



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Archiver, identifier, décrire et citer le code source :

Le dépôt de logiciel de recherche sur l'archive ouverte HAL

Morane Gruenpeter

Software Heritage, Inria

Agenda

- ★ Pourquoi **déposer** et **partager** les codes sources de recherche?
 - Archiver
 - Identifier
 - Décrire
 - Citer
- ★ Outils pour faciliter la démarche de dépôt

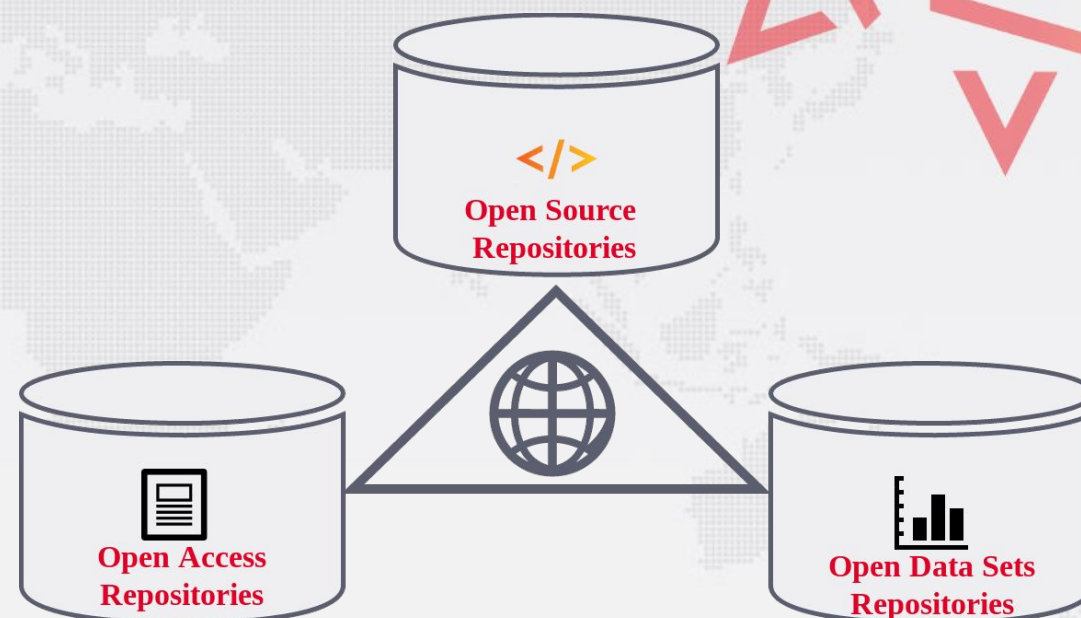
Le logiciel dans la recherche

Les rôles sont multiples:

- Un **outil**
- Un **résultat** de recherche
- Un **objet de recherche**

CoSO- Note d'opportunité sur la valorisation des logiciels issus de la recherche

<https://www.ouvrirlascience.fr/note-dopportunit%C3%A9-sur-la-valorisation-des-logiciels-issus-de-la-recherche/>



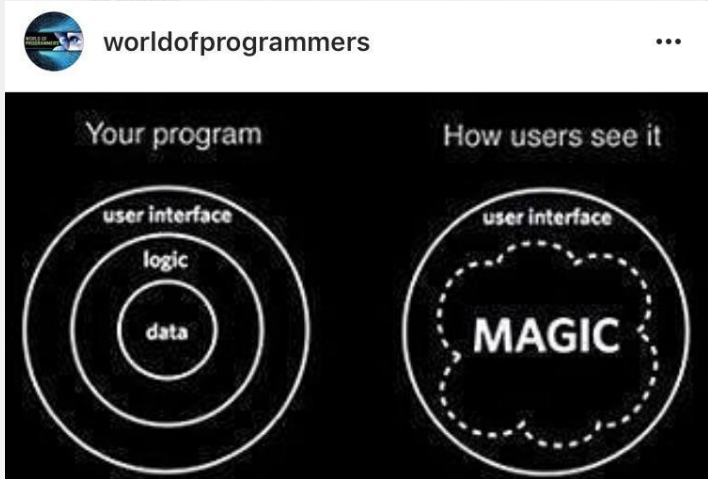
*Three pillars of Open Science, Software Heritage
CC-By 4.0 2019*

La reproductibilité est un élément clé

La Logique de la découverte scientifique, 1934

The Logic of Scientific Discovery, 1934

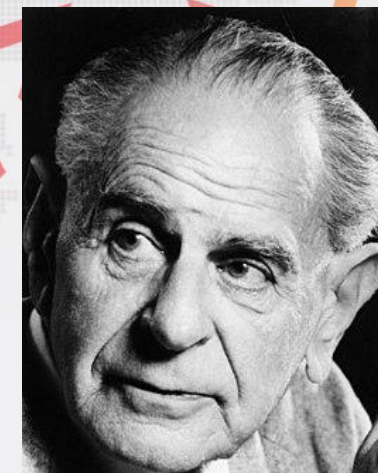
"Non-reproducible single occurrences are of no significance to science"



https://www.reddit.com/r/ProgrammerHumor/comments/70fump/programming_is_magic/



Karl Popper



- ★ La préservation des codes source est à la base de la reproductibilité.
- ★ La lecture du code est essentielle pour permettre la transmission des connaissances scientifiques aux générations futures

Un nouveau type dépôt sur HAL

Les dates clés

- ★ 2017 - Collaboration débute
- ★ mars 2018 - Phase de test sur HAL-Inria
- ★ septembre 2018 - Ouverture sur HAL
- ★ avril 2020 - BibLateX @software

Prochainement:

- ★ été 2021 - Dépôt avec SWHID

COMING SOON

Les acteurs



Pour plus d'information sur l'archive [SoftWare Heritage \(SWH\)](#) :

- Suivez la présentation de Roberto Di Cosmo

Les objectifs

- ★ **Archiver** le logiciel sur HAL et sur SWH
- ★ **Identifier**
 - les objets avec un SWHID (SoftWare Heritage Identifier)
 - la notice et la citation avec un HAL-ID
- ★ **Décrire** le logiciel avec des métadonnées qui sont modéré
- ★ **Citer** le dépôt avec une citation complète



Pourquoi déposer sur HAL

- ★ Grande **visibilité** aux logiciels dans une démarche de science ouverte.
- ★ **Archivage pérenne**, en transférant votre code vers Software Heritage, l'archive universelle du code source.
- ★ **Modération** des métadonnées par l'équipe des documentalistes.
- ★ Différents **formats d'export** pour faciliter la citation.

Le dépôt source

Le dépôt SWHID

COMING SOON

Une question de qualité

Qualité de la curation

- ★ examiner les métadonnées descriptives
- ★ vérifier la cohérence entre la notice et le code (e.g licence)
- ★ crédit correct aux créateurs - tous les auteurs sont reconnus dans la notice

Qualité du code (lisibilité)

- ★ examiner la lisibilité du code source
- ★ évaluer la faisabilité de modification du code

Code Review

Qualité du logiciel

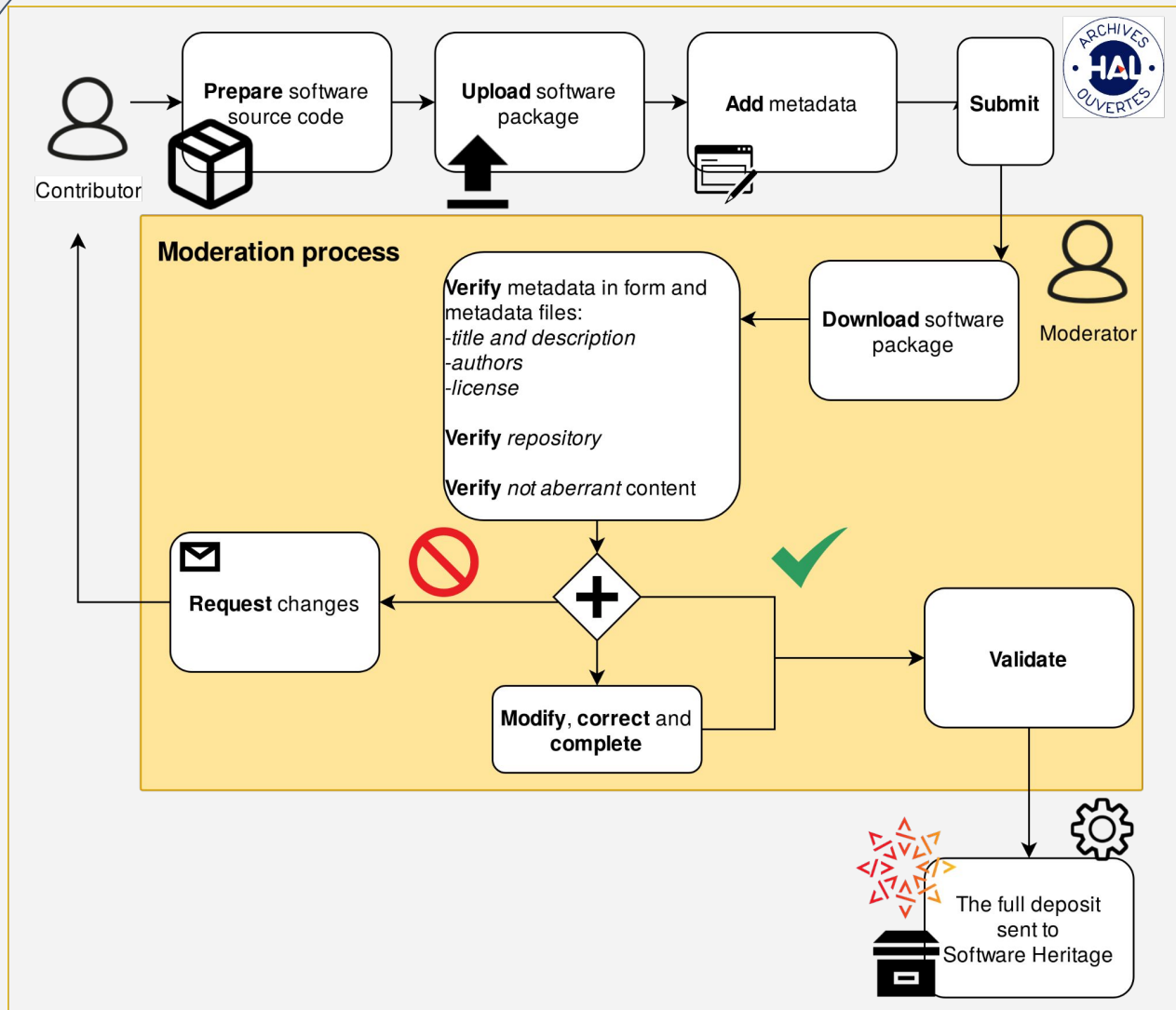
- ★ examiner la fonctionnalité
- ★ compiler et exécuter
- ★ vérifier l'exactitude
- ★ évaluer la reproductibilité

Peer review

La modération

Qualité de la curation

- ★ examiner les métadonnées descriptives
- ★ vérifier la cohérence entre la notice et le code (e.g licence)
- ★ crédit correct aux créateurs - tous les auteurs sont reconnu dans la notice



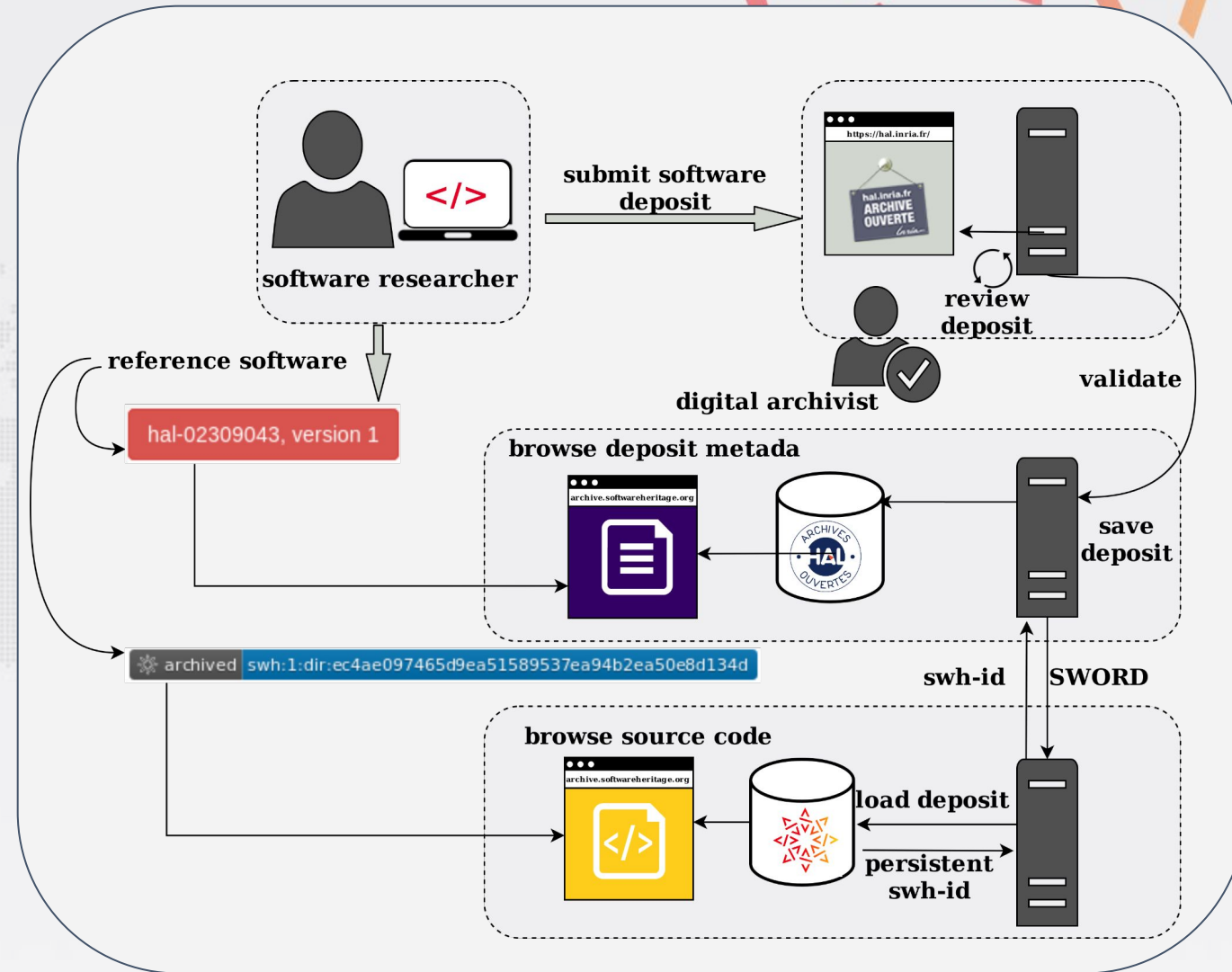
Curated Archiving of Research Software Artifacts:
Lessons Learned from the French Open Archive (HAL) on
IJDC <https://doi.org/10.2218/ijdc.v15i1.698>

Le dépôt source

Les logiciels qui sont développés localement et/ou hébergés sur des sites web personnels ou institutionnels

Il contient :

- une archive compressée
- une collection de métadonnées
 - métadonnées générales
 - métadonnées spécifiques



Curated Archiving of Research Software Artifacts:
Lessons Learned from the French Open Archive (HAL) on
IJDC <https://doi.org/10.2218/ijdc.v15i1.698>

Référence vs. citation

Le SWHID (SoftWare Heritage Identifiers)

- ★ identifiant intrinsèque
- ★ accord sur la méthode de calcul - ne nécessite pas d'autorité
- ★ une empreinte digitale du code
- ★ nécessaire pour :
 - spécificité - reproductibilité
 - archivage

Le HAL-ID

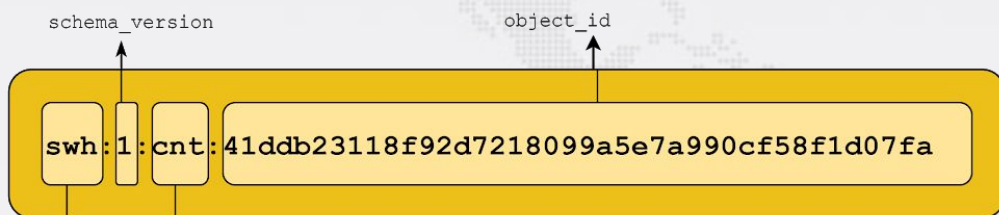
- ★ identifiant extrinsèque
- ★ métadonnées avec le dépôt
- ★ identification d'une notice
- ★ auteurs et contributeurs sont vérifiés en modération
- ★ nécessaire pour :
 - attribution
 - indexation



SWHID + HAL-ID = Citation

Le SWHID

archived swh:1:dir:ec4ae097465d9ea51589537ea94b2ea50e8d134d



- prefix
- object_type
- "snp" - snapshot
- ☆ "rel" - release
- △ "rev" - revision
- 📁 "dir" - directory
- 🗄 "cnt" - content

- origin_ctxt → ;origin=https://github.com/chrislgarry/Apollo-11
- visit_ctxt → ;visit=swh:1:snp:206c27c0c031c6aac6b5feddba8fe082dea9836
- anchor_ctxt → ;anchor=swh:1:rev:3913f198f4383d4d638c0485d6aa902ff2f35828
- path_ctxt → ;path=/Luminary099/BURN_BABY_BURN--MASTER_IGNITION_ROUTINE.agc
- lines_ctxt → ;lines=64-72

Le HAL-ID

hal-02309043, version 1



Le format citation sur HAL

Matthieu Kowalski, Emmanuel Vincent, Rémi Gribonval. *Underdetermined Reverberant Source Separation*. 2019,
[swh:1:dir:ec4ae097465d9ea51589537ea94b2ea50e8d134d](https://hal.archives-ouvertes.fr/hal-02309043);origin=https://github.com/chrislgarry/Apollo-11
;visit=swh:1:snp:206c27c0c031c6aac6b5feddba8fe082dea9836
;anchor=swh:1:rev:3913f198f4383d4d638c0485d6aa902ff2f35828
;path=/Luminary099/BURN_BABY_BURN--MASTER_IGNITION_ROUTINE.agc
;lines=64-72
hal-02309043

1ère étape : la préparation du code

Avant le dépôt sur HAL préparer le code source du logiciel.

Ces éléments sont vérifiés après par les modérateurs.

- ❑ ajouter les fichiers suivants :
 - ❑ README
 - ❑ AUTHORS
 - ❑ LICENSE (à choisir avec les titulaires des droits patrimoniaux du dit logiciel - [liste de référence](#))
 - ❑ **codemeta.json** (facultatif mais pratique)
- ❑ créer archive compressée (.zip, .tar.gz)

2ème étape : le dépôt

- Choisir le type de dépôt - **logiciel**
- Ajouter des métadonnées générales
- Ajouter des métadonnées spécifiques aux logiciels
- Ajouter les **auteurs**
- **Valider** le dépôt

HAL - Inria
Archive ouverte / Open archive

Accueil Déposer Consulter tout HAL Publications Inria Recherche Services Documentation OpenAccess@Inria Mon espace Privileges

Déposer

Déposer le(s) fichier(s)

Afficher la vue détaillée

Vous pouvez déposer un fichier et/ou compléter votre dépôt à partir d'un identifiant externe.

Déposez ou cliquez pour télécharger vos fichiers
Taille maximale du fichier : 200M

Chargez les métadonnées à partir d'un identifiant
Les informations associées à cet identifiant permettront de remplir automatiquement votre dépôt. Par exemple, le DOI est un code qui apparaît généralement sur la page de votre article sur le site de l'éditeur.

DOI 10.xxx

Récupérer les métadonnées

Si vous souhaitez ajouter un embargo, récupérer des fichiers de votre espace FTP ou toute autre action avancée, veuillez afficher la vue détaillée.

Compléter les métadonnées du document

Compléter les données auteur(s) ✓

Afficher toutes les options

Morane Gruenpeter idHAL: morane-gruenpeter Auteur

Software Heritage

Ajouter une affiliation

Ajouter un auteur : Jean Dupont

Options : Une liste d'auteurs Les auteurs

Compléter les données auteur(s)

Morane Gruenpeter

Fonction

- Développement
- Développement
- Maintenance
- Design
- Architecture
- Débogage
- Documentation
- Test
- Support
- Management

Validation des affiliations

Valider le dépôt

Le dépôt avec SWHID

Pour des logiciels qui sont **développés** sur une **plateforme collaborative**
(type GitHub, GitLab, Bitbucket)

Les avantages :

- ★ **Archivage** sur SWH de l'historique du développement
- ★ **Identification** (grâce au SWHID) de la version spécifique
- ★ **Description** du logiciel sur une notice HAL (indexée)
- ★ **Citation** simplifié grâce à divers exports

COMING SOON

1ère étape : la préparation du code

Avant le dépôt sur HAL préparer le code source du logiciel.

Ces éléments sont vérifiés après par les modérateurs.

- ❑ ajouter les fichiers suivants sur le `code repository`:
 - ❑ README
 - ❑ AUTHORS
 - ❑ LICENSE (à choisir avec les titulaires des droits patrimoniaux du dit logiciel - [liste de référence](#))
- ❑ **codemeta.json** (facultatif mais pratique)
- ❑ faire `save code now` sur SWH et récupérer le SWHID complet d'un *directory*

'Save Code Now' sur SWH

Home Archive Development Documentation Donate

Software Heritage Save code now Beta version

Archive Access

- Browse
- Web API
- Features
- Search
- Vault
- Save code now**
- Miscellaneous
- Help

Origin type Origin url

git

Submit

Récupérer le SWHID sur SWH

The screenshot shows the Software Heritage interface for the repository `https://github.com/CGAL/cgal`. The page title is "Browse the archive". A search bar at the top right contains the text "Enter a SWHID to resolve or keyword(s) to search for in origin URLs". The repository page shows a revision `b86a5018c7f5f733c80fe40eee65803c112f2685` dated 23 October 2020, 13:31 UTC. A red vertical box labeled "Permalinks" highlights the "Permalinks" section, which contains the instruction: "To reference or cite the objects present in the Software Heritage archive, permalinks based on SoftWare H copying and pasting the url from the address bar of the browser (as there is no guarantee the current URI)". Below this, a dropdown menu is open, showing options for "directory", "revision", and "snapshot". The "revision" option is selected, and a blue box highlights the generated SWHID: `swh:1:dir:abc0e2cbbfdfee8de52f0842263fbadf65f5b211`. Below the SWHID, a code block displays the full citation string: `swh:1:dir:abc0e2cbbfdfee8de52f0842263fbadf65f5b211; origin=https://github.com/CGAL/cgal; visit=swh:1:snp:78e145aa8174e576786284475a76cf6f187b3475; anchor=swh:1:rev:b86a5018c7f5f733c80fe40eee65803c112f2685; path=/Hyperbolic_triangulation_2/`. A red box highlights the "Add contextual information" checkbox, which is checked. To the right of the code block, there are two buttons: "Copy identifier" and "Copy permalink".

Le format citation sur HAL

Mikhail Bogdanov, Iordan Iordanov, Monique Teillaud. CGAL 2D Hyperbolic Delaunay Triangulations. 2019, [\[swh:1:dir:abc0e2cbbfdfee8de52f0842263fbadf65f5b211;origin=https://github.com/CGAL/cgal;visit=swh:1:snp:78e145aa8174e576786284475a76cf6f187b3475;anchor=swh:1:rev:b86a5018c7f5f733c80fe40eee65803c112f2685;path=/Hyperbolic_triangulation_2/\]. \[hal-02983422\]](https://hal.archives-ouvertes.fr/hal-02983422)

2ème étape : le dépôt

- Mettre le **SWHID** `dans le nuage`
 - ◆ en utilisant un codemeta.json - la plateforme HAL récupère les métadonnées automatiquement
- Compléter les métadonnées
 - ◆ choisir le domaine
 - ◆ bien vérifier les auteurs et affiliations
- Valider le dépôt

CCSD HAL - Episciences.org - Sciencesconf.org Support fr Gruenpeter Morane

Accueil Déposer Consulter tout HAL Publications Inria Recherche Services Documentation OpenAccess Mon espace Privilèges

Déposer

Déposer le(s) fichier(s)

Afficher la vue détaillée

Vous pouvez déposer un fichier et/ou compléter votre dépôt à partir d'un identifiant externe.

Glissez-déposez ou cliquez pour choisir vos fichiers
Taille maximale du fichier : 200M

Chargez les métadonnées à partir d'un identifiant
Les informations associées à cet identifiant permettront de remplir automatiquement votre dépôt. Par exemple, le DOI est un code qui apparaît généralement sur la page de votre article sur le site de l'éditeur.

SWH swh:1.dir.abc0e2cbbfdfee8de52f0842263fbad65f5b211

Récupérer les métadonnées

Si vous souhaitez ajouter un embargo, récupérer des fichiers de votre espace FTP ou toute autre action avancée, veuillez afficher la vue détaillée.

SWHID sans contexte

SWHID avec contexte

3ème étape : citer le logiciel

- Citation accessible sur la notice
- Export BibTeX en utilisant le format **[BibLaTeX](#) @software** ou **@softwareversion** (si un numéro de version a été renseigné)
- Export utilisé sur le rapport d'activité scientifique d'Inria depuis 2020.

Le format citation sur HAL

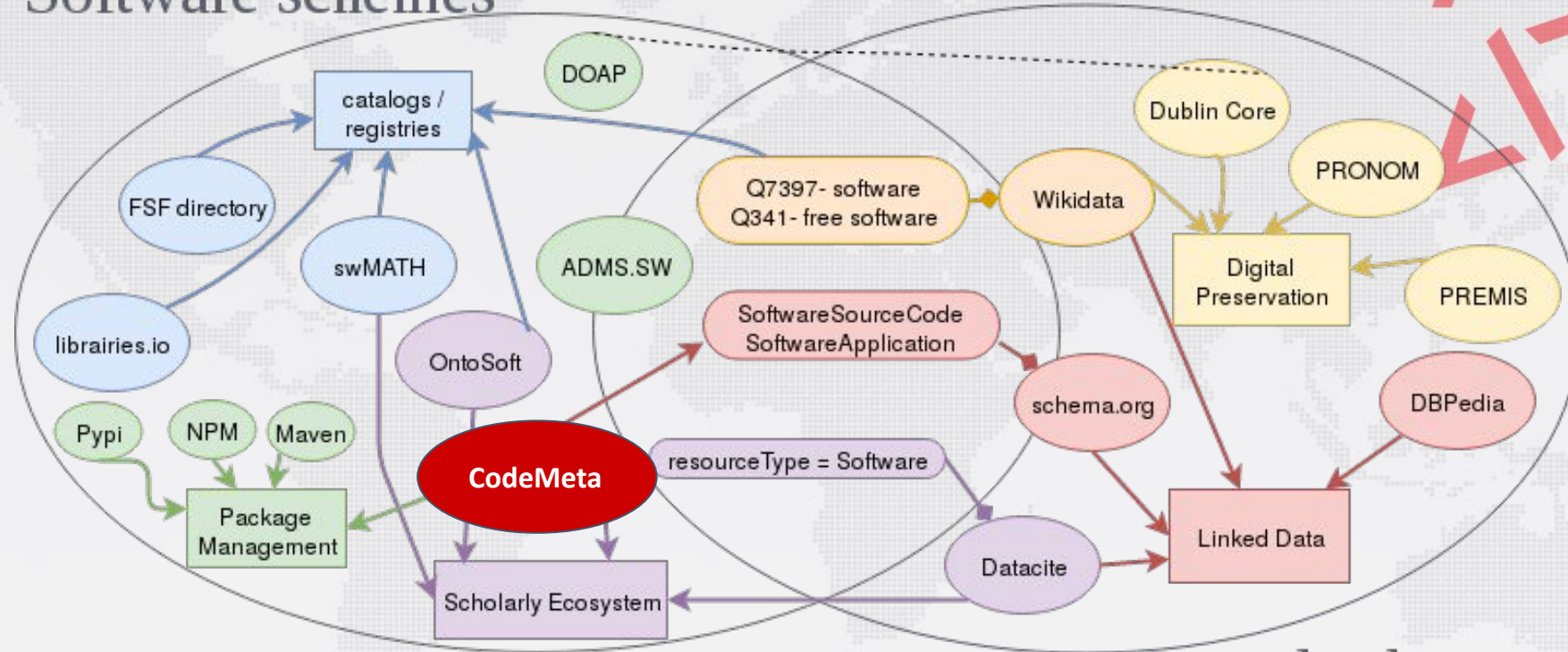
Matthieu Kowalski, Emmanuel Vincent, Rémi Gribonval. *Underdetermined Reverberant Source Separation*. 2019,

[swh:1:dir:ec4ae097465d9ea51589537ea94b2ea50e8d134d;origin=https://hal.archives-ouvertes.fr/hal-02309043;visit=swh:1:snp:e35494fd4cb57af0b22131ab8c4a4d8bd5cffcc6;anchor=swh:1:rev:2d23c3e68b755b720ecca8ddd5e1f8fe99909be2;path=/](https://hal.archives-ouvertes.fr/hal-02309043;visit=swh:1:snp:e35494fd4cb57af0b22131ab8c4a4d8bd5cffcc6;anchor=swh:1:rev:2d23c3e68b755b720ecca8ddd5e1f8fe99909be2;path=/).
[hal-02309043](#)

```
@softwareversion{kowalski:hal-02309043v1,  
  TITLE = {{Underdetermined Reverberant Source  
Separation}},  
  AUTHOR = {Kowalski, Matthieu and Vincent, Emmanuel  
and Gribonval, R{'e}mi},  
  URL =  
{https://hal.archives-ouvertes.fr/hal-02309043},  
  NOTE = {},  
  YEAR = {2019},  
  MONTH = Oct,  
  SWHID =  
{swh:1:dir:ec4ae097465d9ea51589537ea94b2ea50e8d134d;ori  
gin=https://hal.archives-ouvertes.fr/hal-02309043;vis  
it=swh:1:snp:e35494fd4cb57af0b22131ab8c4a4d8bd5cffcc6;  
anchor=swh:1:rev:2d23c3e68b755b720ecca8ddd5e1f8fe99909  
be2;path=/},  
  LICENSE = {CeCILL Free Software License Agreement  
v2.0},  
  FILE =  
{https://hal.archives-ouvertes.fr/hal-02309043/file/UR  
SS_v0.2.zip},  
  HAL_ID = {hal-02309043},  
  HAL_VERSION = {v1},  
}
```

Faciliter le dépôt des métadonnées - CodeMeta qui?

Software schemes

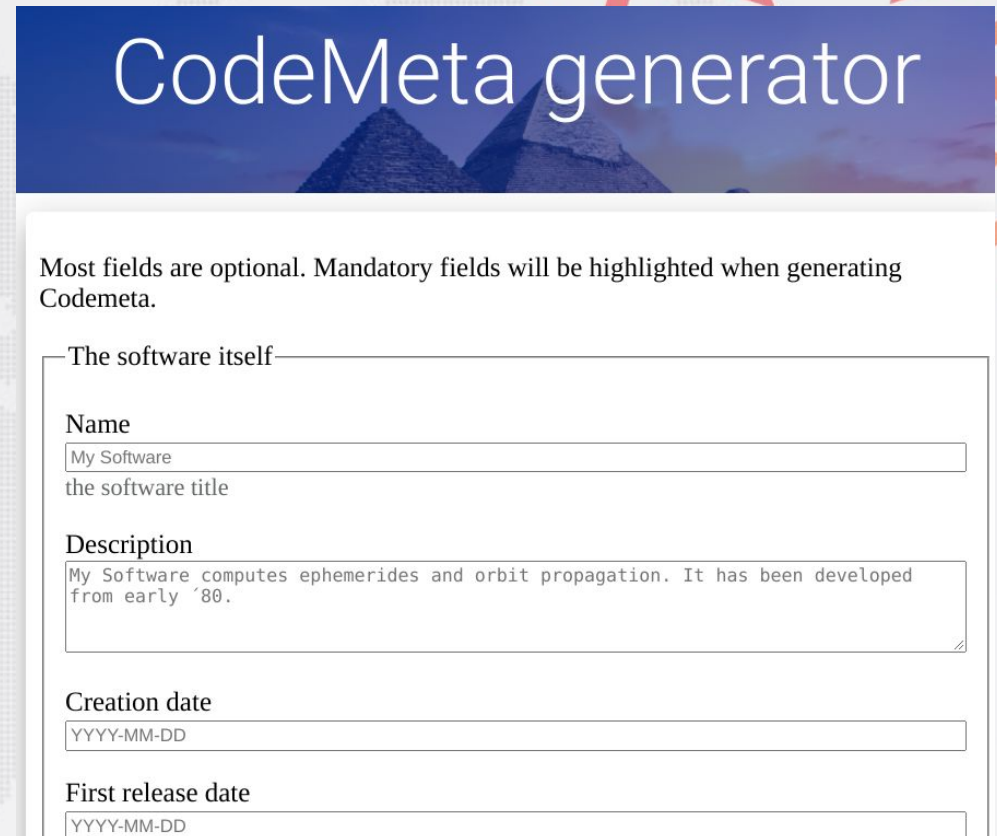


General schemes

Software ontologies landscape from Pathways for Discovery of Free Software (slide deck from LibrePlanet 2018). ([Gruenpeter & Thornton, 2018](#)) CC-by-4

Outil pour créer un codemeta.json

- Un vocabulaire qui étend le vocabulaire schema.org
- Une [communauté académique](#)
- Une [table de correspondances](#) qui permet de traduire une ontologie/vocabulaire vers CodeMeta



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

- [outil en ligne](#)
- [Pour contribuer](#)

Prochaines étapes

Format d'export

- améliorer les formats existant (BibTex, TEI, codemeta.json, etc.)

La collection

- ajouter la possibilité de déposer un module dans une collection (voir cas CGAL)

Créer dépôt à partir d'un repository

- à partir d'un SWHID
- à partir d'une URL (en passant par `Save code now`)

Intégrer le logiciel sur HAL Data

- sur <https://data.archives-ouvertes.fr/>
- avec SPARQL, en utilisant RDF



Un appel aux ambassadeurs SWH

Nous recherchons des organisations et des individus pour se porter volontaires en tant qu'ambassadeurs et ambassadrices de Software Heritage.

- ★ **Présenter** SWH et sa mission à long terme dans votre discipline
- ★ **Présenter** les avantages du dépôt logiciel dans HAL
- ★ **Partager** des informations avec des listes de diffusion
- ★ **Publier des actualités** de SWH sur les blogs et les réseaux sociaux
- ★ **Identifier les plates-formes** d'hébergement de code dans votre domaine est urgent et important d'archiver
- ★ Informer l'équipe de SWH avec **les retours de votre communauté**



Merci de votre attention!

Question?



morane@softwareheritage.org

@moraneottilia, @SWHeritage

<https://www.softwareheritage.org/newsletter/>

Références

- ❖ Y. Barborini, R. Di Cosmo, Antoine R. Dumont, M. Gruenpeter, B. Marmol, A. Monteil, J. Sadowska.. La création du nouveau type de dépôt scientifique - Le logiciel. *JSO 2018 - 7es journées Science Ouverte Couperin : 100 % open access : initiatives pour une transition réussie*, Jan 2018, Paris, France. 2018. ([hal-01688726](#))
- ❖ R. Di Cosmo, M. Gruenpeter, B. Marmol, A. Monteil, L. Romary, J. Sadowska. *Curated Archiving of Research Software Artifacts: lessons learned from the French open archive*. IJDC. 2020 ([10.2218/ijdc.v15i1.698](#)). ([hal-02475835](#))
- ❖ R. Di Cosmo, M. Gruenpeter, S. Zacchioli *Referencing Source Code Artifacts: a Separate Concern in Software Citation*, CiSE, IEEE, pp.1-9. 2020. ([10.1109/MCSE.2019.2963148](#)) ([hal-02446202](#))
- ❖ P. Alliez, R. Di Cosmo, B. Guedj, A. Girault, M.-S. Hacid, et al.. *Attributing and Referencing (Research) Software: Best Practices and Outlook from Inria*. Computing in Science and Engineering, Institute of Electrical and Electronics Engineers, 2019, pp.1-14. ([10.1109/MCSE.2019.2949413](#)). ([hal-02135891](#))