Software Heritage

building an universal, open, shared, long term digital infrastructure

Roberto Di Cosmo

Digital Civil Society Conference, Stanford

November 1st, 2019

Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Outline

1. Introduction
2. Selected remarks
3. Software Heritage: highlights
4. Conclusion

Short Bio: Roberto Di Cosmo

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 structures for the common good

1980 my first computer: ZX80

1998 *Cybersnare* – voice of French FOSS

1999 *DemoLinux* – first live GNU/Linux distro

200x *Software Patents – Electronic voting*

2007 *Free Software Thematic Group*
150 members 50 projects 200Me

2015 *Software Heritage* at INRIA

2018 *National Committee for Open Science*, France

Outline

1 Introduction

2 Selected remarks

3 Software Heritage: highlights

4 Conclusion

Digital infrastructures are *special*

Invisible

Any sufficiently advanced technology is indistinguishable from magic

Arthur C. Clarke

Network effect

Utility grows as $O(n^2)$…

winner takes all → natural monopolies → need to choose: toll booth or commons?

Low upfront investment

remember Monica Lam's talk

For OSS, this may mean *built and deployed* without consensus, governance, sustainability plans

quote from OW2

Tragedy of the commons for OSS

Free riders take time to learn the lesson

Illusion of democracy for OSS

Good technology is *not* democratic

Outline

1. Introduction
2. Selected remarks
3. Software Heritage: highlights
4. Conclusion

Source code: *executable* and *human readable*

"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");
}
```

Source code is 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)

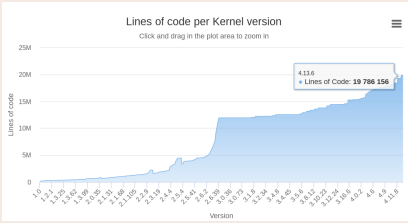
"Source code provides a view into the mind of the designer."

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!)



Software Heritage, in a nutshell

Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Mission

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

Principles and strategic choices

Raise awareness

- make visible the unseen
- actions in the policy space
- build communities

Plan for the long term… not a "project"!

- shared infrastructure
- multiple funders, mirror network
- incubate inside an institution

Technology

- transparency and FOSS
- replicas all the way down

Content

- intrinsic identifiers
- facts and provenance

Organization

- non-profit
- multi-stakeholder

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

An international, non profit initiative

Sharing the vision

United Nations
Educational, Scientific and
Cultural Organization

And many more …

`www.softwareheritage.org/support/testimonials`

Donors, members, sponsors

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Platinum sponsors

intel
Microsoft
SOCIETE GENERALE

Golden sponsor

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

Silver sponsors

CASTFlow
Software Intelligence for Digital Leaders
Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Google
UNIVERSITÀ DI PISA
vmware

Bronze sponsors

DANS
FOSSID
GitHub
NOKIA Bell Labs
UQÀM
Université du Québec à Montréal

Actions in the policy space: zoom on the EU

Europe: Copyright reform — adopted March 28 2019

"Upload filters": a threat to all modern software development

- developing platforms (GitHub, GitLab, Bitbucket, etc.)
- distribution platforms (Maven, Pypi, CRAN, CTAN, etc.)
- archives (Software Heritage)

We got an exclusion for

~~non for profit~~ open source software developing and sharing platforms

Key role of Software Heritage

policy-maker awareness, essential insights for NGOs, government contacts

Bottomline

Building infrastructures like Software Heritage is much more than technology!

Outline

1. Introduction
2. Selected remarks
3. Software Heritage: highlights
4. Conclusion

Come in, we're open!

Software Heritage

www.softwareheritage.org @swheritage

Thank you!

Learn more on www.softwareheritage.org/publications

Jean-François Abramatic, Roberto Di Cosmo, Stefano Zacchiroli
Building the Universal Archive of Source Code, Communications of the ACM, October 2018