

# Software Heritage: a common infrastructure to preserve our Software Commons


Roberto Di Cosmo  
Inria and Université de Paris

19/12/2020, *LibreItalia*



# Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

- 
- 1 Introduction
  - 2 Meet Software Heritage
  - 3 Demo time!
  - 4 Building for the long term ...
  - 5 Everybody is concerned

Computer Science professor in Paris, now working at INRIA

- 30 years of research (Theor. CS, Programming, Software Engineering, Erdos #: 3)
- 20 years of Free and Open Source Software
- 10 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](http://www.mancoosi.org)

2010 *IRILL* [www.irill.org](http://www.irill.org)

2015 *Software Heritage* at INRIA

2018 *National Committee for Open Science*, France

# Software source code: a precious part of our heritage

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

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

## Definition (Commons)

The **commons** is the cultural and natural resources accessible to all members of a society, including natural materials such as air, water, and a habitable earth. These resources are held in common, not owned privately. <https://en.wikipedia.org/wiki/Commons>

## Definition (Software Commons)


The **software commons** consists of all computer software which is available at little or no cost and which can be altered and reused with few restrictions. Thus *all open source software and all free software are part of the [software] commons. [...]*

[https://en.wikipedia.org/wiki/Software\\_Commons](https://en.wikipedia.org/wiki/Software_Commons)

**Source code:** part of our commons ... pillar of Open Science! (*would require another talk*)

Precious, endangered *executable* and *human readable* knowledge

key people **passing away**, platforms (GoogleCode, Gitorious, etc.) closing down ...

- 
- 1 Introduction
  - 2 Meet Software Heritage
  - 3 Demo time!
  - 4 Building for the long term ...
  - 5 Everybody is concerned



# 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

Cultural Heritage



Industry



Research



Education



## Software Heritage



### Technology

- transparency and FOSS
- replicas all the way down

### Content (billions!)

- **intrinsic identifiers**
- facts and provenance


### Organization

- non-profit
- multi-stakeholder




# A dedicated core team



- 
- 1 Introduction
  - 2 Meet Software Heritage
  - 3 Demo time!
  - 4 Building for the long term ...
  - 5 Everybody is concerned

- Browse [the archive](#)
- Get and use SWHIDs ([full specification available online](#))
- cite software [with the biblatex-software style](#) from CTAN
- Example use in a research article: compare Fig. 1 and conclusions
  - in [the 2012 version](#)
  - in [the updated version](#) using SWHIDs and Software Heritage
- Example use in a research article: extensive use of SWHIDs in [a replication experiment](#)
- Trigger archival of your preferred software in a breeze
- curated deposit in SWH via HAL, see for example: [LinBox](#), [SLALOM](#), [Givaro](#), [NS2DDV](#), [SumGra](#), [Coq proof](#), ...
- rescue landmark legacy software, see the [SWHAP process with UNESCO](#)

- 
- 1 Introduction
  - 2 Meet Software Heritage
  - 3 Demo time!
  - 4 Building for the long term ...
  - 5 Everybody is concerned

# An international, non profit initiative...

## Sharing the vision



United Nations  
Educational, Scientific and  
Cultural Organization



And many more ...

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

Roberto Di Cosmo

## Donors, members, sponsors



INVENTEURS DU MONDE NUMÉRIQUE

### Platinum sponsors



MINISTÈRE  
DE L'ENSEIGNEMENT SUPÉRIEUR,  
DE LA RECHERCHE  
ET DE L'INNOVATION



HUAWEI



SOCIÉTÉ  
GÉNÉRALE

### Gold sponsors



Université  
de Paris



SORBONNE  
UNIVERSITÉ

### Silver sponsors



Software Intelligence for Digital Leaders



Liberté • Égalité • Fraternité  
RÉPUBLIQUE FRANÇAISE



UNIVERSITÀ  
DI PISA



### Bronze sponsors



DANS



FOSSID



Thomas Jefferson, February 18, 1791

*...let us save what remains: not by vaults and locks which fence them from the public eye and use in consigning them to the waste of time, but by such a multiplication of copies, as shall place them beyond the reach of accident.*

Welcoming ENEA



Italian National Agency for New Technologies,  
Energy and Sustainable Economic Development

- first **institutional** mirror
- increased resilience
- **infrastructure** for researchers
- stepping stone to  
an international joint effort

# ... raising awareness about Software Source Code

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 ...

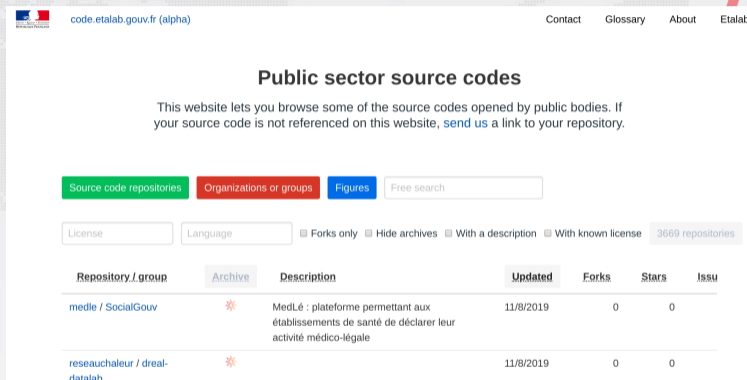


Their call is published on Feb 2019

It's an important *policy tool*, already referenced and used ...

*yes, you can sign it!*

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



The screenshot shows the homepage of code.etalab.gouv.fr (alpha). The page features a navigation bar with links for Contact, Glossary, About, and Etalab. The main heading is "Public sector source codes", followed by a descriptive paragraph. Below this are filter buttons for "Source code repositories", "Organizations or groups", and "Figures", along with a search input field. Further down, there are filters for "License" and "Language", and a list of checkboxes for "Forks only", "Hide archives", "With a description", and "With known license". A "3669 repositories" badge is also present. The main content is a table with columns for "Repository / group", "Archive", "Description", "Updated", "Forks", "Stars", and "Issu". Two entries are visible: "medle / SocialGov" and "reseauchaleur / dreal-datalab", both marked with a star icon.

code.etalab.gouv.fr (alpha) Contact Glossary About Etalab

## Public sector source codes

This website lets you browse some of the source codes opened by public bodies. If your source code is not referenced on this website, [send us](#) a link to your repository.

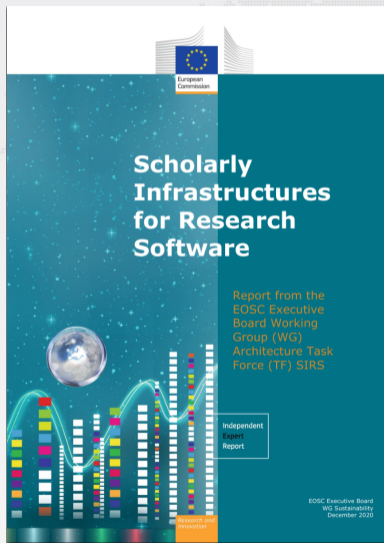
[Source code repositories](#) [Organizations or groups](#) [Figures](#)

Forks only  Hide archives  With a description  With known license 3669 repositories

Repository / group	Archive	Description	Updated	Forks	Stars	Issu
<a href="#">medle / SocialGov</a>	★	MedLé : plateforme permettant aux établissements de santé de déclarer leur activité médico-légale	11/8/2019	0	0	
<a href="#">reseauchaleur / dreal-datalab</a>	★		11/8/2019	0	0	

<https://code.etalab.gouv.fr>






## Important *policy tool* in Open Science (Dec 2020)

- 9 infrastructures
  - 3 archives
  - 3 open access publishers
  - 3 aggregators
- recommendations
  - archive in Software Heritage, use SWHID
  - open non profit
  - default to open source for research software

*"all research software should be made available under an Open Source license by default, and all deviations from this default practice should be properly motivated"*

See <https://doi.org/10.2777/28598>

- 
- 1 Introduction
  - 2 Meet Software Heritage
  - 3 Demo time!
  - 4 Building for the long term ...
  - 5 Everybody is concerned

# It's urgent to expand the archive

## Saving 250.000 endangered repositories...

- summer 2019: BitBucket announce Mercurial VCS phase out
- fall 2019: Software Heritage teams up with Octopus (funded by NLNet, thanks!)
- july 2020: BitBucket erases 250.000 repositories
- august 2020: [bitbucket-archive.softwareheritage.org](https://bitbucket-archive.softwareheritage.org) is live

## ... preserving the web of knowledge

([Tweet is here](#))



Gabriel Altay  
@gabrielaltay

Just realized [@Bitbucket](#) disabled all mercurial repositories when the [@asclnet](#) informed me that a link associated with an old paper of mine was down. Thought all was lost, but someone archived all the repos! very classy move by [@octopus\\_net](#) and [@SWHeritage](#).

[Traduire le Tweet](#)

1:48 AM · 31 août 2020 · Twitter Web App

### Bottomline

*explicit deposit* is important, ...

... and we must promote it...

... but will never be enough.

*(think also of all software dependencies!)*

# So much to do, so many ways to get involved

## Development (selected examples):

**develop** new *listers* and *loaders*

- apply for a Sloan funded minigrant  
<http://bit.ly/swhgrants> (rolling basis)

**rescue** and archive landmark legacy software

- use the SWHAP process  
<https://www.softwareheritage.org/swhap>

## Policy, advocacy (selected examples):

**promote** Software Heritage in (inter)national policy (AgID, EU, OGP, etc.)

**adopt** Software Heritage (article, journals, public administration, etc.)

**advocate** spread the word, make Software Heritage all over the world



# Software Heritage

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

@swheritage

Everybody is concerned, everybody can help build

## The Library of Alexandria of code



- recover the past
- structure the future

## A CERN for Software



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