

# Software Heritage

an archive to enable our digital future

Roberto Di Cosmo

Director, Software Heritage  
Inria and Université de Paris

30 November 2021

#SWH5YEARS at UNESCO



# Software Heritage

THE GREAT LIBRARY OF SOURCE CODE



1 Software and Source Code

2 Software Heritage: a mission at the service of Humankind

3 The road ahead

# Software is all around us



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       GOTOP00H      # 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
```

## Quake III source code (excerpt)

```
float Q_rsqrt( float number )
{
    long i;
    float x2, y;
    const float threehalfs = 1.5F;

    x2 = number * 0.5F;
    y = number;
    i = * ( long * ) &y; // evil floating point bit level hacking
    i = 0x5f3759df - ( i >> 1 ); // what the fuck?
    y = * ( float * ) &i;
    y = y * ( threehalfs - ( x2 * y * y ) ); // 1st iteration
    // y = y * ( threehalfs - ( x2 * y * y ) ); // 2nd iteration, this
    // can be removed

    return y;
}
```

Len Shustek, *Computer History Museum*

2006

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



# Software source code as a key asset of Humankind

Experts call for greater recognition of software source code as heritage for sustainable development

6 November 2018



UNESCO, Inria, Software Heritage invite  
40 international experts meet in Paris ...



The call is published on February 2019

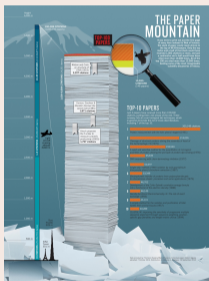
*“Recognise software source code as a fundamental enabler in all aspects of human endeavour”*

# (Open Source) Software is *precious technical and scientific knowledge*

Yuval Noah Harari (on COVID 19)

*"The real antidote [to epidemic] is scientific knowledge and global cooperation."*

## Software powers modern scientific research



The top 100 papers

*[...] the vast majority describe experimental methods or software that have become essential in their fields.*

Nature, October 2014

## We can still talk to the early inventors



*"Telling historical stories is the best way to teach. It's much easier to understand something if you know the threads it is connected to."*

Donald E. Knuth  
Len Shustek

CACM, January 2021

We need a *dedicated infrastructure* to preserve and share *all* this knowledge!

# Enhancing software Reuse, Security and Transparency

Software complexity is growing...

the largest part of modern applications is... just reuse!

... it is important to Know Your SoftWare (KYSW)

Politique publique de la donnée, des algorithmes et des codes sources

... animer les écosystèmes des... réutilisateurs du source code

Circulaire du Premier Ministre, 27 Avril 2021, France



Sec. 4. Enhancing Software Supply Chain Security

*ensuring and attesting, to the extent practicable, to the integrity and provenance of open source software*

May 2021 POTUS Executive Order

We need a *trusted* knowledge base with *software provenance* !

# Software source code is fragile

## Endangered source code ...



- *link rot*: projects are created, moved around, removed
- *data rot*: physical media with legacy software decay
- *platform consolidation*

2015 Google Code and Gitorious.org shutdown:  
~1M endangered repositories

2019 250.000 endangered repositories on  
BitBucket

... is endangered knowledge!

broken links and missing pieces in the *web of knowledge* of humankind

Bottomline: we need a global, long term effort

to build a *universal archive of all software source code*



1 Software and Source Code

2 Software Heritage: a mission at the service of Humankind

3 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 all software source code

## Research infrastructure



enable analysis of all software source code

# Five year later: a *universal* software archive, as a shared infrastructure

Cultural Heritage



Industry



Research



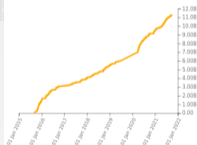
Public Administration



## Software Heritage

Source files

11 311 753 576



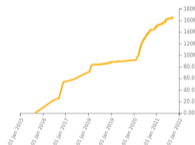
Commits

2 390 337 241



Projects

165 640 910



Directories

9 415 589 293

Authors

45 750 890

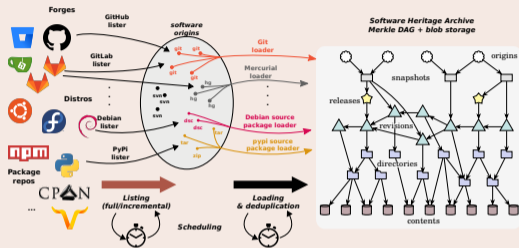
Releases

19 047 072

All the software source code, with *all its development history*

# An operational, evolving infrastructure

## Harvest and archive



- [save.softwareheritage.org](https://save.softwareheritage.org)
- [deposit.softwareheritage.org](https://deposit.softwareheritage.org)

## Reference (20 billion SWHIDs)

Intrinsic, decentralised, cryptographically strong identifiers, SWHIDs



Now supported in **SPDX 2.2**, **Wikidata** etc.

## Adoption in Open Science

reference archive

for research software

## Adoption in Industry and Public Administration

reference archive and knowledge base

for open source software



# A walkthrough

## General

- Browse [the archive](#)
- Get and use SWHIDs, e.g. [Apollo 11 excerpt](#), [Quake III excerpt](#)
- [Trigger archival](#) of a software repository in one click

## Open Science

- [Curated deposit via HAL](#), e.g.: [LinBox](#), [SLALOM](#), [Givaro](#), [SumGra](#), [Coq proof](#), ...
- Cite software [with the biblatex-software style](#), e.g.: [article from IPOL](#)

## History of software

Rescue landmark legacy software, see the [SWHAP process](#)

## Public code

Archived source code from [code.gouv.fr](#)

## Sharing the vision



United Nations  
Educational, Scientific and  
Cultural Organization



And many more ...

[www.softwareheritage.org/support/testimonials](http://www.softwareheritage.org/support/testimonials)

## Donors, members, sponsors

*Inria*

Diamond sponsor



Platinum sponsors



Gold sponsors



Silver sponsors



Bronze sponsors

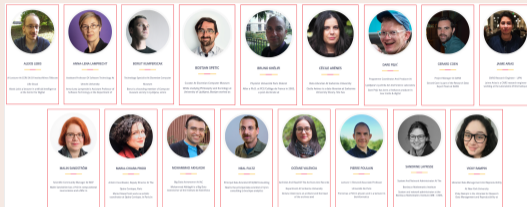


# A growing community

## Team



## Ambassadors



## Foundations and grantees



- Castalia, CottageLabs
- EasterEggs, OcamlPro
- Octobus, Sperling, Tweag.io

## Mirrors and storage partners

*“Let us save what remains: ... by such a multiplication of copies, as shall place them beyond the reach of accident.”*

— Thomas Jefferson

Enea, FossilD

CEA, RedHat



1 Software and Source Code

2 Software Heritage: a mission at the service of Humankind

3 The road ahead

# Listening to key partners

Open Science

Cultural Heritage

Industry

Public Administration

Bring together all the stakeholders, and move to the next level



## Software Heritage

[www.softwareheritage.org](http://www.softwareheritage.org)

[@swheritage](https://twitter.com/swheritage)

### The Library of Alexandria of code



- recover the past
- structure the future

### The Very Large Telescope for Source code



- build better software
  - for industry
  - for society as a whole