

Software Heritage

The universal source code archive

Morane Guenpeter

Metadata specialist, Software Heritage
Crossminer, University of L'Aquila

`morane@softwareheritage.org`

September 20th, 2018



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

- 
- 1 Software Heritage
 - 2 Relevance for research software publishing



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all

Reference catalog



find and reference **all** the
source code



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all

Reference catalog



find and reference **all** the
source code

Universal archive



preserve **all** the source
code



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all

Reference catalog



find and reference **all** the
source code

Universal archive



preserve **all** the source
code

Research infrastructure



enable analysis of **all** the
source code

Cultural Heritage



Industry



Research



Education



Software Heritage



Cultural Heritage



Industry



Research



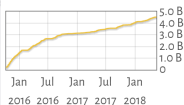
Education



Software Heritage

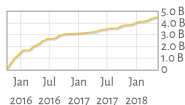
Source files

4,536,067,027



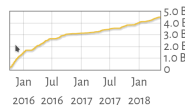
Commits

1,024,675,748



Projects

83,801,775



Cultural Heritage



Industry



Research



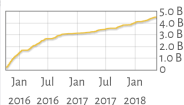
Education



Software Heritage

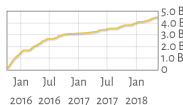
Source files

4,536,067,027



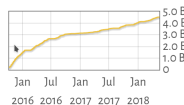
Commits

1,024,675,748



Projects

83,801,775



Open approach

- open source
- transparency

In for the long haul

- non profit, **mirrors**
- **intrinsic** identifiers

Exhaustive

- **all** software
- open to **all** communities

Growing Support

Raising awareness: landmark agreement, 3/4/2017; grand opening, 7/6/2018



Sharing the vision



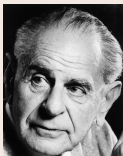
Sponsoring our work



- 
- 1 Software Heritage
 - 2 Relevance for research software publishing

How we built our scientific knowledge

Reproducibility is the key

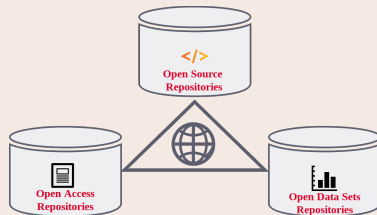


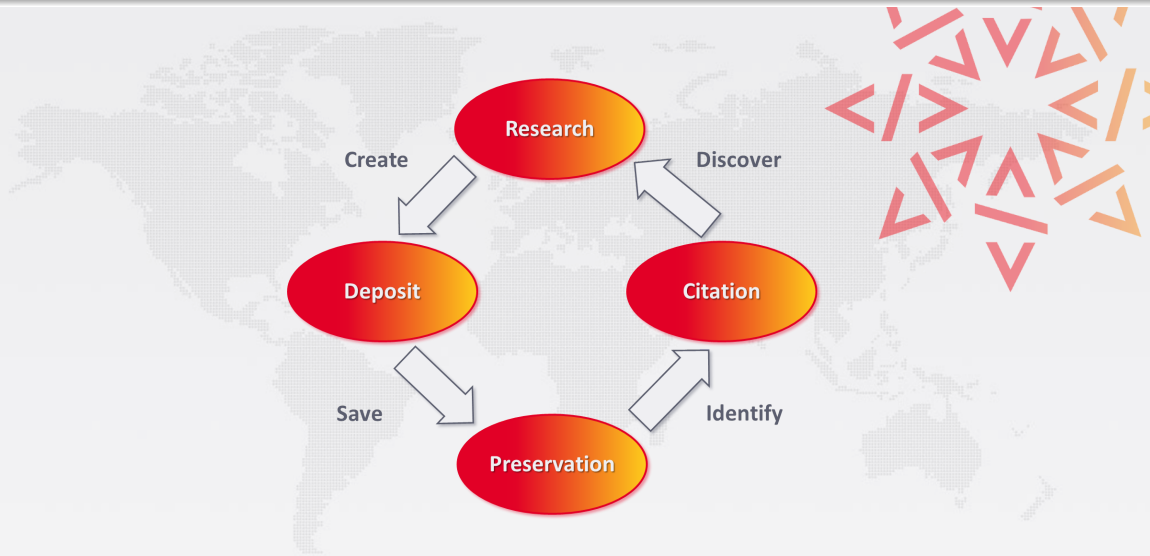
non-reproducible single occurrences are of no significance to science

Karl Popper, The Logic of Scientific Discovery, 1934

the software source code is special

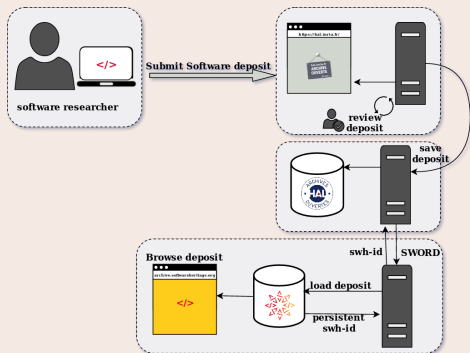
- It **embodies the logic** of the data transformation.
- It **must** be included in strategies for *scientific knowledge preservation*.
- knowing the **exact version** of the software used is essential for reproducibility.





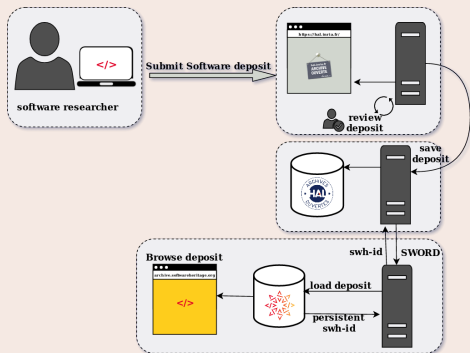
The research software (deposit) use case

Deposit software in HAL



The research software (deposit) use case

Deposit software in HAL

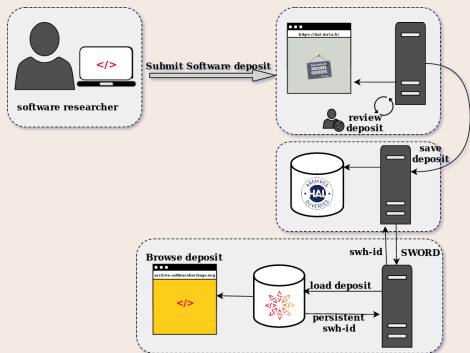


Generic mechanism:

- SWORD based
- review process
- versioning

The research software (deposit) use case

Deposit software in HAL



Generic mechanism:

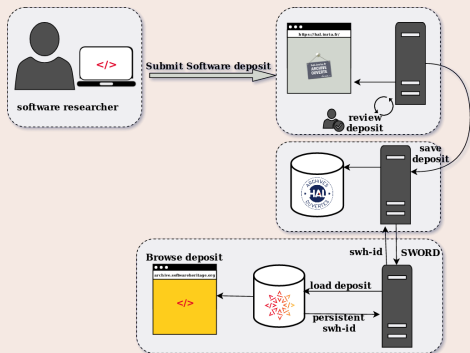
- SWORD based
- review process
- versioning

How to do it:

- **today:** deposit .zip or .tar.gz file
- **tomorrow:**
 - provide *SWH id* and metadata
 - include *metadata file* for automatic metadata extraction
 - ...

The research software (deposit) use case

Deposit software in HAL



Generic mechanism:

- SWORD based
- review process
- versioning

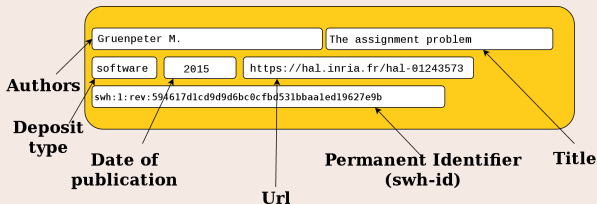
How to do it:

- **today:** deposit .zip or .tar.gz file
- **tomorrow:**
 - provide *SWH id* and metadata
 - include *metadata file* for automatic metadata extraction
 - ...

September 2018: **open to all** on <https://hal.archives-ouvertes.fr/>

Citation is essential

- for promoting the recognition of software as a valuable research output
- for giving/recieving credit for software publication
- for reusage and reproducibiity



Key dates of the deposit feature

Deposit collaboration

2017 deposit specification and implementation on HAL and SWH

January 2018 Poster presentation at the *JSO2018*

Mars 2018 hal.inria.fr accepts software deposits (testing phase)
Poster presentation at the *RDA 11th plenary*


September 2018 HAL opens the software deposit doors on all portals

Institutional & Operational adoption

June 2018 swMath.org points into SWH for the source code

ongoing OpenAire detects source code links in articles, resolves them to SWH

July 2018 Software Heritage is part of the french National Plan for Open Science



Demo Time!

- example deposits in HAL
- example use of `https://archive.softwareheritage.org`