

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

De quoi on ne va PAS parler

- **On ne parle pas des forges logicielles**
 - Ni de développement collaboratif, de gestion de version, ...
- **Archivage != stockage != outil de développement**
 - Archivage = préservation sur du long terme

Pourquoi référencer les codes de recherche ?

■ Pourquoi archiver ?

- Le code source est fragile :
 - Obsolescence des formats, problème matériel, dépendances à des outils (forge par exemple) qui peuvent disparaître ...
 - La perte des codes ayant été utilisés pour de la production scientifique arrive malheureusement régulièrement
- Les logiciels sont un des piliers des processus de recherche, au côté des publications et des données et il est essentiel de les préserver

■ Pourquoi signaler ?

- Assurer la description
- Faciliter la recherche / Rechercher facilement (par domaine scientifique en particulier)
- Permettre la citation
- Valoriser les logiciels

HAL



- **Archive ouverte nationale pluridisciplinaire**
- Initiée en 2000 par le CNRS et exploitée par le CCSD – Centre pour la Communication Scientifique Directe
- Fournit des outils pour l'archivage et la diffusion ouverte des résultats scientifiques.
- Où les chercheurs peuvent déposer leurs résultats académiques dans le respect de leurs droits d'auteur
- Supporte différents types de dépôt :
 - Publications,
 - Documents (par exemple préprints et rapport),
 - 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 codes et 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, puis déployé largement depuis 2021
 - En particulier, le dépôt est possible à partir du SWHID (identifiant unique et pérenne de Software Heritage)
- **Complémentarité des deux plateformes**
 - Archivage pérenne via Software Heritage
 - Grande visibilité des logiciels dans une démarche de science ouverte via HAL
 - Modération des métadonnées
 - Différents formats d'export pour faciliter la citation

Exemple d'un code de recherche référencé



DEMO

Démo : liste des dépôts de type logiciel dans HAL

967 Résultats

Type de document : Logiciel ✕

triés par

Pertinence ▾

Charte graphique et site web International Conférence LCAFOOD 2012
 Sylvaine Bitteur , Haye van Der Werf , . Umr Sas
 2012
Logiciel hal-02806617v1

reporting_rmqs : Scripts de reporting des données de la première campagne du RMQS
 Jean-Philippe Chenu
 2009
Logiciel hal-02813506v1

websol_database_manager : Script d'extraction des Données Donesol3 pour les pousser sur les plateformes Websol
 Benoit B. Toutain
 2009
Logiciel hal-02813105v1

Filtrer vos résultats

Type de dépôt ▾

Type de document ▾

- Logiciel 967
- Article dans une revue 1676930
- Communication dans un congrès 826928
- Chapitre d'ouvrage 300984
- Thèse 149138
- Ouvrages 101518
- Pré-publication, Document de travail 69938
- Autre publication scientifique 61788
- Rapport 61591
- Image 58984



Démo : exemple d'un code de recherche référencé

Liste complète des métadonnées

Format du dépôt	Fichier
Type de document	Logiciel
Titre	en LinBox
Résumé	en LinBox is a C++ template library of routines for solution of linear algebra problems including linear system solution, rank, determinant, minimal polynomial, characteristic polynomial, and Smith normal form. Algorithms are provided for matrices with integer entries or entries in a finite field. A number of matrix storage types is provided, especially for blackbox representation of sparse or structured matrix classes. A few algorithms for rational matrices are available. LinBox also uses underlying data structures and algorithms for integer, rational, polynomial, finite fields and rings, as well as dense and sparse matrix formats coming from the Givaro (https://casys.gricad-pages.univ-grenoble-alpes.fr/givaro) and FFLAS-FFPACK (http://linbox-team.github.io/fflas-ffpack) libraries.
Auteur(s)	The Linbox Group ^{1, 2, 3, 4, 5, 6, 7, 8} 1 - Drexel University (11333) - 3141 Chestnut Street, Philadelphia, PA 19104 - E Unis

Code Repository	https://github.com/linbox-team/linbox
Version	1.7.0
Domaine	<ul style="list-style-type: none"> • Informatique [cs] • Informatique [cs]/Calcul formel [cs.SC]
Projet(s) ANR	<ul style="list-style-type: none"> • Calcul Algébrique Haute-Performance [En savoir plus] — HPAC - ANR-11-BS02-0013 — BLANC - 2011
Projet(s) Européen(s)	<ul style="list-style-type: none"> • OpenDreamKit - Open Digital Research Environment Toolkit for the Advancement of Mathematics — Numéro CORDIS : 676541
Licences	<ul style="list-style-type: none"> • GNU Lesser General Public License v2.1 or later
Langage de programmation	<ul style="list-style-type: none"> • C++
SWHID	swh:1:dir:ab3150274e077cc4c4d18ab7bc074fab0a5c12d3;origin=https://hal.archives-ouvertes.fr/hal-03788347; visit=swh:1:snp:99e4a5b29527e7089567751cd8b75113ecd12c0b; anchor=swh:1:rel:35391e40ce3a7b02180e2dc159c095a89ed26a91;path=/

Démo : archive associée sur Software Heritage



Features

- Search
- Downloads
- Save code now
- Add forge now
- Help

<https://hal.archives-ouvertes.fr/hal-03788347>

04 October 2022, 17:08:04 UTC

<> Code Branches (0) Releases (1) Visits

Release: HEAD ab31502 / linbox-1.7.0 /

Download

File	Mode	Size
------	------	------

benchmarks

- AUTHORS
- COPYING
- COPYING.LESSER
- ChangeLog
- Makefile.am
- Makefile.in
- README.md

README.md

The Linbox Library

Build Status

Purpose

The Linbox library provides functionalities for exact linear algebra. See doc/mainpage.doxy for more info.

Auto-installer for quick install

Download linbox-auto-install.sh, make it executable, and run it!

Préalables

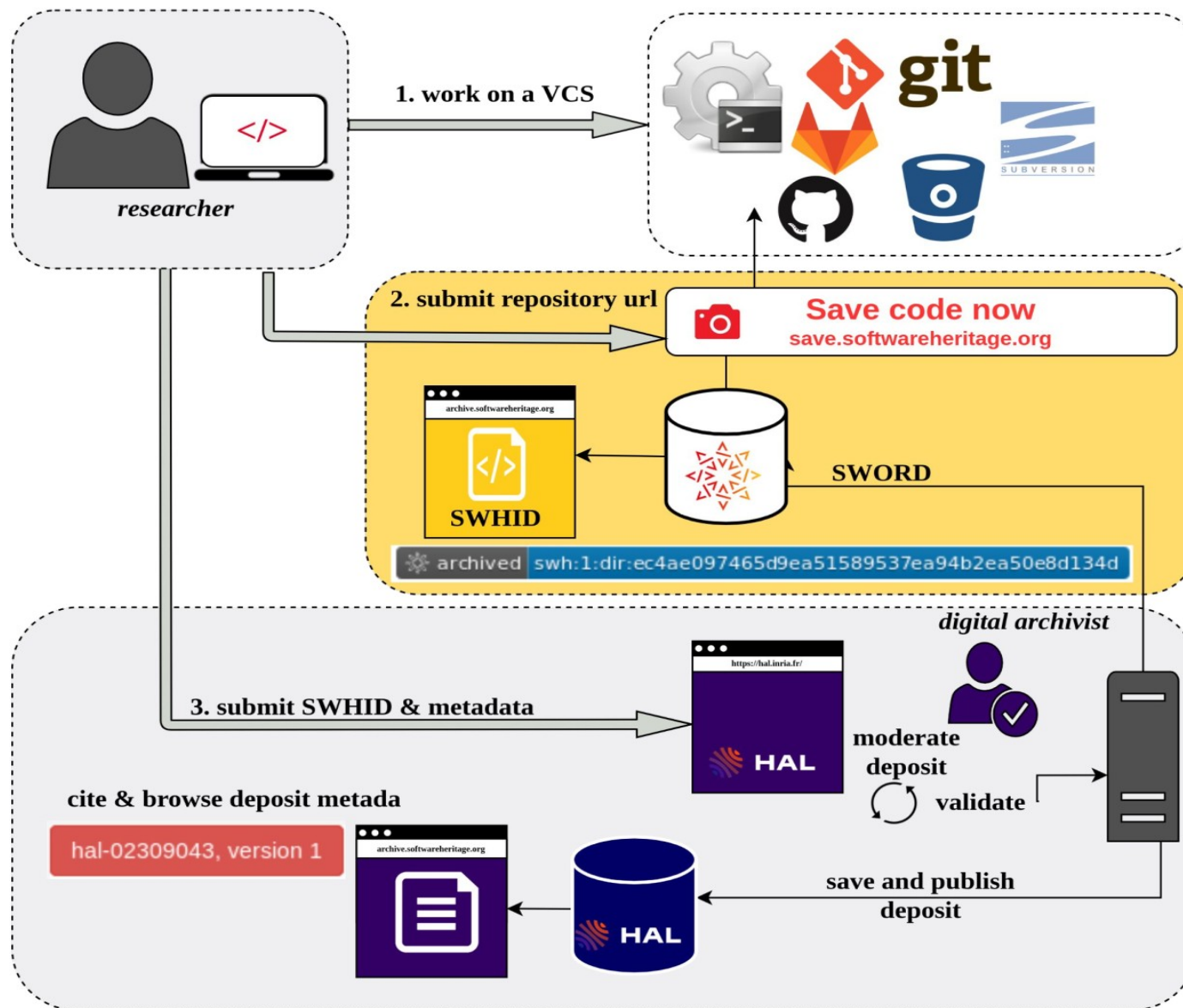
■ On s'intéresse ici au dépôt via SWHID

- 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



Checklist : préparer le logiciel

- Les fichiers indispensables
 - AUTHORS
 - LICENSE (ou COPYING)
 - README

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

DEMO

Démo : fichier « AUTEURS »

AUTHORS

```
1 The Linbox group is composed of the following authors:
2
3 Alexis Breust https://github.com/Breush
4 Brice Boyer <brice.boyer@gmail.com>
5 Jean-Guillaume Dumas <Jean-Guillaume.Dumas@univ-grenoble-alpes.fr>
6 Ahmet Duran
7 Thierry Gautier
8 Pascal Giorgi <Pascal.Giorgi@lirmm.fr>
9 Bradford Hovinen <hovinen@cis.udel.edu>
10 Gavin Harrison <gmh33@drexel.edu>
11 Erich Kaltofen <kaltofen@math.ncsu.edu>
12 Romain Lebreton <romain.lebreton@lirmm.fr>
13 Austin Lobo
14 Clement Pernet https://github.com/ClementPernet
15 Daniel Roche <roche@usna.edu>
16 B. David Saunders <saunders@cis.udel.edu>
17 Rich Seagraves <seagrave@cis.udel.edu>
```

Démo : fichier « LICENSE »

COPYING

```
1          GNU GENERAL PUBLIC LICENSE
2          Version 2, June 1991
3
4  Copyright (C) 1989, 1991 Free Software Foundation, Inc.,
5  51 Franklin Street, Fifth Floor, Boston, MA 02110-1301 USA
6  Everyone is permitted to copy and distribute verbatim copies
7  of this license document, but changing it is not allowed.
8
9          Preamble
10
11  The licenses for most software are designed to take away your
12  freedom to share and change it.  By contrast, the GNU General Public
13  License is intended to guarantee your freedom to share and change free
14  software--to make sure the software is free for all its users.  This
15  General Public License applies to most of the Free Software
16  Foundation's software and to any other program whose authors commit to
17  using it.  (Some other Free Software Foundation software is covered by
18  the GNU Lesser General Public License instead.)  You can apply it to
19  your programs, too.
```




Démo : fichier « README »

README.md

```
1 # The Linbox Library
2
3 [![Build Status](https://ci.inria.fr/linbox/buildStatus/icon?job=LinBox)](https://ci.inria.fr/linbox/job/LinBox/)
4
5 ## Purpose
6
7 The Linbox library provides functionalities for exact linear algebra.
8 See doc/mainpage.doxy for more info.
9
10 ## Auto-installer for quick install
11
12 Download [linbox-auto-install.sh](https://github.com/linbox-team/linbox/raw/master/linbox-auto-install.sh), make it
13
14 To get a list of options:
15 ```
16 ./linbox-auto-install.sh --help
17 ```
18
19 Requirements:
20 - GNU software building tools (e.g. Debian packages `autotools-dev` and `dh-autoreconf`),
21 - possibly the `gfortran` compiler, if Fortran-based BLAS (such as OpenBLAS) are built via this script.
```


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 très pratique pour savoir si le code est déjà archivé ou pas

DEMO

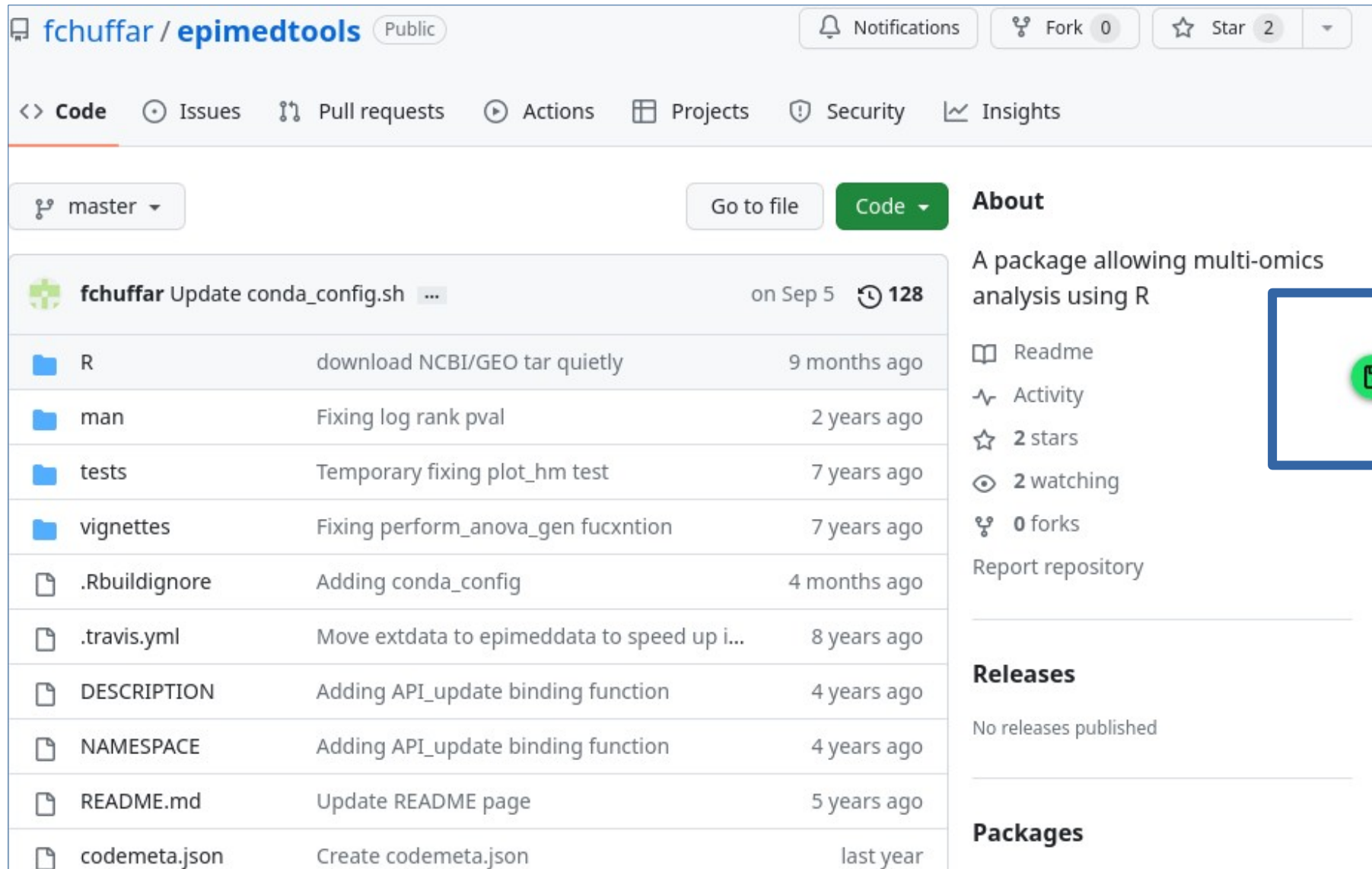
- Save code now

- **DEMO**

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

- **DEMO**

Démo : archiver un code



fchuffar / epimedtools Public

Notifications Fork 0 Star 2

Code Issues Pull requests Actions Projects Security Insights

master

Go to file Code

About
A package allowing multi-omics analysis using R

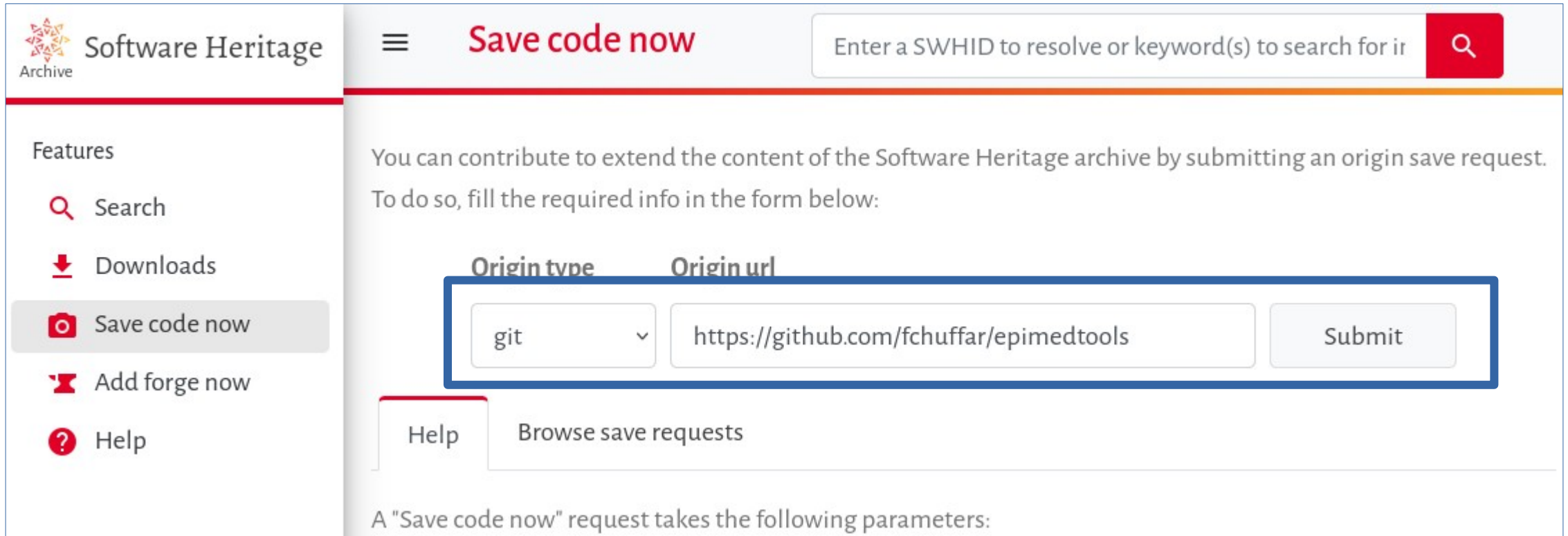
Readme Activity 2 stars 2 watching 0 forks Report repository

Releases
No releases published

Packages

Commit	Message	Time
fchuffar	Update conda_config.sh	on Sep 5 128
R	download NCBI/GEO tar quietly	9 months ago
man	Fixing log rank pval	2 years ago
tests	Temporary fixing plot_hm test	7 years ago
vignettes	Fixing perform_anova_gen fucxntion	7 years ago
.Rbuildignore	Adding conda_config	4 months ago
.travis.yml	Move extdata to epimeddata to speed up i...	8 years ago
DESCRIPTION	Adding API_update binding function	4 years ago
NAMESPACE	Adding API_update binding function	4 years ago
README.md	Update README page	5 years ago
codemeta.json	Create codemeta.json	last year

Démo : archiver avec « save code now »



The screenshot shows the 'Save code now' page on the Software Heritage website. The page has a header with the Software Heritage logo and a search bar. A left sidebar contains navigation links for Features, Search, Downloads, Save code now (highlighted), Add forge now, and Help. The main content area explains that users can contribute to the archive by submitting an origin save request and provides a form with two fields: 'Origin type' (set to 'git') and 'Origin url' (set to 'https://github.com/fchuffar/epimedtools'). A 'Submit' button is next to the url field. Below the form, there are links for 'Help' and 'Browse save requests'. At the bottom, it states that a 'Save code now' request takes the following parameters:

Software Heritage Archive

Save code now

Enter a SWHID to resolve or keyword(s) to search for it

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
git	https://github.com/fchuffar/epimedtools

Submit

Help Browse save requests

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

Démo : le SWHID – sur SWH et HAL

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:36ad628e482793c37591b7f9190e8a508a58dd4f` Iframe embedding

```
swh:1:dir:36ad628e482793c37591b7f9190e8a508a58dd4f;  
origin=https://github.com/fchuffar/epimedtools;  
visit=swh:1:snp:862855031caeadd060f8fd99cfdb28f9600727d7;  
anchor=swh:1:rev:dd8d62d4e20cbdc39dabbefa34b6465d7782f106
```

Add contextual information Copy identifier Copy permalink

Consulter sur  Software Heritage

Dates et versions

hal-03722450 , **version 1**
(13-07-2022)

Identifiants

HAL Id : **hal-03722450** , versio
n 1

SWHID : **swh:1:dir:d8d64b762735d8ec3dc01ff83577b5f9d4c45a96;origin=https://github.com/fchuffar/epimedtools;visit=swh:1:snp:bec1f269a582915f6eb533ad422fb87571bebe28;anchor=swh:1:rev:743ade91a3d634f291d3849b9d2440e7c90cef23**

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

■ DEMO



Démo : fichier « codemeta »

```
codemeta.json
1  {
2    "@context": "https://doi.org/10.5063/schema/codemeta-2.0",
3    "@type": "SoftwareSourceCode",
4    "license": "https://spdx.org/licenses/GPL-3.0",
5    "codeRepository": "https://github.com/fchuffar/epimedtools",
6    "dateCreated": "2016-01-26",
7    "name": "epimedtools",
8    "description": "Useful Tools Allowing Multi-Omics Analysis",
9    "applicationCategory": "Biology",
10   "programmingLanguage": [
11     "R"
12   ],
13   "author": [
14     {
15       "@type": "Person",
16       "givenName": "Florent",
17       "familyName": "Chuffart",
18       "email": "florent.chuffart@gmail.com",
19       "affiliation": {
20         "@type": "Organization",
21         "name": "INSERM"
22       }
23     }
24   ],
25   "contributor": [
```


Démo : CodeMeta generator



CodeMeta generator

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

The software itself	Discoverability and citation	Development community / tools
Name <input type="text" value="epimedtools"/> the software title	Unique identifier <input type="text" value="10.151.xxxxx"/> such as ISBNs, GTIN codes, UUIDs etc.. http://schema.org/identifier	Code repository <input type="text" value="https://github.com/fchuffar/epimedtools"/>
Description <input type="text" value="Useful Tools Allowing Multi-Omics Analysis"/>	Application category <input type="text" value="Biology"/>	Continuous integration <input type="text" value="https://travis-ci.org/You/RepoName"/>
Creation date <input type="text" value="2016-01-26"/>	Keywords <input type="text" value="ephemerides, orbit, astronomy"/>	Issue tracker <input type="text" value="https://github.com/You/RepoName/issues"/>
First release date <input type="text" value="YYYY-MM-DD"/>	Funding <input type="text" value="PRA_2018_73"/> grant funding software development	Related links <input type="text"/>
License(s) <input type="text"/> from SPDX licence list GPL-3.0 <input type="button" value="Remove"/>	Funder <input type="text" value="Università di Pisa"/> organization funding software development	
	Authors and contributors can be added below	

Signalement sur HAL

- **Nouveau dépôt**
 - Ne pas déposer de fichier mais indiquer le SWHID
 - Préférer le **SWHID avec contexte** (pour garder le lien vers l'historique de développement)
 - HAL récupère automatiquement les métadonnées du fichier codemeta
- **Vérifier les métadonnées insérées** automatiquement et **compléter les métadonnées** qui ne sont pas encore remplies dans codemeta ou pas encore importées par HAL
- Ajouter le **domaine**, et vérifier les **auteurs et les affiliations**
- **Valider** le dépôt
- **DEMO**

Démo : référencer un logiciel dans HAL

Je sélectionne un type de document

Choisir le type de document qui correspond à votre publication. L'icône '>' indique que des sous-types de documents sont disponibles et peuvent être sélectionnés ⓘ

 Article dans une revue >	 Chapitre d'ouvrage	 Pré-publication, Document de >	 Vidéo
 Communication dans un congrès	 Article de blog scientifique	 Rapport >	 Son
 Poster de conférence	 Notice d'encyclopédie ou de	 Thèse	 Carte
 Proceedings/Recueil des communications	 Traduction	 HDR	 Logiciel
 N°spécial de revue/special issue	 Brevet	 Cours	
 Ouvrages >	 Autre publication scientifique	 Image >	

Démo : utilisation du SWHID

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 - swh:1:dir:36ad628e482793c37591b7f9190e8a508a5f

Récupérer les métadonnées

⚠ Ce document est
peut-être déjà présent
dans HAL.


Florent Chuffart.
epimedtools. 2022, (swh:1:
dir:d8d64b762735d8ec3dc0
1ff83577b5f9d4c45a96;origi
n=https://github.com/fchuff
ar/epimedtools;visit=swh:1:
snp:bec1f269a582915f6eb5
33ad422fb87571bebe28;an
chor=swh:1:rev:743ade91a3
d634f291d3849b9d2440e7c
90cef23). (hal-03722450)


🔗 Accéder au dépôt

Démo : vérifier l'import automatique

Auteurs

Ajouter un auteur

Florent Chuffart 

Fonction: Développement 

[Ajouter une liste d'auteurs](#)
 [Ajouter les auteurs d'une structure](#)
 [Ajouter mes auteurs](#)
[Affilier les auteurs](#)
 [Supprimer toutes les affiliations](#)

Informations principales *informations nécessaires pour la citation*

- **Domaine** : Cette valeur est obligatoire et ne peut être vide

[Fermer](#)

Nom *
Vous pouvez renseigner le titre en plusieurs langues : choisir la langue et cliquer sur + pour ajouter une nouvelle langue

Anglais ▾

epimedtools

Domaine *

Discipline principale

Le domaine que vous mettrez en premier
domaines peut être changé par glissé/dé

- Chimie
- Chimie analytique
- Catalyse
- Chemo-informatique

Démo : valider le dépôt pour l'envoyer à modération

Validation, Je vérifie mes informations et j'accepte les conditions

Votre dépôt n'est pas encore complet !

Veillez compléter ou corriger les étapes suivantes :

- ✘ Auteurs et affiliations [Accéder](#)
- ✘ Informations principales [Accéder](#)

VÉRIFIER AVANT DE DÉPOSER

Quelques éléments sur le processus de modération

- **Etapes de modération**
 - Vérification de la **validité du SWHID**
 - Vérification sur SWH des **éléments sur le logiciel** (parcourir le dépôt) via les **fichiers** README, LICENSE, AUTHORS, codemeta.json
 - Vérifier les **métadonnées obligatoires** : nom, domaine, licence, auteur
 - Vérifier les **url** (lien vers la forge, lien vers les DOI si indiqués ...)
- **Guide de bonnes pratiques** : 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)

Liens codes, données et publications

■ Référencement

- Des publications sur HAL / autres archives ou éditeurs
- Des données sur Recherche Data Gouv / autres entrepôts
- Des codes sur Software Heritage et HAL

■ Nécessité de **faire le lien** entre toutes ces productions

- Ne pas hésiter à faire des référencements croisés des DOI des différentes productions scientifiques liées

- Au niveau de l'Université Grenoble Alpes, la **Cellule Data Grenoble Alpes** vous accompagne sur toutes ces problématiques :
 - Autour des codes : sos-codes-recherche@univ-grenoble-alpes.fr
 - Autour des données : sos-data@univ-grenoble-alpes.fr

- Au niveau national, il est possible de contacter le **réseau des ateliers de la donnée** :
 - <https://recherche.data.gouv.fr/fr/page/reseau-des-ateliers-de-la-donnee>
 - Un **groupe de travail** est dédié à l'accompagnement au sujet des codes de recherche

Références

- Create software deposit in HAL: User guide and best practices , <https://hal.archives-ouvertes.fr/hal-01872189>
- Modérer un dépôt logiciel dans HAL : dépôt source et dépôt SWHID, <https://hal.inria.fr/hal-01876705v2>
- 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/>

Annexes



Focus : droits d'auteur appliqués au logiciel

- Le logiciel est protégé par le **droit d'auteur** avec des règles spécifiques :
 - **Droits moraux** attachés à l'auteur
 - **Droits patrimoniaux**, qui régissent les modalités d'exploitation, sont propriétés de ou des institutions qui emploient le ou les auteurs
 - Contrairement aux autres droits d'auteurs, il y a une « **dévolution automatique des droits patrimoniaux à l'employeur** »
 - Depuis le 15 décembre 2021, c'est vrai aussi pour les **stagiaires**
 - L'**algorithme**, considéré comme une suite d'idées, ou le modèle mathématique ne peuvent pas être soumis au droit d'auteur
 - La **documentation** est protégée par le droit commun du droit d'auteur

Focus : les deux grands types de licences

- **Licences libres ou Open Source**, termes plus ou moins similaires
 - Les licences dites « libres » (traduction bancale de l'anglais « royalty-free ») viennent de la Free Software Foundation
 - Open Source vient de l'Open Source Society
 - Ils n'ont pas exactement la même philosophie
 - Une licence libre ne veut pas dire libre de droit, bien au contraire !
- **Licences propriétaires**, donc non libre c'est-à-dire que seul l'auteur ou l'ayant droit du logiciel peut le modifier.
- Liste des licences existantes : <https://spdx.org/licenses/> (Software Package Data Exchange)

Focus : les principales licences libres

- Il existe différents types de licences libres :
 - **sans copyleft** : la licence initiale ne s'impose pas. Permission de redistribuer et de modifier, mais aussi d'y ajouter des restrictions.
 - **copyleft faible** : la licence initiale reste, des ajouts peuvent avoir une autre licence.
 - **copyleft fort** : la licence initiale s'impose sur tout. Licence dite contaminante.

Type	Exemple de licences
Sans copyleft	BSD license, Apache License 2 MIT
Copyleft faible	GNU library or «Lesser» General Public License (LGPL)
Copyleft fort	GNU General Public, License EUPL