Towards a Software Pillar for Open Science

from policy to implementation

Roberto Di Cosmo

Chair, Software Chapter, National Committee for Open Science
Director, Software Heritage
Inria and Université de Paris Cité

July 11th 2023

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Outline

1. Introduction
2. Free and Open Source Software
3. Understanding the tidal waves
4. Software as a Pillar of Open Science
5. An emerging policy framework
6. Towards implementation: a perspective from France
7. Need for a global approach: the case of A.R.D.C.
8. Actions

Short bio

Computer Science professor in Paris, now working at INRIA

- *30+ years* of research (Theor. CS, Programming, Software Engineering, Erdos #: 3)
- *20+ years* of Free and Open Source Software
- *10+ years* building and directing structures for the common good

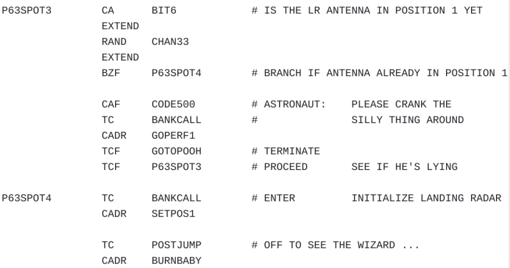
1999	*DemoLinux* – first live GNU/Linux distro
2007	*Free Software Thematic Group* 150 members 40 projects 200Me
2008	*Mancoosi project* www.mancoosi.org
2010	*IRILL* www.irill.org
2015	*Software Heritage* at INRIA
2018	*National Committee for Open Science*, France
2021	*EOSC Task Force on Infrastructures for Software*, European Union

Software *Source Code* is Precious Knowledge

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)



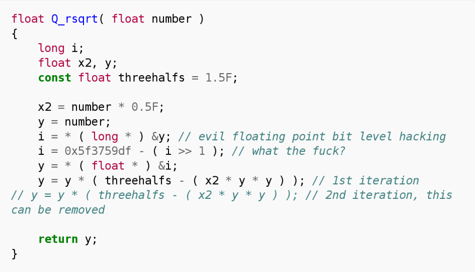
```
P63SPOT3        CA      BIT6            # IS THE LR ANTENNA IN POSITION 1 YET
                EXTEND
                RAND    CHAN33
                EXTEND
                BZF     P63SPOT4        # BRANCH IF ANTENNA ALREADY IN POSITION 1

                CAF     CODE500         # ASTRONAUT:    PLEASE CRANK THE
                TC      BANKCALL        #               SILLY THING AROUND
                CADR    GOPERF1
                TCF     GOTOPOOH        # 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 threehalfs = 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;
    y = y * ( threehalfs - ( x2 * y * y ) ); // 1st iteration
// y = y * ( threehalfs - ( x2 * y * y ) ); // 2nd iteration, this
can be removed

    return y;
}
```

Len Shustek, Computer History Museum 2006

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

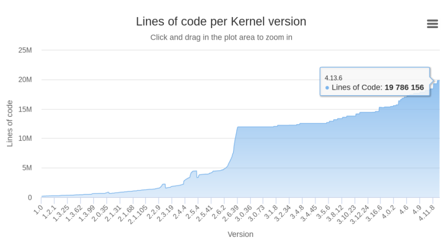
~ 50 years, a lightning fast growth

Apollo 11 Guidance Computer (~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

Linux Kernel



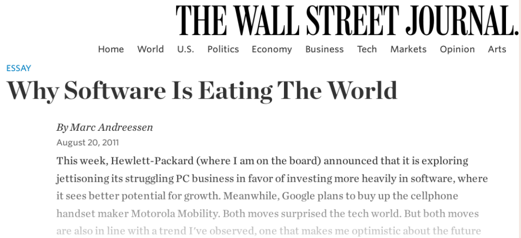
… now in your pockets!

Outline

1. Introduction
2. Free and Open Source Software
3. Understanding the tidal waves
4. Software as a Pillar of Open Science
5. An emerging policy framework
6. Towards implementation: a perspective from France
7. Need for a global approach: the case of A.R.D.C.
8. Actions

Software is eating the world…

Business



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

Software companies

outperform or buy out

hardware companies

Marc Andreesen, 2011

Technology

Software Defined Everything

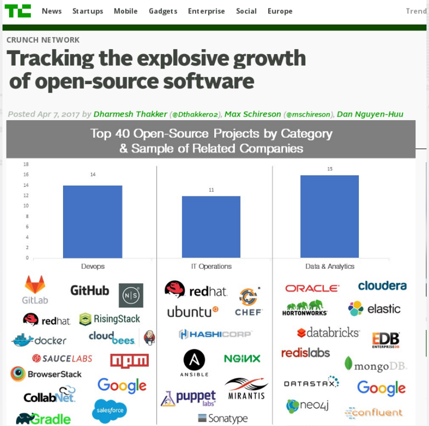
Hardware gets commoditised

Software becomes the new value!

```
<td class="menuItem" colspan="4" style="background: url(images/www_0
</tr>
<tr>
<td colspan="4" style="background: url(images/www_06.jpg); vertical-align: top;">
<div>
<div>
<table border="0" width="423" cellspacing="0" cellpadding="0"
<tr class="pfmargins">
<td width="195" valign="bottom" align="right">
<a href="portfolio_logo.html">
<img src="images/log.png" alt="" border="0"
</a>
</td>
<td width="106" valign="bottom">
<tr>
<td colspan="3" align=
<div class="photosPortfolioMargins">
```

Worldwide Software Defined Everything (SDE) Market to grow at a CAGR of 27.9% over the period 2016–2022 to aggregate $143.35 billion by 2022

… Open Source is eating the Software World



News Startups Mobile Gadgets Enterprise Social Europe Trend

CRUNCH NETWORK

Tracking the explosive growth of open-source software

Posted Apr 7, 2017 by Dharmesh Thakker (@Dthakkero2), Max Schireson (@mschireson), Dan Nguyen-Huu

Open Source Software

can be openly (re)used, modified, (re)distributed, *with full access to its source code!*

Outline

1. Introduction
2. Free and Open Source Software
3. **Understanding the tidal waves**
4. Software as a Pillar of Open Science
5. An emerging policy framework
6. Towards implementation: a perspective from France
7. Need for a global approach: the case of A.R.D.C.
8. Actions

Free Software: 40 years, 4 layers, in a nutshell

First 15 years: 1984-... The early revolution

focus	*freedom* for users and (especially) developers
keyword	free software

The second wave: 1999-... Progressive industry adoption

focus	software quality and reduced cost
keyword	open source (~25th anniversary!)

The third wave: 2010-... Ecosystems, strategic alignment

focus	community, organisation, foundations
keyword	governance

The fourth wave: 2015-... Industry consolidation

focus	mergers and acquisitions
keyword	control

Key resources : competency, and adoption

We really are in a *knowledge* economy!

- competencies
- talent
- network
- adoption
- mindshare

Bottomline

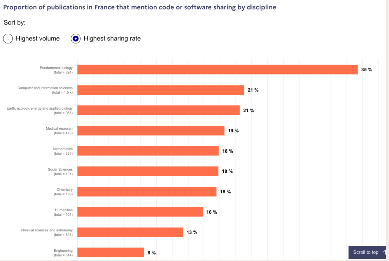
The infrastructure for (open) collaboration is the new competitive advantage!

Outline

1. Introduction
2. Free and Open Source Software
3. Understanding the tidal waves
4. Software as a Pillar of Open Science
5. An emerging policy framework
6. Towards implementation: a perspective from France
7. Need for a global approach: the case of A.R.D.C.
8. Actions

Software is a pillar of Open Science

Software powers modern research



Proportion of publications in France that mention code or software sharing by discipline

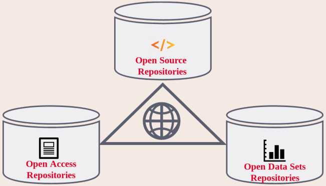
Sort by:

○ Highest volume ● Highest sharing rate

Over 20% of articles across all disciplines share software

2023 French Open Science Monitor

Key pillar: software



Links are important

Nota Bene

software may be a *tool*, a *research outcome* and a *research object*

access to the *source code* is essential!

Preserving (the history of) source code is necessary for *reproducibility*

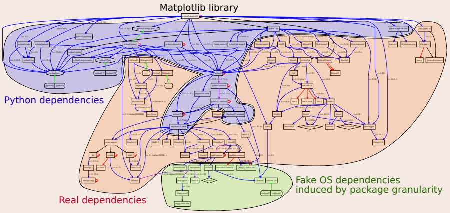
Source code is *special* (software is *not* data)

Software *evolves* over time

- projects may last decades
- the *development history* is key to its *understanding*

Complexity

- *millions* of lines of code
- large *web of dependencies*
  - easy to break, difficult to maintain
  - *research software* a thin top layer
- sophisticated *developer communities*



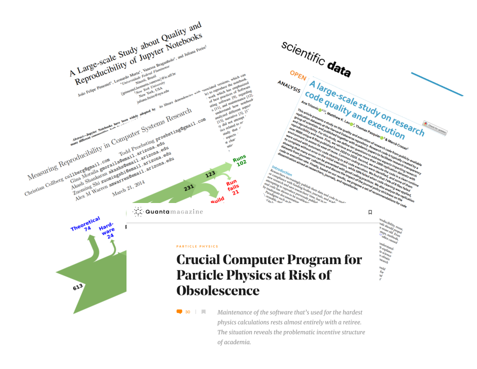
The human side

design, algorithm, code, test, documentation, community, funding

and so many more facets . . .

How are we managing our software ?

Reproducibility, maintenance in Academia



A Large-scale Study about Quality and Reproducibility of Jupyter Notebooks

João Felipe Pimentel*, Leonardo Murta*, Vanessa Braganholo*, and Juliana Freire†
*Universidade Federal Fluminense
Niterói, Brazil
{jpimentel,leomurta,vanessa}@ic.uff.br
†New York University
New York, USA
juliana.freire@nyu.edu

Abstract—Jupyter Notebooks have been widely adopted by many different communities

Measuring Reproducibility in Computer Systems Research

Christian Collberg collberg@gmail.com
Todd Proebsting proebsting@gmail.com
Gina Moraila gmoraila@email.arizona.edu
Akash Shankaran akashs@email.arizona.edu
Zuoming Shi zuomingshi@email.arizona.edu
Alex M Warren amwarren@email.arizona.edu
March 21, 2014

scientific data

OPEN

ANALYSIS

A large-scale study on research code quality and execution

Ana Trisovic, Matthew K. Lau, Thomas Pasquier & Mercè Crosas

This article presents a study on the quality and execution of research code from publicly-available replication datasets at the Harvard Dataverse repository. Research code is typically created by a group of scientists and published together with academic papers to facilitate research transparency and reproducibility. For this study, we define ten questions to address aspects impacting research reproducibility and reuse. First, we retrieve and analyze more than 2000 replication datasets with over 9000 unique R files published from 2010 to 2020. Second, we execute the code in a clean runtime environment to assess its ease of reuse. Common coding errors were identified, and some of them were solved with automatic code cleaning to aid code execution. We find that 74% of R files failed to complete without error in the initial execution, while 56% failed when code cleaning was applied, showing that many errors can be prevented with good coding practices. We also analyze the replication datasets from journals' collections and discuss the impact of the journal policy strictness on the code re-execution rate. Finally, based on our results, we propose a set of recommendations for code dissemination aimed at researchers, journals, and repositories.

Introduction

Quanta magazine

PARTICLE PHYSICS

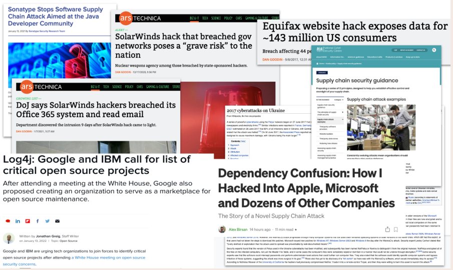
Crucial Computer Program for Particle Physics at Risk of Obsolescence

30

Maintenance of the software that's used for the hardest physics calculations rests almost entirely with a retiree. The situation reveals the problematic incentive structure of academia.

(articles: here, here, here and here)

Security, integrity, traceability in Industry



Sonatype Stops Software Supply Chain Attack Aimed at the Java Developer Community

January 13, 2021 By Sonatype Security Research Team

ars TECHNICA BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE STORE

ALERT —

SolarWinds hack that breached gov networks poses a "grave risk" to the nation

Nuclear weapons agency among those breached by state-sponsored hackers.

DAN GOODIN - 12/17/2020, 9:56 PM

Equifax website hack exposes data for ~143 million US consumers

Breach affecting 44 pe

DAN GOODIN - 9/8/2017, 12:31 AM

ars TECHNICA BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE STORE

GROWING LIST —

DoJ says SolarWinds hackers breached its Office 365 system and read email

Department discovered the intrusion 9 days after SolarWinds hack came to light.

DAN GOODIN - 1/7/2021, 3:27 AM

National Cyber Security Centre

Supply chain security guidance

Proposing a series of 12 principles, designed to help you establish effective control and oversight of your supply chain.

Supply chain attack examples

Constantly evolving attacks mean organisations should

2017 cyberattacks on Ukraine

From Wikipedia, the free encyclopedia

Contents [hide]
1 Approach
2 Attack
3 Attribution
4 Affected companies

Log4j: Google and IBM call for list of critical open source projects

After attending a meeting at the White House, Google also proposed creating an organization to serve as a marketplace for open source maintenance.

Written by Jonathan Greig, Staff Writer on January 13, 2022 | Topic: Open Source

Google and IBM are urging tech organizations to join forces to identify critical open source projects after attending a White House meeting on open source security concerns.

Dependency Confusion: How I Hacked Into Apple, Microsoft and Dozens of Other Companies

The Story of a Novel Supply Chain Attack

Alex Birsan 14 hours ago · 11 min read

Can they track the software that they

- ship, use, acquire
- has that bug or vulnerability

awareness is raising at the level of public policy

Outline

1 Introduction

2 Free and Open Source Software

3 Understanding the tidal waves

4 Software as a Pillar of Open Science

5 An emerging policy framework

6 Towards implementation: a perspective from France

7 Need for a global approach: the case of A.R.D.C.

8 Actions

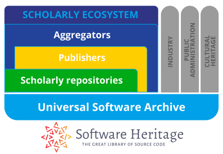
International highlights

Paris Call on Software Source code (2019, UNESCO)

40 international experts call to *“promote software development as a valuable research activity, and research software as a key enabler for Open Science/Open Research, […] recognising in the careers of academics their contributions to high quality software development, in all their forms”*

☞ Open Source in UNESCO recommendations for Open Science, 2021

Software in the EOSC



SCHOLARLY ECOSYSTEM

Aggregators

Publishers

Scholarly repositories

INDUSTRY

PUBLIC ADMINISTRATION

CULTURAL HERITAGE

Universal Software Archive

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

2020 EOSC SIRS connect scholarly ecosystem via Software Heritage
2021 EOSC Task Force on Infrastructures for Research Software
2022 FAIRCORE4EOSC project WP6 implements SIRS report
2023 INFRAEOSC call on quality of scientific software

And much more

Software track in OSEC 2022, Software working group launched in Science Europe, DFG adds software to model CV (9/22), NASA unveils Open Science policy (12/22), …

What is at stake

ARDC

- Archive for retrieval (*reproducibility*)
- Reference for identification (*reproducibility*)
- Describe for discovery and reuse
- Cite/Credit for credit and evaluation

Before ARDC

- Development practices and tools (VCS, build system, test suites, CI, code quality, ...)
- Opening up towards a community (documentation, organization, communication)

Need training, tooling, infrastructures, best practices

Beyond ARDC

- Policies (dissemination, reuse, careers, ...)
- Sustainability (legal, financial, etc.)
- Technology transfer
- Advanced technologies and tools (quality, traceability, etc.)

a humbling challenge, and a complex one (we are not in a vacuum)

Outline

1. Introduction
2. Free and Open Source Software
3. Understanding the tidal waves
4. Software as a Pillar of Open Science
5. An emerging policy framework
6. **Towards implementation: a perspective from France**
7. Need for a global approach: the case of A.R.D.C.
8. Actions

French National plan for Open Science, 2021-2024

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION
Liberté
Égalité
Fraternité

SECOND FRENCH PLAN FOR OPEN SCIENCE

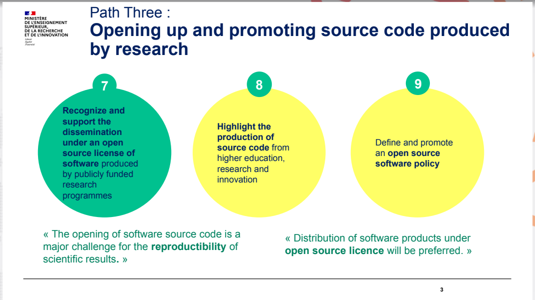
Generalising open science in France 2021-2024

Second French Plan for Open Science

GENERALISING OPEN SCIENCE IN FRANCE 2021-2024

Launch on 6 July 2021 by Frédérique Vidal, Minister for Higher Education, Research and Innovation

- Multiplying the **levers for change** in order to **generalise open science practices**
- Structuring the **policy for opening up or sharing research data**
- New commitments to the **opening of source code** produced by research
- **European and international inclusion** in the context of the French Presidency of the European Union
- **Disciplinary and thematic variations**: open science policies must be adapted to disciplinary specificities



Path Three : Opening up and promoting source code produced by research

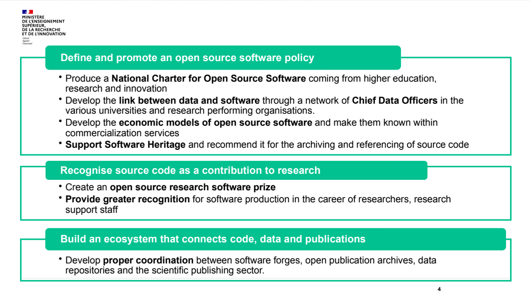
7 **Recognize and support the dissemination under an open source license of software** produced by publicly funded research programmes

8 **Highlight the production of source code** from higher education, research and innovation

9 Define and promote an **open source software policy**

« The opening of software source code is a major challenge for the **reproductibility** of scientific results. »

« Distribution of software products under **open source licence** will be preferred. »



Define and promote an open source software policy

- Produce a **National Charter for Open Source Software** coming from higher education, research and innovation
- Develop the **link between data and software** through a network of **Chief Data Officers** in the various universities and research performing organisations.
- Develop the **economic models of open source software** and make them known within commercialization services
- **Support Software Heritage** and recommend it for the archiving and referencing of source code

Recognise source code as a contribution to research

- Create an **open source research software prize**
- **Provide greater recognition** for software production in the career of researchers, research support staff

Build an ecosystem that connects code, data and publications

- Develop **proper coordination** between software forges, open publication archives, data repositories and the scientific publishing sector.

Software College in the CoSO

Five action lines

- Identifying and highlighting research software production
- Technical tools and best practices
- Translation and sustainability
- National, European, and International coordination
- Recognition, evaluation and careers

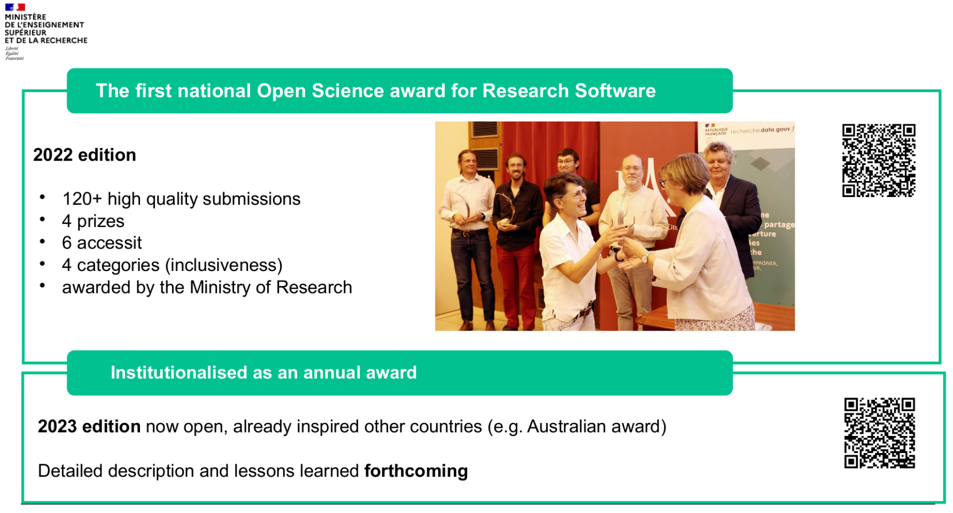
Leveraging experience and connections

- Open Source thematic group in Systematic (since 2007, more on demand)
- Collaboration with DINUM, Eclipse Foundation, OW2, …

Coordination with other colleges

- The Open Science passport software booklet

Recognizing Research Software: french national open science award



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
Liberté Égalité Fraternité

The first national Open Science award for Research Software

2022 edition

- 120+ high quality submissions
- 4 prizes
- 6 accessit
- 4 categories (inclusiveness)
- awarded by the Ministry of Research

Institutionalised as an annual award

2023 edition now open, already inspired other countries (e.g. Australian award)

Detailed description and lessons learned **forthcoming**

Software Chapter in the CoSO, cont'd

Twenty-three active members

Chairs: Roberto Di Cosmo and François Pellegrini

- Florent CHUFFART (Univ Grenoble Alpe)
- Mélanie CLÉMENT-FONTAINE (Univ Paris-Saclay - Versailles Saint-Quentin)
- Laurent COSTA (UMR 7041 ArScAn)
- Ludovic COURTÈS (Inria)
- Sébastien GÉRARD (Univ Paris-Saclay, CEA, List)
- Mathieu GIRAUD (CNRS, Univ Lille)
- Timothée GIRAUD (CNRS)
- Jean-Yves JEANNAS (Univ Lille, AFUL)
- Nicolas JULLIEN (IMT Atlantique)
- Daniel LE BERRE (Univ Artois, CNRS)
- Violaine LOUVET (CNRS / GRICAD - Univ Grenoble Alpes)
- Camille MAUMET (Inria, Univ Rennes, CNRS, Inserm)
- Clémentine MAURICE (CNRS)
- Grégory MIURA (Univ Bordeaux Montaigne)
- Raphaël MONAT (LIP6, Sorbonne Université)
- Patrick MOREAU (CNRS)
- Sophie RENAUDIN (AP-HP)
- Nicolas ROUGIER (Inria, Univ Bordeaux, CNRS)
- François SABOT (IRD)
- Sylvie TONDA-GOLDSTEIN (Inria)
- Samuel THIBAULT (Univ Bordeaux) (Univ Paris-Saclay)

Outline

1. Introduction
2. Free and Open Source Software
3. Understanding the tidal waves
4. Software as a Pillar of Open Science
5. An emerging policy framework
6. Towards implementation: a perspective from France
7. Need for a global approach: the case of A.R.D.C.
8. Actions

How (not) to preserve and share research software

A - Since the ~~1970's~~ 1990's

.zip or .tar file on:

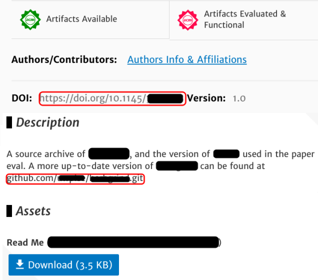
- ~~ftp server~~ (e.g. gnu)
- web page (example)
- document archive (+ DOI sample)

B - Since the 2000's

Rely on *software forges*

- institutional/project (e.g. example)
- free commercial ones: BitBucket, GitHub, GitLab, … (e.g. parmap)

C: a mix of the two



Artifacts Available

Artifacts Evaluated & Functional

Authors/Contributors: Authors Info & Affiliations

DOI: https://doi.org/10.1145/[redacted] **Version:** 1.0

Description

A source archive of [redacted], and the version of [redacted] used in the paper eval. A more up-to-date version of [redacted] can be found at github.com/[redacted]/[redacted].git

Assets

Read Me [redacted]

Can get no satisfaction…

A *Poor user experience*

B *No preservation guarantee*

C Can do *so much* better

Forges are *not* archives!

2015: the first big bad news

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

- broken links in the web of knowledge (my papers too)

Big bad news keep coming in

- summer 2019: BitBucket announces Mercurial VCS sunset
- july 2020: BitBucket erases *250.000+* repositories (including research software)
- summer 2022: GitLab.com considers erasing all projects that are inactive for a year

In Academia too!

- 2021: Inria's old gforge is unplugged… breaks the Opam build chain for OCaml

We need a universal archive of software source code: now we have one!

Software Heritage in a nutshell

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

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

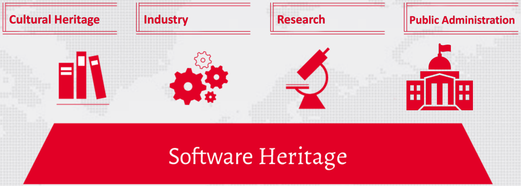
preserve and share all software source code

Research infrastructure

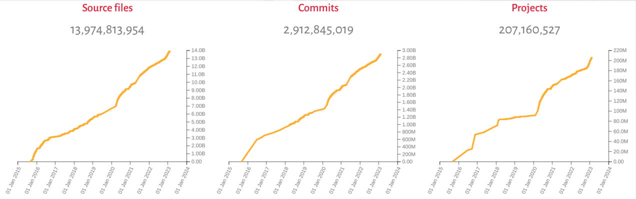
enable analysis of all software source code

The largest software archive, a shared infrastructure

One infrastructure
open and shared



The largest archive ever built


Source files

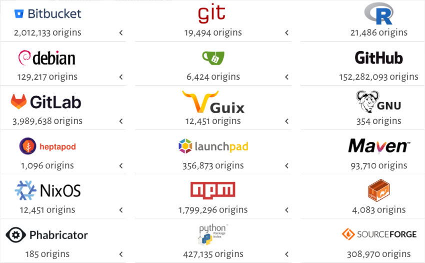
13,974,813,954

Commits

2,912,845,019

Projects

207,160,527



Bitbucket 2,012,133 origins	git 19,494 origins	21,486 origins
debian 129,217 origins	6,424 origins	GitHub 152,282,093 origins
GitLab 3,989,638 origins	Guix 12,451 origins	GNU 354 origins
heptapod 1,096 origins	launchpad 356,873 origins	Maven 93,710 origins
NixOS 12,451 origins	npm 1,799,296 origins	4,083 origins
Phabricator 185 origins	python 427,135 origins	SOURCEFORGE 308,970 origins

An international, non profit initiative built for the long term

Sharing the vision

UNESCO
United Nations
Educational, Scientific and
Cultural Organization

IEEE Software
GitLab
ENGINEERING
eclipse
OW2
fsfe
nlnet FOUNDATION
SIGSOFT
Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
INAPI
THE LINUX FOUNDATION
ADULLACT
INFORMATICS EUROPE
Computer History Museum
software freedom conservancy
Software Freedom Law Center
AdaCore The GNAT Pro Company
gandi.net
TDF
FREE SOFTWARE FOUNDATION
creative commons
SIF Société Informatique de France
open source initiative
openinventionnetwork

And many more ...
www.softwareheritage.org/support/testimonials

Donors, members, sponsors

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Diamond sponsor
cea

Platinum sponsors
cnrs
SOCIETE GENERALE
intel
Microsoft
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
HUAWEI

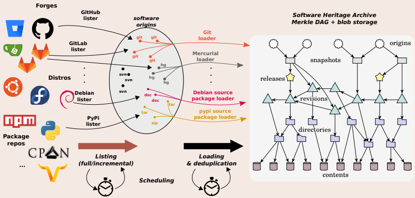
Gold sponsors
openinventionnetwork
SORBONNE UNIVERSITÉ
Université de Paris

Silver sponsors
AdaCore
MINISTÈRE DES ARMÉES
DGA
RÉPUBLIQUE FRANÇAISE
GitHub
Google
UNIVERSITÀ DI PISA

Bronze sponsors
Red Hat
SCANOSS
SCUOLA NORMALE SUPERIORE
ALMA MATER STUDIORUM

Addressing the four needs (see ICMS 2020 for details)

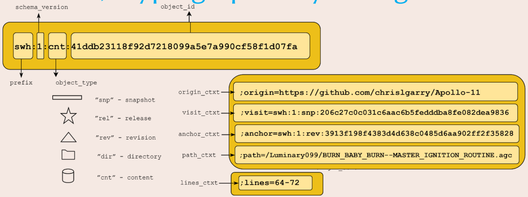
Archive (15B+ files, 240M+ projects)



- save now, updateswh, webhooks
- deposit.softwareheritage.org

Reference (30 billion SWHIDs)

Intrinsic, cryptographically strong IDs


schema_version
object_id
swh:1:cnt:41ddb23118f92d7218099a5e7a990cf58f1d07fa
prefix
object_type
"snp" - snapshot
"rel" - release
"rev" - revision
"dir" - directory
"cnt" - content
origin_ctxt
visit_ctxt
anchor_ctxt
path_ctxt
lines_ctxt
;origin=https://github.com/chrislgarry/Apollo-11
;visit=swh:1:snp:206c27c0c031c6aac6b5fedddba8fe082dea9836
;anchor=swh:1:rev:3913f198f4383d4d638c0485d6aa902ff2f35828
;path=/Luminary099/BURN_BABY_BURN--MASTER_IGNITION_ROUTINE.agc
;lines=64-72

Now in SPDX 2.2, Wikidata

Specification: `https://swhid.org`

Describe

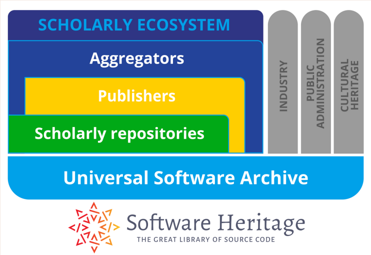
- *Intrinsic metadata* from source code
- Contributed the Codemeta generator

Cite/Credit

- Contributed *software citation* style
  biblatex-software, v 1.2-2 now on CTAN

Mutualization and standardisation at work

One archive, multiple infrastructures



universal software archive *Software Heritage* connects with the global software ecosystem

scholarly repositories institutional and disciplinary archives

publishers journals, proceedings, preprints

aggregators disciplinary catalogues, meta-portals, …

Building interconnection and interoperability FAIRCORE4EOSC HE (2022-2025)

Beta release: EOY 2023

Interconnection with SWH

repositories HAL, InvenioRDM, Dataverse

publishers Dagstuhl, episciences

agregators swMath, OpenAire

Interoperability

metadata schema *CodeMeta*

intrinsic identifier *SWHID*

specifications open/public

Outline

1. Introduction
2. Free and Open Source Software
3. Understanding the tidal waves
4. Software as a Pillar of Open Science
5. An emerging policy framework
6. Towards implementation: a perspective from France
7. Need for a global approach: the case of A.R.D.C.
8. Actions

Call to action: best practices for ARDC are available… today!

Archiving and referencing

For all source code used in research (*yes, even small scripts!*)

- ensure it is archived in Software Heritage (see save code now)
- get the proper SWHID for your software (see detailed HOWTO)
- add it to research articles for reproducibility (see detailed HOWTO)

Describing and Citing/Crediting

For software you want to put forward (*mention in your CV, reports, etc., get citations and credit for it*), do the following extra steps:

- add codemeta.json with description (see the codemeta generator)
- reference in the HAL portal (french partners, see online HAL documentation)
- cite software using the biblatex-software package (in CTAN and TeXLive)

- train students and colleagues
- engage journals, conferences, learned societies

Call to action: policy making

A working agenda

- avoid proprietarisation: set the default to open
  - *publicly funded research software should be open source*, exceptions must be justified
  - set up institutional support
  - build common knowledge base for technology transfer offices
- establish intelligent and effective incentives
  - count quality software contributions in careers, avoid purely numerical indicators, keep the human in the loop (mind Goodhart's law)
- avoid balkanisation, support mutualised common infrastructures
  - build on common, shared, open, non profit infrastructures, like Software Heritage
  - acknowledge the predominant human component of digital infrastructures
    - recurrent funding of their cost
    - proper evaluation of their service

The floor is yours

it's a long road, but together we can make it

Questions?

References

- UNESCO, *Draft recommendations on Open Science*
  2021, (online)

- French Ministry of Research, *Second National Plan for Open Science*
  2021, (online)

- EOSC SIRS Task Force, *Scholarly Infrastructures for Research Software*
  2020, Publications office of the European Commission, (10.2777/28598)

- R. Di Cosmo, *Archiving and Referencing Source Code with Software Heritage*
  International Conference on Mathematical Software 2020 (10.1007/978-3-030-52200-1_36)

- J.F. Abramatic, R. Di Cosmo, S. Zacchiroli, *Building the Universal Archive of Source Code*
  CACM, October 2018 (10.1145/3183558)