Software Heritage
an archive to enable our digital future

Roberto Di Cosmo
Director, Software Heritage
Inria and Université de Paris

30 November 2021
#SWH5YEARS at UNESCO
1. Software and Source Code

2. Software Heritage: a mission at the service of Humankind

3. The road ahead
Software is all around us
Software is built from **Source Code**

**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.”

**Apollo 11 source code** (excerpt)

```
P63SPOT3 CA BIT6 # IS THE LR ANTENNA IN POSITION 1 YET
     EXTEND
     RAND CHAN33
     EXTEND
     BZF P63SPOT4 # BRANCH IF ANTENNA ALREADY IN POSITION 1
     CAF CODE580 # ASTRONAUT: PLEASE CRANK THE
     TC BANKCALL # SILLY THING AROUND
     CADR GOPERF1
     TCF GOTOP00M # TERMINATE
     TCF P63SPOT3 # PROCEED SEE IF HE’S LYING

P63SPOT4
     TC BANKCALL # ENTER INITIALIZE LANDING RADAR
     CADR SETPOS1
     TC POSTJUMP # OFF TO SEE THE WIZARD ...
     CADR BURNBABY
```

**Quake III source code** (excerpt)

```
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);  // 2nd iteration, this can be removed
    y = y * (threehalves - (x2 * y * y)); // 1st iteration
    return y;
}
```

**Len Shustek, Computer History Museum** 2006

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

R. Di Cosmo roberto@dicosmo.org @rdicosmo (CC-BY 4.0)
UNESCO, Inria, Software Heritage invite 40 international experts meet in Paris …

The call is published on February 2019

“Recognise software source code as a fundamental enabler in all aspects of human endeavour"
Yuval Noah Harari (on COVID 19)

“The real antidote [to epidemic] is scientific knowledge and global cooperation.”

Software powers modern scientific research

The top 100 papers

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

Nature, October 2014

We can still talk to the early inventors

"Telling historical stories is the best way to teach. It’s much easier to understand something if you know the threads it is connected to."

Donald E. Knuth
Len Shustek
CACM, January 2021

We need a dedicated infrastructure to preserve and share all this knowledge!
Software complexity is growing…the largest part of modern applications is…just reuse!

…it is important to Know Your SoftWare (KYSW)

Politique publique de la donnée, des algorithmes et des codes sources
...animer les écosystèmes des...réutilisateurs du source code

Circulaire du Premier Ministre, 27 Avril 2021, France

Sec. 4. Enhancing Software Supply Chain Security

ensuring and attesting, to the extent practicable, to the integrity and provenance of open source software

May 2021 POTUS Executive Order

We need a trusted knowledge base with software provenance!
Software source code is fragile

Endangered source code …

- **link rot**: projects are created, moved around, removed
- **data rot**: physical media with legacy software decay
- **platform consolidation**

2015 Google Code and Gitorious.org shutdown:
~1M endangered repositories

2019 250,000 endangered repositories on BitBucket

… is endangered knowledge!

broken links and missing pieces in the web of knowledge of humankind

Bottomline: we need a global, long term effort

to build a universal archive of all software source code
Outline

1. Software and Source Code

2. Software Heritage: a mission at the service of Humankind

3. The road ahead
Software Heritage in a nutshell

Unveiled in 2016

Collect, preserve and share all software source code

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

Reference catalog

find and reference all software source code

Universal archive

preserve all software source code

Research infrastructure

enable analysis of all software source code
Five year later: a universal software archive, as a shared infrastructure

All the software source code, with all its development history
An operational, evolving infrastructure

Harvest and archive

- save.softwareheritage.org
- deposit.softwareheritage.org

Reference (20 billion SWHIDs)

Intrinsic, decentralised, cryptographically strong identifiers, SWHIDs

Now supported in SPDX 2.2, Wikidata etc.

Adoption in Open Science

reference archive for research software

Adoption in Industry and Public Administration

reference archive and knowledge base for open source software

Reference

swb:1:cnt:41db23118f92d721b099a5e7a990cf58f1d07fa

- origin=https://github.com/chrislgarry/Apollo-11
- visit=swb:1:cmp:206c27e00315a8ac0b7f6eddabb9e0b28d8a9836
- anchors=swb:1:rev:3913f18f4383dd638e0485d6a902ff2f35820
- path=/luminary099/BUSH_BUSH--MASTER_IGNITION_ROUTINE.agc
- limes=64-72
A walkthrough

General

- **Browse** the archive
- Get and use SWHIDs, e.g. Apollo 11 excerpt, Quake III excerpt
- **Trigger archival** of a software repository in one click

Open Science

- **Curated deposit** via HAL, e.g.: LinBox, SLALOM, Givaro, SumGra, Coq proof, ...
- **Cite software** with the biblatex-software style, e.g.: article from IPOL

History of software

Rescue landmark legacy software, see the SWHAP process

Public code

Archived source code from code.gouv.fr
An international, non profit initiative
for the long term

Sharing the vision

United Nations Educational, Scientific and Cultural Organization

Donors, members, sponsors

Diamond sponsor

Platinum sponsors

Gold sponsors

Silver sponsors

Bronze sponsors

And many more ...

www.softwareheritage.org/support/testimonials

R. Di Cosmo roberto@dicosmo.org @rdicosmo (CC-BY 4.0)
A growing community

Team

Ambassadors

Foundations and grantees

- Castalia, CottageLabs
- EasterEggs, OcamlPro
- Octobus, Sperling, Tweag.io

Mirrors and storage partners

“Let us save what remains: … by such a multiplication of copies, as shall place them beyond the reach of accident.”

— Thomas Jefferson

Enea, FossId
CEA, RedHat

R. Di Cosmo roberto@dicosmo.org @rdicosmo (CC-BY 4.0)
Outline

1. Software and Source Code

2. Software Heritage: a mission at the service of Humankind

3. The road ahead
Listening to key partners

- Open Science
- Cultural Heritage
- Industry
- Public Administration
Building a global movement

Bring together all the stakeholders, and move to the next level.

Software Heritage

www.softwareheritage.org  @swheritage

The Library of Alexandria of code

- recover the past
- structure the future

The Very Large Telescope for Source code

- build better software
  - for industry
  - for society as a whole