Software Heritage
joining forces with GitHub to preserve humankind’s source code

Roberto Di Cosmo
Director, Software Heritage

Nov 13th, 2019
Software source code: human readable and executable 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.”

Apollo 11 source code (excerpt)

Quake III source code (excerpt)

Len Shustek, Computer History Museum (2006)

“Source code provides a view into the mind of the designer.”
Precious asset, *endangered heritage*

We are at a turning point
key people *are passing away* …
software development *skyrockets* …
and yet *we are missing* …

**A catalog**

**An archive**

**A research infrastructure**

it is *urgent* to take action!
Our mission

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

Past, present and future

Preserving the past, enhancing the present, preparing the future
An international, non profit initiative built for the long term

Sharing the vision

Donors, members, sponsors

And many more ...

www.softwareheritage.org/support/testimonials

Platinum sponsors

Golden sponsor

Silver sponsors

Bronze sponsors
Global development history permanently archived in a *unique* git-like Merkle DAG

- ~400 TB (uncompressed) blobs, ~20 B nodes, ~280 B edges
Milestones

**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

**December 7th 2018**

Starting the mirror network

**February 26th 2019**

Publication of the expert meeting Paris Call on Software Source Code
Today: joining forces with GitHub

Paris Call on Software Source Code

UNESCO, February 2019

[We call to] support:

- “efforts to gather and preserve the artifacts and narratives of the history of computing, while the earlier creators are still alive”
- “a universal archive, as part of a broad effort at digital preservation, that will ensure persistence of and universal access to software source code”

Support SWH Acquisition Process

Better GitHub archival in SWH

https://www.softwareheritage.org/swhap
Join the effort, support the mission!

Software Heritage

www.softwareheritage.org  @swheritage

Library of Alexandria of code

- Working together to
  - recover the past
  - preserve our heritage
  - share the knowledge
  - prepare the future

Learn more  www.softwareheritage.org/publications

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

Roberto Di Cosmo, Morane Gruenpeter, Stefano Zacchirolı
Identifiers for Digital Objects: the Case of Software Source Code Preservation, iPRES 2018: Intl. Conf. on Digital Preservation

Roberto Di Cosmo, Stefano Zacchirolı
Software Heritage: Why and How to Preserve Software Source Code, iPRES 2017: Intl. Conf. on Digital Preservation