

Software as a Pillar of Open Science

Roberto Di Cosmo

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

October 29, 2025



Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

- 1 Software as a pillar of Open Science
- 2 Software Heritage: supporting Software in Open Science
- 3 Preview from the Software Heritage very large telescope
- 4 The road ahead

Short Bio: Roberto Di Cosmo

Computer Science professor in Paris, now working at INRIA

- 35+ years of research (Theor. CS, Programming, Software Engineering, Erdos #: 3)
- 25+ years of Free and Open Source Software
- 15+ years building and directing structures for the common good



1999 *DemoLinux* – first live GNU/Linux distro

2007 *Free Software Thematic Group*

150 members 40 projects 200Me

2008 *Mancoosi project* www.mancoosi.org

2010 *IRILL* www.irill.org

2015 *Software Heritage* at INRIA

2018 *National Committee for Open Science*, France

2021 *EOSC Task Force on Infrastructures for Software*,
European Union

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)

```
P63SP0T3      CA      BIT6      # IS THE LR ANTENNA IN POSITION 1 YET
              EXTEND
              RAND      CHAN33
              EXTEND
              BZF      P63SP0T4      # BRANCH IF ANTENNA ALREADY IN POSITION 1

              CAF      CODE500      # ASTRONAUT: PLEASE CRANK THE
              TC      BANKCALL      # SILLY THING AROUND
              CADR      GOPERF1
              TCF      GOTOPOOH      # TERMINATE
              TCF      P63SP0T3      # PROCEED SEE IF HE'S LYING

P63SP0T4      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;
};
```

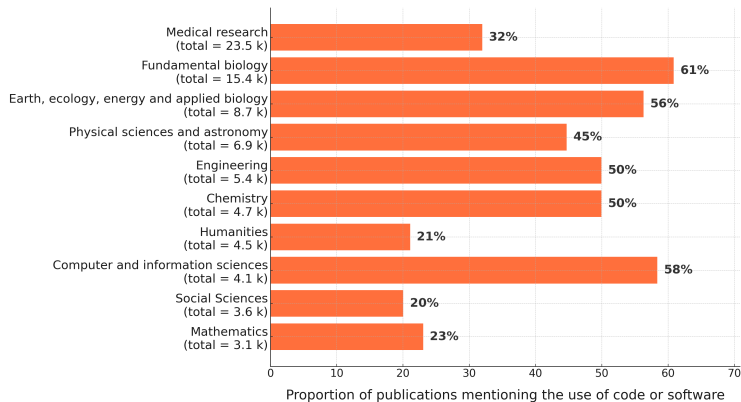
Len Shustek, Computer History Museum

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*



nature
COMMENT | 07 October 2025

Stop treating code like an afterthought: record, share and value it

Scientists, research institutions, funders, libraries and publishers must all improve software practices.



Software Heritage



Software and Source Codes College

Software & source codes

Software has now become essential in all areas of scientific research, both as a research tool, as a research product and as a research object.

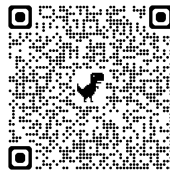
Code is a key research output — let's preserve, reference, and value it

 **Roberto Di Cosmo** 
Director at Software Heritage

October 7, 2025



Nature link



LinkedIn post

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 Software Heritage: supporting Software in Open Science
- 3 Preview from the Software Heritage very large telescope
- 4 The road ahead

Unveiled in 2016



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share *all* software source code

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

Reference catalog



find and reference all
software source code

Universal archive



preserve and share all
software source code

Research infrastructure



enable analysis of all
software source code

A universal software archive, as a shared infrastructure

One infrastructure
non profit
open and shared

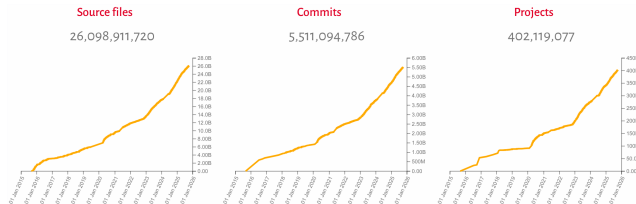


Unveiled in 2016

Inria



The largest archive ever built



Diamond sponsors



Platinum sponsors



Gold sponsors



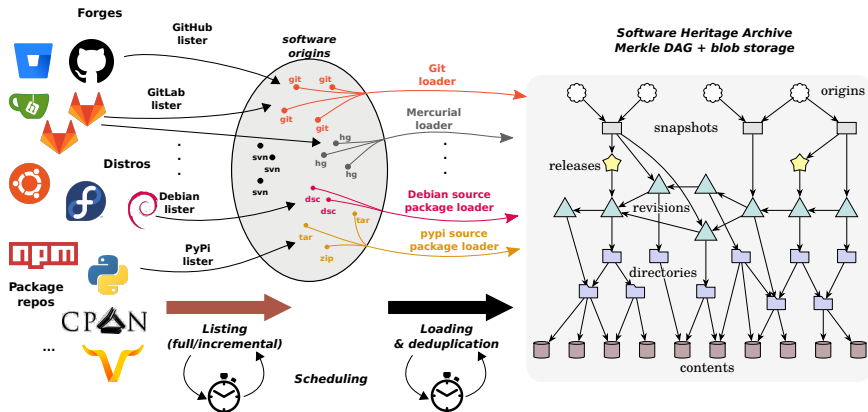
Silver sponsors



Bronze sponsors



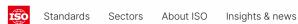
The archive under the hood



Global development history permanently archived in a uniform data model

- over 26 billion unique source files from over 400 million software projects
- ~2PB (compressed) blobs, ~50 B nodes, ~800 B edges

Software Hash Identifiers (ISO 18670)



**ISO/IEC
18670:2025**

Information technology —
SoftWare Hash Identifier
(SWHID) Specification V1.2

[Read sample](#)

Published (Edition 1, 2025)

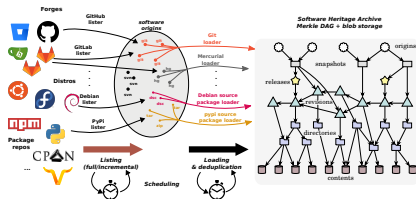
Cryptographically strong identification
for 50 billion software artifacts

Integrity and traceability support reproducibility
and transparency



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](#)

Projects



FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC



FAIR-IMPACT
Expanding FAIR solutions across EOSC

The CodeMeta Project

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



Zoom on UNESCO's recommendations for Open Science



Ref.: CL/4363

Subject: Draft text of the UNESCO Recommendation on Open Science

Madam/Sir,

At its 40th session in November 2019, the UNESCO General Conference decided to elaborate a draft Recommendation on Open Science.

This was a major decision, which has since mobilized the entire Organization and all of its Member States in the development of this new standard-setting instrument.

After two years of joint work, this process is now entering its final phase, following the consensus reached on the draft text during the intergovernmental meeting of experts held from 6 to 11 May 2021.

I have the pleasure to submit to you this draft recommendation, which will be put forward for adoption at our next General Conference in November 2021.

The definitions and principles that it contains constitute a common – and currently unprecedented – framework to support scientific cooperation and make science more transparent, more accessible, more equitable and more inclusive.

For any further information, Sharmila Nair-Bedouelle, Assistant Director-General for Natural Sciences, is at your disposal at the following email address: openscience@unesco.org.

Thanking you for your commitment, please accept, Madam/Sir, the assurances of my highest consideration.


Audrey Azoulay
Director-General

Enclosure: 1

1. Draft text of the UNESCO Recommendation on Open Science

cc: Permanent Delegations to UNESCO
National Commissions for UNESCO

7, place de Fontenay
92526 Paris 15^e France
Tel.: +33 (0)1 55 36 36 32
www.unesco.org

To Ministers responsible for relations with UNESCO

Selection from the recommendations

- Open Source for Open Science

"The source code must be included in the software release and made available on openly accessible repositories and the chosen license must allow modifications, derivative works and sharing under equal or compatible open terms and conditions"

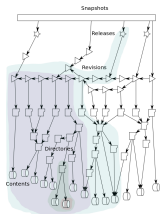
- Infrastructures

"Open science infrastructures should be organized and financed upon an essentially not-for-profit and long-term vision, which enhance open science practices and guarantee permanent and unrestricted access to all, to the largest extent possible."

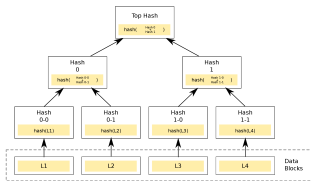
Much more than an archive: a revolutionary infrastructure for Software

Modern "Library of Alexandria", *international, non profit, long term* initiative
addressing the needs of *industry, research, culture and society as a whole*

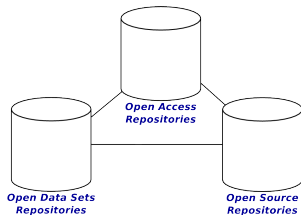
Software Graph



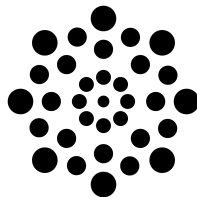
Software Blockchain



Open Science pillar



Big Code



One infrastructure, shared: more efficient, less waste ...
... addressing a broad spectrum of needs!

- 1 Software as a pillar of Open Science
- 2 Software Heritage: supporting Software in Open Science
- 3 Preview from the Software Heritage very large telescope
- 4 The road ahead

An unprecedented view of software institutional contributions



Contributions



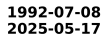
Software Projects

Active last 12 months: 3425

Still crawled: 44874

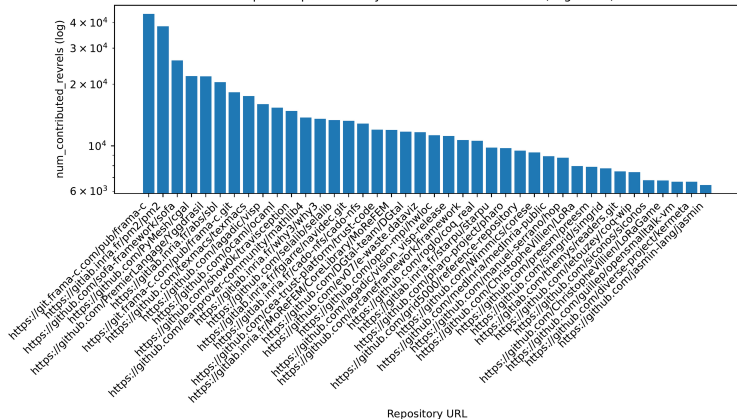


Contributors



Time Span

Top 40 Repositories by number of Contributions (Log Scale)



domain	count
imira.fr	3833
univ-lyon2.fr	1806
vica.fr	1371
univ-lyon1.fr	1216
utsr.fr	1216
cea.fr	1187
insa-toulouse.fr	747
ens-lyon.fr	706
univ-artois.fr	696
univ-lens.fr	690
crs.fr	682
insa.fr	675
univ-fcomte.fr	674
unilim.fr	641
insa-nantes.fr	493
univ-lyon3.fr	459
uttr.fr	437
polytechnique.fr	432
uha.fr	422
pasteur.fr	420
telecom-paris.fr	412
univ-lyon4.fr	352
univ-smb.fr	352
univ-tours.fr	350
u-picardie.fr	349
emc.fr	328
sciencexpo.fr	327
univ-st-etienne-paris4.fr	323
univ-poitiers.fr	319
u-poitiers.fr	319
univ-angers.fr	267
univ-inserm.fr	264
centralesupelec.fr	212
ind.fr	171
antilles-quebec.org	159
univ-tlse3.fr	133
insa.fr	130
emac.fr	126
univ-eiffel.fr	125
agroparistech.fr	105
cranm.fr	104
univ-lyon5.fr	95
uphil.fr	94
univ-tln.fr	89

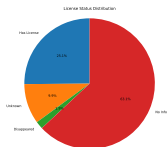
Repository URL

A picture of the French academic software sky

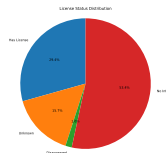
New frontiers: licences, sovereignty and resilience

Licence status of projects with at least N contributions

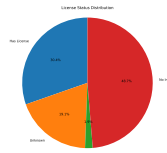
N=100



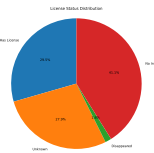
N=500



N=1000

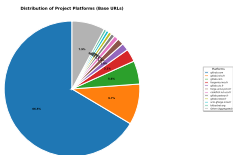


N=2500

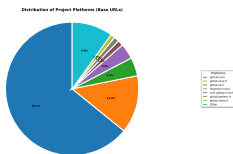


Platforms used by projects with N contributions

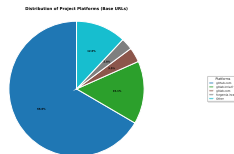
N=100



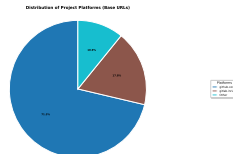
N=500



N=1000



N=2500



- 1 Software as a pillar of Open Science
- 2 Software Heritage: supporting Software in Open Science
- 3 Preview from the Software Heritage very large telescope
- 4 The road ahead

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](#))
- cite using the [biblatex-software](#) package (in CTAN and TeXLive)
- *french partners*: register in HAL (see [online HAL documentation](#))



Identifying, monitoring, and much more!

Let's grow and use the very large telescope of software source code

- 
- R. Di Cosmo, S. Granger, K. Hinsén, 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>