

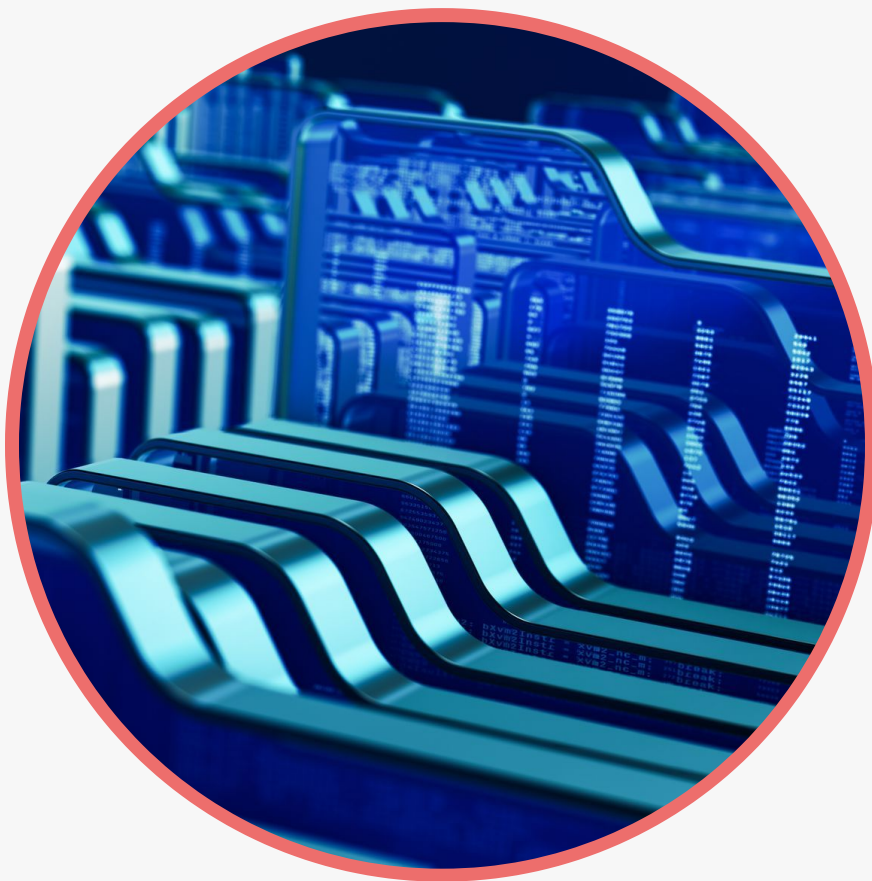


ARCHIVER

En amont de chaque projet, il convient
d'anticiper les questions d'archivage



COMMENCER





SOMMAIRE

Cliquez pour accéder directement aux parties.

L'archivage pérenne : qu'est-ce que
c'est ?



Pourquoi archiver ?



Pourquoi archiver aussi dans HAL INRAE ?



Comment archiver : quels outils, quelles
méthodes et comment préparer ?





Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

Outils, méthodes,
préparation

QU'EST-CE QUE L'ARCHIVAGE PÉRENNE ?



L'archivage numérique pérenne consiste à conserver les fichiers de données, les logiciels et leurs métadonnées :

Dans leur aspect **physique et intellectuel**

Sur le **moyen et long terme**
(+ de 30 ans)

De manière à ce qu'ils soient en
permanence **accessibles et
compréhensibles**



Mais que garantit le dépôt dans une archive
ouverte institutionnelle ?



Où sont archivés les logiciels et codes
sources produits par la recherche française ?





Le dépôt dans une archive ouverte institutionnelle garantit **la visibilité d'un document et son accès pérenne pour l'ensemble des acteurs et actrices de la société.**

L'enjeu de l'archivage pérenne n'est pas seulement le stockage et la préservation des données mais également **l'assurance que ces données seront toujours disponibles et consultables quelle que soit la période écoulée, avec les évolutions techniques que cela implique.**



Les logiciels et codes sources produits par la recherche française sont archivés dans **Software Heritage** et **dans HAL**, qui proposent des **fonctionnalités complémentaires**.





Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

Outils, méthodes,
préparation

POURQUOI ARCHIVER UN LOGICIEL ?

L'archivage d'un logiciel est essentiel car il garantit **la reproductibilité des résultats** : sans accès au code source, il est impossible de vérifier ou de reproduire les analyses, même avec les données et le protocole en main.

L'archivage assure également la pérennité des objets numériques. Les liens vers des logiciels hébergés sur des sites personnels, des forges, ou des serveurs institutionnels peuvent disparaître ou devenir obsolètes. Une infrastructure dédiée à l'archivage du « patrimoine logiciel mondial de l'humanité », comme [Software Heritage](#), protège contre ces risques en offrant un **accès durable**.

De plus, le code source contient bien plus que des instructions pour la machine : il inclut des commentaires, des explications et l'historique des modifications, qui sont cruciaux pour comprendre la logique du logiciel et son évolution. Ces informations, absentes dans la version exécutable, sont préservées grâce à l'archivage.

L'archivage facilite aussi la citation et la reconnaissance des logiciels dans les travaux scientifiques. Des **identifiants pérennes, comme les SWHID**, permettent de référencer précisément une version ou un composant spécifique, ce qui est souvent impossible avec un simple lien ou un article associé.

Enfin, archiver un logiciel dans Software Heritage favorise **la réutilisation et l'innovation**. Un code bien préservé peut être réutilisé, adapté ou amélioré par d'autres chercheurs et chercheuses, contribuant ainsi à l'avancée collective des connaissances.





Source : Sabrina Granger, Baptiste Mèlès, et Frédéric Santos, «Préserver et rendre identifiables les logiciels de recherche avec Software Heritage», Programming Historian en français 6 (2024), <https://doi.org/10.46430/phfro034>.



Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

Outils, méthodes,
préparation

POURQUOI ARCHIVER DANS HAL INRAE ?

Déposer un logiciel dans l'archive ouverte HAL INRAE participe à la mise en

œuvre des principes FAIR et permet notamment la production d'une notice en

vue de sa citabilité. Cela **garantit un traitement qualité des métadonnées**

facilitant la visibilité et l'accessibilité (Findable) des logiciels et codes développés
par les personnels de recherche et d'appui. Ces métadonnées sont indexées et
peuvent faire l'objet d'une recherche, l'accès est libre et gratuit. Elles permettent

également la **citation du document**. De plus, **HAL est la source principale**

d'alimentation du [catalogue national](#) pour le référencement de logiciels libres
développés dans la recherche académique.

Par ailleurs, tout logiciel déposé dans **HAL INRAE** reçoit un **identifiant unique et**
pérenne. Le formulaire de dépôt permet également de renseigner d'autres

identifiants du logiciel s'il a été archivé sur d'autres plateformes. Des liens
accessibles dans la notice permettent de naviguer vers ces différents

emplacements, notamment vers l'archive **Software Heritage**.

Depuis 2018, **HAL est interfacé avec Software Heritage**
pour garantir un archivage optimal des logiciels et codes
sources. Certaines forges sont moissonnées
automatiquement par Software Heritage. Le logiciel ainsi
archivé reçoit un identifiant (SWHID) qui permet d'effectuer
un dépôt dans HAL INRAE et garantit également un lien
entre les deux archives.

Exemple de forges

Choix de la licence

Elle doit être clairement
mentionnée dans les fichiers
déposés afin de valider le
champ dédié dans HAL qui
est obligatoire pour ce type
de dépôt.



**Valoriser le lien entre votre dépôt HAL et un jeu
de données ou un logiciel**



La structure de HAL est conçue pour **reproduire le triptyque publications-données-logiciels** et établit le lien naturel réunissant ces trois éléments.

Le formulaire de dépôt dans HAL comporte un champ *Ressources associées* permettant de **renseigner les identifiants des jeux de données et des publications (article, software paper, etc.) en rapport avec le logiciel déposé**, tout en précisant la relation entre ces différents éléments, qui apparaîtront sous forme de liens sur la notice. Si la ressource se trouve dans un entrepôt associé (recherche.data.gouv et Nakala), ces informations, voire la ressource elle-même, seront visibles dans le dépôt HAL.

[Documentation complète du CCSD pour lier vos ressources dans HAL INRAE](#)



Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

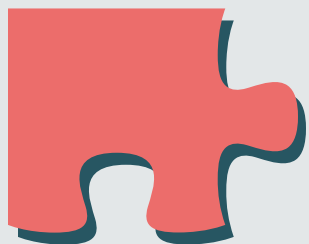
Outils, méthodes,
préparation



COMMENT ARCHIVER ?

Cliquez sur les différentes cartes pour en savoir plus sur les outils et les méthodes d'archivage

Hal et Software Heritage



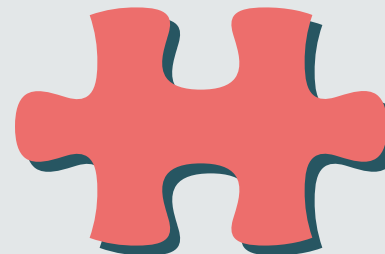
Deux archives ouvertes
au service de la
communauté scientifique

Préparer les informations



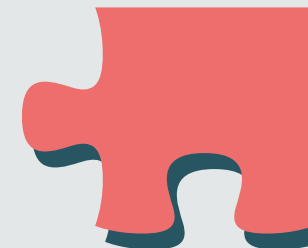
Préparer les informations
nécessaires pour
l'archivage du logiciel ou
du code source

Le dépôt



Le dépôt du logiciel dans
HAL et dans Software
Heritage

Dépôt source ou dépôt SWHID ?



Deux méthodes
complémentaires



Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

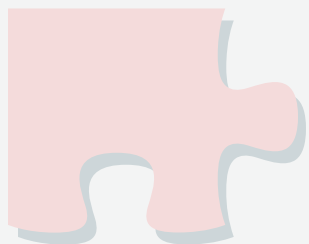
Outils, méthodes,
préparation



COMMENT ARCHIVER ?

Cliquez sur les différentes cartes pour en savoir plus sur les outils et les méthodes d'archivage.

Hal et Software
Heritage



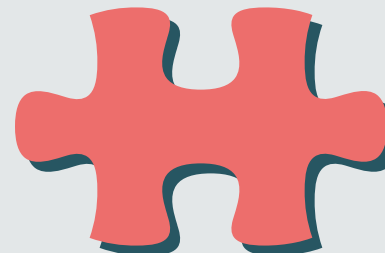
Deux archives ouvertes
au service de la
communauté scientifique

Préparer les
informations



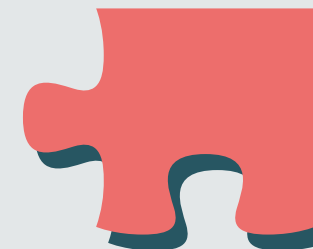
Préparer les informations
nécessaires pour
l'archivage du logiciel ou
du code source

Le dépôt



Le dépôt du logiciel dans
HAL et dans Software
Heritage

Dépôt source ou dépôt
SWHID ?



Deux méthodes
complémentaires



Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

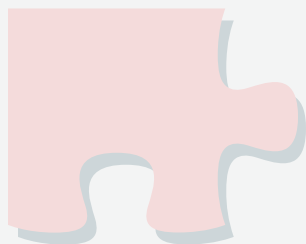
Outils, méthodes,
préparation



COMMENT ARCHIVER ?

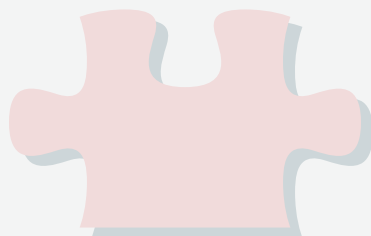
Cliquez sur les différentes cartes pour en savoir plus sur les outils et les méthodes d'archivage.

Hal et Software Heritage



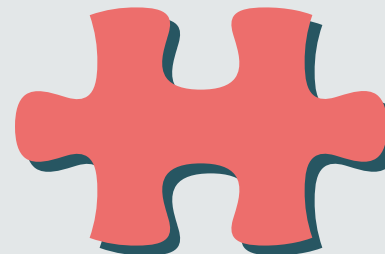
Deux archives ouvertes
au service de la
communauté scientifique

Préparer les informations



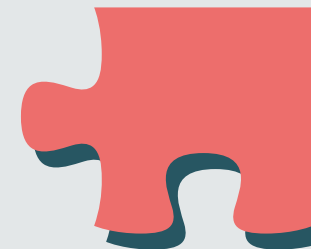
Préparer les informations
nécessaires pour
l'archivage du logiciel ou
du code source

Le dépôt



Le dépôt du logiciel dans
HAL et dans Software
Heritage

Dépôt source ou dépôt SWHID ?



Deux méthodes
complémentaires



Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

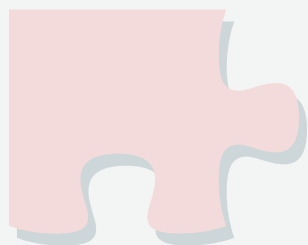
Outils, méthodes,
préparation



COMMENT ARCHIVER ?

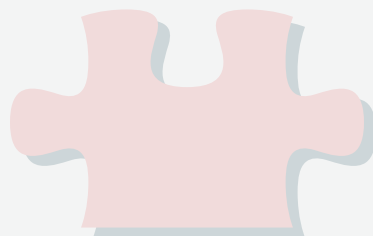
Cliquez sur les différentes cartes pour en savoir plus sur les outils et les méthodes d'archivage.

Hal et Software Heritage



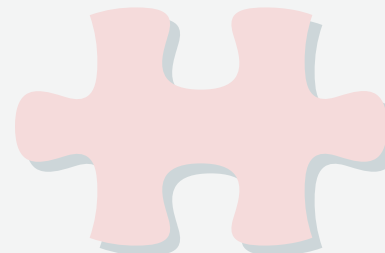
Deux archives ouvertes
au service de la
communauté scientifique

Préparer les informations



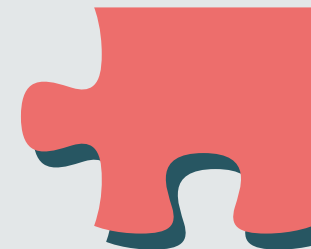
Préparer les informations
nécessaires pour
l'archivage du logiciel ou
du code source

Le dépôt



Le dépôt du logiciel dans
HAL et dans Software
Heritage

Dépôt source ou dépôt SWHID ?



Deux méthodes
complémentaires



Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

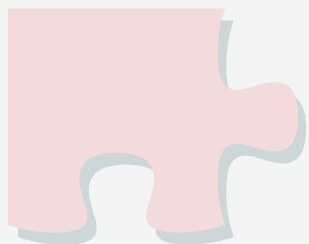
Outils, méthodes,
préparation



COMMENT ARCHIVER ?

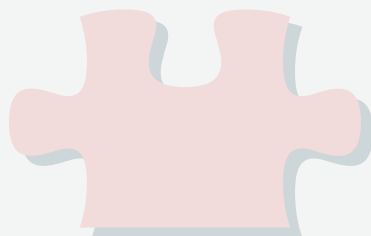
Cliquez sur les différentes cartes pour en savoir plus sur les outils et les méthodes d'archivage.

Hal et Software Heritage



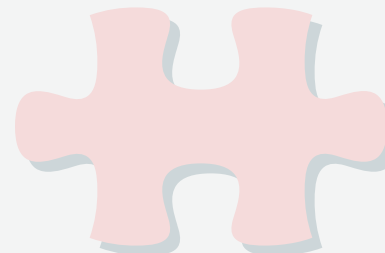
Deux archives ouvertes
au service de la
communauté scientifique

Préparer les informations



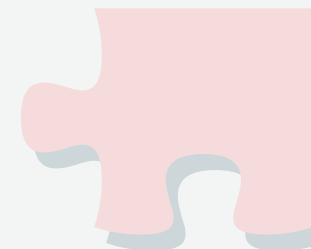
Préparer les informations
nécessaires pour
l'archivage du logiciel ou
du code source

Le dépôt



Le dépôt du logiciel dans
HAL et dans Software
Heritage

Dépôt source ou dépôt SWHID ?



Deux méthodes
complémentaires



Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

Outils, méthodes,
préparation

HAL ET SOFTWARE HERITAGE

Deux archives ouvertes complémentaires. Cliquez sur chaque archive pour en savoir plus.



Software Heritage



Software Heritage (SWH)

[Software Heritage](#) (SWH) est une **archive ouverte** dont la mission est de **collecter, préserver et partager les codes sources de tous les logiciels (c'est à dire la partie du logiciel compréhensible par l'humain) publiquement disponibles.**

Créé en 2015 sous l'impulsion de l'Inria, SWH est une organisation à but non lucratif, inscrite dans le Plan national pour la science ouverte et soutenue par l'UNESCO et le Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI).



HAL et HAL INRAE



[HAL](#) est l'archive ouverte multidisciplinaire choisie par l'ensemble de la communauté scientifique et universitaire française pour la diffusion des savoirs.

Elle est à la fois une application, un entrepôt unique et une **plateforme mutualisée pour les archives institutionnelles** telle que [HAL INRAE](#), les archives ouvertes thématiques et les thèses.

La plateforme garantit l'accès et la préservation à long terme des documents et propose un ensemble de services qui contribuent à leur valorisation .



Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

Outils, méthodes,
préparation

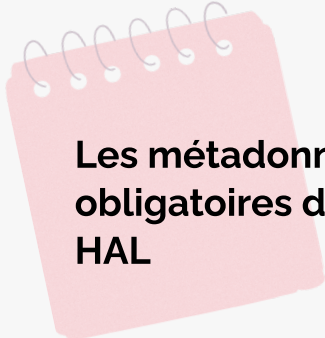
PRÉPARER LES INFORMATIONS NÉCESSAIRES

Pour l'archivage du logiciel ou du code source dans HAL INRAE



L'archivage logiciel implique la présence, conjointement au code et à la documentation, de **métadonnées décrivant l'objet déposé**. Cette assertion est vraie que l'on dépose dans HAL ou dans SWH. **Lorsque le dépôt est effectué dans HAL INRAE**, la mise en ligne passe par **une étape de modération** visant à vérifier que les éléments attendus figurent bien dans les fichiers. Ces éléments doivent correspondre aux métadonnées saisies dans le formulaire de dépôt. Le code lui-même relève de la responsabilité de ses auteurs ou autrices et ne fait pas l'objet d'une modération. Il est important en revanche de joindre les éléments permettant d'établir la paternité du logiciel ou du code et les conditions de partage, c'est à dire les auteurs ou autrices et une ou plusieurs licences.

La modération et l'archivage du fichier seront facilités si les éléments attendus sont présents dans les fichiers dès le dépôt initial.



Les métadonnées
obligatoires dans
HAL



Les informations
importantes dans les
fichiers



Les fichiers



Consultez la
documentation complète
du CCSD pour archiver
votre dépôt dans HAL
INRAE et Software
Heritage [↗](#)

Les fichiers de format .txt ou .md pour transmettre les informations de mon logiciel

- README (décrit le logiciel déposé)
- AUTHORS (contient la liste des auteurs ou autrices et des contributeurs ou contributrices éventuelles)
- LICENSE (décrit les droits d'utilisations du code source déposé, à choisir avec les titulaires des droits patrimoniaux dudit logiciel)
- Un fichier codemeta.json permet de compléter automatiquement les métadonnées sans avoir à les saisir dans chaque champ (facultatif mais pratique). Le site CodeMeta Generator propose un [formulaire en ligne](#) qui permet de générer ces fichiers.

S'il s'agit d'un dépôt dans HAL INRAE, les fichiers, souvent présents à la racine du projet, doivent être compressés dans un dossier .zip ou .tar.gz avant d'être déposés.



Les métadonnées obligatoires

Pour valider un dépôt de logiciel dans HAL INRAE, vous devez renseigner les métadonnées suivantes :

- Auteurs
- Affiliations
- Nom du logiciel
- Domaine
- Public visé
- Licence(s) du logiciel

Les informations à intégrer aux fichiers



Les fichiers déposés dans HAL INRAE ou importés dans Software Heritage devront donc comporter a minima les informations suivantes :

- Liste exhaustive des auteurs et autrices et au moins une affiliation.
- Nom du logiciel
- Date de production
- Numéro de version le cas échéant
- Mention de la ou les licence(s) choisie(s) en précisant, si possible, la version de la licence



Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

Outils, méthodes,
préparation

L'ARCHIVAGE DU LOGICIEL DANS HAL ET SWH

Cliquez sur chaque type de dépôt pour en savoir plus.



Deux possibilités pour le
dépôt de logiciel



Dépôt au format zip. dans HAL



Archivage dans SWH



L'archivage dans HAL INRAE

Vous déposez **dans HAL INRAE** un **dossier au format zip contenant votre code et les fichiers README, AUTHORS et LICENSE** en ayant préalablement sélectionné le type de document « Logiciel ». Au moment de valider la saisie, les fichiers sont transférés vers SWH où ils seront archivés et accessibles.

Cette opération génère un identifiant (SWHID) qui sera intégré à la notice HAL pour assurer le lien entre les deux archives.

Ce type de dépôt est également appelé **dépôt source**.

Attention le dépôt dans HAL est limité à 200 Mo. Si votre .zip dépasse cette limite, chaque utilisateur dispose d'un espace de 5 Go sur le [serveur FTP du CCSD](#)



Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

Outils, méthodes,
préparation

L'ARCHIVAGE DU LOGICIEL DANS HAL ET SWH

Cliquez sur chaque type de dépôt pour en savoir plus.



Deux possibilités pour le
dépôt de logiciel



Dépôt au format zip. dans HAL



Archivage dans SWH



L'archivage dans HAL INRAE

Vous déposez **dans HAL INRAE** un **dossier au format zip contenant votre code et les fichiers README, AUTHORS et LICENSE** en ayant préalablement sélectionné le type de document « Logiciel ». Au moment de valider la saisie, les fichiers sont transférés vers SWH où ils seront archivés et accessibles.

Cette opération génère un identifiant (SWHID) qui sera intégré à la notice HAL pour assurer le lien entre les deux archives.

Ce type de dépôt est également appelé **dépôt source**.

Attention le dépôt dans HAL est limité à 200 Mo. Si votre .zip dépasse cette limite, chaque utilisateur dispose d'un espace de 5 Go sur le [serveur FTP du CCSD](#)



L'ARCHIVAGE DANS SOFTWARE HERITAGE



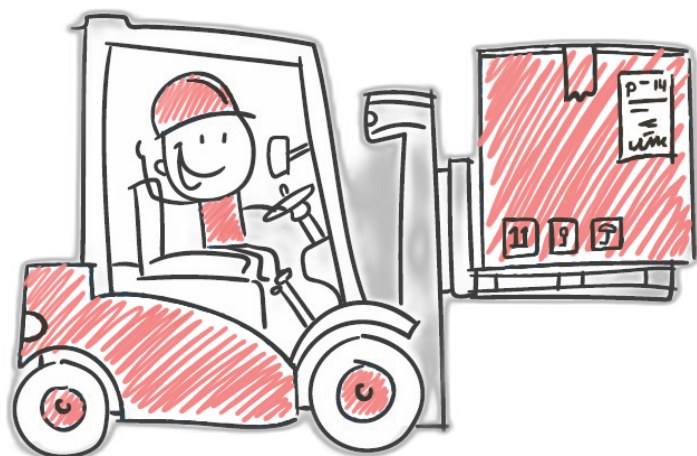
L'archivage dans Software Heritage peut être effectué à partir de la forge dans laquelle le logiciel a été développé, ce qui en simplifie l'usage pour les déposants et déposantes.

Son architecture est adaptée au dépôt de logiciel.

Elle permet une gestion simplifiée du versioning et la conservation de l'historique de développement.



Software Heritage



Le transfert vers HAL n'est pas automatique : une fois les fichiers archivés dans Software Heritage, il est nécessaire de faire un second dépôt dans HAL à l'aide du SWHID généré lors de l'opération .

Ce double dépôt est important car ce sont les métadonnées saisies dans HAL INRAE qui permettront d'avoir une visibilité sur la production scientifique de l'établissement, de faire des recherches et de citer les documents.





Recommandation INRAE

Penser à synchroniser Software Heritage au moment de la release par sécurité. Le moissonnage automatique des forges par SWH ne garantit pas que l'identifiant SWH pointe sur la version relasée si on ne le fait pas de façon explicite.



Archivage
pérenne

Pourquoi
archiver

Pourquoi
archiver
dans HAL

Outils, méthodes,
préparation

DÉPÔT SOURCE OU DÉPÔT SWHID ?



Cliquer sur l'image pour l'agrandir

Afin d'effectuer un archivage optimal du logiciel, **il est recommandé de faire un double archivage dans HAL INRAE et dans SWH**. Les deux archives proposent des fonctionnalités permettant de faciliter le transfert des informations de l'une à l'autre afin de faciliter la démarche de l'utilisateur et l'utilisatrice. Le point d'entrée sera l'une ou l'autre des archives car les deux dépôts seront liés. Consulter [la documentation complète du CCSD](#) sur les deux modalités de dépôt.

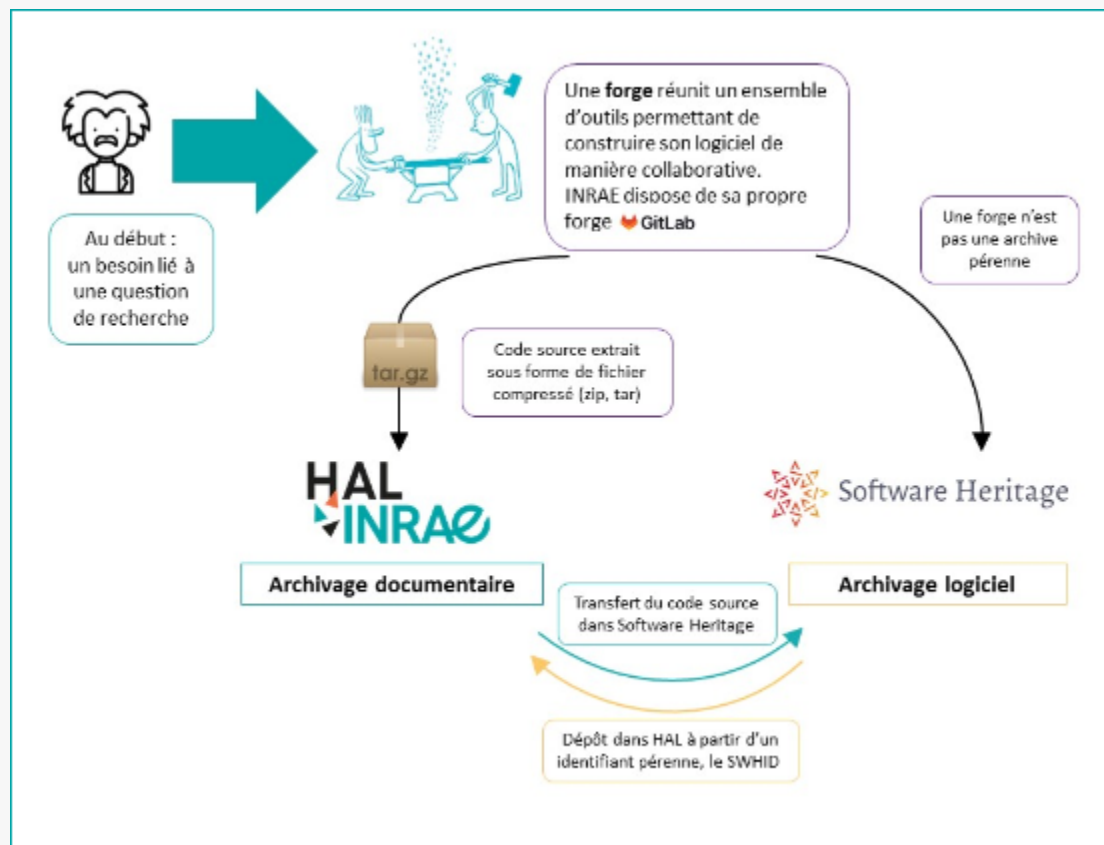


Dépôt source



Dépôt SWHID

**Collaboration entre SWH
et HAL**



La collaboration entre SWH et HAL permet d'établir des liens entre les deux archives afin de combiner les avantages de chacune.



La collaboration entre SWH et HAL permet d'établir des liens entre les deux archives afin de combiner les avantages de chacune :

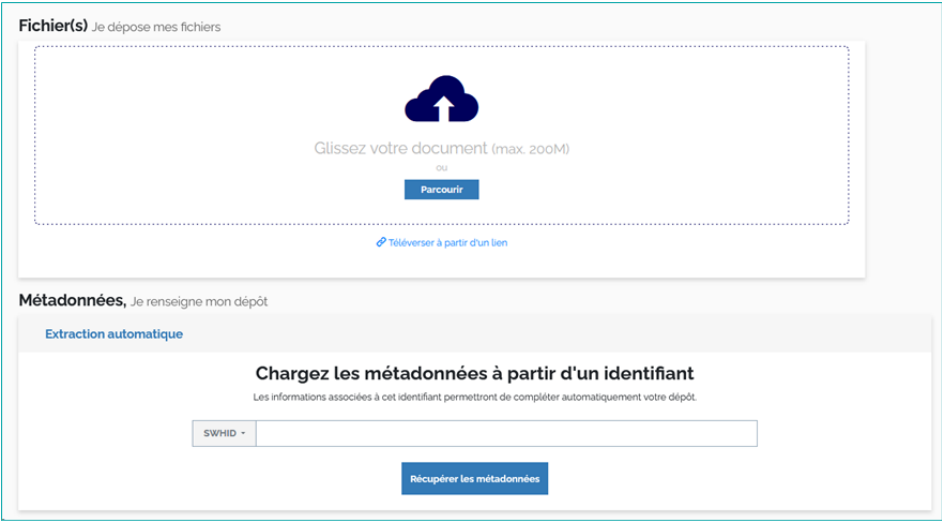
Software Heritage dispose d'une architecture et de fonctionnalités adaptées au dépôt et à l'archivage de logiciels et de codes sources, permettant notamment de conserver l'historique de développement, les branches, les releases, etc.

HAL met l'accent sur la qualité des métadonnées qui vont permettre d'indexer le document, de favoriser sa visibilité et sa valorisation et également de préciser les droits relatifs au logiciel ou au code déposé. De plus, HAL est la source principale d'alimentation du [catalogue national](#) pour le référencement de logiciels libres développés dans la recherche académique.

[Le dépôt dans HAL](#) permet de créer une **description citable du logiciel**, alors que l'archivage dans [Software Heritage](#) est principalement dédié à **rendre identifiable des composants techniques du logiciel**. Multidisciplinaire, HAL permet de partager en libre accès les résultats de recherche, publiés ou non.

Le dépôt SWHID

Si vous avez privilégié l'archivage dans SWH, il est recommandé de signaler également votre logiciel dans HAL INRAE au moyen de l'identifiant attribué par SWH. Ce type de dépôt dans HAL est appelé dépôt SWHID. Le formulaire de dépôt dans HAL est doté d'un champ permettant de renseigner le SWHID et de faire ainsi le lien entre les deux archives. Si vous avez intégré un fichier [codemeta.json](#) lors de l'archivage dans SWH, les métadonnées seront importées dans HAL automatiquement. Une fois le dépôt effectué, un lien dans la notice HAL permet de naviguer d'une archive à l'autre.



The screenshot shows the HAL INRAE deposit interface. The top section, titled 'Fichiers) Je dépose mes fichiers', contains a large dashed box for file upload with a cloud icon and the text 'Glissez votre document (max. 200M)'. Below this is a 'Parcourir' button and a link 'Téléverser à partir d'un lien'. The bottom section, titled 'Métadonnées, Je renseigne mon dépôt', has a sub-header 'Extraction automatique' and a main heading 'Chargez les métadonnées à partir d'un identifiant'. A subtext explains that associated information will automatically complete the deposit. Below this is a text input field with 'SWHID' in a dropdown menu and a 'Récupérer les métadonnées' button.

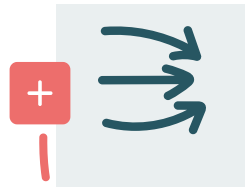
L'import automatique de données dans HAL INRAE

Comment et pourquoi déposer son code dans HAL ? 🔗

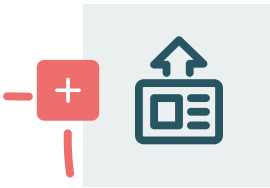
Préserver et rendre identifiable les logiciels de recherche avec Software Heritage 🔗

LE DÉPÔT SOURCE

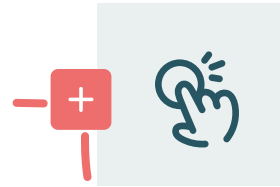
Cliquez sur chaque étape pour en savoir plus.



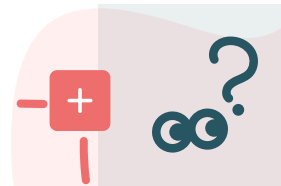
Rassembler les fichiers



Remplir le formulaire de dépôt



Validation du dépôt



La modération



L'archivage



REEMPLIR LE FORMULAIRE

Le dépôt dans HAL INRAE ou « dépôt source »
s'effectue au moyen d'un formulaire de dépôt
accessible à partir du portail

[HAL INRAE](#)



RASSEMBLER LES FICHIERS

Les fichiers doivent être rassemblés en un seul dossier zippé. .



ARCHIVER



Si le dépôt est mis en ligne, les fichiers sont archivés dans Software Heritage et seront accessibles à partir de la notice HAL grâce au SWHID.

Cet identifiant unique attribué par Software Heritage pour chaque dépôt se présente dans la notice sous forme de lien. Il est possible de déposer dans HAL plusieurs versions d'un même logiciel, ce qui donne lieu à plusieurs versions numérotées d'une même notice.

Chaque version se verra attribuer un SWHID différent dans Software Heritage, qui figurera dans la version correspondante de la notice HAL.

LE DÉPÔT

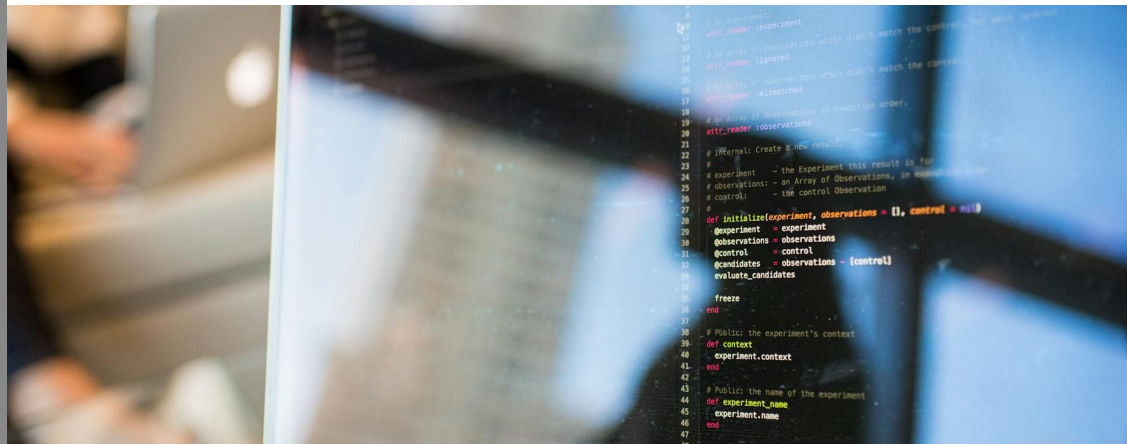


Lors de la validation du dépôt, un transfert vers Software Heritage est proposé par défaut.

 Voir

MODÉRATION

Le dépôt est transmis à l'équipe HAL INRAE pour être modéré, ce qui implique **l'analyse des fichiers et le contrôle qualité des métadonnées.**





RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

INRAE



Ce support a été créé dans le cadre de la formation à la science ouverte



développée par la Direction pour la science ouverte d'INRAE



A réutiliser sans modération

