

# Preserving our Software Heritage and its Stories

Roberto Di Cosmo  
Journées du CIS

Director, Software Heritage  
Inria and Université de Paris Cité

Octobre 2nd 2024



Software Heritage  
THE GREAT LIBRARY OF SOURCE CODE

- 
- 1 Software as Heritage
  - 2 Meet Software Heritage
  - 3 Preserving the past
  - 4 Conclusions

# 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     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.”*

# Calling for preservation: UNESCO

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 ...

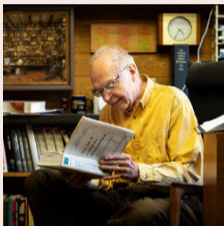
The call is published on Feb 2019

“[We call to] support efforts to gather and preserve the artifacts and narratives of the history of computing, while the earlier creators are still alive”

<https://en.unesco.org/foss/paris-call-software-source-code>



Communications of the ACM, February 2021



*"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."*

*Let's Not Dumb Down the History of Computer Science*

Donald E. Knuth, Len Shustek

<https://doi.org/10.1145/3442377>

A unique opportunity

most of the creators are still here: we can talk to them!

but the clock is ticking...

We need a universal archive of software source code: now we have one!

- 
- 1 Software as Heritage
  - 2 Meet Software Heritage
  - 3 Preserving the past
  - 4 Conclusions



# 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

One infrastructure  
open and shared



Largest archive

## Technology

- transparency and FOSS
- replicas all the way down

## Content (billions!)

- **intrinsic identifiers**
- facts and provenance

## Organization

- non-profit
- multi-stakeholder



## 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



Diamond sponsor



Platinum sponsors



Gold sponsors



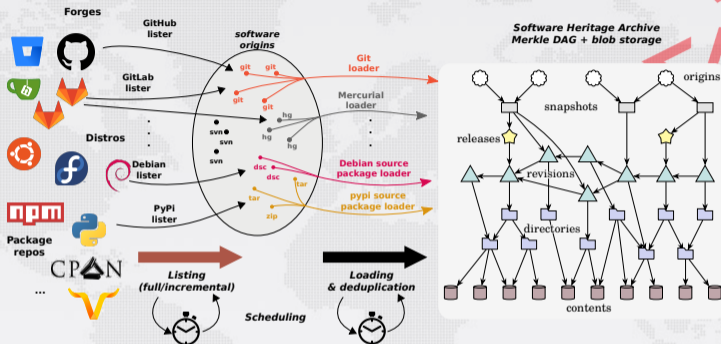
Silver sponsors



Bronze sponsors



# A peek under the hood: a universal archive



Global development history permanently archived in a uniform data model

- over 20 billion unique source files from over 300 million software projects
- ~1.5PB (compressed) blobs, ~50 B nodes, ~700 B edges

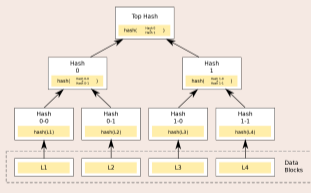
# A revolutionary infrastructure

Modern "Library of Alexandria", *international, non profit, long term* initiative addressing the needs of *industry, research, culture and society as a whole* One infrastructure, *shared*: more efficient, less waste ...

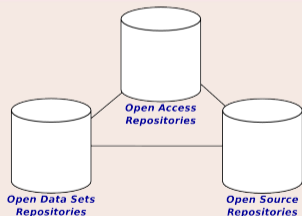
## Software Graph



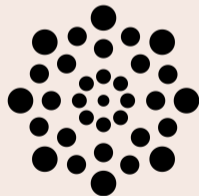
## Software Blockchain



## Open Science pillar



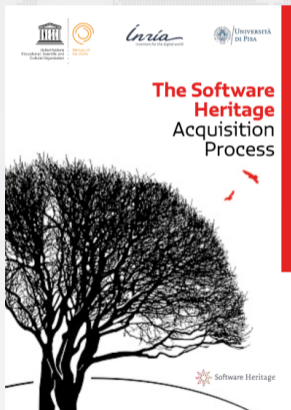
## Big Code



- 
- 1 Software as Heritage
  - 2 Meet Software Heritage
  - 3 Preserving the past
  - 4 Conclusions

## Paris Call on Software Source Code

“[We call to] support efforts to gather and preserve the artifacts and narratives of the history of computing, while the earlier creators are still alive”



- **Rescue** Legacy Software from different media
  - physical
  - digital
    - legacy / unsupported
    - recent / supported
- **Curate** the code
  - reconstructing the development history
  - collecting metadata
- And **illustrate** with dedicated presentations

## Paris Call on Software Source Code

“[We call to] support efforts to gather and preserve the artifacts **and narratives** of the history of computing, while the earlier creators are still alive”



- **Expand** the SWHAP scope to
  - documents
  - media (videos, pictures, images, etc.)
  - oral history
- **Preserve and Present** all this material
- **Share** process and tools (all open source!)
  - with museums, archives and all interested parties

see this live on [the Software Stories website](#), and get [the guide](#) and [the SWHAP Days hybrid event](#), 19 and 20 october 2022



## Ongoing works

- Inria's landmark legacy software
- [SCOSS and ALIG for Open Science](#)

## New opportunities for research

- <https://opensource.cern/sw-heritage-impact-opening>

## A proposal for a working agenda

**Search and find** software source code associated to *landmark ACM articles*

**Reconstruct** development history, *archive in SWH*

**Link** publications to the source code using the SWHID identifier






**Collect** oral and documentary history around it, and build a Software Story

**Connect** with all the relevant history collections



- 
- 1 Software as Heritage
  - 2 Meet Software Heritage
  - 3 Preserving the past
  - 4 Conclusions

References (see <https://www.softwareheritage.org/publications>)

-  Morane Gruenpeter, Roberto Di Cosmo, Katherine Thornton, Kenneth Seals-Nutt, Carlo Montangero, Guido Scatena  
*Software Stories for landmark legacy code*, Inria TR, 2022  
<https://hal.archives-ouvertes.fr/hal-03483982>
-  EOSC SIRS Task Force, *Scholarly Infrastructures for Research Software*  
2020, European Commission, ([10.2777/28598](https://doi.org/10.2777/28598))
-  R. Di Cosmo, *Archiving and Referencing Source Code with Software Heritage*  
ICMS 2020 ([10.1007/978-3-030-52200-1\\_36](https://doi.org/10.1007/978-3-030-52200-1_36))
-  Laura Bussi, Roberto Di Cosmo, Carlo Montangero, Guido Scatena  
*The software heritage acquisition process*  
UNESCO, Università di Pisa, Inria, 2019
-  J.F. Abramatic, R. Di Cosmo, S. Zacchiroli, *Building the Universal Archive of Source Code*,  
CACM, October 2018 ([10.1145/3183558](https://doi.org/10.1145/3183558))