The creation of a new type of scientific deposit: Software

Yannick Barborini, Roberto Di Cosmo, Antoine R. Dumont, Morane Gruenpeter, Bruno Marmol, Alain Monteil, Jozefina Sadowska, Stefano Zacchiroli

The creation of a new type of scientific deposit:

Software

CCSD[1], HAL-Inria[2], Software Heritage[3]

Y. Barborini[1], R. Di Cosmo[3], A.R. Dumont[3], M. Gruenpeter[3], B. Marmol[1], A. Monteil[2], J. Sadowska[2], S. Zacchiroli[3]

Software preservation: a scientific challenge

Software has become an indissociable support of **technical and scientific knowledge**. The preservation of this universal body of knowledge has become as essential as preserving research articles and data sets. Software preservation is a pillar of **reproducibility**.

In the quest for making scientific results reproducible, and pass the knowledge over to future generations, the three main pillars are: **scientific articles**, that describe the results, the **data sets** used or produced, and the **software** that embodies the logic of the data transformation[1].

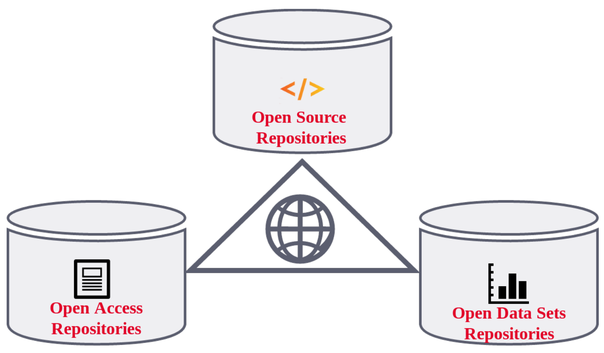

Figure 1: The pillars of knowledge preservation

Software deposit

The collaboration between **Software Heritage (SWH), Hal-Inria and the CCSD** has resulted with a new type of scientific deposit in the national open archive.
Researchers have now the possibility to deposit *software* source code on Hal-Inria.

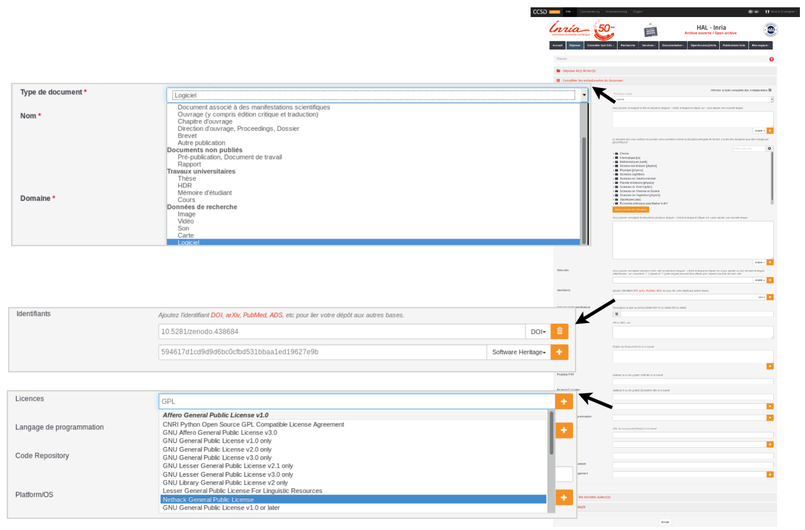

Figure 2: The form dedicated to software deposits

The steps for a software deposit:

- deposit a source code archive (.zip)
- choose deposit type: *software*
- add associated metadata
- add the software authors
- accept the archival of the deposit on SWH

The actors

Software Heritage took the challenge to collect, preserve and share all software that is publicly available in source code form. **Hal-Inria** is the open archive of Inria- The French Institute for Research in Computer Science and Automation. Hal-Inria provides, since 2005, access to the Hal platform, developed by the **CCSD**- The Center for Direct Scientific Communication. Its main mission is to provide tools, in the respect of open access principles, for archiving and dissemination of scientific publications and data.

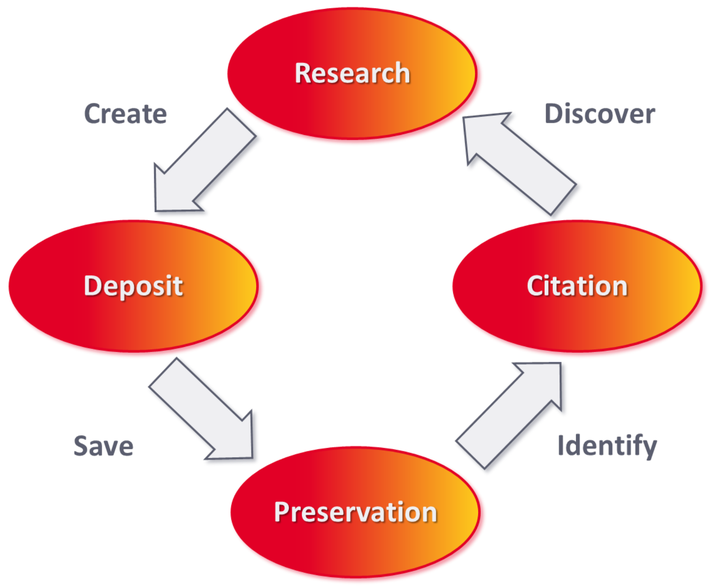

Figure 3: The life cycle of research software

The descriptive metadata

To ensure an accurate description of the software, different metadata are available on the deposit form and are preserved with the software in the SWH archive. An example:

Provided by the system:	***MUST:***
- Hal identifier	- title
- publication date	- description
- swh-id	- authors
SHOULD:	***MAY:***
- license	- dependencies
- keywords	- platform/OS
- repository	- funding

The intrinsic and persistent identifier

To be able to reproduce an experiment, knowing the exact version of the software used is essential. Software Heritage will provide the *swh-id*, intrinsically bound to software components, ensuring persistent traceability across future development and organizational changes. The *swh-id*, like a fingerprint of the Software is specific, persistent and unique. It does not depend on an ID resolver.

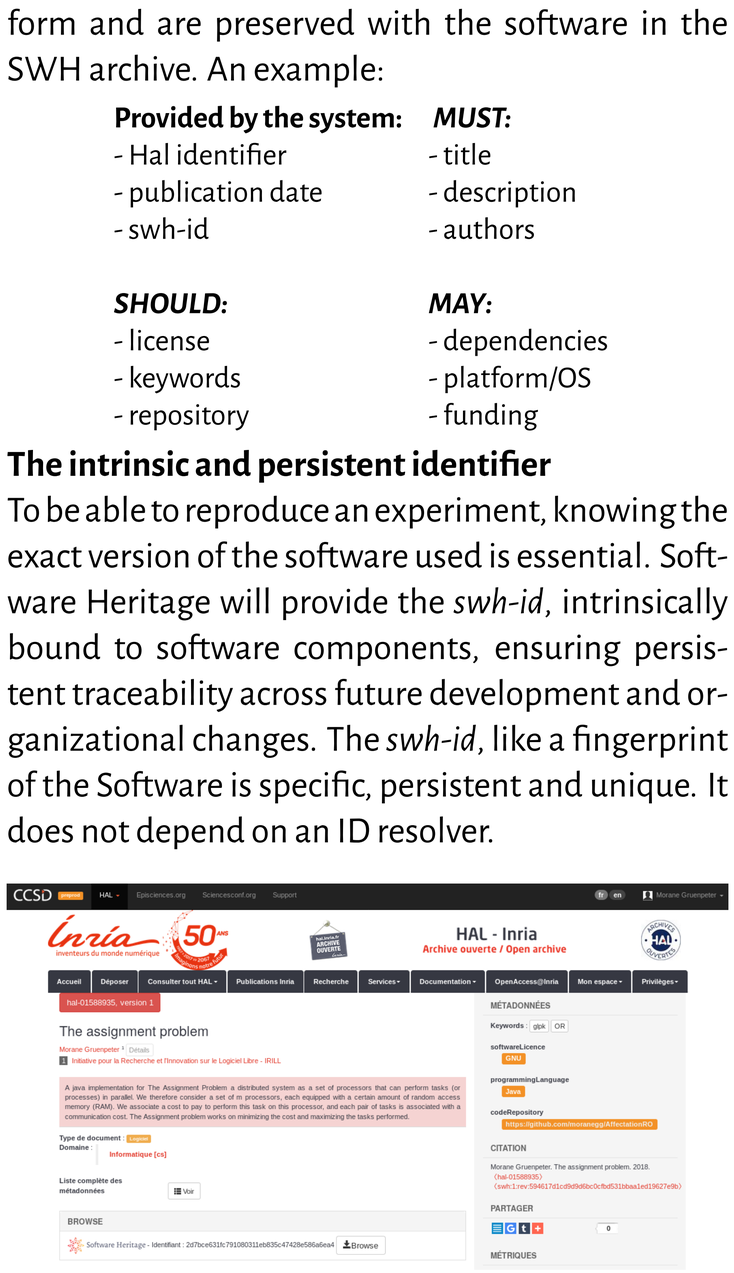

Figure 4: The deposit on Hal-Inria

Transfer deposit to SWH

Once the deposit is validated, it is pushed to SWH using SWORD protocol. SWH will proceed with the injection of the source code into **Alexandria's Library of Software** and will generate the intrinsic identifier - the *swh-id*. Hal retrieves the *swh-id* to use in the citation format.

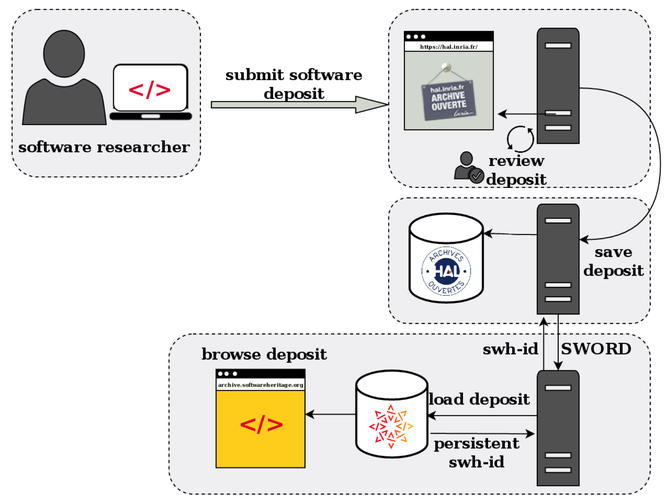

Figure 5: The software deposit workflow

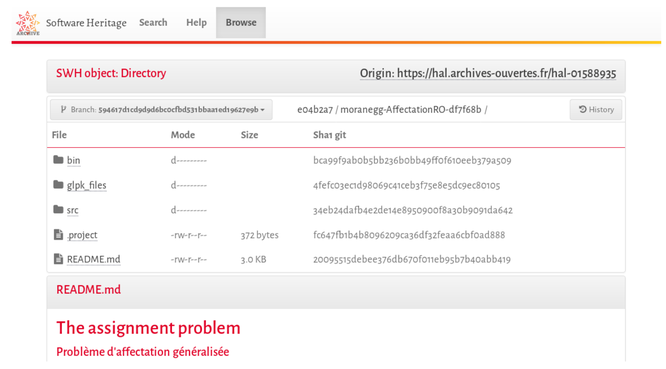

Figure 6: Browse the deposit on Software Heritage

Software citation

Following the software citation principles[2] and thus considering that software is a legitimate and citable product of research, we have proposed a citation format containing metadata submitted with the software.

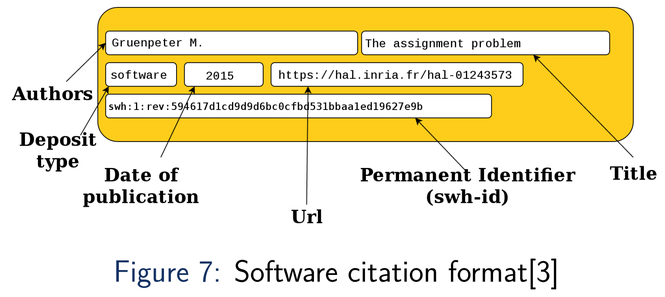

Figure 7: Software citation format[3]

Citation is essential for promoting the recognition of software as a valuable research output, and ensuring that the authors have their contributions recognised and rewarded[4].

Références

1.Roberto Di Cosmo, Stefano Zacchiroli (2017) Software Heritage: Why and How to Preserve Software Source Code. iPRES 2017. https://hal.archives-ouvertes.fr/hal-01590958

2.Smith et al. (2016), Software citation principles. PeerJ Comput. Sci. 2:e86; DOI 10.7717/peerj-cs.862.

3.Yolanda Gil (2015) Documenting Software through Metadata. Geosoft.

4.Mike Jackson (2014) How to cite and describe software. The Software Sustainability Institute https://www.software.ac.uk/how-cite-and-describe-software

Software Heritage sponsors : Microsoft, Societe Generale, intel, Huawei, Data Archiving and Networked Services DANS, GitHub, Alma Mater Studiorum A.D. 1088, Nokia Bell Labs

Contacts :

hal.inria.fr
josephine.sadowska@inria.fr

hal.archives-ouvertes.fr
ccsd@ccsd.cnrs.fr

www.softwareheritage.org
morane@softwareheritage.org