

Software as a Pillar of Open Science

Roberto Di Cosmo

Director, Software Heritage
Chair, Software Chapter, CoSO
Inria and Université de Paris Cité

October 1st, 2025



Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

1 Software as a pillar of Open Science

2 Meet Software Heritage

3 Conclusion

Software *Source Code* is Precious Knowledge

Harold Abelson, *Structure and Interpretation of Computer Programs* (1st ed.) 1985

“Programs must be written for people to read, and only incidentally for machines to execute.”

Apollo 11 source code (excerpt)

```
P63SPOT3      CA      BIT6      # IS THE LR ANTENNA IN POSITION 1 YET
              EXTEND
              RAND     CHAN33
              EXTEND
              BZF      P63SPOT4      # BRANCH IF ANTENNA ALREADY IN POSITION 1

              CAF      CODE500      # ASTRONAUT: PLEASE CRANK THE
              TC       BANKCALL      # SILLY THING AROUND
              CADR     GOPERF1
              TCF      GOTOP00H      # TERMINATE
              TCF      P63SPOT3      # PROCEED SEE IF HE'S LYING

P63SPOT4      TC       BANKCALL      # ENTER INITIALIZE LANDING RADAR
              CADR     SETPOS1

              TC       POSTJUMP      # OFF TO SEE THE WIZARD ...
              CADR     BURNBABY
```

Covid Sim (excerpt)

```
/**
 * @brief The basic unit of the simulation and is associated to a geographical location.
 *
 * Interventions (e.g., school closures) are tracked at this level. It contains a list of its
 * members (people), places (schools, universities, workplaces etc.), road networks, links to
 * airports etc.
 */
struct Microcell
{
    /* Note use of short int here limits max run time to USHRT_MAX*ModelTimeStep - e.g. 65536*0.25=16384 days=44 yrs.
     * Global search and replace of 'unsigned short int' with 'int' would remove this limit, but use more memory.
     */

    int n; // Number of people in microcell
    int adunit; // admin unit microcell belongs to
    int* members; // array of members/hosts of microcell

    int* places[MAX_NUM_PLACE_TYPES]; // list of places (of various place types) within microcell
    unsigned short int NumPlacesByType[MAX_NUM_PLACE_TYPES]; // number of places (of various place types) within microcell
    unsigned short int keyworkerproph, move_trig, place_trig, socdist_trig, keyworkerproph_trig;
    unsigned short int move_start_time, move_end_time;
    unsigned short int place_end_time, socdist_end_time, keyworkerproph_end_time;
    TreatStat moverest, treat, vacc, socdist, placeclose;
    unsigned short int treat_trig, vacc_trig;
    unsigned short int treat_start_time, treat_end_time;
    unsigned short int vacc_start_time;
    IndexList* AirportList;
};
```

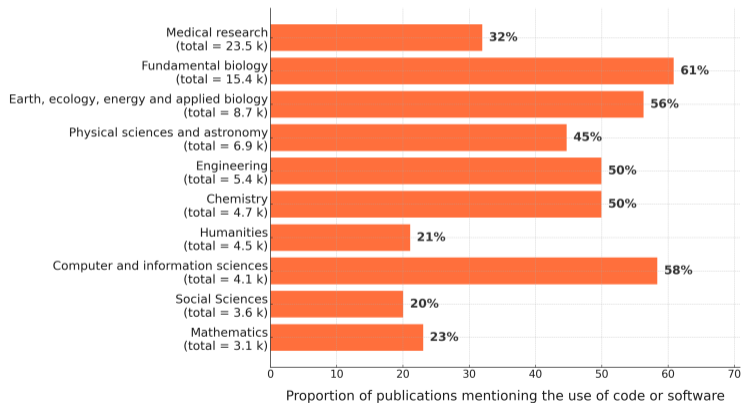
Len Shustek, *Computer History Museum*

2006

“Source code provides a view into the mind of the designer.”

(Open) Source Code is a pillar of Open Science

Software powers modern research



French Open Science
Monitor 2025



From national to global:
Open Science
Monitoring Framework
Initiative



**MINISTÈRE
CHARGÉ
DE L'ENSEIGNEMENT
SUPÉRIEUR
ET DE LA RECHERCHE**

*Liberté
Égalité
Fraternité*



A bird's eye view of arising policies: reproducibility, open science, career

Publishers

"share the software associated to your articles" mandates

Artifact Evaluation Committees

"review the software artifacts associated to articles" initiatives (ACM badges, etc.)

Funders

"software developed under grant XXXX must be made available under an open license"

Institutions

"software in CV" (e.g. DFG model CV) and *"in lab/institute/organization reports"*

Communities

"software must be cited and credited"

Key issues: how to Archive, Reference, Describe, Cite and Credit (research) software?

we need a dedicated infrastructure... now we have one!

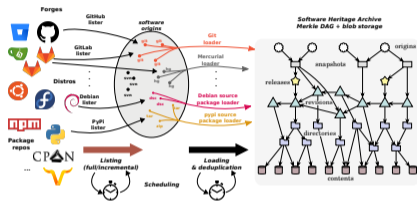
1 Software as a pillar of Open Science

2 Meet Software Heritage

3 Conclusion

How Software Heritage addresses Open Science's needs

Archive (25B+ files, 400M+ projects)



- save now, updateswh, webhooks
- deposit.softwareheritage.org

Describe

- *Intrinsic metadata* from source code
- Contributed the [Codemeta generator](#)

Reference (50 billion SWHIDs)

Intrinsic, cryptographically strong IDs



Now in SPDX 2.2, Wikidata
<https://swhid.org> - ISO/IEC 18670

Cite/Credit

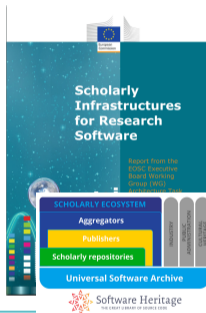
- Contributed [biblatex-software style](#)
- Software Citation from the archive!

- Browse + Reference [DIS 18670] (Apollo 11 [excerpt], your work may be already there !)
- Trigger archival, use the [updateswh](#) browser extension, configure the webhooks
- Cite [from the archive](#) with [biblatex-software](#) (CTAN, [ACMART](#))
- Describe with Codemeta (use [codemeta generator](#))
- Curated deposit in SWH via HAL, see for example: [LinBox](#), [SLALOM](#), [Givaro](#), [NS2DDV](#), [SumGra](#), [Coq proof](#), ...
- Extracting all the software products for Inria, for CNRS, for CNES, for LIRMM or for Rémi Gribonval using [HalTools](#)
- Example with Parmap: [devel on Github](#), [archive in SWH](#), [curated deposit in HAL](#)
- Example research articles:
 - compare Fig. 1 and conclusions in [the 2012 version](#) and [the updated version](#)
 - SWHID in [a replication experiment](#)

A few adoption indicators



Policy



- [Recommendations in ANR 2023 guidelines \(p. 17\)](#)
- HAL+SWH in [the Open Science software booklet](#)

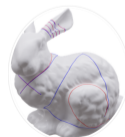
Users and collaborations



What are they “referencing”?

source	n	percentage
Not available	2868	46.22
GitHub	1151	18.55
software heritage	387	6.24
zenodo	142	2.29
r package	70	1.13
cran	56	0.90
r package version	54	0.87
gitlab	35	0.56

Graphics Replicability Stamp Initiative



b/Surf: Interactive Bézier Splines on Surface Meshes

Claudio Mancinelli, Giacomo Nazzaro, Fabio Pellacini, Enrico Puppo
IEEE Transactions on Visualization and Computer Graphics (TVCG)



Repository



Projects



FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC

The CodeMeta Project



FAIR-IMPACT
Expanding FAIR solutions across EOSC

- 1 Software as a pillar of Open Science
- 2 Meet Software Heritage
- 3 Conclusion

Call to action: adopt and share best practices ... today!

Archive and reference

All **source code** used in research (*yes, even small scripts!*)

for reproducibility

- save in Software Heritage
- add **SWHID** in articles

See [detailed HOWTO online](#)



Describe and Cite/Credit

For **software you want to put forward** (*mention in your CV, reports, etc., get citations and credit for it*), do the following **extra steps**:

video tutorials

- add **codemeta.json** (see the [codemeta generator](#))
- reference in HAL (*french partners*, see [online HAL documentation](#))
- cite using the [biblatex-software](#) package (in CTAN and TeXLive)



-  R. Di Cosmo, S. Granger, K. Hinsen, N. Jullien, D. Le Berre, V. Louvet, et al.
Stop treating code like an afterthought: record, share and value it.
Nature, 646(8084):284–286, October 2025.
-  L. Courtès, T. Sample, S. Tournier, S. Zacchioli
Source Code Archiving to the Rescue of Reproducible Deployment
ACM Conference on Reproducibility and Replicability, 2024.
-  R. Di Cosmo, M. Gruenpeter, B. Marmol, et al.
Curated Archiving of Research Software Artifacts: Lessons Learned from the French Open Archive
International Journal of Digital Curation, 2020.
<https://doi.org/10.2218/ijdc.v15i1.698>
-  R. Di Cosmo, J. B. G. Lopez, J.-F. Abramatic, et al.
Scholarly Infrastructures for Research Software
EOSC report, 2020. <https://data.europa.eu/doi/10.2777/28598>
-  J.-F. Abramatic, R. Di Cosmo, S. Zacchioli
Building the Universal Archive of Source Code
Communications of the ACM, 2018. <https://cacm.acm.org/magazines/2018/10/231366-building-the-universal-archive-of-source-code/fulltext>