



# Roberto DI COSMO

## Inria and University of Paris



INSTITUT DE FRANCE  
Académie des sciences



OUVRIR  
LA SCIENCE !

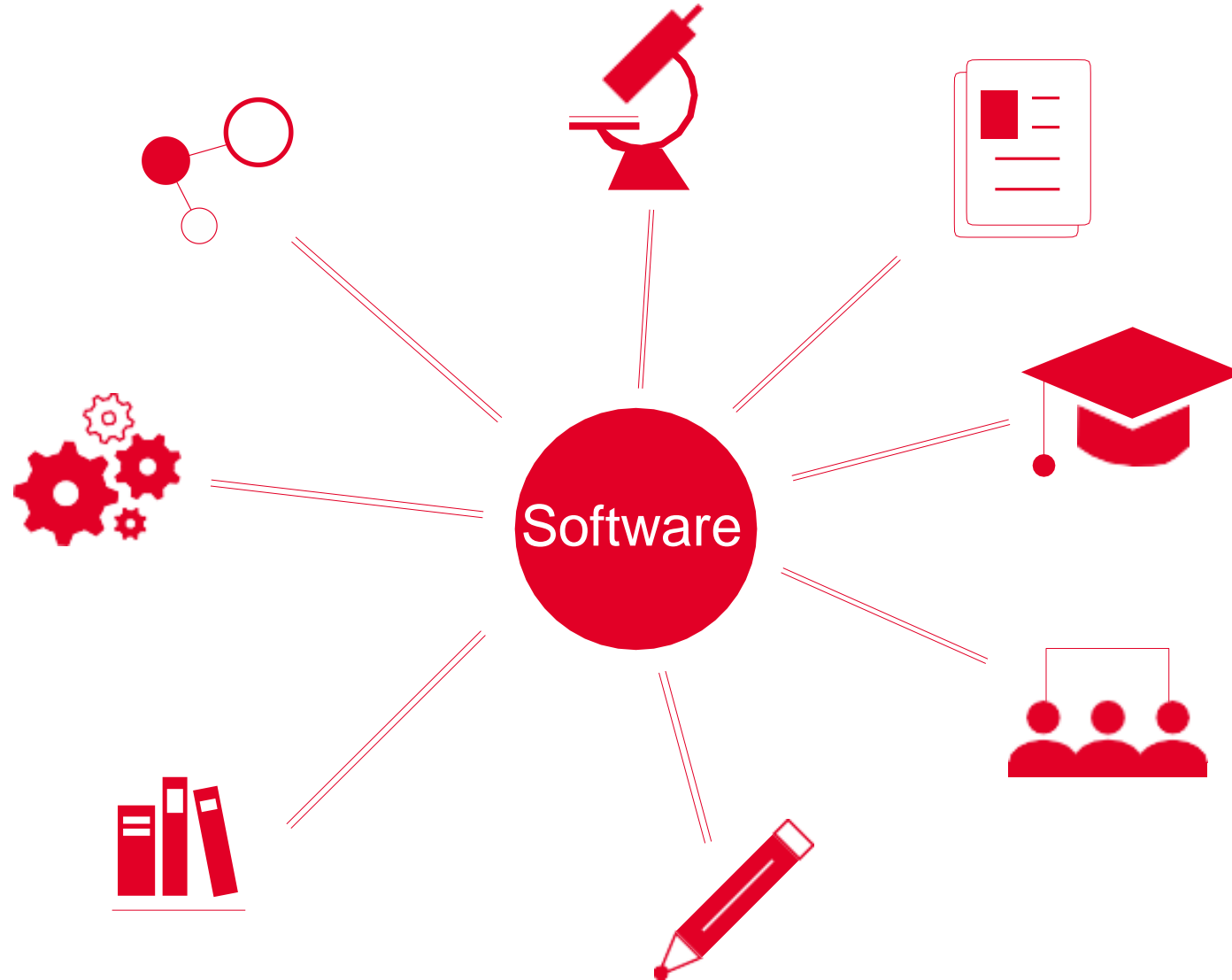




# Building the Software Pillar of Open Science



# Software is everywhere



# Software *Source Code*



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

GPL Licence

Hello World

# 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.”*

# A lightning fast growth

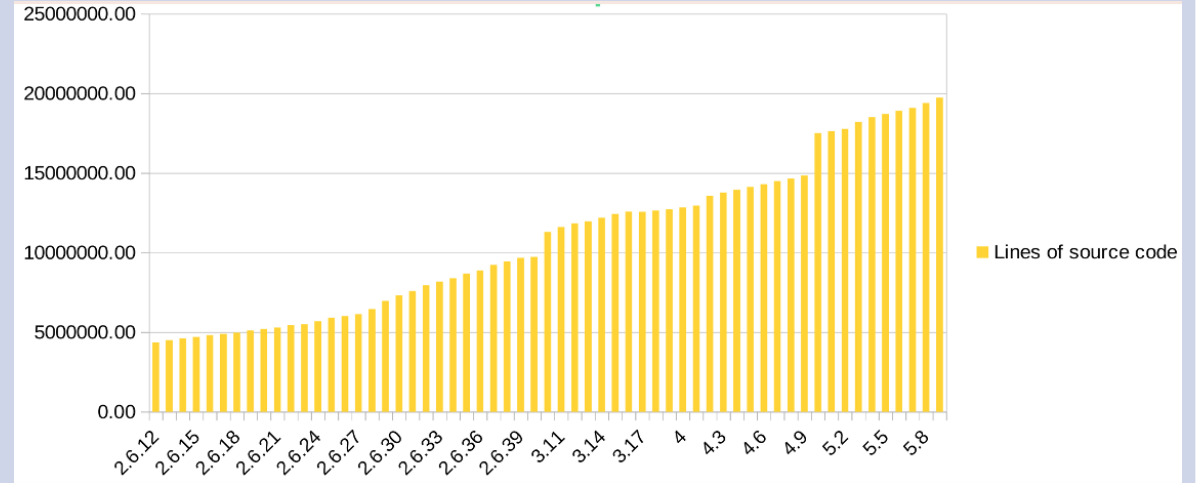
Apollo 11 (~60.000 lines), 1969



"When I first got into it, nobody knew what it was that we were doing. It was like the Wild West."

Margaret Hamilton

Linux Kernel : 20 million lines. . .



. . . now in your pockets!

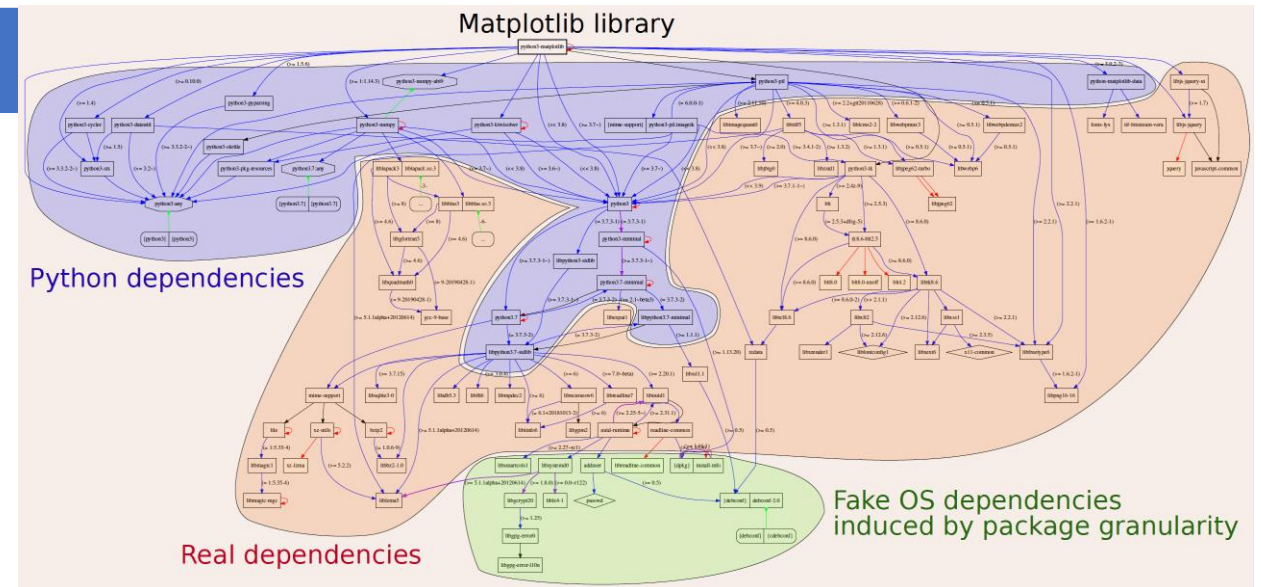
# Source code is *special*: software is *not* data

## 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*



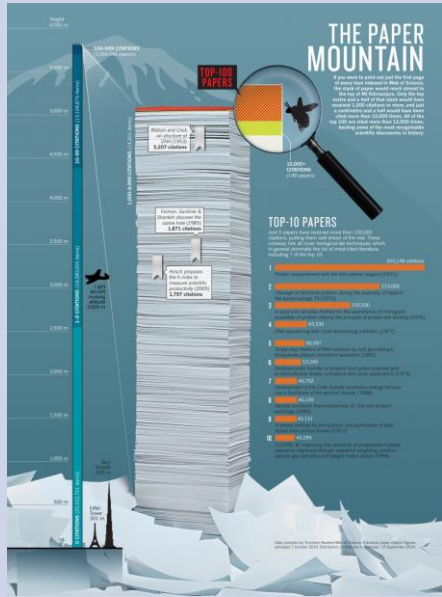
## The human side

copyright law applies!

fruit of human ingenuity: design, algorithm, code, test, documentation, community, funding, and so many more facets...

# A long overlooked pillar of Open Science

Software powers modern research



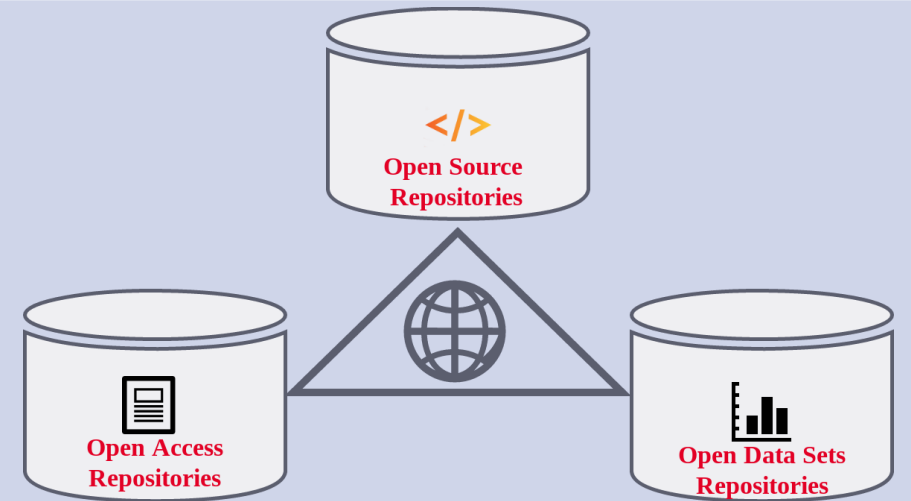
*[...] software [...] essential in their fields.*

*Top 100 papers (Nature, 2014)*

*Sometimes, if you don't have the software, you don't have the data*

*Christine Borgman, Paris, 2018*

Missing pillar: software (source code)



The links in the picture are essential

Software may be a **tool**, an **outcome** and a **research object**

Open source (open access to the source code) is necessary

- avoid reinventing the wheel, accelerate scientific discoveries
- **preserving source code** and its history is necessary for *reproducibility*



# A plurality of needs

## Researchers

- **archive** and **reference** software used in articles
- **find** useful software
- get **credit** for software contributions
- verify, **reproduce**, improve results

## Laboratories/teams

- **track** software contributions
- produce reports
- maintain webpage

## Research Organization

know **its software assets**

- technology **transfer**
- impact **metrics**
- funding **strategy**
- career **evaluation**

## Archive

Research software artifacts must be properly **archived**

make sure we can *retrieve* them (*reproducibility*)

## Reference

Research software artifacts must be properly **referenced**

make sure we can *identify* them (*reproducibility*)

## Describe

Research software artifacts must be properly **described**

make it easy to *discover* and reuse them (*visibility*)

## Cite/Credit

Research software artifacts must be properly **cited** (*not the same as referenced!*)

to give credit to authors (*evaluation!*)

# What is at stake: beyond ARDC

## Policy framework for dissemination, reuse, evaluation and recognition

open source policy for publicly funded research software  
incentives and recognition for researchers and engineers

## Sustainability

Organisational schemas, legal tools, economic models, processes and policies  
to ensure research software can be maintained and sustained over time

## Technology transfer and industry collaboration

Approaches, support, methods, processes to establish connections with industry  
in order to foster uptake and transfer of research software

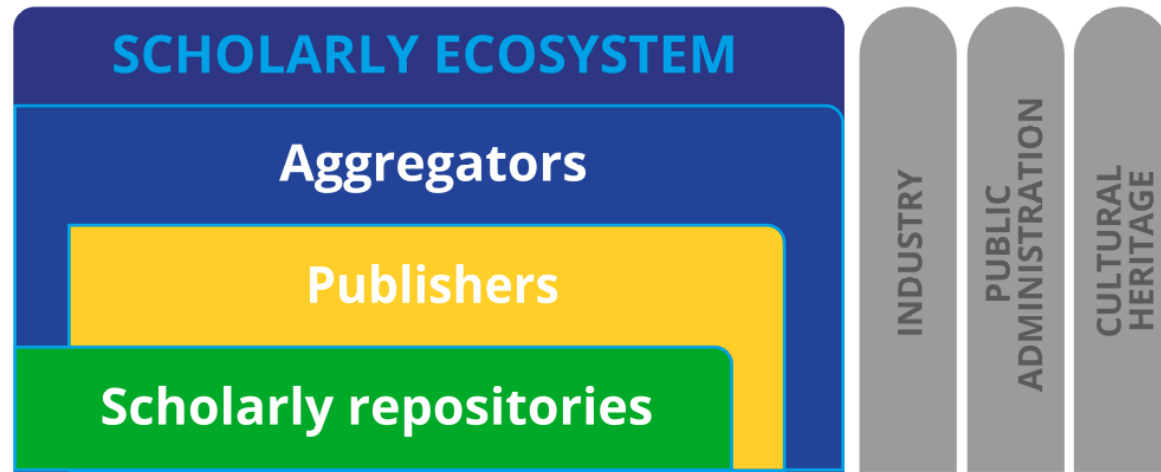
## Advanced technologies and tools

software quality, reproducibility, and traceability

# Focus on infrastructures for ARDC

# The big picture (EOSC SIRS 2020 report)

## Research Software Infrastructures: Overall Architecture

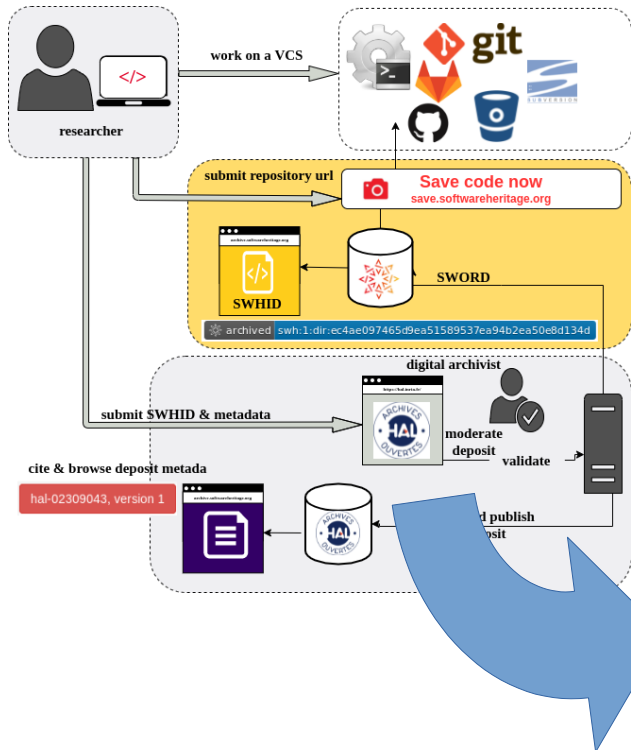


### ★ Scholarly ecosystem

- Aggregators collecting data from...
- Academic publishers
- Scholarly repositories

# The HAL – Software Heritage success story

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



**LinBox**  
The LinBox Group 1, 2, 3, 4, 5, 6, 7, 8, 9

- 1 ECO - Exact Computing
- 2 LIRMM - Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier
- 3 IRISA - Inria Grenoble - Rhône-Alpes, LIP - Laboratoire de l'Informatique du Parallélisme
- 4 ARIC - Arithmetic and Computing
- 5 AVALON - Algorithms and Software Architectures for Distributed and HPC Platforms
- 6 Inria Grenoble - Rhône-Alpes, LIP - Laboratoire de l'Informatique du Parallélisme
- 7 CIS - Department of Computer and Information Sciences [Newark]
- 8 Drexel University
- 9 NCSU - Department of Mathematics [Raleigh]
- 10 United States Naval Academy
- 11 SCG - Symbolic Computation Group
- 12 CASC - Calcul Algébrique et Symbolique, Sécurité, Systèmes Complexes, Codes et Cryptologie
- 13 LJK - Laboratoire Jean Kuntzmann

**Abstract :** LinBox is a C++ template library of routines for solution of linear algebra problems including linear system solution, rank, determinant, minimal polynomial, characteristic polynomial, and Smith normal form. Algorithms are provided for matrices with integer entries or entries in a finite field. A number of matrix storage types is provided, especially for blackbox representation of sparse or structured matrix classes. A few algorithms for rational matrices are available. LinBox also uses underlying data structures and algorithms for integer, rational, polynomial, finite fields and rings, as well as dense and sparse matrix formats coming from the Givaro (<https://casys.griacad-pages.univ-grenoble-alpes.fr/givaro>) and FFLAS-FFPACK (<http://linbox-team.github.io/fflas-ffpack>) libraries.

**Document type :** Software  
**Domain :** Computer Science [cs]  
Computer Science [cs] / Symbolic Computation [cs.SC]

**Complete list of metadata** [Display]

**BROWSE**

Software Heritage swh:1:dir:393b611a1424f032e83569bf6762502371cfcf65;origin=https://hal.archives-ouvertes.fr/hal-02130801;visit=swh:1:snp:19c29b988fe02623c70c7dc8bc97c42481eb691b;anchor=swh:1:rev:e8e18328952266b7875c692963b11963b1496107;path=/

[Browse]

**Browse the archive** Enter a SWHID to resolve or keyword(s) to search for

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

14 June 2019, 13:43 UTC

<> Code Branches (1) Releases (0) Visits

Revision: e8e18328952266b7875c692963b11963b1496107 393b611 / linbox-1.6.3 / linbox / config-blas.h

Tip revision: e8e18328952266b7875c692963b11963b1496107 authored by Software Heritage on 11 June 2019, 08:12 UTC  
hal: Deposit 297 in collection hal

**config-blas.h**

```

1 /* config-blas.h
2  * Copyright (C) 2005 Pascal Giorgi
3  * Copyright (C) 2007 Clement Pernet
4  * Written by Pascal Giorgi <pgiorgi@uwaterloo.ca>
5  *
6  * =====LICENCE=====
7  * This file is part of the library LinBox.
8  *
9  * LinBox is free software: you can redistribute it and/or modify
10 * it under the terms of the GNU Lesser General Public
11 * License as published by the Free Software Foundation; either
12 * version 2.1 of the License, or (at your option) any later version.
13 *
14 * This library is distributed in the hope that it will be useful,
15 * but WITHOUT ANY WARRANTY; without even the implied warranty of
16 * MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU
17 * Lesser General Public License for more details.
18 *
19 * You should have received a copy of the GNU Lesser General Public
20 * License along with this library; if not, write to the Free Software
21 * Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301 USA
22 * =====LICENCE=====
23
24
25
26 #ifndef LINBOX config blas H

```

[swh:1:dir:393b611a1424f032e83569bf6762502371cfcf65](https://hal.archives-ouvertes.fr/hal-02130801)

# Strategic remarks

## Significant risk factors that we need to take into account

- **balkanisation**
  - proliferation of **infrastructure silos**
  - **duplicated** contents with different identifiers
  - **costly** efforts to federate *after-the-fact*
- **closed** (code, gouvernance) and/or **for profit**
- **operations** funded with **project money**

# Focus on broader policy issues



# What is needed

## Policy for dissemination and reuse

- **Set the default to open source** for research software
- Open source **creates value** : adapt technology transfer and industry collaboration to it

## Framework for evaluation and recognition

- Make software development count in a career (not the case in many countries)
- Avoid purely quantitative indicators

## Sustainability

- Technical** improve quality of key research software
- Organisational** professional practices for governance and maintenance
- Financial** make open source research software as easy to fund as buying a license

# Good news: awareness is raising

## Paris Call on Software Source code (2019)

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

☞ <https://en.unesco.org/foss/paris-call-software-source-code>

## EOSC SIRS report (2020) and EOSC TF on infrastructures for research software

*"all research software should be made available under an Open Source license by default, and all deviations from this default practice should be properly motivated »*

☞ See <https://doi.org/10.2777/28598>

## UNESCO Open Science Recommendations (2021)

*"Open science infrastructures should be organized and financed upon an essentially not-for-profit and long-term vision, which enhance open science practices and guarantee permanent and unrestricted access to all."*



## 2nd National Plan for Open Science (6/7/2021)

### Open and promote research software source code

#### ○ actions (selection)

- charter for research software policy
- recognize software development (see [the 2021 prize](#))
- coordinate communities of practice
- build a connected ecosystem of research outputs

#### ○ recommendations (selection)

- **Archive source code in Software Heritage**
- **Standardise and use SWHID**
- **Build a national catalog of research software**

☞ See [official announcement](#)

# Conclusions

# The road ahead

## Infrastructures for research software

Recognise software as a key enabler of research, establish an international network of infrastructures for research software, adapt funding instruments to digital infrastructures where human cost is predominant.

See Ulrike Luke's talk today

## Linking software, publications and data

Key role of the publishers and institutional archives: recognize software as a noble research output on its own, **not just a piece of data**. Software source code is special: it needs specific infrastructures and identifiers.

See Melissa Harrison's talk today

# The road ahead, cont'd

## Institutional representation and support

Establish an office in charge of (open source) research software, to help with funding, open licenses, governance, translation, etc.

See Sayeed Choudhury's talk today

## Incentives, recognition and evaluation

Value quality research software on a par with publications for careers in academia and give them recognition and visibility. **Beware of quantitative indicators** in evaluation, they are much more damaging for software than for publications.

See the Open Science Free Software Awards ceremony today

Let's build together the software pillar of open science  
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](#))