

La création du nouveau type de dépôt scientifique

Le logiciel

CCSD¹, HAL-Inria², Software Heritage³

Y. Barborini¹, R. Di Cosmo³, A.R. Dumont³, M. Gruenpeter³, B. Marmol¹, A. Monteil², J. Sadowska², S. Zacchiroli³



Préserver les logiciels: un enjeu scientifique

Les logiciels sont devenus le support indissociable des **connaissances techniques et scientifiques** de l'humanité tout entière. Ils constituent un véritable socle pour les sciences. La préservation des logiciels est à la base de **la reproductibilité**. L'utilisation des logiciels intervient à toutes les étapes de la recherche, dans tous les domaines scientifiques, et se révèle essentielle à de multiples égards.

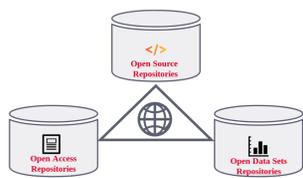


Figure 1: Les piliers de la préservation de la connaissance

Dans la quête pour la reproductibilité des résultats scientifiques et la transmission des connaissances scientifiques aux générations futures, les trois piliers principaux sont: des articles scientifiques, décrivant les résultats, les ensembles de données et les logiciels utilisés ou produits[1].

Le dépôt logiciel

La collaboration entre **Software Heritage (SWH)**, **Hal-Inria** et le **CCSD** a permis l'ouverture de Hal vers le nouveau type de données scientifiques: *le logiciel*.

Les chercheurs ont désormais la possibilité de déposer le code source via Hal-Inria tout en contribuant au patrimoine logiciel constitué par SWH.

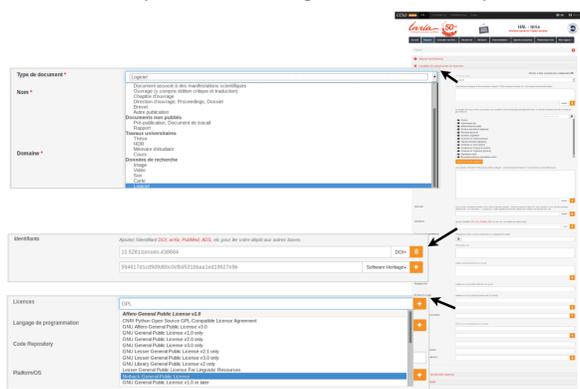


Figure 2: Le formulaire dédié au dépôt de logiciel

Les étapes pour le dépôt de logiciel:

- déposer une archive du code source (.zip)
- choisir le type de dépôt : *Logiciel*
- ajouter les métadonnées associées
- ajouter les auteurs du logiciel
- valider l'archivage du logiciel sur SWH

Les acteurs

Software Heritage a relevé le défi de collecter, préserver et partager les logiciels sous leur forme fondamentale : leur code source. **Hal-Inria** est l'archive ouverte d'Inria- Institut national de recherche dédié au numérique. Elle permet, depuis 2005, de déposer des publications et d'accéder à l'ensemble du portail Hal développé par le **CCSD - Centre pour la Communication Scientifique Directe** dont la mission principale est de fournir, dans l'esprit du libre accès, des outils pour l'archivage, la diffusion et la valorisation des publications et des données scientifiques.

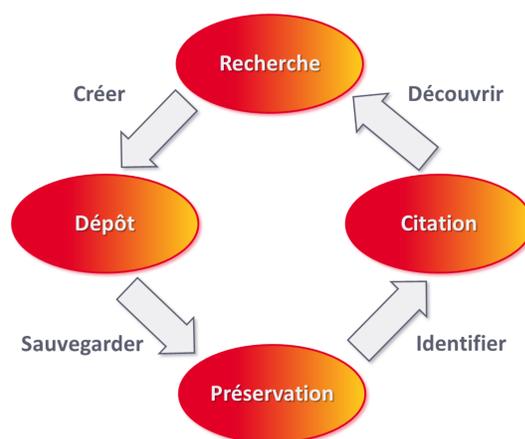


Figure 3: Le cycle de vie du logiciel de recherche

Les métadonnées

Des métadonnées pour bien décrire le logiciel sont disponibles sur le formulaire de dépôt et sont préservées avec le logiciel dans l'archive SWH.

- | | |
|--------------------------------|---------------------|
| Fournie par le système: | Obligatoire: |
| - identifiant Hal | - titre |
| - date de publication | - description |
| - SWH-ID | - auteurs |
| Recommandée: | Possible: |
| - licence | - dépendances |
| - mots-clés | - platform/OS |
| - repository | - financement |

L'identifiant intrinsèque et pérenne

Software Heritage fournit l'identifiant **SWH-ID**, intrinsèquement lié aux composants logiciels, garantissant une traçabilité pérenne à travers les futures évolutions du développement logiciel et de l'organisation. Le **SWH-ID** comme une empreinte digitale du logiciel est spécifique, pérenne et unique et ne dépend pas d'un résolveur.

Pour reproduire une expérience, il est indispensable de connaître avec exactitude la version du logiciel.

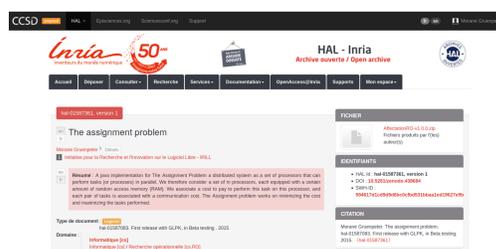


Figure 4: Le dépôt final sur Hal-Inria

L'archivage du dépôt

Une fois le dépôt validé il est poussé grâce au protocole **SWORD** sur Software Heritage. SWH procède à l'injection du logiciel dans la Bibliothèque d'Alexandrie du code source et génère l'identifiant intrinsèque - le **SWH-ID**. Hal revient récupérer le **SWH-ID** qui est mis à disposition sur la page du dépôt et dans le format de citation.

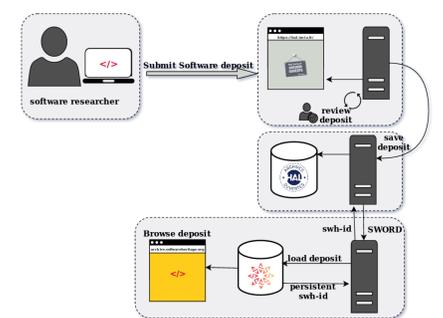


Figure 5: Le processus de dépôt et d'archivage

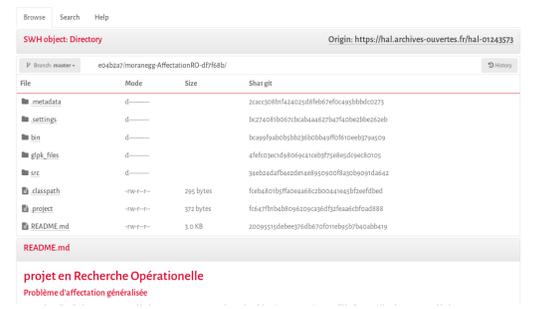


Figure 6: Feuilletter le dépôt sur Software Heritage

La citation du logiciel

Suivant les principes de citation de logiciel [2] et en considérant ainsi que le logiciel est **un produit légitime et citable de recherche**, le format de citation proposé sur Hal contient une partie des métadonnées obligatoires soumises avec le logiciel.

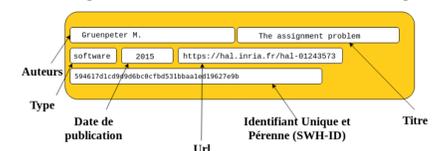


Figure 7: Le format de citation d'un logiciel[3]

La citation est indispensable pour promouvoir le rôle du logiciel dans la recherche et s'assurer que les développeurs de logiciels de recherche voient leurs contributions reconnues et récompensées[4].

Références

1. Roberto Di Cosmo, Stefano Zacchiroli (2017) Software Heritage: Why and How to Preserve Software Source Code. iPRES 2017. <https://hal.archives-ouvertes.fr/hal-01590958>
2. Smith et al. (2016), Software citation principles. PeerJ Comput. Sci. 2:e86; DOI 10.7717/peerj-cs.862.
3. Yolanda Gil (2015) Documenting Software through Metadata. Geosoft.
4. Mike Jackson (2014) How to cite and describe software. The Software Sustainability Institute <https://www.software.ac.uk/how-cite-and-describe-software>



Contacts :
 hal.inria.fr
 josephine.sadowska@inria.fr
 www.softwareheritage.org
 morane@softwareheritage.org