Software Heritage
A revolutionary infrastructure for Open Source

Roberto Di Cosmo

June 17th, 2020
OW2Con - Paris
Outline

1. Knowing Open Source Software
2. The Software Heritage initiative
3. A revolutionary infrastructure
4. Zoom on selected industry use cases
5. Conclusion
Software is eating the world

**Why Software Is Eating The World**

By Marc Andreessen  
August 20, 2011

This week, Hewlett-Packard (where I am on the board) announced that it is exploring jettisoning its struggling PC business in favor of investing more heavily in software, where it sees better potential for growth. Meanwhile, Google plans to buy up the cellphone handset maker Motorola Mobility. Both moves surprised the tech world. But both moves are also in line with a trend I’ve observed, one that makes me optimistic about the future.

---

**Software companies outperform or buy out traditional companies**

Marc Andreessen, 2011

---

Reuse is the new rule

80% to 90% of a new application is ... just reuse!

(Sonatype survey, 2017)
Reuse is the new rule …

Where does reused software come from?

Do you know where it comes from?

- the software you ship
- the software you use
- the software you acquire
- the software that
  - has that bug
  - has that vulnerability

KYSW: Know Your SoftWare

Like KYC in banking, KYSW is now essential all over IT

… we need a common infrastructure to track all Open Source software!

Roberto Di Cosmo
www.softwareheritage.org

Software Heritage: key infrastructure CC-BY 4.0

June 17th, 2020
Outline

1. Knowing Open Source Software
2. The Software Heritage initiative
3. A revolutionary infrastructure
4. Zoom on selected industry use cases
5. Conclusion
Software Heritage, in a nutshell

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
An international, non profit initiative built for the long term

Sharing the vision

United Nations Educational, Scientific and Cultural Organization

Donors, members, sponsors

Platinum sponsors

Gold sponsor

Silver sponsors

Bronze sponsors

And many more ...

www.softwareheritage.org/support/testimonials

www.softwareheritage.org
Outline

1. Knowing Open Source Software
2. The Software Heritage initiative
3. A revolutionary infrastructure
4. Zoom on selected industry use cases
5. Conclusion
full development history permanently archived!

over 8 billions unique source files from 120+ million origins
A revolutionary infrastructure for software source code

The graph of Software Development

All software development with its history, in a single graph ...

The blockchain of Software Development

... a single Merkle graph, with intrinsic ids for traceability

A pillar of Open Science

Reference archive of Research Software

Reference platform for Big Code

One uniform data structure enables massive machine learning for quality, cybersecurity, etc.
Outline

1. Knowing Open Source Software
2. The Software Heritage initiative
3. A revolutionary infrastructure
4. Zoom on selected industry use cases
5. Conclusion
Software Heritage Identifiers (SWHID)

An emerging standard
- in Linux Foundation’s SPDX 2.2
- IANA registered, WikiData property P6138

Examples:
- Apollo 11 AGC excerpt,
- Quake III rsqrt
## Industry use cases (selection)

<table>
<thead>
<tr>
<th>Open Source complete and corresponding source code distribution</th>
<th>(Intel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software Heritage members can:</td>
<td></td>
</tr>
<tr>
<td>- archive source code in Software Heritage, distribute only the SWHID</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Traceability and integrity</th>
<th>(OIN for the <em>Linux System Definition</em>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software Heritage members can:</td>
<td></td>
</tr>
<tr>
<td>- archive source code in Software Heritage</td>
<td></td>
</tr>
<tr>
<td>- track it and verify its integrity using its SWHID</td>
<td></td>
</tr>
</tbody>
</table>

### And much more!

- compliance (collaborations with Intel, FossId, CAST, …)
- security (ongoing collaboration, US Department of Commerce)
- supply chain management, long term archive

*add your use case here*
Outline

1. Knowing Open Source Software
2. The Software Heritage initiative
3. A revolutionary infrastructure
4. Zoom on selected industry use cases
5. Conclusion
Join the revolution!

Software Heritage

www.softwareheritage.org  @swheritage

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

Becoming a sponsor

https://sponsorship.softwareheritage.org