

Lessons Learned from the Research Software Journey



Morane Gruenpeter, Director of Scholarly Ecosystem



Software Heritage

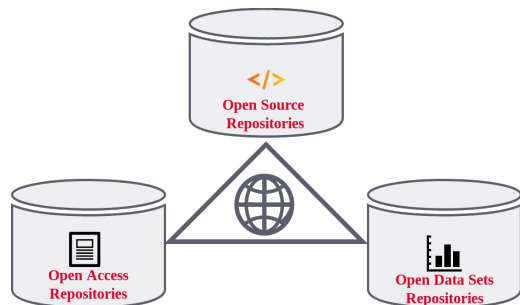
THE GREAT LIBRARY OF SOURCE CODE



Inria

Research Software in EOSC: From SRIA (2021) and SIRS (2020) to OA7 (2025)

A key pillar: software (source code)



← The links between research outputs are super important for reproducibility!



FAIRCORE4EOSC

Core Components Supporting a FAIR EOSC



FAIR-IMPACT

Expanding FAIR solutions across EOSC

Lesson #1: Cross-Infrastructure coordination starts with clear **policy decisions**

Lesson #2: Streamlined integration needs dedicated **support and resources**

Lesson #3: Beyond researchers – strengthening **institutional engagement and training**

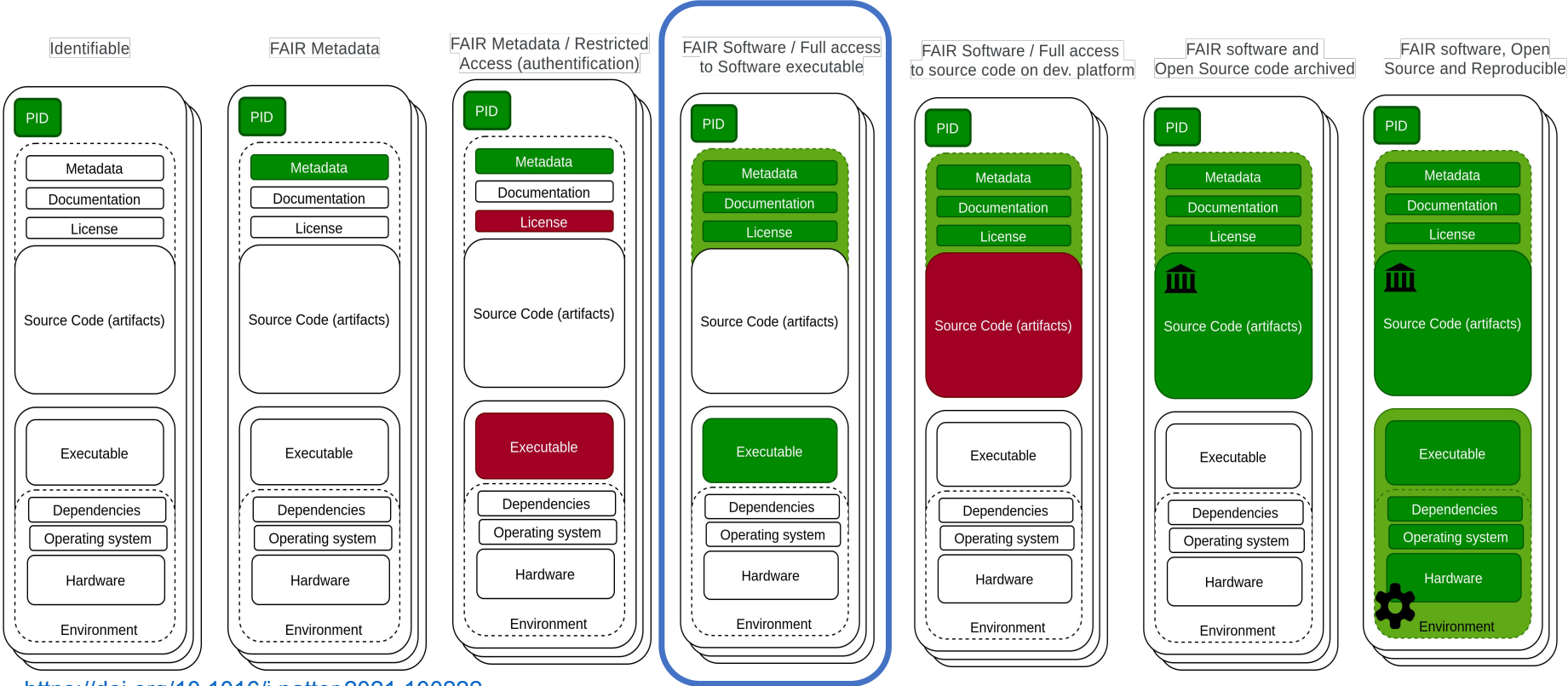
Collaborating with



OA Expert Group:
Research Software



FAIR is not enough for Software - is EOSC enough?



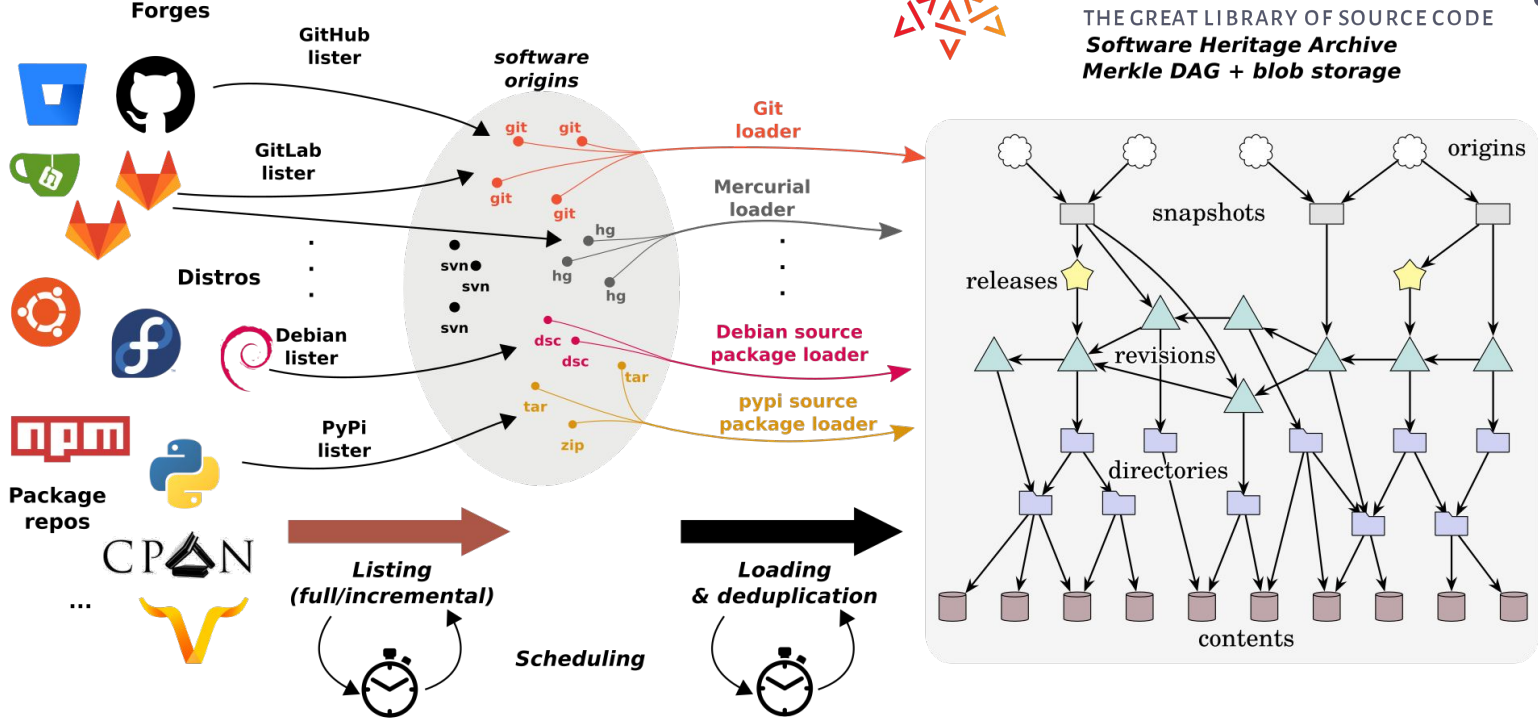
<https://doi.org/10.1016/j.patter.2021.100222>

A resilient solution: from forges to the universal archive



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE
Software Heritage Archive
Merkle DAG + blob storage





Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

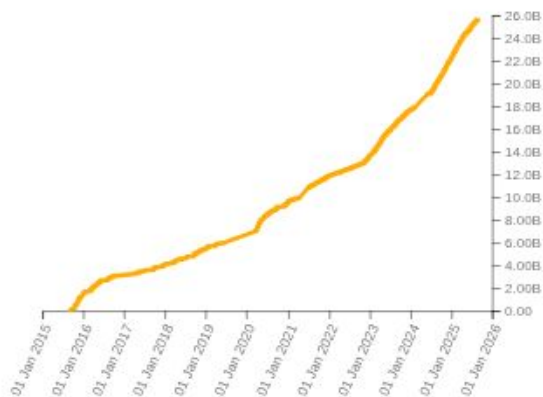
An open infrastructure

to **collect, preserve and share** *all source code*



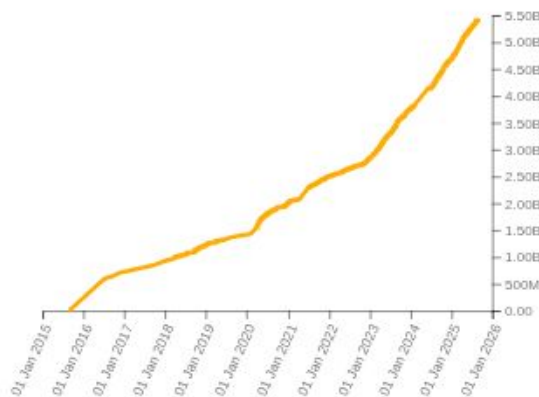
Source files

25,762,930,139



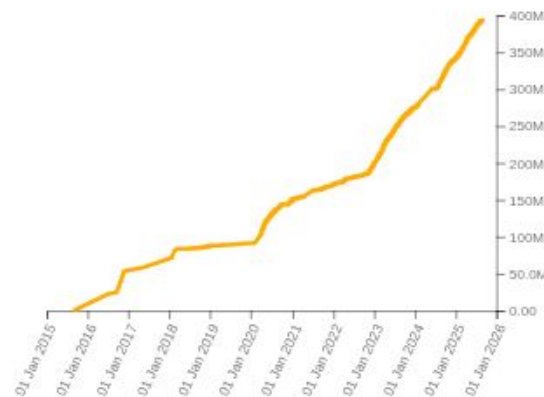
Commits

5,443,004,374



Projects

395,971,819



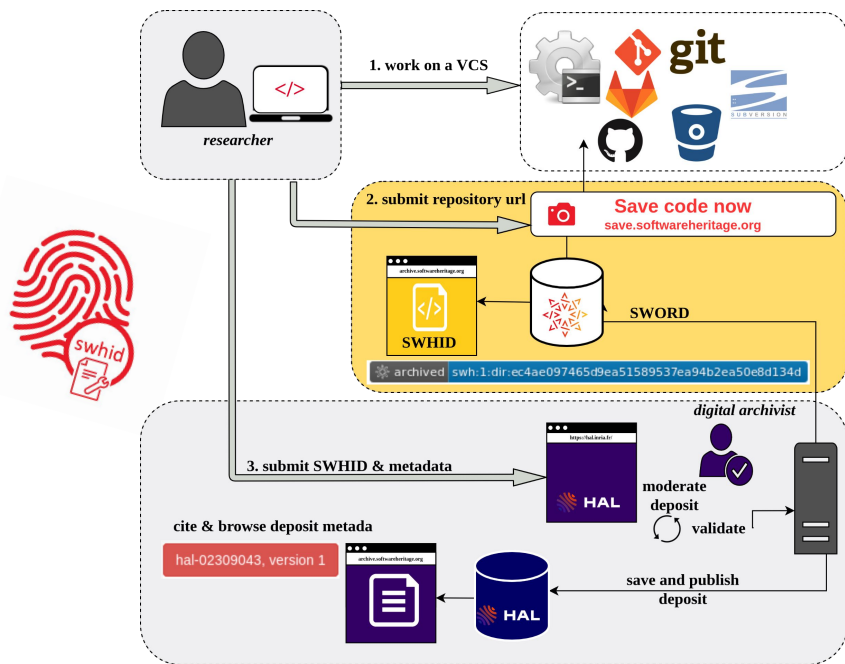
Snapshot: Sept. 2nd 2025

<https://archive.softwareheritage.org/>

HAL <> SWH - a French success story



- Collaboration started in 2017 by the CCSD, IES-INRIA & Software Heritage
- A streamlined easy workflow for researchers to get acknowledgement



CCSD
Centre pour la Communication
Scientifique Directe

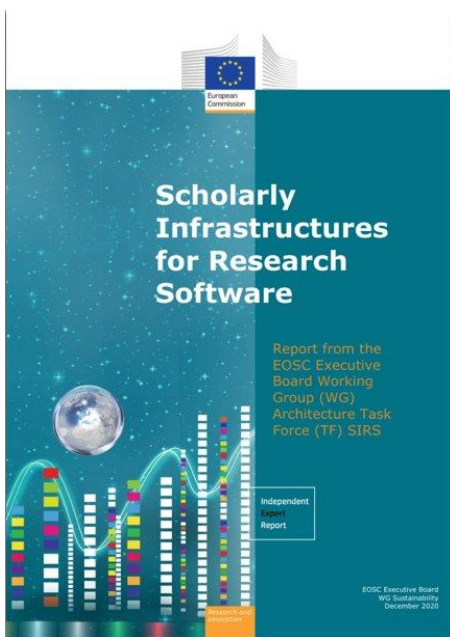
HAL
science ouverte

**RÉPUBLIQUE
FRANÇAISE**
Liberté
Égalité
Fraternité

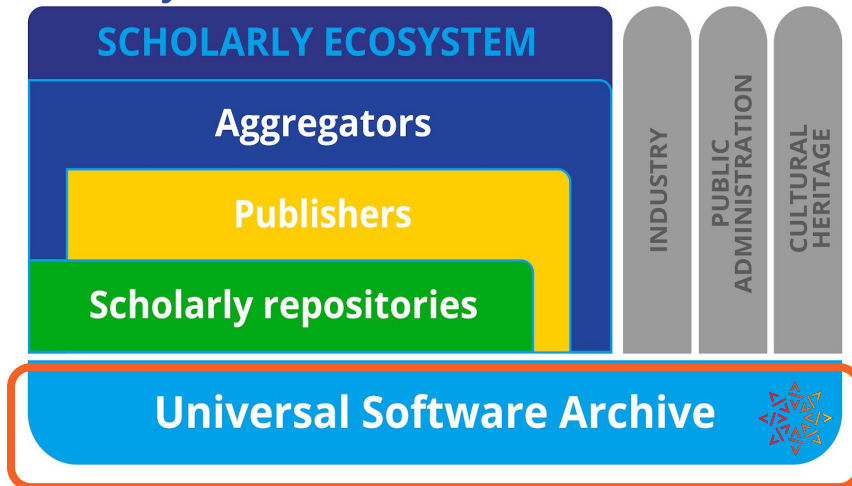
Inria

 Software Heritage

Lesson #1: Cross-Infrastructure Coordination Starts with Clear Policy Decisions



Scholarly Infrastructures for Research Software (SIRS)



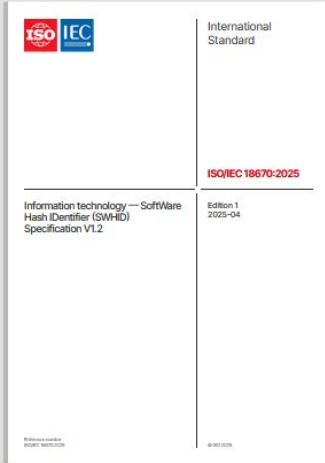
Short term recommendations

- Strengthening interactions between
 - Aggregators, publishers, scholarly repositories, and Software Heritage
- Metadata standards & tools
- Generalizing the use of PIDs (extrinsic & intrinsic) => **SWHID**

SIRS report: European Commission, Directorate-General for Research and Innovation, *Scholarly infrastructures for research software : report from the EOSC Executive Board Working Group (WG) Architecture Task Force (TF) SIRS*, Publications Office, 2020, <https://data.europa.eu/doi/10.2777/2859>



The SWHID is now an ISO/IEC standard (18670:2025)



[Read sample](#)

ISO/IEC 18670:2025

Information technology — Software
Hash Identifier (SWHID)
Specification V1.2

Published (Edition 1, 2025)

<https://www.iso.org/standard/89985.html>

<https://www.swhid.org/>

<https://www.softwareheritage.org/software-hash-identifier-swhid/>



ISO/IEC 18670:2025

 eosoc

FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC

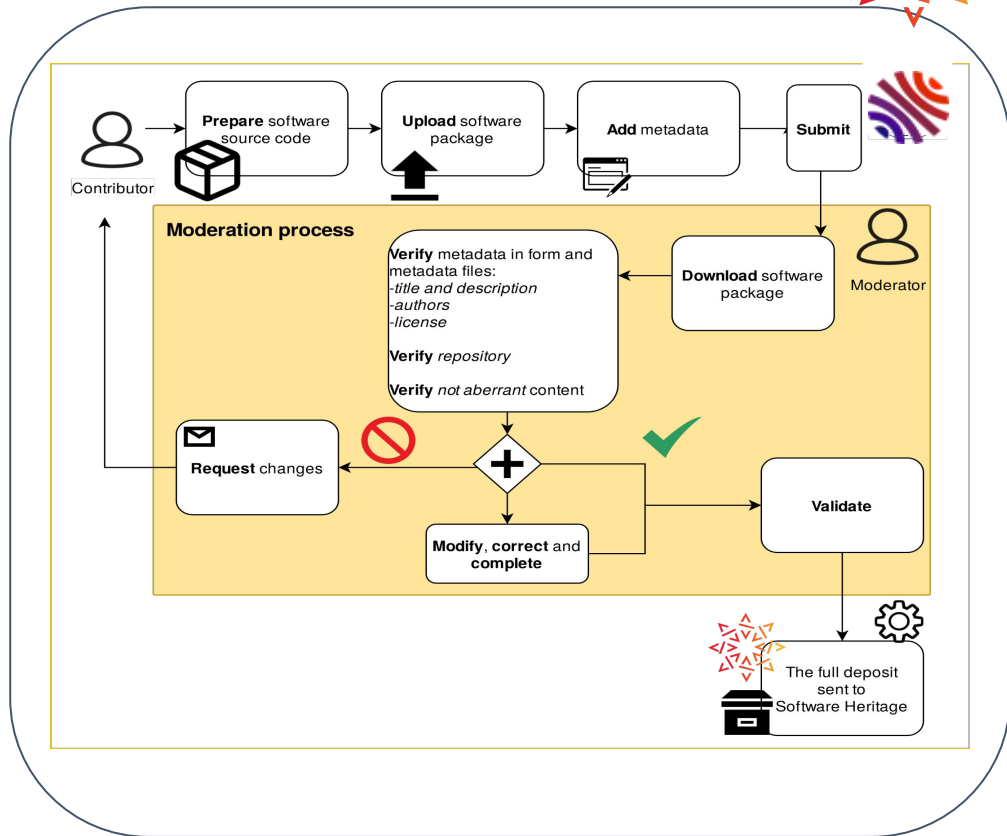
Lesson #2: Streamlined integration needs dedicated support and resources



Checking metadata quality in HAL



Sabrina Granger et al. Modérer un dépôt logiciel dans HAL : dépôt source et dépôt SWHID. [Rapport Technique] Inria; CCSD; Software Heritage. 2022. (hal-01876705v2)



The Research Software MetaData guidelines

The RSMD seven Aspects

1. General Metadata Requirements

2. Accessibility & preservation

3. Reference & identification

4. Description & classification

5. Attribution & credit

6. Reuse, licensing & legal aspects

7. Re-execute: Dependencies & execution environment

SIRS report

FAIR4RS

A = Archive

A = Accessible

R = Reference

F = Findable

D = Describe

I = Interoperable

C = Cite

R = Reusable

 <https://github.com/FAIR-IMPACT/RSMD-guidelines>
<https://fair-impact.github.io/RSMD-guidelines/>

The Software Metadata Curation Roadmap



Lesson #3: Beyond researchers – strengthening **institutional engagement** and training

Short-term (0-2 years)

1. **Support and training:** Academic institutions should invest in curation and training activities for researchers and support staff, acknowledging the significant effort required.
2. **Infrastructure curation capabilities:** Infrastructures encompass a wide range of platforms, including aggregators, publishers and scholarly repositories.
3. **Adopting and adapting metadata guidelines:**
4. **Community effort:** Infrastructures should actively engage in community-driven efforts (
5. **Recognition & acknowledgement:**
 - **Career evaluation:** Institutions should integrate curated metadata records into activity reports to highlight software as a recognized and valued research output...
 - **Citation standard:** Strengthening the connection between researchers and their software outputs



[10.5281/zenodo.14509418](https://doi.org/10.5281/zenodo.14509418)

Read the full **D6.1**
report

Call to action: Code and AI



Software recognition requires a **supported** and **coordinated** effort in **EOSC**

We need to treat code not just as a tool, but as a core pillar of our intellectual heritage.

- **Open foundational models:** *If a foundational model is trained on public code, it “must be given back,” at least through open or accessible licensing. AI should not privatize the commons that were built openly.*
- **Transparency:** *There must be full transparency regarding which code files from the archive are used in training. Thanks to the Software Heritage archive and its intrinsic identifiers (SWHIDs), each source file can be precisely referenced, ensuring traceable datasets.*
- **Opt-out mechanism:** *Code owners must retain agency. A robust opt-out mechanism should allow maintainers to request that their code be excluded from future model training.*

Mutualisation

Curation

Sustainability

Open by default

Recognition



Join the effort

morane@softwareheritage.org

<https://www.softwareheritage.org/>

Become an **ALIG member**

<https://www.softwareheritage.org/support/members-alig/>

Volunteer as a **SWH ambassador**

<https://www.softwareheritage.org/ambassadors/>

Subscribe to the **newsletter**

<https://www.softwareheritage.org/newsletter/>

