Archiving, referencing and describing software with **HAL** and **Software Heritage**

EOSC France 2024 : National Tripartite Event,
12-13 Sep 2024 Paris

Yannick Barborini - **CCSD / CNRS**

Agenda

- The **HAL+ infrastructure**

- **Collaboration** between CCSD - Inria - Software Heritage

- **Archive** software in HAL and in SWH

- **Describe** and **Reference** source code

- Software in **publication workflows**

- The **next steps**

HAL+ : French Research Infrastructure for Open Science

HAL is the **multidisciplinary open archive** chosen by the entire French scientific and academic community for the dissemination of knowledge

Diamond open access journal platform: a complete solution for editing and publishing overlay journals

Platform dedicated to **organisers of conferences**, workshops or scientific meetings

CCSD
Centre pour la Communication Scientifique Directe

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
Liberté
Égalité
Fraternité

The **collaboration**

Stakeholders

CCSD · · · · · · · · · · · Inria · · · · · · · · Software Heritage

First contact between stakeholders

Service available for all HAL users
Local source code deposit to HAL

SWHID with metadata deposit
HAL-Inria portal

2017

Autumn 2018

Spring 2022

Spring 2018

2020

2023

Preliminary phase with HAL-Inria

Export available
BibLaTeX & codemeta.json export

SWHID with metadata deposit
For all HAL users

4

Archiving software - Source code method

Software **developed locally** and shared on a personal or institutional website

One compressed archive containing the following files :
- **source code files**
- README
- AUTHORS
- LICENSE
- codemeta.json
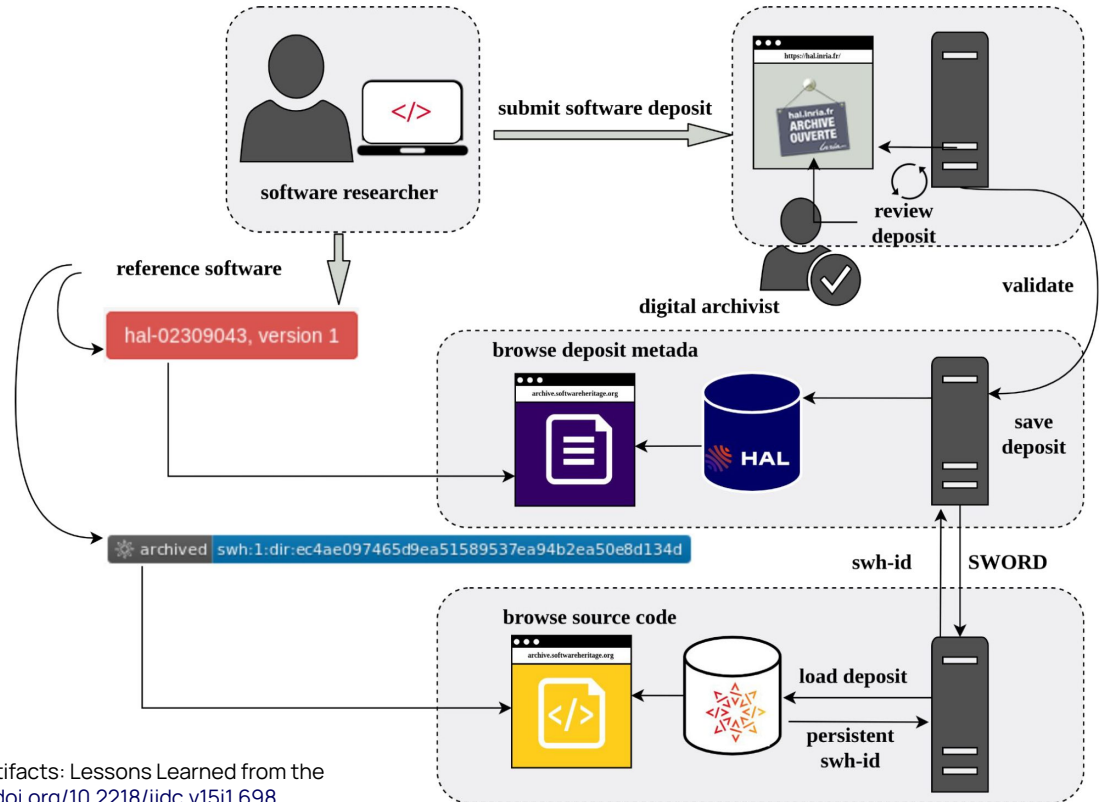
Archiving software - SWHID method

Software **developed on a collaborative plateforme** using a version control system (on GitHub, GitLab, Bitbucket…)

Best practices on repository VCS
- AUTHORS, LICENSE & README file
- codemeta.json

Save code now with Software Heritage
- you have a SWHID

Use SWHID to create HAL entry complete metadata extract from codemeta.json
- complete author affiliation
- add references publications

researcher

1. work on a VCS

2. submit repository url

Save code now
save.softwareheritage.org

archive.softwareheritage.org

SWHID

SWORD

archived   swh:1:dir:ec4ae097465d9ea51589537ea94b2ea50e8d134d

digital archivist

https://hal.inria.fr/

HAL

3. submit SWHID & metadata

moderate deposit

validate

cite & browse deposit metada

hal-02309043, version 1

archive.softwareheritage.org

HAL

save and publish deposit

6

Describing software

General properties common to the other types of deposit
- Title,
- Description
- Funding information
- Thesaurus (ACM, JEL, …)
- …

Information about author(s)
- specific role (coding, management, …)
- affiliations

Software specific metadata
- Licences **based on the SPDX reference list (https://spdx.org/licenses/)**
- Programming language
- Code repository
- Platform/OS - environment
- Version
- Development status
- Runtime Platform

7
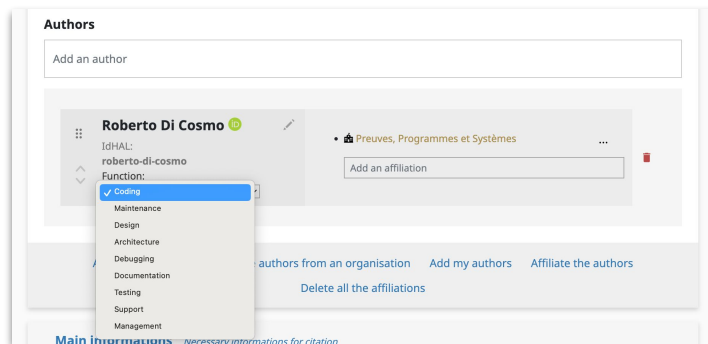
The Moderation process

Librarians curators

⇒ control metadata and sources files
⇒ add or corrects metadata
⇒ dialogue with contributor to improve deposit
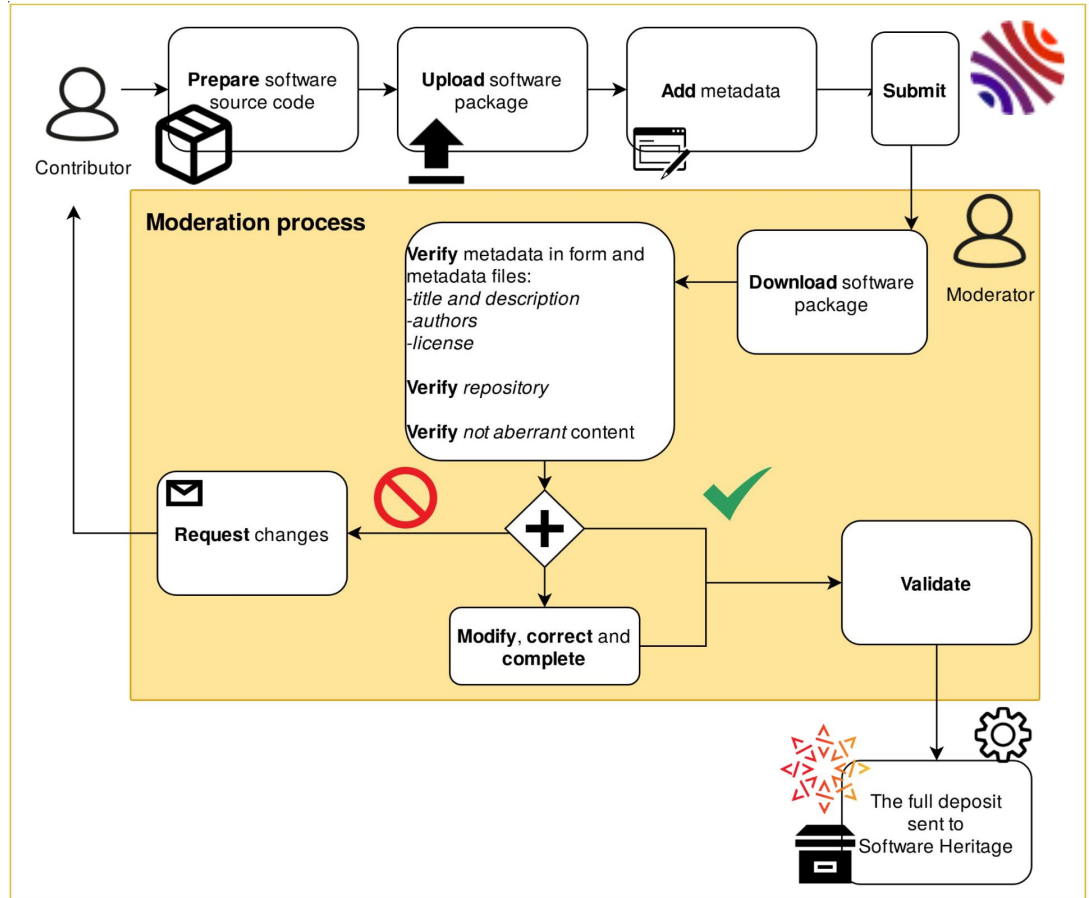
Moderation guide :
Morane Gruenpeter, Jozefina Sadowska. La modération d'un dépôt logiciel : Bonnes pratiques et guide pour le modérateur. [Rapport Technique] Inria; CCSD; Software Heritage. 2018. ⟨hal-01876705⟩

8

Software deposit in HAL

Link to the content on SWH

Software specific metadata

Citation

Export

https://hal.science/hal-04367491v1

Citation and **Exports**

- Citation available directly from records
  - **SWHID** : Gives access to the source code files
  - **HAL-ID** : Refers to the software as a work

- BibTeX Export with BibLaTeX **@software** or **@softwareversion**

- **codemeta.json** export

Cite

Alain Bonardi, Paul Goutmann, David Fierro, Adrien Zanni. abclib. 2023, ⟨swh:1:dir:1cc37ea 1204712905a0f25fd0ce309eea9b25af0;origin =https://hal.archives-ouvertes.fr/hal-0436749 1;visit=swh:1:snp:670b1be9dc54c1765fc49ae9 51a22c9109a6e76b;anchor=swh:1:rel:42914f2 eced0e28fbb498e3c948075deba586350;path =/⟩. ⟨hal-04367491⟩

Exporter

CodeMeta	BibTeX	XML-TEI
Dublin Core	DC Terms	EndNote
DataCite		

```
@softwareversion{bonardi:hal-04367491v1,
  TITLE = {{abclib}},
  AUTHOR = {Bonardi, Alain and Goutmann, Paul and Fierro, David and Zanni, Adrien},
  URL = {https://hal.science/hal-04367491},
  NOTE = {},
  INSTITUTION = {{Centre de recherche Informatique et cr{\'e}ation musicale}},
  YEAR = {2023},
  SWHID =
{swh:1:dir:1cc37ea1204712905a0f25fd0ce309eea9b25af0;origin=https://hal.archives-
ouvertes.fr/hal-
04367491;visit=swh:1:snp:670b1be9dc54c1765fc49ae951a22c9109a6e76b;anchor=swh:1:rel:42914f
2eced0e28fbb498e3c948075deba586350;path=/},
  VERSION = {1.0.3},
  REPOSITORY = {https://github.com/alainbonardi/abclib},
  LICENSE = {GNU LESSER GENERAL PUBLIC LICENSE Version 3},
  KEYWORDS = {spatial sound processing ; spatial sound synthesis ; ambisonics ; multi-
channel sound processing ; utility objects for mixed music ; Faust language},
  HAL_ID = {hal-04367491},
  HAL_VERSION = {v1},
}
```

The identifiers of the software deposit

Reference with the **SoftWare Hash IDentifier (SWHID)**	Cite with the **HAL-ID**

archived swh:1:dir:1cc37ea1204712905a0f25fd0ce309eea9b25af0

hal-04367491v1

- Identification of the software source code artefact
- a digital fingerprint specific source code content

- Identification of the software record
- Metadata of the deposit
- Authors and contributors are verified in the moderation process

Needed to :
- Identify - reproduce
- Archive

Needed to :
- Give credit to the authors
- Index

Software in **publication workflows**

Possibility of **linking a publication submitted to episciences to software** (with SWHID or HALId)
⇒**The link is displayed on the article landing page during the evaluation phases and after publication.**

Link also available :
- in Episciences APIs (REST, OAI-PMH)
- in export formats (XML, …)

⇒ **Possibility of submitting and evaluating software** in Episciences previously deposited in HAL

Next steps

- Joint program in 2024-2025 to **support research software preservation** in university libraries

- Continue to organise **software moderation** at community level

- Export Format
    - improve **XML-TEI** format

- Integration in **Hal Data Triplesore**
    - https://data.hal.science/
    - with SPARQL and RDF

- **Make links between publications data and software**

Links between publications data and software

Improving the creation of links between publications, data and software source codes and exchange the relationships between repositories

⇒HALiance project - WP7

HAL

Links between publications data and software

Software related to the publication

Benefits of depositing in **HAL**

- **Archive** software in HAL and in SWH
  - **Long term preservation** : HAL and SWH have a long term preservation service

- **Ease** the software **identification**
  - the software artifacts with a **SWHID** (SoftWare Heritage Identifier)
  - the metadata record and citation with the **HAL-ID**

- **Describe** source code with verified metadata
  - **Moderation** and control of the metadata by librarians and/or curators

- **Cite** the software deposit with a complete citation
  - Several **exports format** make it easier for citation

- **Great visibility** for software in open science context
  - Softwares integrates the scientific output of researchers (**HAL CV**), research units, universities and institutions
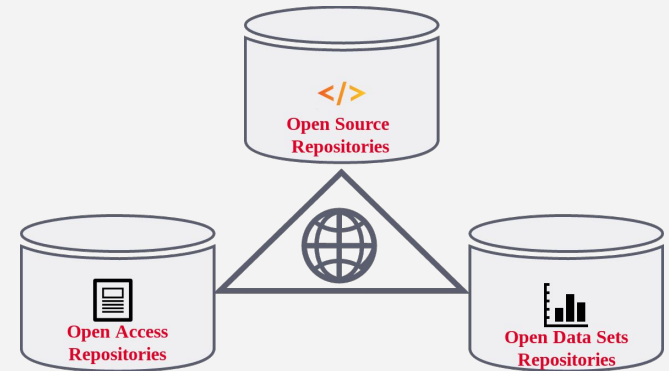
The three pillars of Open Science, Software Heritage CC-By 4.0 2019

Thank you !

Yannick Barborini
yannick.barborini@ccsd.cnrs.fr

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR
ET DE LA RECHERCHE
*Liberté
Égalité
Fraternité*

cnrs

Inria

INRA℮