

# Software Heritage

Collecting, preserving and sharing the software source code of Mankind

**Davide Storti**, Roberto Di Cosmo  
Unesco Inria

`roberto@dicosmo.org`

March 2018



# Software Heritage

THE GREAT LIBRARY OF SOURCE CODE



United Nations  
Educational, Scientific and  
Cultural Organization

## At the heart of our society



- communication, entertainment
- administration, finance
- health, energy, transportation
- education, research, politics
- ...

## Key mediator for accessing *all* information (c) Banski



Information is a main pillar of our modern societies.

*Absent an ability to correctly interpret digital information, we are left with [...] "rotting bits" [...] of no value.*

Vinton G. Cerf IEEE 2011

## Software is *an essential component* of modern scientific research

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

Top 100 papers (Nature, October 2014)

Software embodies our collective **Knowledge** and **Cultural 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.”*

## Quake 2 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;
}
```

## Net. queue in Linux (excerpt)

```
/*
 * SFB uses two B[1][n] : L x N arrays of bins (L levels, N bins per level)
 * This implementation uses L = 8 and N = 16
 * This permits us to split one 32bit hash (provided per packet by rxhash or
 * external classifier) into 8 subhashes of 4 bits.
 */
#define SFB_BUCKET_SHIFT 4
#define SFB_NUMBUCKETS (1 << SFB_BUCKET_SHIFT) /* N bins per Level */
#define SFB_BUCKET_MASK (SFB_NUMBUCKETS - 1)
#define SFB_LEVELS (32 / SFB_BUCKET_SHIFT) /* L */

/* SFB also uses a virtual queue, named "bin" */
struct sfb_bucket {
    u16      qlen; /* length of virtual queue */
    u16      p_mark; /* marking probability */
};
```

Len Shustek, Computer History Museum

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

# Software is spread all around

Debian CPAN  
Sourceforge Gitorious  
Maven Inria  
Bitbucket  
Git GitHub  
BerliOs CTAN  
GoogleCode GitLab Adullact CRAN



A word cloud of terms related to software fragility, set against a light gray world map background. The words are arranged in a roughly circular pattern. The largest word is 'damage' in a dark red font. Other prominent words include 'disaster' in purple, 'malicious' in brown, 'deletion' in blue, 'obsolete' in purple, 'attack' in blue, 'format' in green, 'reference' in blue, 'storage' in blue, 'dangling' in blue, 'wear' in blue, 'corruption' in blue, 'encryption' in blue, 'dependencies' in blue, 'aging' in blue, 'media' in blue, and 'tear' in blue. The background also features a decorative pattern of red and orange triangles on the right side.

damage  
disaster  
malicious  
deletion  
obsolete  
attack  
format  
reference  
storage  
dangling  
wear  
corruption  
encryption  
dependencies  
aging  
media  
tear

# Software lacks its own research infrastructure



Photo: ALMA(ESO/NAOJ/NRAO), R. Hills





## Software Heritage

THE GREAT LIBRARY OF SOURCE CODE



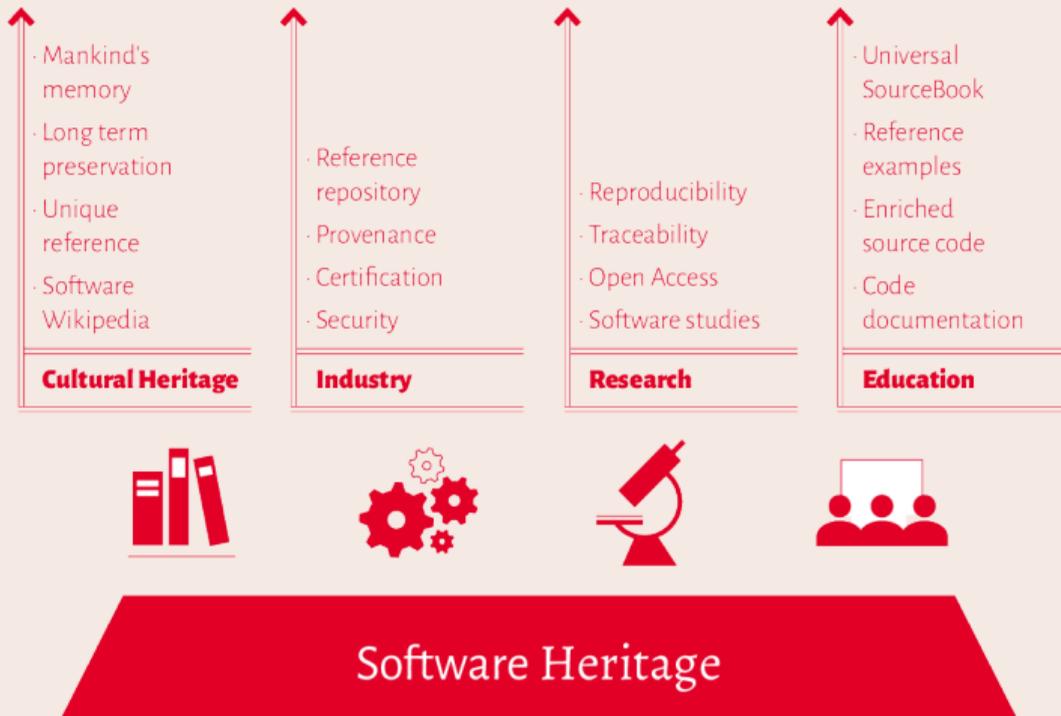
### Our mission

**Collect**, **preserve** and **share** the *source code* of *all the software* that is publicly available.

### Past, present and future

*Preserving the past, enhancing the present, preparing the future.*

# We are working on the foundations



# The archive is alive and growing



## Our current sources

- GitHub
- Debian, GNU
- Gitorious, Google Code, ...

# Our principles

Cultural Heritage



Industry



Research



Education



Software Heritage

Open approach

open source, transparency

In for the long haul

non profit, replication

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.

# Sharing the Software Heritage vision



See more

<http://www.softwareheritage.org/support/testimonials>

# Sponsoring Software Heritage work

*Inria*  
INVENTEURS DU MONDE NUMÉRIQUE



NOKIA Bell Labs



ALMA MATER STUDIORUM  
UNIVERSITA' DI BOLOGNA  
DIPARTIMENTO DI INFORMATICA - SCIENZA E INGEGNERIA

GitHub

April 3rd, 2017: landmark Inria Unesco agreement...

*Inria*  
INVENTEURS DU MONDE NUMÉRIQUE



<https://www.softwareheritage.org/blog>

September 27-28: Mauritius Call

Come in, we're open!



# Software Heritage

`www.softwareheritage.org`

`@swheritage`

Everybody can participate

strategy

sponsoring / partnership

own code

working groups, leads

`mailto:roberto@dicosmo.org`

`sponsorship.softwareheritage.org`

`forge.softwareheritage.org`

`wiki.softwareheritage.org`