

Software Heritage

Opening the doors to the public

Roberto Di Cosmo

`roberto@dicosmo.org`

June 7th, 2018



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Software is everywhere



Software is everywhere



Software embodies a growing part of...

... our scientific, technical and *Cultural Heritage!*

Source Code: *executable* and *human readable* knowledge



“The source code for a work means the preferred form of the work for making modifications to it.”

GPL Licence





"The source code for a work means the preferred form of the work for making modifications to it."

GPL Licence

```
Hello World
```



“The source code for a work means the preferred form of the work for making modifications to it.”

GPL Licence

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
```



"The source code for a work means the preferred form of the work for making modifications to it."

GPL Licence

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
```

Program (source code)

```
/* Hello World program */

#include<stdio.h>

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

~ 50 years, a lightning fast growth

Apollo 11 (~60.000 lines), 1969



"When I first got into it, nobody knew what it was that we were doing. It was like the Wild West."

Margaret Hamilton



~ 50 years, a lightning fast growth

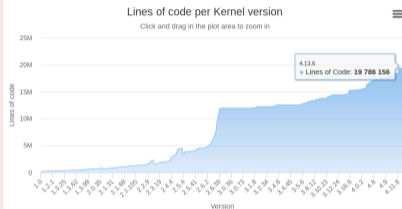
Apollo 11 (~60.000 lines), 1969



"When I first got into it, nobody knew what it was that we were doing. It was like the Wild West."

Margaret Hamilton

Linux Kernel (in your pockets!)



~ 50 years, a lightning fast growth

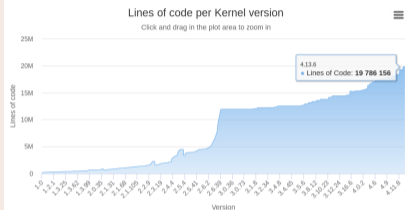
Apollo 11 (~60.000 lines), 1969



"When I first got into it, nobody knew what it was that we were doing. It was like the Wild West."

Margaret Hamilton

Linux Kernel (in your pockets!)



Harold Abelson, Structure and Interpretation of Computer Programs

(1985)

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

~ 50 years, a lightning fast growth

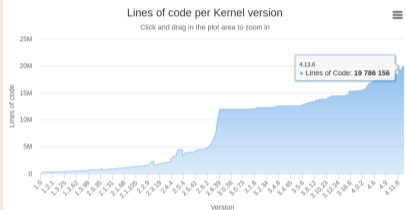
Apollo 11 (~60.000 lines), 1969



"When I first got into it, nobody knew what it was that we were doing. It was like the Wild West."

Margaret Hamilton

Linux Kernel (in your pockets!)



Harold Abelson, Structure and Interpretation of Computer Programs

(1985)

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

Len Shustek, Computer History Museum

(2006)

"Source code provides a view into the mind of the designer."

We are at a turning point

Preserve the past

Only a few years left to recover the history of software technology

We are at a turning point

Preserve the past

Only a few years left to recover the history of software technology

Improve the future

We need a **universal** platform for all the future software source code



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all

Reference catalog



find and reference **all** the
source code



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all

Reference catalog



find and reference **all** the source code

Universal archive



preserve **all** the source code



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all

Reference catalog



find and reference **all** the source code

Universal archive



preserve **all** the source code

Research infrastructure



enable analysis of **all** the source code



Milestones



Inria

INVENTEURS DU MONDE NUMÉRIQUE

initial push

core team and infrastructure

long term commitment

open approach, and continued support



© June 30, 2016

Our first press release





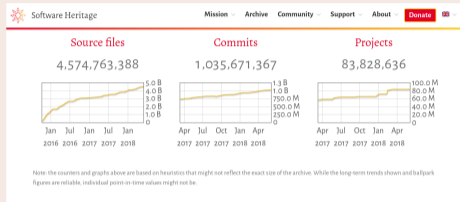
Raising awareness for Software Source Code preservation and access



The team



The archive



GitHub, Debian, Gitorious, GoogleCode, ...

Sponsors

Microsoft Intel Société Générale
Huawei, Google
DANS, Nokia Bell Labs, University of Bologna,
GitHub, UQAM, FOSSID

Collaborations

- Qwant (search engine)
- Crossminer
- HAL, Adullact



© June 7, 2018

Opening the door of the Software Heritage archive



Software Heritage

Mission ▾ Archive Community ▾ Support ▾ About ▾ [Donate](#)

[Browse the archive](#)

See visits

Orig.

Visit date: 05 Dec.

```
P Branch: HEAD 3c235a1 / Comanche055 / ALARM_Ag
1 # Copyright: Public domain.
2 # Filename: ALARM_AND_ABORT.agc
3 # Purpose: Part of the source code for Comanche.
4 # is part of the source code for the Co.
5 # (CH) Apollo Guidance Computer (AGC), Ap.
6 # Assembler: yaYUL
7 # Reference: pp. 1493-1496
8 # Contact: Ron Burkey <info@sandroid.org>
9 # Website: http://www.ibiblio.org/apollo.
10 # Mod history: 2009-05-07 RSB Adapted from Colossus249 file o
11 # name, and page images. Corrected
12 # typos in the transcription of progr
13 # comments, and these should be back-por
```

Visiting the archive: the Apollo 11 source code

Margaret Hamilton today



The Apollo 11 source code in SWH

The screenshot shows the Software Heritage (SWH) archive interface for the Apollo 11 source code. The origin is listed as <https://github.com/chrisgarry/Apollo-11>. The interface includes a search bar, a list of files, and a table of file details.

File	Mode	Size
ADD_BLOCK_TWO_SELF_CHECK.agc	mp-r--	15.4 KB
ACS_INITIALIZATION.agc	mp-r--	8.0 KB
ALARM_AND_ABORT.agc	mp-r--	5.0 KB
ADSTAR_AND_ADSTAR.agc	mp-r--	28.0 KB
ADTMARK.agc	mp-r--	16.8 KB
ASCENT_GUIDANCE.agc	mp-r--	10.9 KB
ASSEMBLY_AND_OPERATION_INFORMATION.agc	mp-r--	30.1 KB
ATTITUDE_MANEUVER_ROUTINE.agc	mp-r--	26.5 KB
BURN_BABY_BURN_MASTER ignition_ROUTINE.agc	mp-r--	21.8 KB
CDMIC_SUBROUTINES.agc	mp-r--	46.5 KB
CONTROLLED_CONSTANTS.agc	mp-r--	14.5 KB
DAPIOLER_PROGRAM.agc	mp-r--	12.7 KB

Some pointers

- Entry point
- Burn, baby, burn!

Visiting the archive: the Quake 3 source code

John Carmack



The Quake 3 source code in SWH

A screenshot of the Software Heritage (SWH) archive interface. The page title is "Browse the archive". The origin is listed as "https://github.com/id-Software/Quake-III-Arena". The current branch is "HEAD" with commit "c6f02c2". A table lists the files and directories in the archive:

File	Mode	Size
code	d-----	
common	d-----	
lec	d-----	
libs	d-----	
q3asm	d-----	
q3map	d-----	
q3gradient	d-----	
ui	d-----	
COPYING.txt	-rwxr-xr-x	14.8 KB
README.txt	-rwxr-xr-x	8.8 KB

Below the table, there is a red link for "README.txt" and a red heading "Quake III Arena GPL source release".

Some pointers

- Entry point
- What the f...

Yes, now you can!

Wayback machine for source code

retrieve the source code as it was

Open science

deposit scientific software (via HAL)

Reference catalog

use intrinsic identifiers for software

Universal knowledge base

store the knowledge about source code

And much, much more is in store!

see the strategic plan...



Software Heritage

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