

Software Heritage

Collecting, Preserving and Sharing all the Source Code

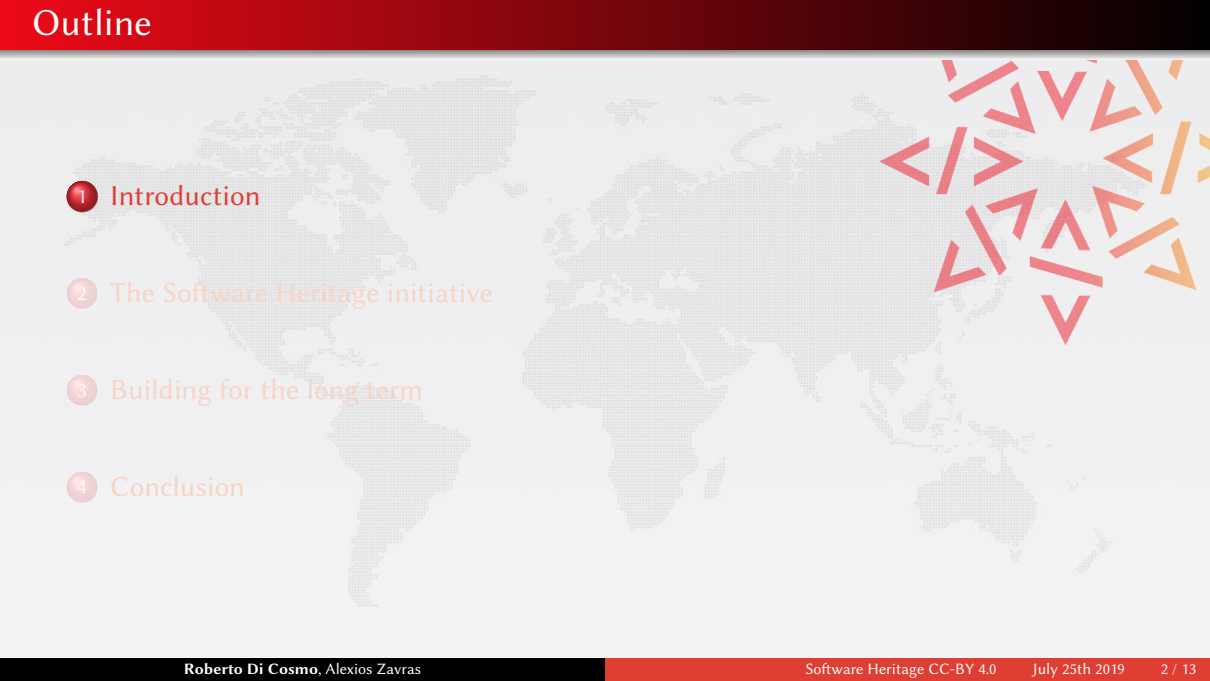
Roberto Di Cosmo, Alexios Zavras

July 25th, 2019



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

- 
- 1 Introduction
 - 2 The Software Heritage initiative
 - 3 Building for the long term
 - 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");
}
```

An essential part of our knowledge

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

Today...

THE WALL STREET JOURNAL.

Home World U.S. Politics Economy Business Tech Markets Opinion Arts

ESSAY

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

We are at a turning point

Losing our heritage

- the **key people** are passing away (J.M. Hullot last month!)
- we miss **a catalog**
- we miss **an archive**

Research lagging behind

- **software** undervalued in research
- **source code** not a first class citizen
- **computing** needs *historians and archivists*

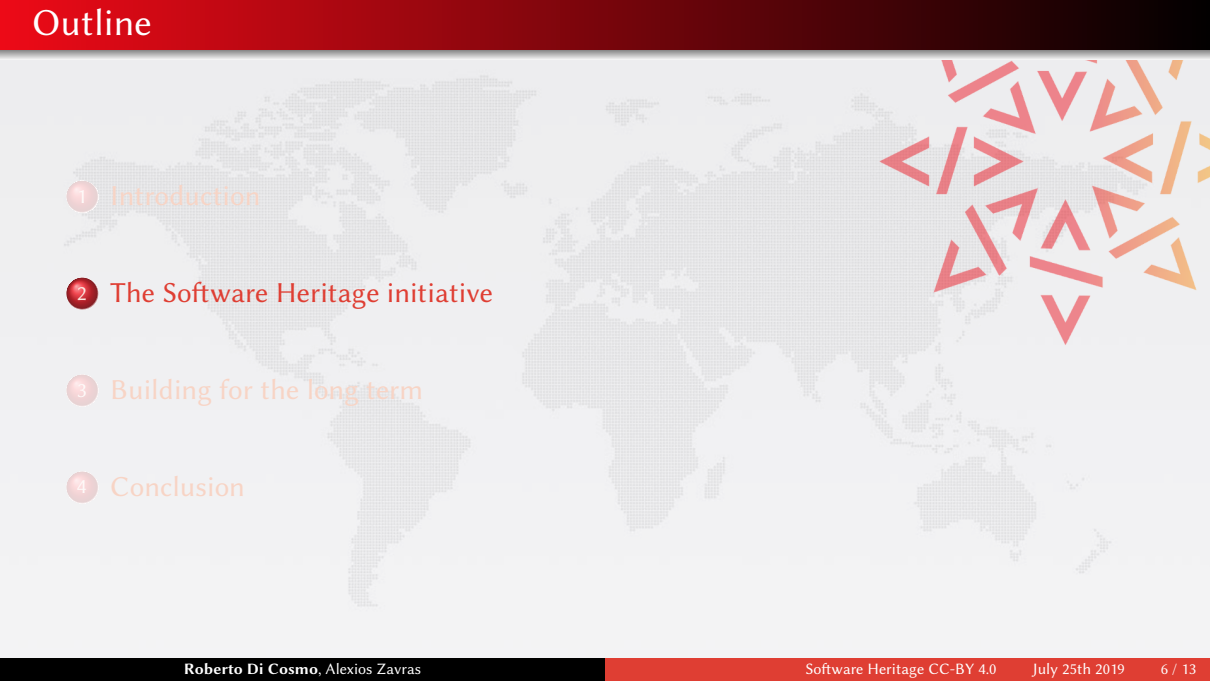
Open source is 'only reliable way' to preserve human history, argues Vatican

Proprietary solutions are not the Holy Grail, warns Vatican Library CIO

LAS VEGAS: OPEN SOURCE is key for humanity to preserve its history in the digital age, Vatican Library CIO Luciano Ammenti has argued.

Ammenti made the bold claim at at EMC World in Las Vegas, attended by *The INQUIRER*, citing his experience digitising the Vatican's 82,000 manuscripts as proof.

"The Vatican Library is a conservation library. We try to preserve our history. We tried to expand the number of reading rooms available for people that want to use our library," he said.

- 
- 1 Introduction
 - 2 The Software Heritage initiative
 - 3 Building for the long term
 - 4 Conclusion



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

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



Technology

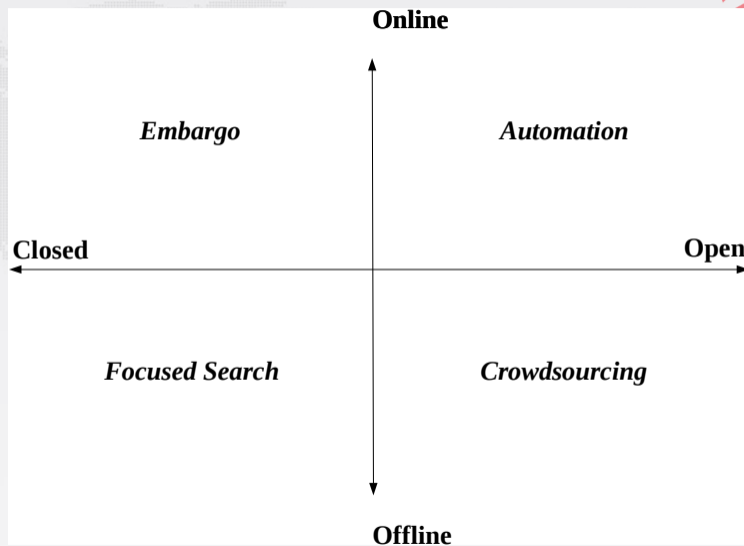
- transparency and FOSS
- replicas all the way down

Content

- intrinsic identifiers
- facts and provenance

Organization

- non-profit
- multi-stakeholder



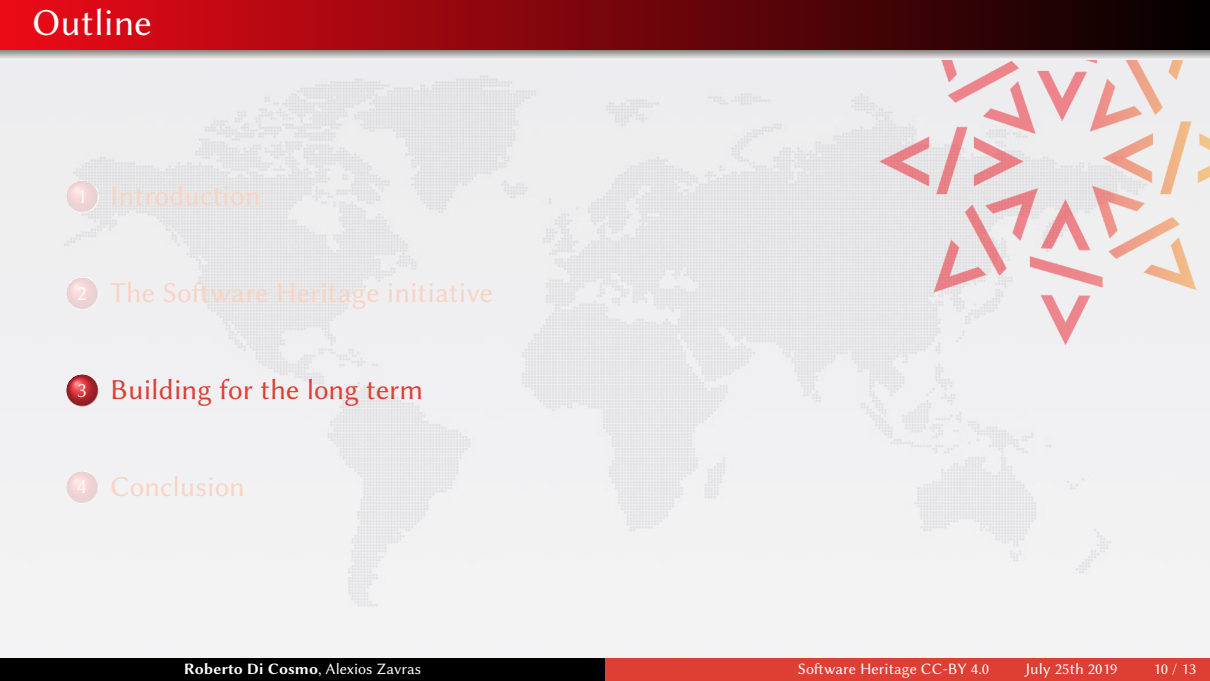
You get intrinsic persistent identifiers ...

bit.ly/swhpidpaper

- example: `swh:1:cnt:41ddb23118f92d7218099a5e7a990cf58f1d07fa;lines=53-82`

Advantages

- no link rot
- no middle man
- browsable and downloadable
- stable references for historical research
- see for example: [blog post on 50th Apollo 11 anniversary](#)

- 
- 1 Introduction
 - 2 The Software Heritage initiative
 - 3 Building for the long term
 - 4 Conclusion

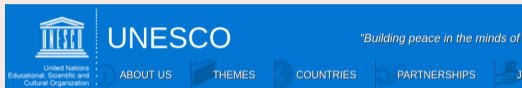
Raising Awareness

April 3rd 2017, Unesco Inria agreement



Roberto Di Cosmo, Alexios Zavras

November 2018, Unesco Inria expert call



Home > All News > Experts call for greater recognition of software source code as heritage for sustainable development

Experts call for greater recognition of software source code as heritage for sustainable development

16 November 2018

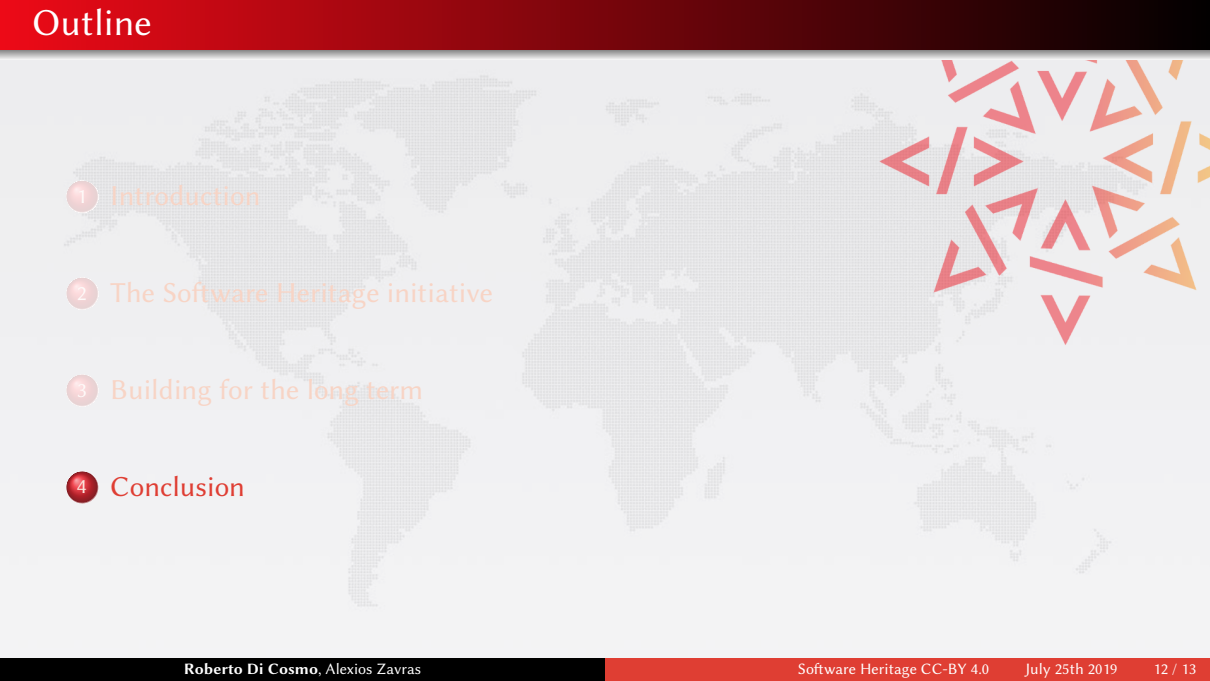


Sharing the vision



Donors, members, sponsors



- 
- 1 Introduction
 - 2 The Software Heritage initiative
 - 3 Building for the long term
 - 4 Conclusion

You can help!

Collect

- help build a *shared process*
- contribute to the *focused search*
- collect *oral history* while we can!

Preserve

- host *ancillary material*
- join *the SWH Foundation*
- become a *mirror*

Share

- build *on top of* SWH
- use it for your own research



Software Heritage

www.softwareheritage.org

[@swheritage](https://twitter.com/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 Zacchioli
Building the Universal Archive of Source Code, Communication of the ACM, October 2018



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



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