Software Heritage
Inria Developer Meetup

Nicolas Dandrimont
Software Heritage
nicolas@dandrimont.eu

March 19th, 2019
Software is everywhere

Software embodies a growing part of our scientific, technical and cultural heritage!

Nicolas Dandrimont
www.softwareheritage.org March 19th, 2019
Software is everywhere

So/f_tware embodies a growing part of... … our *scientific*, technical and Cultural Heritage!
The source code for a work means the preferred form of the work for making modifications to it.
"The source code for a work means the preferred form of the work for making modifications to it."

Hello World
“The source code for a work means the preferred form of the work for making modifications to it.”

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 ff ff ff
4004f9: 90
4004fa: 5d
4004fb: c3
Source Code: executable and human readable knowledge

“The source code for a work means the preferred form of the work for making modifications to it.”

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");
}

Nicolas Dandrimont

www.softwareheritage.org March 19th, 2019
We are at a turning point

Preserve the past

Only a few years left to recover the history of software technology
We are at a turning point

Preserve the past

Only a few years left to recover the history of software technology

Improve the future

We need a universal platform for all the future software source code
Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all
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
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
Software Heritage, in a nutshell

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

Research infrastructure
enable analysis of all the source code

Nicolas Dandrimont
www.softwareheritage.org  March 19th, 2019 5 / 10
Collect

Scheduling
Listing (full/incremental)

...
Collect

Software Heritage Archive
Merkle DAG + blob storage

Loading & deduplication

Scheduling

Listing (full/incremental)

Package repos

CPAN

Distros

Debian lister

PyPi lister

Forges

GitHub lister

GitLab lister

Mercurial lister

Debian source package loader

tar loader

Debian

PyPi

Releases

Revisions

Directories

Contents

Snapshots

Nicolas Dandrimont

www.softwareheritage.org

March 19th, 2019
- all the software source code
- full development history permanently archived
- push deposit for scientific software source code
Preserve

Build a mirror network

- resilience
- biodiversity
- work in progress with FOSSID

“Let us save what remains: not by vaults and locks which fence them from the public eye and use in consigning them to the waste of time, but by such a multiplication of copies, as shall place them beyond the reach of accident.”

Thomas Jefferson
Open collaborations

- Open Data publications ("Make your own queries on our dataset")
- Research collaborations (e.g. on efficient representations of our graph)
- Industry partnerships (e.g. with Qwant on making a search engine scale to our needs)

Open to the World

- FLOSS development workflow (copyleft Free Software, open Forge, Google Summer of Code participation)
- Public wayback machine for Software Source Code
- Public exports of the dataset
Yes, now you can!

Wayback machine for source code

*retrieve* the source code as it was

Reference catalog

*use intrinsic identifiers for software*

Open science

*deposit* scientific software (via HAL)

Universal knowledge base

*store the knowledge about source code*

And much, much more is in store!

With your help?

Nicolas Dandrimont
Come in, we’re open!

Software Heritage

www.softwareheritage.org  @swheritage

Everybody is concerned, everybody can help build

The 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

Nicolas Dandrimont  www.softwareheritage.org  March 19th, 2019