Software [is our] Heritage
Collecting, preserving and sharing the software source code of Mankind

Roberto Di Cosmo
Inria

roberto@dicosmo.org

January 10th 2017
Santiago
Software is everywhere
"The source code for a work means the preferred form of the work for making modifications to it."

Hello World

Program (source code)

```c
/* Hello World program */

#include<stdio.h>

void main()
{
    printf("Hello World");
}
```

Program (excerpt of binary)

```
4004e6:  55
4004e7:  48  89  e5
4004ea:  bf  84  05  40  00
4004ef:  b8  00  00  00  00
4004f4:  e8  c7  fe  ff  ff
4004f9:  90
4004fa:  5d
4004fb:  c3
```
So/ftware Source Code is special

Harold Abelson, Structure and Interpretation of Computer Programs

“Programs must be written for people to read, and only incidentally for machines to execute.”

Quake 2 source code (excerpt)

```c
float Q_rsqrt( float number )
{
    long i;
    float x2, y;
    const float threehalves = 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 * ( threehalves - ( x2 * y + y ) ); // 1st iteration
    // y = y * ( threehalves - ( x2 * y + y ) ); // 2nd iteration, this can be removed
    return y;
}
```

Network queue in Linux (excerpt)

```c
/*
 * SF8 uses two B[size][size] arrays of bins (size levels, size 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 SF8_BUCKET_SHIFT 4
#define SF8_NUMBuckets (1 << SF8_BUCKET_SHIFT) /* N bins per level */
#define SF8_BUCKET_MASK (SF8_NUMBuckets - 1)
#define SF8_LEVELS (32 / SF8_BUCKET_SHIFT) /* L */

/* SF8 algo uses a virtual queue, named "Bln" */
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.”
Our mission

Collect, preserve and share the source code of all the software that lies at the heart of our culture and our society.

Past, present and future

Preserving the past, enhancing the present, preparing the future.
The reference repository of all Source Code

Fashion victims

- many disparate development platforms
- a myriad places where distribution may happen
- projects tend to migrate from one place to the other over time

One place...

... in Software Heritage you can find and search all the source code
like all digital information, Software Source Code is fragile

- inconsiderate and/or malicious code loss (e.g., Code Spaces)
- business-driven code loss (e.g., Gitorious, Google Code)
- for obsolete code: physical media decay (data rot)

If (a repository on) GitHub goes away …

… Software Heritage will have a copy of it!
A wealth of software research on crucial issues...

- safety, security; test, verification, proof;
- software engineering, software evolution;
- empirical and big data studies;

If you study the stars, you go to Atacama...

... Software Heritage is the very large telescope of source code
Our principles

**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.
Status

The archive (we collect!)

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

Sponsors and supporters

Microsoft, Huawei, Société Générale, Nokia Bell Labs, DANS, ACM, Creative Commons, Eclipse, FSF, OSI, GitHub, GitLab, IEEE, OIN, OW2, SFC, SFLC, The Document Foundation, The Linux Foundation, …

Software Heritage for society as a whole

- preservation of knowledge embedded in software
- access to the knowledge embedded in software
Come in, we’re open!

Software Heritage

www.softwareheritage.org  @swheritage

Everybody can participate

sponsoring / partnership  sponsorship.softwareheritage.org
working groups, leads  wiki.softwareheritage.org
our own code  forge.softwareheritage.org