Software Heritage
Building the Library of Alexandria/the CERN of software source code

Roberto Di Cosmo
Director, Software Heritage

Oct 24th, 2019
1. Software all around us
2. The Software Heritage initiative
3. Building for the long term
At the heart of our society

Software embodies our collective Knowledge and Cultural Heritage

A pillar of modern research

Open Access Repositories

Open Data Sets Repositories

Open Source Repositories
Source code: executable and human readable

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

Program (source code)

/* 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
An essential part of our knowledge

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.”

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!)

graph showing lines of code per kernel version
Outline

1. Software all around us
2. The Software Heritage initiative
3. Building for the long term
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
The largest software archive, a shared infrastructure

Software Heritage

- Cultural Heritage
- Industry
- Research
- Education

Source files: 6,197,000,081
Commits: 1,379,380,527
Projects: 90,231,104
Let’s see this live!
Outline

1. Software all around us
2. The Software Heritage initiative
3. Building for the long term
Software Heritage’s key dates

**Summer 2015**
The collection starts: first server, (very) early prototype

**June 30th 2016**
Public unveiling, with the first sponsors: Microsoft and DANS

**April 3rd 2017**
Unesco - Inria agreement on software access and preservation.

**June 7th 2018**
Opening the archive to the world

**February 26th 2019**
Expert meeting Paris Call on Software Source Code

**June 10th 2019**
Software Heritage Acquisition Process with UNESCO and UniPi