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

To cite this version:
Yannick Barborini, Roberto Di Cosmo, Antoine R. Dumont, Morane Gruenpeter, Bruno Marmol, et al.. The creation of a new type of scientific deposit: Software. RDA Eleventh Plenary Meeting, Berlin, Germany, Mar 2018, Berlin, Germany. 2018. hal-01738741

HAL Id: hal-01738741
https://hal.inria.fr/hal-01738741
Submitted on 20 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
The creation of a new type of scientific deposit: **Software**
CCSD¹, HAL-Inria², Software Heritage³

Y. Barborini¹, R. Di Cosmo³, A.R. Dumont³, M. Gruenpeter³, B. Marmol¹, A. Monteil², J. Sadowska², S. Zacchiroli³

**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].

**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.

**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.

**Transfer deposit to SWH**

Once the deposit is validated, it is pushed to SWH using SWORD protocol. SWH will proceed with the inclusion 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.

**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.

**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:

<table>
<thead>
<tr>
<th>Provided by the system</th>
<th>MUST:</th>
<th></th>
<th>SHOULD:</th>
<th>MAY:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Hal identifier</td>
<td>title</td>
<td>- license</td>
<td>dependencies</td>
<td></td>
</tr>
<tr>
<td>- publication date</td>
<td>description</td>
<td>- keywords</td>
<td>platform/OS</td>
<td></td>
</tr>
<tr>
<td>- swh-id</td>
<td>authors</td>
<td>- repository</td>
<td>funding</td>
<td></td>
</tr>
</tbody>
</table>

**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.

**Citation**

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**

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