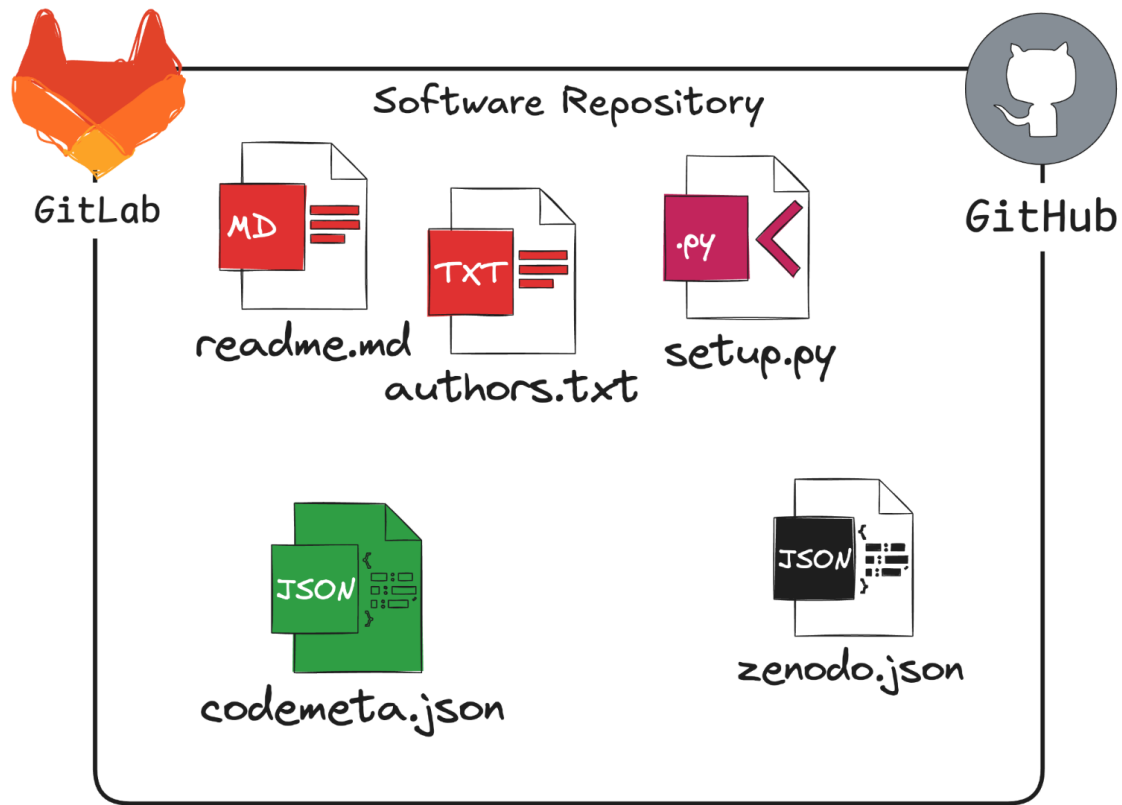
Step 5
Demos in OSSR, workflows, actions

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

Organisations involved: Ontology Engineering Group   LAPP Laboratoire d'Annecy de Physique des Particules   cnrs

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.
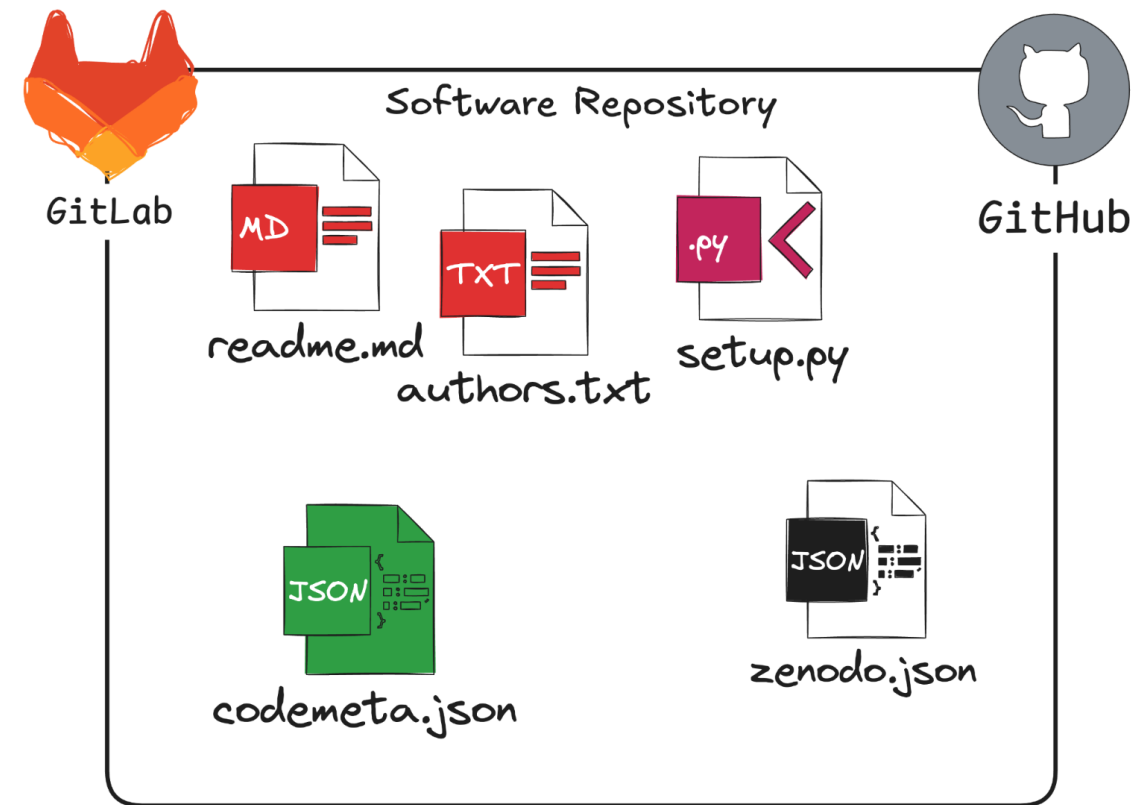
- Software metadata is currently disseminated in heterogeneous files and documentation

- Lack of automated suggestions and enrichment for improving software metadata
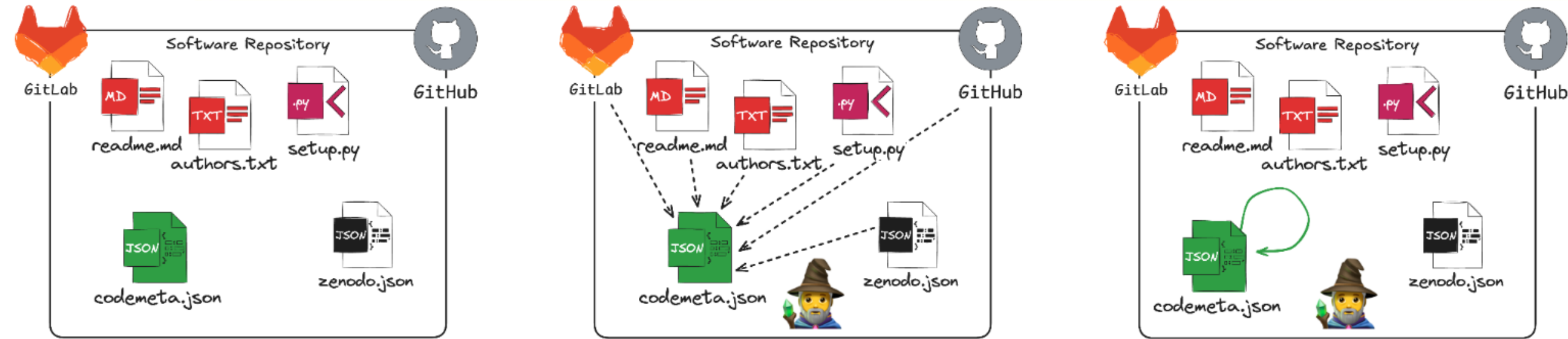
- Integrate and enrich Research Software (RS) metadata records

- 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/

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

Completeness of Codemeta Records Over Time
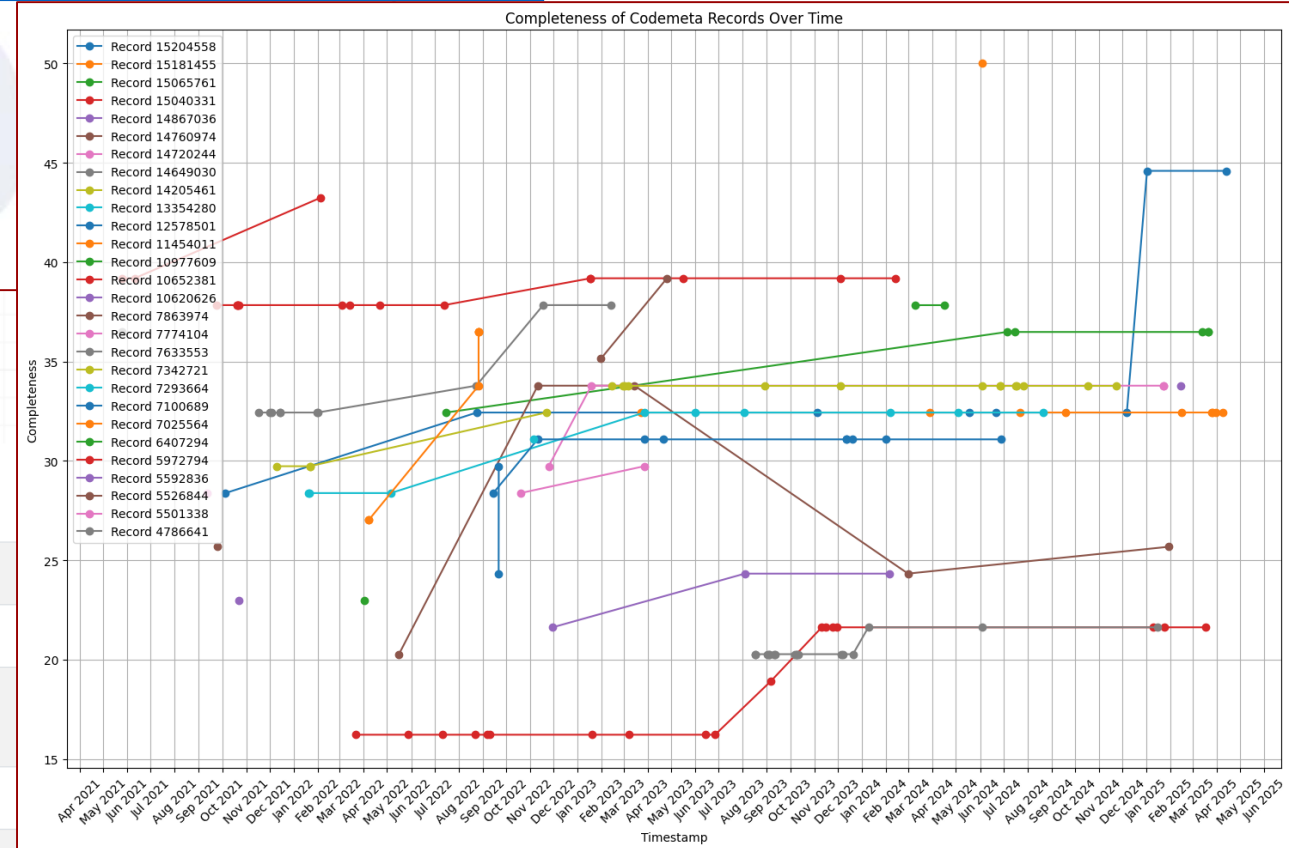
Metadata pitfall catalog (in progress)
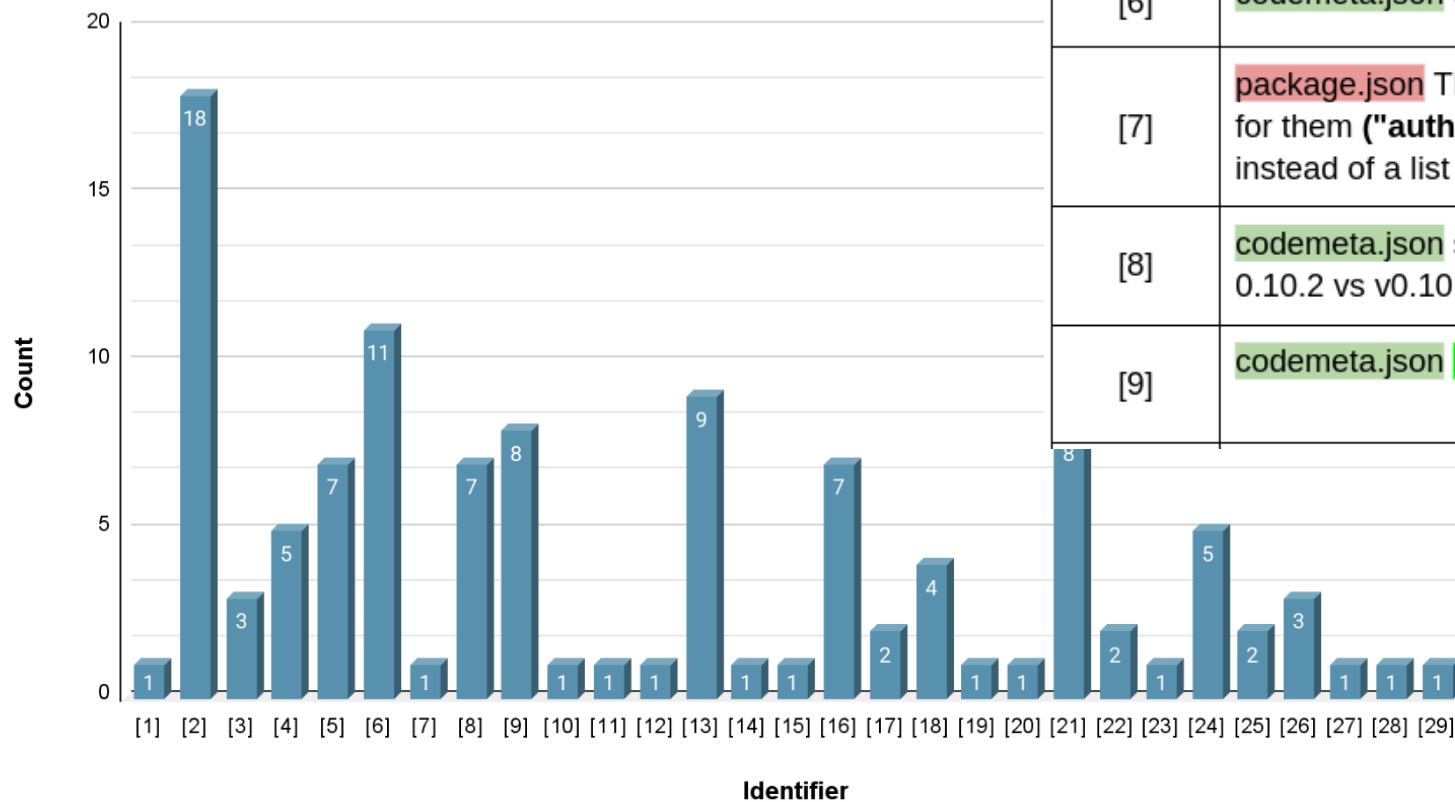
[4]	codemeta.json SoftwareRequirements don't have versions	5
[5]	codemeta.json version is outdated	7
[6]	codemeta.json dateModified is outdated	11
[7]	package.json There are more than two authors but only one section for them (**"author": "Thomas Boch and Matthieu Baumann"**) instead of a list	1
[8]	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

Pitfalls Count Chart

Count

20

15

10

5

0

18

11

9

8

8

7

7

7

5

5

4

3

3

2

2

2

1 1 1 1 1 1 1 1 1 1 1

[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29]

Identifier

- PaN software catalogue https://software.pan-data.eu/software
  - Reused for landscape analysis of best practices [1]

- To integrate: **European Photon and Neutron open data search portal**
  - Will be used for expanding landscape analysis

- Next steps:
  - Extend pitfall analysis to PaN catalogs
  - Extend metadata completeness analysis to PaN catalogs

[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

- ## 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