

Référencer les logiciels de la recherche grâce à Software Heritage et HAL

Mathrice/RNBM/Groupe Calcul, 2024

Pourquoi référencer les codes de recherche ?

- **S'inscrire dans la démarche Science ouverte !**
- **Pourquoi archiver les codes ?**
 - Risque d'obsolescence (formats, environnements, etc.) ou de perte ..
 - Les logiciels sont un des piliers des processus de recherche, au côté des publications et des données
- **Pourquoi signaler les codes ?**
 - Assurer la description
 - Rechercher facilement (par domaine scientifique en particulier)
 - Permettre la citation
 - Valoriser les logiciels
- **Comment faire ? HAL et Software Heritage peuvent vous aider ...**

HAL, Hyper Articles en Ligne

- **Archive ouverte pluridisciplinaire**
- Créée en 2000 par le CNRS et exploitée par le CCSD – Centre pour la Communication Scientifique Directe
- Où les chercheurs peuvent déposer leurs résultats académiques dans le respect de leurs droits d'auteur
- Supporte différents types de productions :
 - Publications
 - Documents (par exemple pré-prints et rapports),
 - Thèses ...
- Pour rendre la **recherche aussi accessible et ouverte que possible**



Software Heritage



- Initiative dont l'objectif est de construire une **archive universelle des codes sources**
- En les collectant, les préservant et les partageant sur le long terme
- Lancée en 2016 par INRIA et soutenue par l'UNESCO
- Collecte de l'**intégralité des logiciels disponibles publiquement** sous forme de code source.
- Depuis des plateformes d'hébergement de code, comme GitHub, GitLab.com ou Bitbucket, et des archives de paquets, comme Npm ou Pypi ...



HAL + Software Heritage

■ Un nouveau type de dépôt sur HAL

- Collaboration initiée en 2018 entre HAL et SWH
- Phase de test avec INRIA
- En production depuis 2023

■ Complémentarité des deux plateformes

- Grande visibilité des logiciels dans une démarche de science ouverte via HAL
 - En lien avec les publications
- Archivage pérenne via Software Heritage
- Modération des métadonnées
- Différents formats d'export pour faciliter la citation

Exemple d'un code de recherche référencé sur SWH

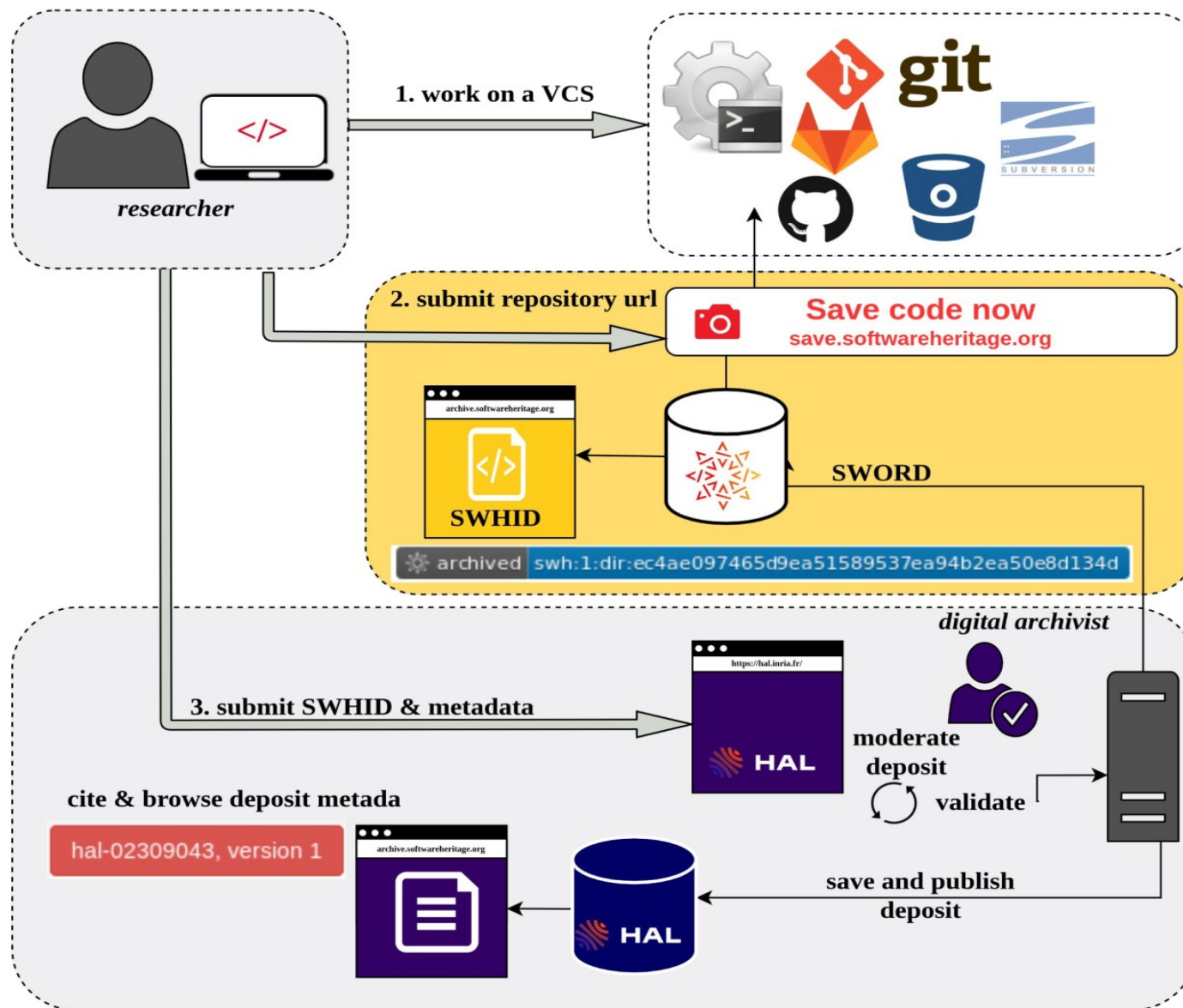


Indice Biologique Diatomées en Lac (IBDL)

Préalables

- **On s'intéresse ici au dépôt via SWHID**
 - Cela suppose l'utilisation d'une forge (qui soit moissonnée par SWH) : c'est **LA bonne pratique pour le développement des codes**
 - **Processus SWH puis HAL** : Les **informations sur la forge sont toujours à jour** donc l'archive sur SWH l'est aussi (possibilité de forcer la mise à jour) et il reste à signaler sur HAL
 - A noter qu'il est aussi possible de déposer via une archive zip sur HAL. Dans ce cas les fichiers des codes sont automatiquement archivés sur SWH.
- **Que déposer ?**
 - Le code source
 - D'un logiciel développé dans un milieu académique

Etapes du processus



Archivage sur SWH

■ Code déjà présent

- Cas où le code est hébergé sur une forge moissonnée par SWH : par exemple gricad-gitlab
- Il existe un plugin de navigateur (Updateswh) très pratique pour savoir si le code est déjà archivé ou pas



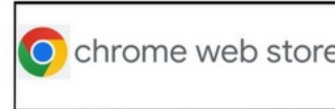
■ Save code now

- L'archivage conduit à l'existence d'un identifiant unique : le **SWHID**

This handy browser extension allows you to seamlessly check if a repository that you are browsing is archived and up to date in Software Heritage.







Getting the extension

You can find and install the Updateswh extension by clicking on the image of the browser you are interested in among the following



Understanding the color code

The color coded button on the right of the browser carries information on archival status and allows to trigger appropriate actions, as follows:

- : the archived copy is up to date, and clicking on the button opens a tab on the corresponding page of the archive (very practical if you want to get a SWHID permalink)
- : the archived copy is not up to date, and clicking on the button triggers a save code now request to update the archive; right clicking brings you to the page of the last successful archival
- : the archived copy is not up to date, and the last try to archive it failed; clicking on the button triggers a save code now request to update the archive, but there may be technical issues that prevent this repository from being archived at the moment; right clicking brings you to the page of the last archival, that may be failed
- : this repository is not found in the archive (at least not from this origin), and clicking on the button triggers a save code now request to archive it for the first time
- : you have used up your quota of API calls for the moment, either on the SWH API or on the forge API: wait a bit before trying again, or get a token to raise the rate limit (more info below)
- : the API requests to Software Heritage did not succeed. This may happen for several reasons:
 - **in most of the cases**, the repository you are visiting is private, and hence cannot be archived
 - nobody is perfect, you tripped on an error in the extension: you may enable debug mode in the options panel, and [report the issue on the extension development repository](#)



Enter a SWHID to resolve or keyword(s) to search for in origin URLs

Features

🔍 Search

⬇️ Downloads

📁 Save code now

🔧 Add forge now

❓ Help

You can contribute to extend the content of the Software Heritage archive by submitting an origin save request. To do so, fill the required info in the form below:

Origin type	Origin url	
<input type="text" value="git"/>	<input type="text" value="https://github.com/SebastienBoutry/IBDL"/>	<input type="button" value="Submit"/>

Help

Browse save requests

A "Save code now" request takes the following parameters:

- **Origin type:** the type of version control system the software origin is using. Currently, the supported types are:
 - **git**, for origins using Git
 - **hg**, for origins using Mercurial
 - **svn**, for origins using Subversion
 - **cv**s, for origins using CVS
 - **bzr**, for origins using Bazaar
- **Origin url:** the url of the remote repository for the software origin.
In order to avoid saving errors from Software Heritage, you should provide the clone/checkout url as given by the provider hosting the software origin. It can easily be found in the web interface used to browse the software origin.
For instance, if you want to save a **git** origin into the archive, you should check that the command `$ git clone <origin_url>` does not return an error before submitting a request.

Once submitted, your save request can either be:

- **accepted:** a visit to the provided origin will then be scheduled by Software Heritage in order to load its content into the archive as soon as possible
- **rejected:** the provided origin url is blacklisted and no visit will be scheduled
- put in **pending** state: a manual review will then be performed in order to determine if the origin can be safely loaded or not into the archive

Once a save request has been accepted, you can follow its current status in the **submitted save requests list**.

If you submitted requests while **authenticated**, you will be able to only display your own requests.

Checklist : préparer le logiciel

- **Les fichiers indispensables**
 - AUTHORS
 - LICENSE
 - README

- Optionnel mais **fortement conseillé**
 - Codemeta.json

Checklist : préparer le logiciel

```
1  
2 #Auteurs  
3  
4 ###Contact  
5 Sébastien Boutry (sebastien.boutry@inrae.fr)  
6  
7 ###Affiliations  
8 UR EABX - Ecosystèmes aquatiques et changements globaux  
9 ECLA - Pôle Écla - écosystèmes lacustres  
10  
11 ###Auteurs  
12 Sébastien Boutry (Orcid = ...)  
13 Soizic Morin  
14 Vincent Bertrin  
15 Juliette Roseberry
```

```
1  
2 #License  
3 Le code source et les fichiers de ce dépôt sont sous license GPL-3  
4 (https://www.gnu.org/licenses/gpl-3.0.fr.html#license-text)
```

12 October 2023, 09:09:29 UTC

<> Code Branches (2) Releases (0) Visits

Revision: eb33c3cd68932571745d2758f6b1e8dbcb75a47d

Tip revision: eb33c3cd68932571745d2758f6b1e8dbcb75a47d
ajout_list_flor

File

- R
- docs
- inst
- man
- pkgdown
- vignettes

.Rbuildignore	-rwxr-xr-x	76 bytes
.gitignore	-rwxr-xr-x	77 bytes
CITATION.cff	-rw-r--r--	277 bytes
DESCRIPTION	-rwxr-xr-x	1.1 KB
IBDL.Rproj	-rwxr-xr-x	414 bytes
NAMESPACE	-rwxr-xr-x	461 bytes
README.md	-rwxr-xr-x	11.1 KB
README.rmd	-rwxr-xr-x	17.6 KB
devtools_history.R	-rwxr-xr-x	794 bytes

Permalinks

To reference or cite the objects present in the Software Heritage archive, permalinks based on SoftWare Heritage persistent IDentifiers (SWHIDs) must be used.

Select below a type of object currently browsed in order to display its associated SWHID and permalink.

directory revision snapshot

archived repository archived swh:1:dir:170c9dbd9a2880cbd5b99ce39c9718295205c1a1

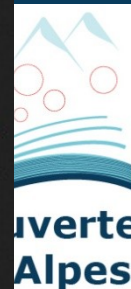
Iframe embeddi

```
swh:1:dir:170c9dbd9a2880cbd5b99ce39c9718295205c1a1;  
origin=https://github.com/SebastienBoutry/IBDL;  
visit=swh:1:snp:bfeb4ca40d7a39c643b4cfd4806038a18a708839;  
anchor=swh:1:rev:eb33c3cd68932571745d2758f6b1e8dbcb75a47d
```

Add contextual information

Copy identifier

Copy permalin



Métadonnées

■ Automatiser au maximum

- Utiliser le fichier **codemeta.json** : métadonnées ajoutées ensuite automatiquement sur HAL
- Vocabulaire CodeMeta qui étend le vocabulaire schema.org
- Existence d'un outil web pour faciliter la création du fichier : CodeMeta generator

■ Métadonnées obligatoires

- Nom du logiciel
- Licences
- Auteurs

CodeMeta generator

Most fields are optional. Mandatory fields will be highlighted when generating Codemeta.

The software itself

Name

My Software

the software title

Description

My Software computes ephemerides and orbit propagation. It has been developed from early '80.

Creation date

YYYY-MM-DD

First release date

YYYY-MM-DD

License(s)

from [SPDX licence list](#)

Run-time environment

Programming Language

C#, Java, Python 3

Runtime Platform

.NET, JVM

Operating System

Android 1.6, Linux, Windows, macOS

Other software requirements

Python 3.4
<https://github.com/psf/requests>

Discoverability and citation

Unique identifier

10.151.xxxxx

such as ISBNs, GTIN codes, UUIDs etc.. <http://schema.org/identifier>

Application category

Astronomy

Keywords

ephemerides, orbit, astronomy

Funding

PRA_2018_73

grant funding software development

Funder

Università di Pisa

organization funding software development

Authors and contributors can be added below

Current version of the software

Version number

1.0.0

Release date

YYYY-MM-DD

Download URL

<https://example.org/MySoftware.tar.gz>

Release notes

Change log: this and that;
Bugfixes: that and this.

Development community / tools

Code repository

[git+https://github.com/You/RepoName.git](https://github.com/You/RepoName.git)

Continuous integration

<https://travis-ci.org/You/RepoName>

Issue tracker

<https://github.com/You/RepoName/issues>

Related links

Additional Info

Reference Publication

<https://doi.org/10.1000/xyz123>

Development Status

see www.repostatus.org for details

Is part of

<http://The.Bigger.Framework.org>

Signalement sur HAL

- **Nouveau dépôt >>> type « logiciel »**
 - Ne pas déposer de fichier mais indiquer le SWHID
 - Prendre le SWHID avec contexte (pour garder le lien vers l'historique de développement)
- **Vérifier les métadonnées insérées** automatiquement via le fichier codemeta et **compléter les métadonnées** qui n'ont pas été récupérées
 - Par ex, ajouter le domaine
- Vérifier les **auteurs et les affiliations**
- **Ajouter, le cas échéant, la référence**
 - à une publication
 - à un jeu de données
 - à un projet (ANR...)
- **Valider** le dépôt

Fichier(s)

Métadonnées

Extraction automatique

Auteurs et affiliations

Informations principales

Identifiants

Financements

Autres Informations

Validation



Glissez votre document (max. 200M)

ou

[Parcourir](#)

[Téléverser à partir d'un lien](#)

Métadonnées, Je renseigne mon dépôt

Extraction automatique

Chargez les métadonnées à partir d'un identifiant

Les informations associées à cet identifiant permettront de compléter automatiquement votre dépôt.

SWHID ▾

06038a18a708839;anchor=swh:1:rev:eb33c3cd68932571745d2758f6b1e8dbcb75a47d

[Récupérer les métadonnées](#)

Auteurs et affiliations

Auteurs

Ajouter un auteur

VÉRIFIER AVANT DE DÉPOSER

Récupérer les métadonnées

Fichier(s)

Métadonnées

Extraction automatique

Auteurs et affiliations

Informations principales

Identifiants

Financements

Autres Informations

Validation

Auteurs et affiliations

Auteurs

Ajouter un auteur

⋮ Sébastien Boutry ^{ID}

IdHAL:
sebastien-boutry

Fonction: Développement

• Ecosystèmes aquatiques et changements globaux

Ajouter une affiliation

⋮ Soizic Morin ^{ID}

IdHAL:
soizic-morin

Fonction: Développement

• Ecosystèmes aquatiques et changements globaux

Ajouter une affiliation

⋮ Vincent Bertrin ^{ID}

IdHAL:
vincent-bertrin

Fonction: Développement

• Ecosystèmes aquatiques et changements globaux

• Pôle Écla - écosystèmes lacustres

Ajouter une affiliation

⋮ Juliette Tison-rosebery ^{ID}

IdHAL:
juliette-rosebery

Fonction: Développement

• Ecosystèmes aquatiques et changements globaux

Ajouter une affiliation

[Ajouter une liste d'auteurs](#)

[Ajouter les auteurs d'une structure](#)

[Ajouter mes auteurs](#)

[Affilier les auteurs](#)

[Supprimer toutes les affiliations](#)

VÉRIFIER AVANT DE DÉPOSER



Éditeur

+ Ajouter

Licences *

La ou les licences sous lesquelles est publié ce logiciel (vous pouvez utiliser l'autocomplétion)

Supprimer

+ Ajouter

Langage de programmation

Supprimer

+ Ajouter

Code Repository

Lien où se trouve le développement du code (SVN, github, gitlab, CodePlex).

Platform/OS

Le système d'exploitation compatible avec le logiciel

+ Ajouter

Version

la version du logiciel (peut être différente de la version publiée sur HAL)

Etat du développement

L'état du développement du logiciel au moment du dépôt (Concept, WIP, Suspendu, Actif, Inactif)

Outils de développement

Les outils de développement associés au logiciel (Framework, middleware, plateforme logicielle)

+ Ajouter

VÉRIFIER AVANT DE DÉPOSER



Français ▾

{IBDL} Outil de calcul l'Indice Biologique Diatomées en Lac

Supprimer

Anglais ▾

+ Ajouter

Domaine *

Discipline principale

Informatique [cs] ▾

Le domaine que vous mettez en premier, sera considéré comme la discipline principale du dépôt (article, logiciel ou autre). L'ordre des domaines peut être changé par glissé/déposé.

Description

Cliquer sur + pour ajouter une nouvelle langue

Français ▾

Ce package regroupe toutes les fonctions nécessaires afin de calculer l'Indice Biologique Diatomées en Lac et déterminer la classe d'état écologique du plan d'eau.

+ Ajouter

VÉRIFIER AVANT DE DÉPOSER

Identifiants



Science ouverte
Grenoble Alpes

Identifiants

Ajoutez les identifiants du dépôt dans d'autres contextes: DOI, arXiv, PubMed, ADS, Software Heritage-hash, et autres.

DOI ▾

+ Ajouter

Données associées

Ajoutez l'identifiant DOI fourni par l'entrepôt où vos données sont archivées.

+ Ajouter

Publications associées

Ajoutez les identifiants des publications associées à votre logiciel

HAL ▾

+ Ajouter



Financements

Projet(s) ANR

Veillez saisir le code décision (ANR-19-ASMA-0007), l'acronyme du projet ou le titre du projet. Vous pouvez indiquer plusieurs projets ANR.

Projet(s) Européen(s)

Indiquez le ou les projets Européens liés à ce travail

Financement

Origine du financement lié à ce travail

[+ Ajouter](#)

Liens codes, données et publications

■ Référencement

- Des publications sur HAL / autres éditeurs
- Des données sur les entrepôts thématiques ou généralistes (par ex Recherche Data Gouv)
- Des codes sur Software Heritage et HAL

■ Nécessité de **faire le lien** entre toutes ces productions grâce aux identifiants

Liens codes, données et publications

Dans HAL côté publications :



Je dépose sur HAL



Article dans une revue

Identifiants

Identifiants

Ajoutez l'identifiant [DOI](#), [arXiv](#), [PubMed](#), [ADS](#), etc pour lier votre dépôt aux autres bases.

SWHID ▾

+ Ajouter

Données associées

Ajoutez l'identifiant [DOI](#) fourni par l'entrepôt où vos données sont archivées.

+ Ajouter

Signalement sur HAL



<https://hal.science/hal-04239218v1>

Ou <https://hal.science/hal-04418015>

Références

- Fiche pratique (UGA) : diffusion des codes
- Documentation dans HAL : <http://doc.hal.science/deposer/deposer-le-code-source/>
- Create software deposit in HAL: User guide and best practices ,
<https://hal.archives-ouvertes.fr/hal-01872189>
- HOWTO archive and reference your code,
<https://www.softwareheritage.org/howto-archive-and-reference-your-code/>
- Depositing scientific software into Software Heritage,
<https://www.softwareheritage.org/2018/09/28/depositing-scientific-software-into-software-heritage/>
- Tutoriaux vidéos : <https://youtube.com/playlist?list=PLD2VqrZz2-u3bOWtoCoBIh5Flt6iYXsq3>
- Save code now : <https://archive.softwareheritage.org/save/>
- Merci à Morane Gruenpeter et à Sabrina Granger de Software Heritage !