Moderate software deposit

A guide and best practices for the digital librarian v1.0

Auteurs:

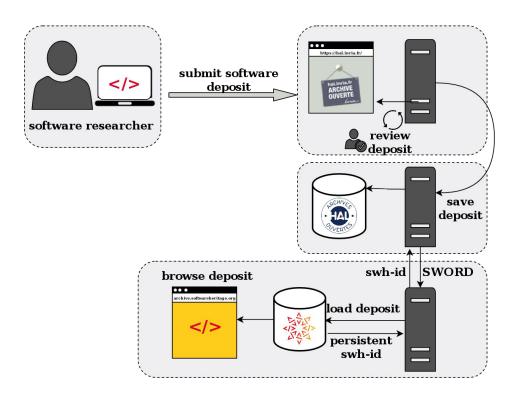
- Morane Gruenpeter morane@softwareheritage.org
- Jozefina Sadowska josephine.sadowska@inria.fr





A software deposit

- One compressed archive containing source code (mostly text files)
- A collection of metadata
 - Generic metadata
 - Software specific metadata

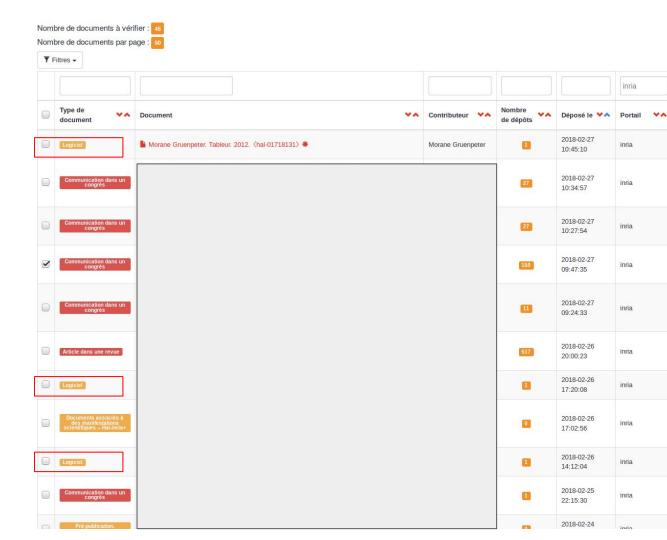


Here is the link for the user guide:

https://hal.inria.fr/hal-01872189

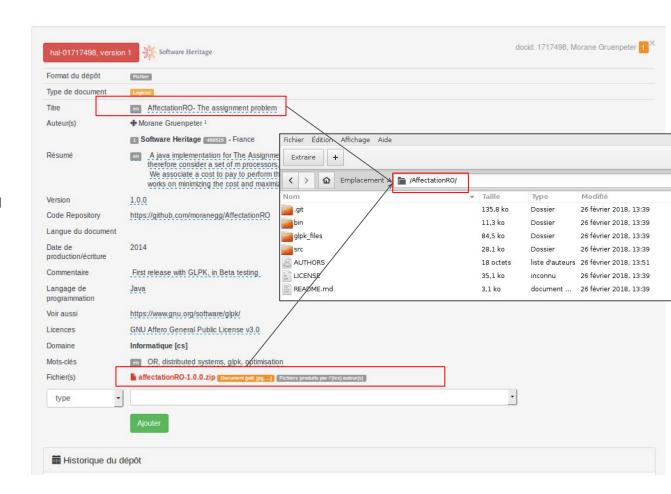
Moderate a software artifact

Choose a deposit with 'software' as the deposit type.



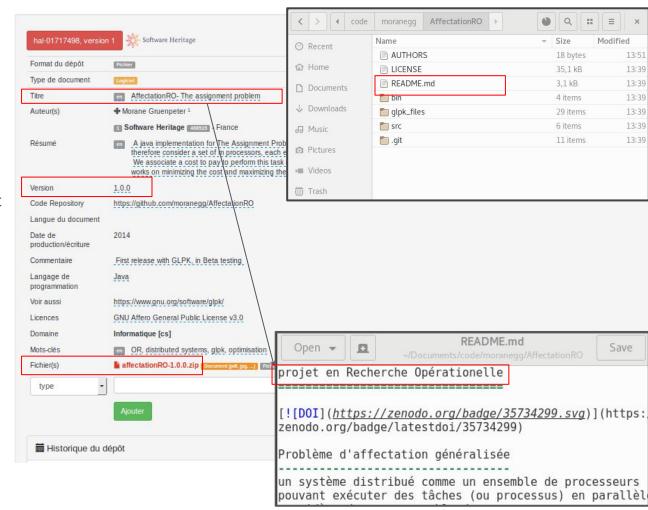
First step

- Download the archive
- Extract files
- Check that the software title matches the content, by finding the name present in one of the files or as the name of a directory:
 - If it is validate
 - If not verify with the contributor



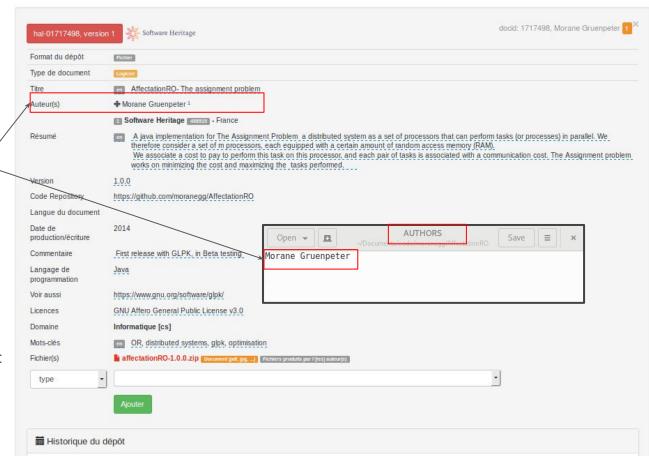
Title and description

- Check if there is README file
 - If not check with the contributor
- Open the README file in a text editor (simple-text editor, notepad, etc.)
- Check if the content is coherent with the metadata:
 - the title
 - the description
 - the version
- If it is validate



The authors

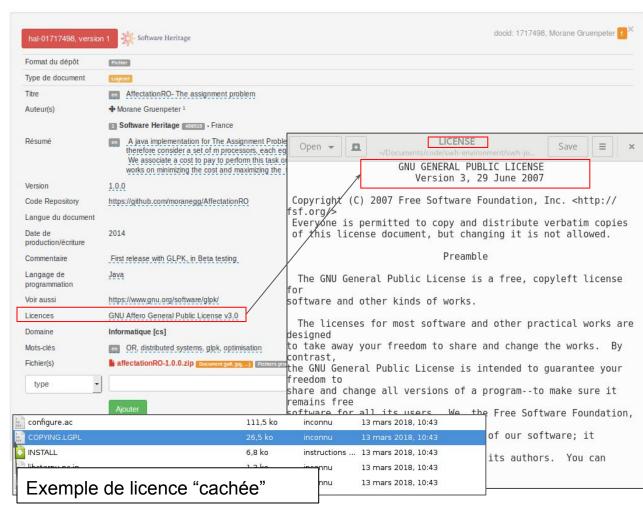
- Check if there's an:
 - AUTHOR file
 - or/and a CONTRIBUTORS
 - or/and a CREDITS file
 - or/and a CITATION file
 - Or the README contains all authors by name.
 - if there is validate
- If there is an AUTHOR file without one of the authors in the metadata:
 - check with the contributor
- If there is an AUTHOR file with names that are not in the metadata:
 - check with the contributor



The license/s

- Check there is at least one license in the metadata
- Check for a LICENSE file (COPYING, or any other name for a file containing the license).
- Check that the license/s in the metadata matches the license/s in the content
 - If it is validate
 - If not Check with the contributor

N.B. In HAL we accept only content with an open license. For your information here is the spdx reference list of open source licenses.



Citation file or codemeta.json file

- Check for a file containing the software citation or any other metadata files to verify the contributed metadata;
 - CITATION
 - CITATION.cff
 - codemeta.json



Fichier de citation utilisé pour le logiciel doi2cff:

https://github.com/citation-file-format/doi2cff/blob/master/CITATION.cff

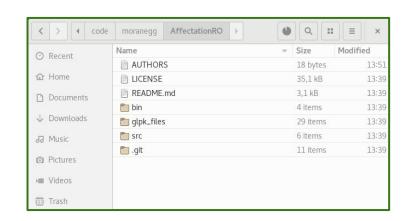
The content

- Check that the content is not aberrant:
 - If it's a zip containing another zip
 - Or If one file's size is greater than
 1GB and if the file extension is part of this list
 - Check with the contributor

You should know that all downloaded files may contain malware, even pdf files. To protect yourself:

- Open the files only with up-to-date tools and try opening only text files with the text editor.
- Never execute code explicitly (do not open files ending with .exe)





Film

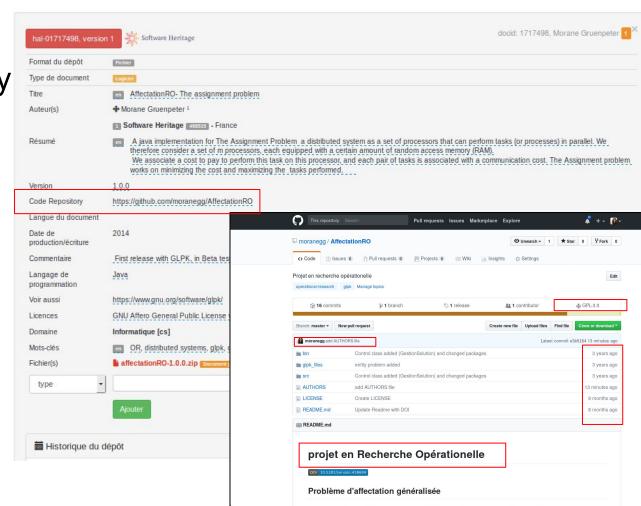
Name ▼	Description	Size	Modified
La.Classe.americaine.1993[2010	Matroska video	2.6 GiB	02/23/2018 11:34
La.Classe.americaine.1993[2010	SubRip subtitles	88.6 KiB	02/23/2018 11:27
README.nfo.txt	plain text document	10.0 KiB	02/23/2018 11:27

The code repository

- Visit the code repository associated with a deposit and verify that it is the same software:
 - title,
 - actors,
 - general coherence

There can be some differences that can be accepted:

- Pseudonymes for the authors names
- The content can be different than the deposit (The developpement may have evolved since the deposit).



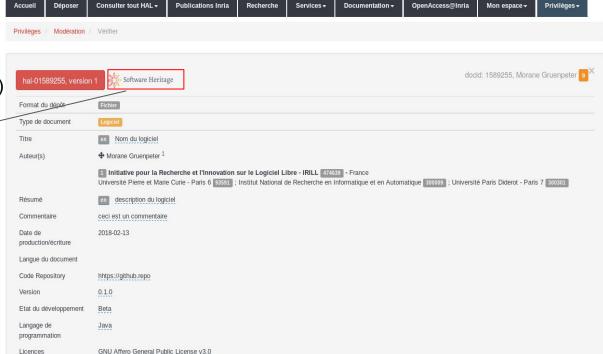






Preserve the deposit in the Software Heritage archive (SWH)

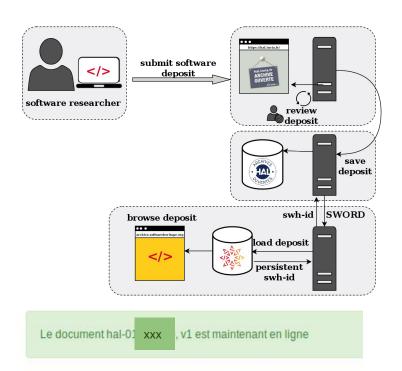
- If the contributor accepts the software transfer to SWH it will appear with this logo.
- After the deposit was validated the transfer can take a few minutes (it depends on the size of the deposit and other technical factors)
- Do not leave the page.
- If an error occurs, contact the CCSD technical service.



The transfer to SWH-How it works?

The deposit is pushed to SWH via SWORD protocol the same way as ArXiv. The deposit can be:

- 1. Accepted by SWH \rightarrow
- - It disappears from the deposits list and it is published on HAL
 - A few hours later the swh-id is integrated to the published depos
- 2. Refused by SWH \rightarrow
 - It stays in the deposits list
 - Write a comment that the deposit shouldn't be re-moderated
 - Contact the CCSD and SWH technical service



SWORD SWH response: rejected https://deposit.softwareheritage.org/1/hal/147/status/

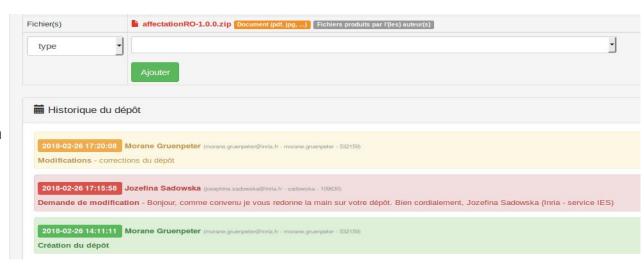
SWORD SWH response: Max recursion time...

The deposit's moderation history

In the deposit's moderation history the following messages can be found:

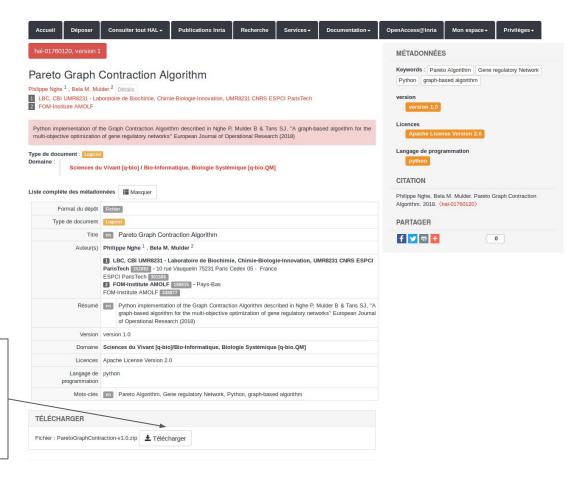
- Interaction between the moderator and the contributor
- Messages about the moderation process
- Error messages received from SWH.

Please read carefully the history before validating a deposit.

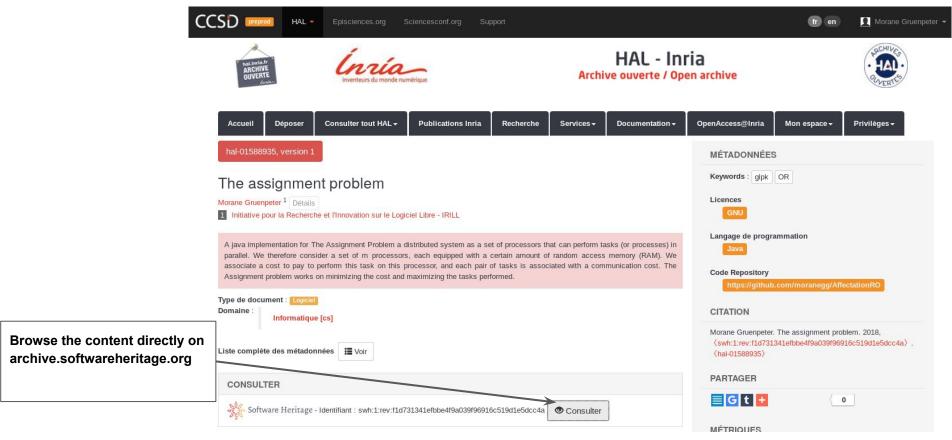


The published deposit before the swh-id integration

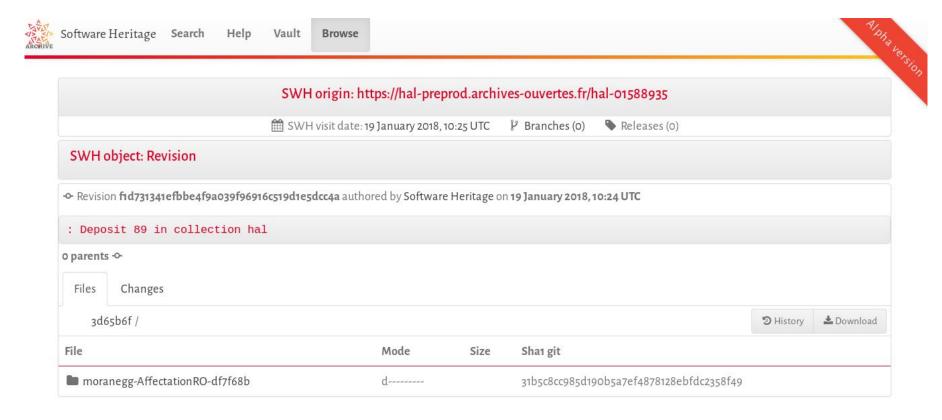
It can take a few hours or even a day before the swh-id is available on the published deposit.



The finalized software artifact deposit on hal.inria.fr



The software artifact on archive.softwareheritage.org



Particular points of vigilance

Authors



Scenario: the AUTHOR* file contains 36 names, but the metadata has only 3 authors (*AUTHORS, CREDITS, CITATION files)

- Verify the situation with the contributor
- Propose to the contributor two separated lists or two separated files for:
 - authors
 - contributors

Scenario: The metadata indicates the software team or the software project as the author and the AUTHORS file captures a complete list of authors:

 Verify the situation with the contributor that all authors are aware and acknowledge that they are not in the metadata.

```
Simon Archipoff <simon.archipoff@etu.u-bordeaux1.fr>
Cédric Augonnet <cedric.augonnet@inria.fr>
Olivier Aumage <olivier.aumage@inria.fr>
William Braik <wbraik@gmail.com>
Berenger Bramas <berenger.bramas@inria.fr>
Alfredo Buttari <alfredo.buttari@enseeiht.fr>
Adrien Cassagne <adrien.cassagne@inria.fr>
Jérôme Clet-Ortega <jerome.clet-ortega@labri.fr>
Nicolas Collin <nicolas.collin@inria.fr>
Ludovic Courtès < ludovic.courtes@inria.fr>
Yann Courtois <vann.courtois33@gmail.com>
Jean-Marie Couteven < jm.couteven@gmail.com>
Nathalie Furmento <nathalie.furmento@labri.fr>
David Gómez <david gomez1380@yahoo.com.mx>
Sylvain Henry <sylvain.henry@inria.fr>
Andra Hugo <andra.hugo@inria.fr>
Mehdi Juhoor <mjuhoor@gmail.com>
Xavier Lacoste <xavier.lacoste@inria.fr>
Benoît Lizé <benoit.lize@gmail.com>
Antoine Lucas <antoine.lucas.33@gmail.com>
Brice Mortier <bri>drice.mortier@etu.u-bordeaux1.fr>
Stojce Nakov <stojce.nakov@inria.fr>
Joris Pablo <joris.pablo@orange.fr>
Damien Pasqualinotto <dam.pasqualinotto@wanadoo.fr>
Nguyen Quôc-Dinh <nguyen.quocdinh@gmail.com>
Cyril Roelandt <cyril.roelandt@inria.fr>
Anthony Roy <theanthony33@gmail.com>
Corentin Salingue <corentin.salingue@gmail.com>
Marc Sergent <marc.sergent@inria.fr>
Anthony Simonet <anthony.simonet@etu.u-bordeaux.fr>
Luka Stanisic < luka.stanisic@imag.fr>
Ludovic Stordeur < ludovic.stordeur@inria.fr>
Guillaume Sylvand <quillaume.sylvand@airbus.com>
François Tessier <françois.tessier@inria.fr>
Samuel Thibault <samuel.thibault@labri.fr>
Pierre-André Wacrenier <wacrenier@labri.fr>
```

The license

Scenario: a LICENSE, LICENCE.txt, LICENSE.txt, LICENCE.md, LICENSE.md, Licence.txt or COPYRIGHT, COPYING etc. is included but do not match the license in the metadata

Do you validate?

Verify the situation with the contributor

Software: Caml Dim, version of february 1998, hereinafter referred to as "the software".

The software has been designed and produced by Bruno Blanchet at the Ecole Normale Supérieure - 45 rue d'Ulm - 75005 PARIS - France, from Caml Light 0.74.

It is distributed under the same license as Caml Light 0.74, which legal notice follows.

.....

CAML LIGHT 0.74 LEGAL NOTICE

Software: Caml Light, version 0.74 of November 1997, hereinafter referred to as "the software".

The software has been designed and produced by Xavier Leroy,
Damien Doligez, Francois Rouaix, Jerome Vouillon and Pierre Weis,
research workers for the Institut National de Recherche en Informatique et
en Automatique (INRIA) - Domaine de Voluceau - Rocquencourt - 78153 Le
Chesnay Cedex - France.

INRIA holds all ownership rights to Caml Light version 0.74.

The software has been registered at Agence pour la Protection des Programmes (APP).

Preamble:

The content- Only one zipped file

Only one zipped file is expected in a software deposit

Scenario: There are more files in the deposit.

Do you validate?

- Verify the situation with the contributor what happened and if they can deposit only one zipped file.
- If the problem persist, verify with the CCSD technical support.



Admin view of all software deposits on SWH (private)

