



De l'archivage à la citation : Software Heritage et HAL, une solution intégrée pour le logiciel de recherche

Libre ABC

Centre pour la Communication Scientifique Directe, Software Heritage

CC-BY 4.0 | Sabrina Granger | Sébastien Mazzaresse | 09-09-2025

HAL & SWH, des infrastructures complémentaires



Archive ouverte nationale pluridisciplinaire de la recherche française

Au service des chercheurs affiliés à des institutions académiques

Objectif : partager en accès libre et gratuit les résultats de recherche



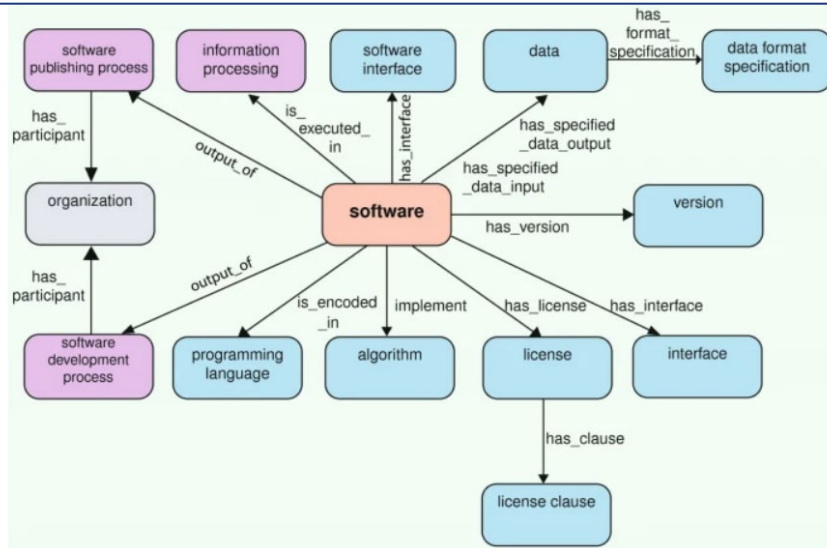
Archive mondiale du code source (Free and Open Source Software)

Pour toutes les personnes qui rédigent et utilisent du code source: administration publique, entreprises, secteur culturel, recherche

Objectif: rendre accessible et identifiable sur le long terme le code source des logiciels

1. Les enjeux documentaires associés aux logiciels de recherche

Le logiciel, une ressource très spécifique ...



The SWO's 'schema'.

Un logiciel n'est pas un jeu de données :

architecture,
cycle de vie,
"comportement"

Malone, J., Brown, A., Lister, A. L., Ison, J., Hull, D., Parkinson, H., & Stevens, R. (2014). The Software Ontology (SWO): A resource for reproducibility in biomedical data analysis, curation and digital preservation. *Journal of Biomedical Semantics*, 5(1), 25.
<https://doi.org/10.1186/2041-1480-5-25>



... mais des enjeux bien connus des bibliothèques

RENDRE IDENTIFIABLE

RENDRE DES RELATIONS VISIBLES

PRÉSERVER, DONNER ACCÈS À LONG TERME

2. Les solutions déployées dans HAL, basées sur SWH

2.1 Améliorer l'identification des logiciels académiques, sans réinventer la roue

Décrire et référencer sont 2 besoins distincts

DÉCRIRE

Biographie	
Naissance	29 mai 1830  Vroncourt-la-Côte 
Décès	9 janvier 1905  (à 74 ans) Marseille 
Sépulture	Cimetière de Levallois-Perret (depuis le 22 janvier 1905) 
Nom de naissance	Louise Michel 
Pseudonymes	Enjolras, La Vierge Rouge 
Nationalité	française 
Activités	Enseignante, communarde,

RÉFÉRENCER



Photo de [Immo Wegmann](#) sur [Unsplash](#)

```
// state.js (for the game, see:
// https://book.southyt.ru/thalacronomata/state.html#state)
function state(self, data) {
  let world = data.world;

  // let the screen dimensions so we can initialize the camera and
  // place our sprites correctly later, we'll store this since we'll
  // pass the world mutably to the following functions.
  let dimensions = world.read_resource('screenDimensions').clone();

  // Place the camera
  init_camera(world, dimensions);

  // load our sprites and display them
  let sprites = load_sprites(world);
  init_sprites(world, sprites, dimensions);
}

function render() {
  let self;
  self.data = state.data;
  event: StateEvent,
  } -> SimpleFrame {
    if let StateEvent(render) = event {
      // Check if the window should be closed
      if let CloseRequested(event) || let KeyDown(event, KeyCode.Escape) {
        return Trans::Quit;
      }
    }

    // Update the map key events
    let key_events = get_key_events();
  }
```

Différents niveaux de granularité



S'appuyer sur les forces de l'existant pour répondre aux besoins

Format du dépôt	Fichier
Type de document	Logiciel
Titre	en Parmap
Résumé	en Parmap is a minimalistic library allowing to exploit happens to be a map, fold or map/fold (map-reduce) subprocesses to use by the optional parameter "re"
Auteur(s)	Roberto Di Cosmo ^{1,2} , Marco Danelutto ³ ¹ DSD-1, Direction générale déléguée à l'innovation (D • Institut National de Recherche en Informatique et en ² - Software Heritage (498515) - France ³ - University of Pisa - Università di P
Code Repository	https://github.com/rdicosmo/parmap
Date de production/écriture	2011-07-18
Version	1.2.4
Etat du développement	active
Licences	• GNU Library General Public License v2 only
Platform/OS	• Linux • MacOS
Domaine	• Informatique [cs]/Langage de programmation [cs]
Langage de programmation	• OCaml
Classification ACM 2012	• acm2012_ACM2012/Computing methodologies/Par
Publications associées	• 10.1016/j.procs.2012.04.202 
Mots-clés	en parallel computing, functional programming, C
SWHID	swh:1:dir:ec88e5b901c034d5a91aa133e824d65cff3788a3;origin =https://github.com/rdicosmo/parmap;visit=swh:1:snp:25490d451af2414 b2a08ece0df643dfdf2800084;anchor =swh:1:rev:db44dc9cf7a6af7b56d8eb da8c75be3375c89282

DÉCRIRE

CodeMeta

BibTeX

XML-TEI

Dublin Core

DC Terms

EndNote

DataCite

Identifiants

HAL Id : **hal-03516539** , version 1

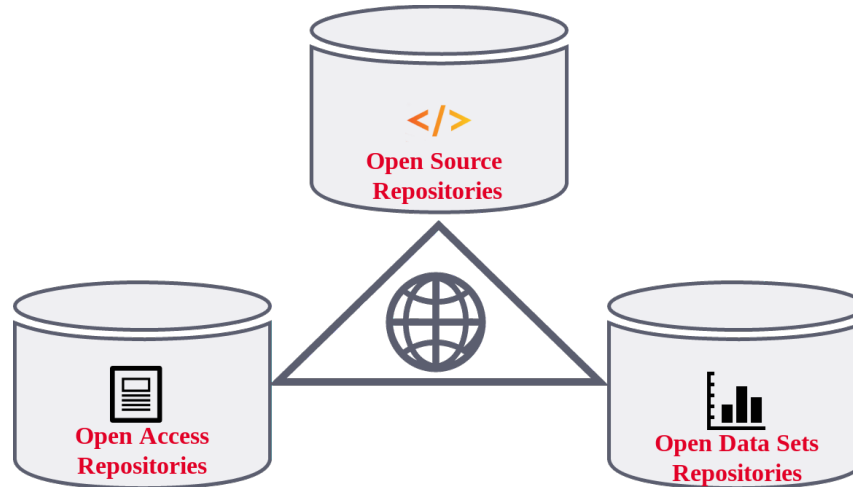
SWHID : **swh:1:dir:ec88e5b901c034
d5a91aa133e824d65cff3788a3;origin
=https://github.com/rdicosmo/parmap;
visit=swh:1:snp:25490d451af2414
b2a08ece0df643dfdf2800084;anchor
=swh:1:rev:db44dc9cf7a6af7b56d8eb
da8c75be3375c89282**

RÉFÉRENCER



2.2 Rendre visibles les relations entre productions académiques

Le logiciel est inséparable des données et des publications, mais rendre visibles ces liens ne va pas (encore) de soi



*Three pillars of Open Science
Software Heritage CC-BY 4.0 2019*

Associer les publications et les logiciels dans HAL

Parmap

Roberto Di Cosmo

Fonction : **Auteur**
PersonId : **21808**
IdHAL : **roberto-di-cosmo**
ORCID : **0000-0002-7493-5349**
IdRef : **057604916**

Marco Danelutto

Fonction : **Auteur**
PersonId : **1140217**
 **University of Pisa - Università di Pisa**

Notice descriptive d'un logiciel archivé dans HAL

Ressources associées

Relation

Cite (Cites) ▼

Entrepôt

Software Heritage ▼

Identifiant SWHID (commençant par swh:)

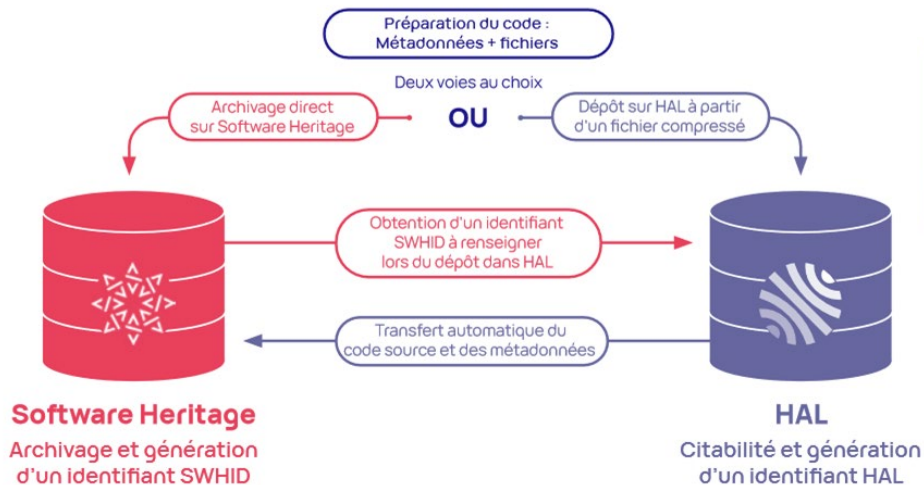
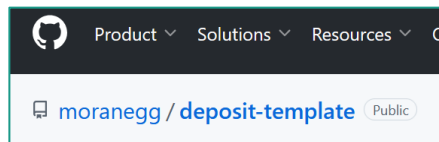
Commentaire

Annuler

Ajouter

2.3 Préserver les logiciels de recherche, en s'adaptant aux pratiques des développeuses et développeurs

2 voies de dépôts pour une collecte plus exhaustive



D'après *Le passeport de la science ouverte : codes et logiciels*, 2022

Zoom arrière : quelques définitions



Quelques éléments de définition

“Logiciel de recherche” = ... De quoi parle-t-on ?

- Acteurs
- Objectifs
- Formes

Code source et logiciel : quelle articulation, quelles différences ?

3. Au-delà des aspects documentaires : accompagner et convaincre

De nombreuses actions concrètes à mener ...

Incentives (Rewarding Open and Reproducible Practices)	Citation standards	Helping create citation standards for data, code, research materials, etc.
	Education	Teaching faculty, researchers, and students about how different citation metrics work and the costs and benefits of each, as well as the longevity of scholar identity (e.g., ORCID)
	Citation data provider to Tenure and Promotion (T&P) Committees	Providing citation data for data, code, software, and materials to tenure and promotion committees and advocating for changes to academic incentives.

Sayre, F., & Riegelman, A. (2019). Replicable Services for Reproducible Research: A Model for Academic Libraries. College & Research Libraries, 80(2), Article 2. <https://doi.org/10.5860/crl.80.2.260>

... mais ce qui semble évident dans le cas des articles ne l'est pas (encore) pour les logiciels

“👁️ The authors had cited an OGSA-DAI paper that should have meant they were using a version of the software between OGSA-DAI 1 and 6.
😞 Later in their paper, the authors mentioned a component that was specific to OGSA-DAI versions 2.5 to 6.
😞 However, the authors then talked of another component and a toolkit, which was only available with a completely different version of the software.

Without my highly detailed knowledge of the OGSA-DAI project, it would have been impossible to determine what software was used.”

Jackson, M. (s.d.). *How to cite and describe software*. Software Sustainability Institute. <https://www.software.ac.uk/how-cite-software>

Le cas de la **citation** :

Des pratiques très hétérogènes

Un travail de prise de conscience à mener

Conclusion : se mettre en ordre de marche

“Make it possible” ... sans réinventer la roue

Rôle clé de l'accompagnement
Une évolution des compétences des bibliothécaires ?
Ce que nous (CCSD, SWH) faisons



Strategy for Culture Change, B. Nosek

Merci pour votre attention

S'informer

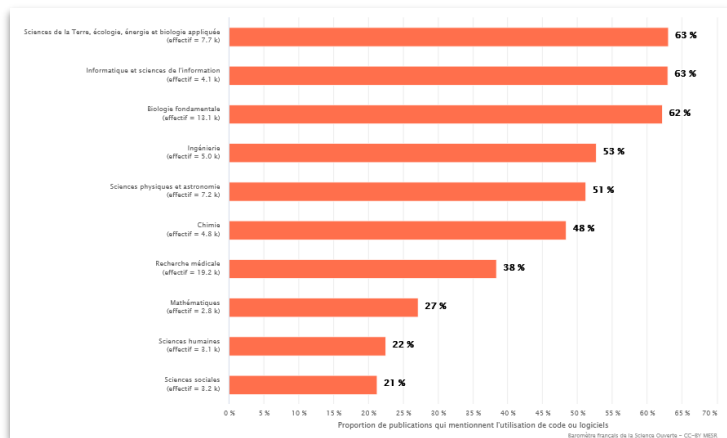
Commonalities Between Major Reproducibility Guidelines



	Transparency and Openness Promotion (TOP)	American Statistical Association (ASA)	National Science Foundation (NSF)	National Institute of Health (NIH)
Data Transparency (Sharing)	Yes	Yes (+ assessment of DMPs by funders)	Yes	Yes
Software Transparency (Sharing)	Yes	Yes	Yes	Yes
Materials Transparency (Sharing)	Yes			Yes
Methods Transparency	Yes		Yes	Yes
Preregistration	Yes (study + analysis plans)			
Supporting Replications	Yes	Yes (directly fund)	Yes (+ publishing negative results)	Yes
Best Practices for Methodology and Analysis		Computational methods / Scripting analysis		

Sayre, F., & Riegelman, A. (2019). Replicable Services for Reproducible Research: A Model for Academic Libraries. College & Research Libraries, 80(2), Article 2. <https://doi.org/10.5860/crl.80.2.260>

De nombreuses disciplines sont concernées au-delà de l'informatique : cf. Baromètre français de la science ouverte



Proportion de publications françaises mentionnant l'utilisation de code ou de logiciels, par discipline (données 2023 issues de <https://barometredelascienceouverte.esr.gouv.fr/>)

ekdosis

Robert Alessi (1, 2)

Afficher plus de détails

1 OM - ORIENT ET MÉDITERRANÉE : Textes, Archéologie, Histoire

2 OM-MG - Équipe Médecine grecque et littérature technique

Résumé

en

Mots clés

en

LaTeX

LuaLaTeX

TEI xml

Critical edition

Domaines

Etudes classiques

Liste complète des métadonnées

ekdosis is a LuaLaTeX package designed for multilingual critical editions. It can be used to typeset texts and different layers of critical notes in any direction accepted by LuaTeX. Texts can be arranged in running paragraphs or on facing pages, in any number of columns which in turn can be synchronized or not. In addition to printed texts, ekdosis can convert .tex source files so as to produce TEI xml-compliant critical editions. Database-driven encoding under LaTeX then allows extraction of texts entered segment by segment according to various criteria: main edited text, variant readings, translations or annotated borrowings between texts. It is published under the terms of the OpenBSD license.

Un exemple de logiciel en LSH: ekdosis (R. Alessi)

24

La collaboration CCSD-Software Heritage (SWH)

🔗 Doter les bibliothèques universitaires et leurs équipes d'outils et connaissances nécessaires pour participer à la préservation des logiciels de recherche.

Un programme d'accompagnement:

- **Des formations** pour maîtriser les enjeux de la préservation et de la mise à disposition des logiciels
- **Des ressources clés en main** pour sensibiliser vos communautés
- **Des opportunités pour échanger autour de retours d'expérience** d'intégration d'un volet "Logiciels" dans des stratégies Science Ouverte

Les bénéfices pour les bibliothèques:

- **Pérenniser les résultats de la recherche** en garantissant l'accès durable aux logiciels
- **Renforcer la reproductibilité** des travaux scientifiques
- **Rendre plus visibles** les productions de votre institution
- **Favoriser l'innovation** en facilitant la réutilisation des codes
- **Positionner votre bibliothèque** comme un acteur majeur de la Science Ouverte

Métadonnées pour le logiciel : recommandations



Gruenpeter, M., Granger, S., Monteil, A., Chue Hong, N., Breitmoser, E., Antonioletti, M., Garijo, D., González Guardia, E., Gonzalez Beltran, A., Goble, C., Soiland-Reyes, S., Juty, N., & Mejias, G. (2023). D4.4—**Guidelines for recommended metadata standard for research software within EOSC.**

<https://doi.org/10.5281/ZENODO.8199104>

“Following the first introduction and state of the art, T4.3 introduces the RSMD guidelines proposal **to collect and curate research software metadata**. These guidelines are directly addressed to end users, including software creators and curators in their quest to improve the FAIRness of their software.”



Focus sur le SoftWare Hash Identifier (SWHID)

- Un identifiant intrinsèque : pas d'abonnement à prévoir pour votre institution
- https://www.softwareheritage.org/faq/#3_Referencing_and_identification
- [https://www.swhid.org/faq/Un_standard_ISO_\(18670:2025\)](https://www.swhid.org/faq/Un_standard_ISO_(18670:2025))
- <https://www.swhid.org/specification/v1.2/>
- <https://www.softwareheritage.org/software-hash-identifier-swhid/>



Quelles sont les solutions apportées par Software Heritage ?

- Une infrastructure portée par l'Inria et l'Unesco
- Une équipe dédiée, pilotée par des chercheurs en informatique
- [Un point d'accès unique au code source de millions de logiciels](#)
- Une [fourniture gratuite d'identifiants pérennes](#)
- [Un service d'archivage transparent pour les auteurs de logiciels](#)
- Le [service "Save code now"](#) pour archiver un logiciel
- Le [service "Add forge now"](#) pour archiver une forge institutionnelle
- Un [gisement de métadonnées](#) constamment enrichi
- Des ressources documentaires couvrant différents besoins : [documentation technique](#) pour construire une offre de services personnalisée (API, etc.) ; [tutoriels](#) et [supports de formation](#) ; [billets de blogs](#) et [newsletter](#)
- Un [réseau d'ambassadeurs](#) pouvant assurer présentations et formations dans votre institution

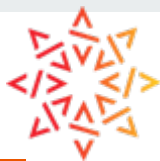


Rejoindre l'Interest Group ALIG (Archives and Libraries Interest Group) pour bénéficier d'un niveau de services supérieur

- Bénéficier d'un **soutien personnalisé** pour vos projets : formations, guichet de renseignement
- Rejoindre une communauté internationale pour **partager les bonnes pratiques**, les expériences
- Rendre visible et **valoriser l'implication de votre service** : bénéficier des canaux de diffusion Software Heritage pour promouvoir vos actions
- <https://www.softwareheritage.org/support/members-alig/>

Ils ont rejoint l'Interest Group:

Universität Bern, ETH Zürich, Universität Luzern, Université de Neuchâtel, CERN, Memorial University of Newfoundland, Université de Montréal, University of Toronto, York University, University of Saskatchewan, University of New Brunswick, Western University.



COAR Notify : signaler et échanger des relations entre infrastructures



<https://coar-notify.net/catalogue/workflows/repository-relationship-repository/>

<https://docs.softwareheritage.org/user/coarnotify/explanations.html>



Quelles sont les solutions apportées par le CCSD ?

- Une infrastructure portée par le MESR, CNRS, Inria et Inrae (HAL+ inscrite dans la feuille de route des IR du MESR en 2021)
- Une [gouvernance partagée](#) avec les utilisateurs de la plateforme
- Une [équipe](#) dédiée, pluridisciplinaire
- Un service d'accompagnement couvrant différents besoins : [formation](#), [documentation](#), [assistance utilisateurs](#), [accompagnement à l'ouverture de portails HAL](#), service continu de validation technique des dépôts
- Des outils pour [rester informé](#) et pour mieux utiliser la plateforme HAL : newsletter actuHAL, boîte à outils [usuHAL](#), fiches pratiques les [Essentiels de HAL](#), [tutoriels](#), listes de diffusion
- Un réseau d'[ambassadeurs et ambassadrices de HAL](#) pouvant assurer présentations dans votre institution

Suivre l'actualité du CCSD, HAL et Software Heritage



linkedin.com/company/ccsd-science-ouverte



hal_fr@social.numerique.gouv.fr



x.com/ccsd_fr



ccsd.cnrs.fr/inscription-actuhal



ccsd.cnrs.fr/blog

linkedin.com/company/swheritage

mstdn.social/@swheritage

softwareheritage.org/newsletter

softwareheritage.org/blog





Vos contacts



Hélène Bégnis | Animatrice des réseaux d'utilisateurs de HAL :
helene.begnis@ccsd.cnrs.fr

Sébastien Mazzaresse | Coordinateur des portails HAL et Responsable de
l'équipe d'assistance aux utilisateurs : sebastien.mazzaresse@ccsd.cnrs.fr



Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Sabrina Granger | Software Heritage Open science community manager :
sabrina.granger@inria.fr