

Logiciel (libre): pilier de la Science Ouverte

Roberto Di Cosmo

Pilote du Collège Logiciel, SPSO
Directeur, Software Heritage
Inria and Université de Paris Cité

November 22nd 2023



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

- 1 Introduction
- 2 Software as a Pillar of Open Science
- 3 An emerging policy framework
- 4 France can lead the way
- 5 Need for a global approach



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)

```
P63SP0T3      CA      BIT6      # IS THE LR ANTENNA IN POSITION 1 YET
              EXTEND
              RAND   CHAN33
              EXTEND
              BZF    P63SP0T4      # BRANCH IF ANTENNA ALREADY IN POSITION 1

              CAF    CODE500      # ASTRONAUT:  PLEASE CRANK THE
              TC     BANKCALL      #              SILLY THING AROUND
              CADR   GOPERF1
              TCF    GOTOP00H      # TERMINATE
              TCF    P63SP0T3      # PROCEED    SEE IF HE'S LYING

P63SP0T4      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.”

Software Source Code: ~ 60 years, a lightning fast growth

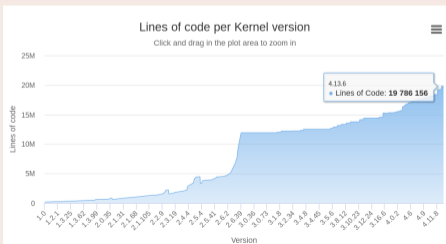
Apollo 11 Guidance Computer (~60.000 lines), 1969



"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!

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

Free Software, AKA: *Open Source*, FOSS, FLOSS,...

Software that offers to *its users* the freedom to:

- use the software
- study and adapt the software
- distribute software copies
- distribute modified copies

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

focus *freedom* (users, **developers**)

keyword free software (individual)

1999-... Progressive industry adoption

focus software quality, reduced cost

keyword open source (entities)

2010-... Ecosystems, strategic alignment

focus organisation, foundations

keyword governance and funding

2015-... Industry consolidation

focus mergers and acquisitions

keyword control

we faced many common issues *way before Open Science was on the radar*

- 1 Introduction
- 2 Software as a Pillar of Open Science
- 3 An emerging policy framework
- 4 France can lead the way
- 5 Need for a global approach



Why Open Science?

Open Science ([Second National Plan for Open Science](#), France, 2021)

Unhindered dissemination of results, methods and products from scientific research. It draws on *the opportunity provided by recent digital progress* to develop *open access to publications* and – as much as possible – *data, source code and research methods*.

Jean-Eric Paquet (EU DGRI, [on the objective of Open Science](#))

“Increase scientific quality, the pace of discovery and technological development, as well as societal trust in science.”

Mariya Gabriel ([EU Commissioner](#) for Research)

The COVID-19 crisis has also shown that cooperation at international level in research and innovation is more important than ever, including through *open access to data and results*. *No nation, no country can tackle any of these global challenges alone.*

Yuval Noah Harari (on COVID 19)

“The real antidote [to epidemic] is scientific knowledge and global cooperation.”

Two well known pillars of Open Science

Open Access (a long, painful, unfinished story)

- 19XX's compulsory exclusive copyright transfer to publishers (unlawful?) (notable exceptions: [US federal agencies](#) and [UK Crown Copyright](#))
 - 1990's Internet, Web and ArXiv break the [marriage of convenience of researchers with publishers](#)
 - 2000's declarations (Budapest, 2001; Berlin 7, 2009) and actions (LIPIcs, 2009)
 - 2010's reactions (SciHub, 2011; [Plan S](#), 2018) and transformations ([not so easy](#))
- TL;DR: see [my viewpoint in 2005](#) and [the SIGPLAN blog in 2020](#)

Open Data (less painful, but still unfinished story)

- 1957-1958: International Geophysical Year shows the way
- 2006 (and 2021): OECD recommendation on [publicly funded research data](#)
- 2016 and later: FAIR terminology (*focus on metadata, sort of forgets open...*)

Risk factors, mistakes to avoid

- legal and economic framework
 - closed, for profit infrastructures with unaligned goals may lead to
 - proprietarization of public research results
 - creation of dysfunctional markets
 - operation of open non profit infrastructure funded with project money
- operational balkanisation
 - proliferation of infrastructure silos
 - duplicated contents with different identifiers
 - costly efforts to federate after-the-fact
 - uneven quality of information

Taking notice

2021: [exemplarity criteria for the french national open science fund](#)

French National plan for Open Science, 2018-2021



Launch on 4 July 2021 by Frédérique Vidal, Minister for Higher Education, Research and Innovation

- First Commitment : **Generalise open access to publications**
- Second Commitment : **Structure research data and make it available through open access**
- Third Commitment : **Be part of a sustainable european and international open science dynamic**

1



Main achievements

- Creation of the **National Fund for Open Science** (11.6 M€ of resources over 3 years), a dedicated funding instrument for open science policy in France
- **Two calls for projects for "open access publishing, open publication and its ecosystem"**
- **Support for the national open archive HAL**: exceptional financial support, audit and ongoing renovation of the technical base, definition of a sustainable financing model and shared governance
- The **French National Research Agency** and other funding agencies request the **deposit of publications in an open archive** and the **drafting of data management plans** for the projects they fund
- **ANR flash call for projects on open science**: €2.3M, 25 projects supported to accelerate the maturation of disciplinary communities in the face of data management issues
- Creation of the **position of chief data officer** at the Ministry of Higher Education, Research and Innovation

3



Governance : The Open Science Committee

Open science steering committee

Ministry of Higher Education, Research and Innovation (MESRI), research performing organisations, Universities, National Research Agency (ANR), Couperin consortium, High Council for Evaluation of Research and Higher Education (Hcéres)

Makes decisions, arbitrates the use of funds from the National Open Science Fund

Permanent secretariat for Open Science

MESRI, research performing organisations, Universities, ANR, Couperin consortium, Hcéres, ADBU, EPRIST, Colleges

Prepares decisions, proposes guidelines, monitors work

Colleges

Publications, Research Data, Skills and Training, Europe and International
(72 members from 44 institutions)

Investigate issues, propose guidelines, initiate and manage projects

2



Main achievements

- Creation of the **open science barometer**: measurement of the rate of French scientific publications in open access
- Progressive deployment of **open science strategies within research performing organisations and universities**, creation of a network of open science referents
- **Training actions targeted at doctoral students**: Passport to open science, practical guide for PhD students
- Contribution to the **structuring and governance of the EOCS**: structuring of EOCS France, French presence on the board and other EOCS governance bodies
- **Support to international open science infrastructures**: SCoSS labelled projects (DOAB, PKP, OpenCitations), RDA, Software Heritage

4

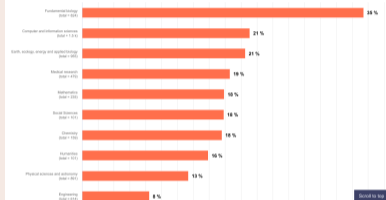
Software is a pillar of Open Science

Software powers modern research

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

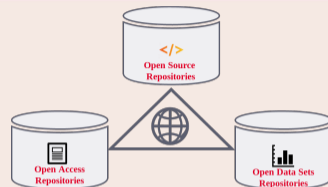
Sort by:

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*

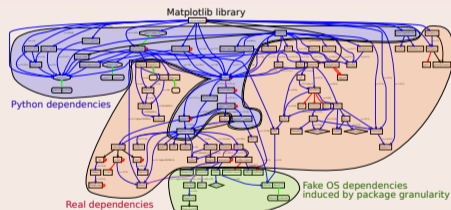
Source code is *special* (software is *not* data)

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



(articles: [here](#), [here](#), [here](#) and [here](#))

Security, integrity, traceability in Industry



Can they track the software that they

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

awareness is raising at the level of public policy

- 
- 1 Introduction
 - 2 Software as a Pillar of Open Science
 - 3 An emerging policy framework**
 - 4 France can lead the way
 - 5 Need for a global approach

International highlights

Paris Call on Software Source code (2019, UNESCO)

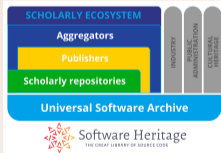


40 international experts call to “promote software development as a valuable research activity, and research software as a key enabler for Open Science/Open Research, [...] recognising in the careers of academics their contributions to high quality software development, in all their forms”



Open Source in UNESCO [recommendations](#) for Open Science, 2021

Software in the EOSC



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)

- 1 Introduction
- 2 Software as a Pillar of Open Science
- 3 An emerging policy framework
- 4 France can lead the way
- 5 Need for a global approach



French National plan for Open Science, 2021-2024



SECOND FRENCH PLAN FOR OPEN SCIENCE

Generalising open science in France 2021-2024



1

Second French Plan for Open Science



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

2

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

7

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

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

8

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

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

9

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

3

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.

4

Five action lines (see [details online](#))

- Identifying and highlighting research software production
- Technical and social tools and best practices
- Valorization and sustainability
- Liaison and animation at national, European, and international levels
- Recognition and careers

Source Code primer



key concepts

- for students
- for teachers
- for researchers



Report on software forges



in academia (FR):

- needs
- options
- limitations



Annual award

Establishing a national research software award.

Open Research Europe

2023



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**



20+ 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)
- Sophie RENAUDIN (AP-HP)
- Jeanne ROBINEAU (IRD)
- Nicolas ROUGIER (Inria, Univ Bordeaux, CNRS)
- François SABOT (IRD)
- Sylvie TONDA-GOLDSTEIN (Inria)
- Samuel THIBAUT (Univ Bordeaux) (Univ Paris-Saclay)

Software Heritage: *one software archive, a shared infrastructure ...*

One infrastructure
open and shared



Inria



The largest archive ever built



Diamond sponsor



Platinum sponsors



Gold sponsors



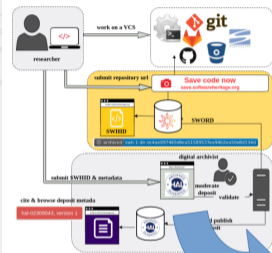
Silver sponsors



Bronze sponsors



... best experience for researchers, worldwide



<https://hal.archives-ouvertes.fr/hal-02130801>

The screenshot shows the HAL website interface. At the top, it says 'Free and accessible knowledge'. Below that, there are navigation links: 'Home', 'Submit', 'Browse', 'Search', and 'Documentation'. The main content area is titled 'LinBox' and includes a list of contributing institutions such as IRCC - Exact Computing, ANIC - Arithmetic and Computing, Inria Grenoble - Rhône-Alpes, LIP - Laboratoire de l'Informatique du Parallélisme, HALDRON - Algorithms and Software Architectures for Distributed and HPC Platforms, Inria Grenoble - Rhône-Alpes, LIP - Laboratoire de l'Informatique du Parallélisme, CIS - Department of Computer and Information Sciences (Newark), Ghent University, NCSU - Department of Mathematics (Raleigh), United States Naval Academy, ICCD - Symbolic Computation Group, CASC - Calcul Algébrique et Symbolique, Sécurité, Systèmes Complexes, Codes et Cryptologie, and LJK - Laboratoire Jean Kuntzmann. The page also features an abstract, document type, domains, and a complete list of metadata.

The screenshot shows the Software Heritage archive page for the LinBox repository. The URL is <https://hal.archives-ouvertes.fr/hal-02130801>. The page displays the revision ID 'e8e18328952266b775c692963b1b963b1496107', the date '14 June 2019, 13:43 UTC', and the file path '393b611/linbox-1.6.3/linbox/config-bias.h'. The page also shows the license 'GNU Lesser General Public License v2.1 or later', the programming language 'C++', and the code repository 'https://github.com/linbox-team/linbox'. The main content is the source code for 'config-bias.h', which includes copyright information for Pascal Giorgi and Clement Perret, and a license notice for the GNU Lesser General Public License. The code ends with a preprocessor directive: '#ifndef LINBOX_CONFIG_BIAS_H'.

sw.h:1.dir:393b611a1424f032e83569bf6762502371cfcf65

Learn more at <https://softwareheritage.org/>

- 1 Introduction
- 2 Software as a Pillar of Open Science
- 3 An emerging policy framework
- 4 France can lead the way
- 5 **Need for a global approach**



Call to action: push adoption of best practices for ARDC

Archive and reference

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

for reproducibility

- save in Software Heritage
- add **SWHID** in articles

See [detailed HOWTO online](#)



Describe and Cite/Credit

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

video tutorials

- add **codemeta.json** (see the [codemeta generator](#))
- reference in HAL (*french partners*, see [online HAL documentation](#))
- cite 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

- **support** open source research software
 - clear **policy** and institutional home (see OSPOs in the US)
 - common **knowledge base** for technology transfer
 - technical and financial **sustainability**
 - modern, efficient, scalable, maintained **collaboration infrastructures**
- establish *intelligent and effective* **incentives** (mind Goodhart's law)
 - count *quality software contributions* in careers
 - avoid *purely numerical indicators*, keep the human in the loop
 - respect software complexity, *dont treat it as data*
- **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 the cost, for proper *evaluation of the service*
- **lead, dont follow** the international conversation

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

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](#))