

Software Heritage

revolutionary infrastructure for our digital future

Roberto Di Cosmo

Director, Software Heritage
Inria and Université Paris Cité

January 28 2026
UNESCO



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Invisible fabric of digital society



Knowledge is in the source

```
/**
 * @brief The basic unit of the simulation and is associated to a geographical location.
 *
 * Interventions (e.g., school closures) are tracked at this level. It contains a list of its
 * members (people), places (schools, universities, workplaces etc.), road networks, links to
 * airports etc.
 */
struct Microcell
{
    /* Note use of short int here limits max run time to USHRT_MAX*ModelTimeStep - e.g. 65536*0.25=16384 days=
     * Global search and replace of 'unsigned short int' with 'int' would remove this limit, but use more mem
     */

    int n; // Number of people in microcell
    int adunit; // admin unit microcell belongs to
    int* members; // array of members/hosts of microcell

    int* places[MAX_NUM_PLACE_TYPES]; // list of places (of various place types) within microcell
    unsigned short int NumPlacesByType[MAX_NUM_PLACE_TYPES]; // number of places (of various place types) with
    unsigned short int keyworkerproph, move_trig, place_trig, socdist_trig, keyworkerproph_trig;
    unsigned short int move_start_time, move_end_time;
    unsigned short int place_end_time, socdist_end_time, keyworkerproph_end_time;
    TreatStat moverest, treat, vacc, socdist, placeclose;
    unsigned short int treat_trig, vacc_trig;
    unsigned short int treat_start_time, treat_end_time;
    unsigned short int vacc_start_time;
    IndexList* AirportList;
};
```

Covid Sim ([excerpt](#))

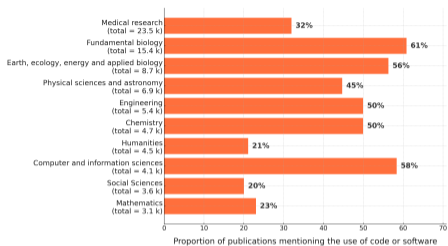
Len Shustek, Computer History Museum

2006

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

A Global Undertaking

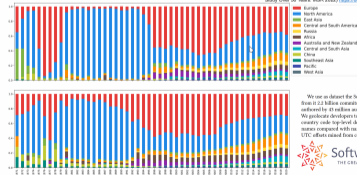
Pillar of Science across all research areas



From all continents

NSR '22, May 21-24, 2022, Pittsburgh, PA, USA

David R. Rossi and Stefano Zachariu
Geographic Diversity in Public Code Contributions: An Exploratory Large-Scale Study Over 50 Years. MSR 2022 (<https://doi.org/10.1145/3534462.3534471>)



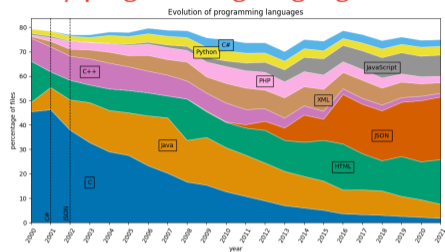
We use as dataset the Software Heritage archive [1] and extract from it 2.2 billion commits archived from 500 million projects and authored by 41 million authors during the 1971-2021 time period. We geocode developers in 12 world regions, using as signal each country code top-level domain (ccTLDs) and author (first last) names compared with name distributions around the world, and UTC offsets raised from commit metadata.

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

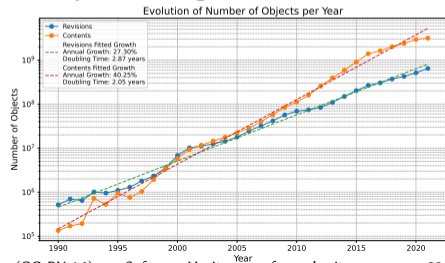
Figure 3: Ratio of commits (above) and active authors (below) by world zone over the 1971-2020 period.



Many programming languages



An exponential growth



Unveiled in 2016



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

Software Heritage : a decade of commitment (2015-2025)

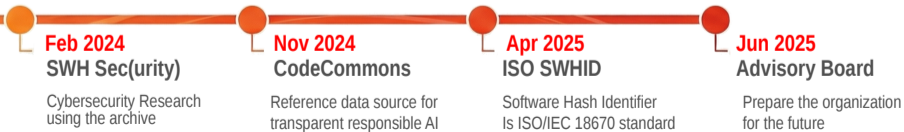
Laying the foundations (2015-2019)



Building infrastructure and raising awareness (2019-2023)



Building solutions and strategic initiatives (2023-2025)



Today: a *universal* software archive, as a shared infrastructure

One infrastructure
open and shared

Cultural Heritage



Industry



Research

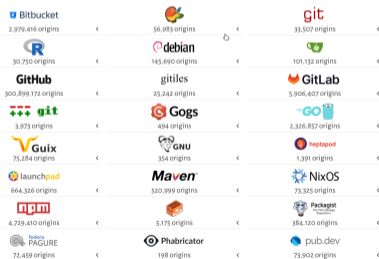


Public Administration



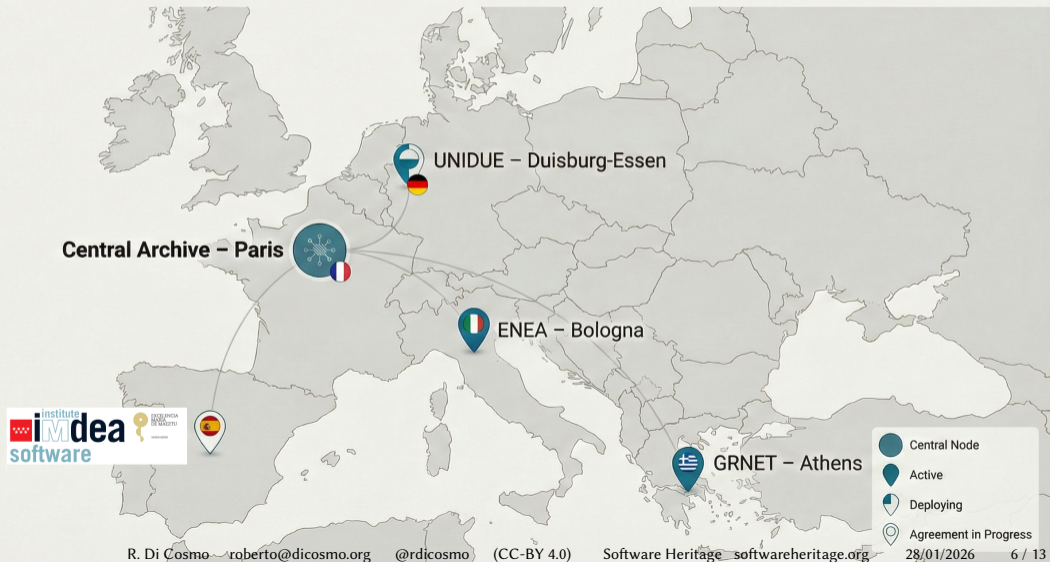
Software Heritage

The largest archive ever built

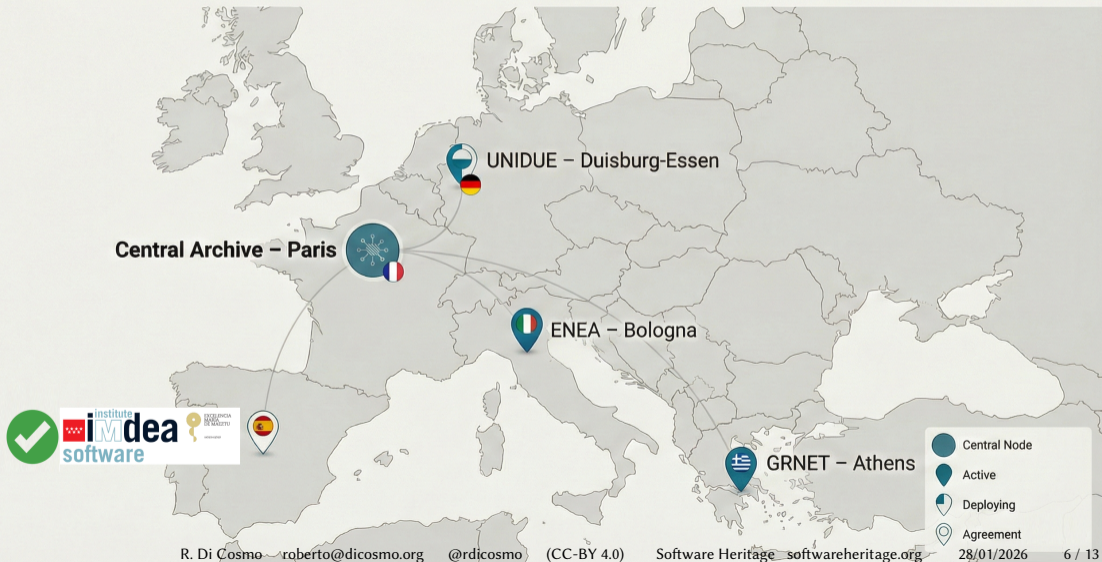


figures as of January 9 2026

The Software Heritage European Network: A Distributed Infrastructure for Global Source Code Preservation



The Software Heritage European Network: A Distributed Infrastructure for Global Source Code Preservation



Sharing the vision



And many more ...

www.softwareheritage.org/support/testimonials

Members and sponsors

Inria



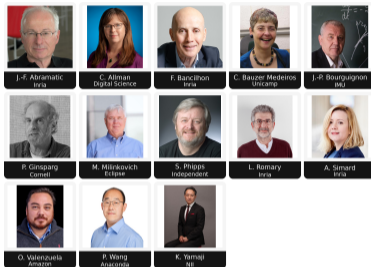
*academia, industry, public sector, society
we are all concerned*

Today: a committed growing community

Core Team



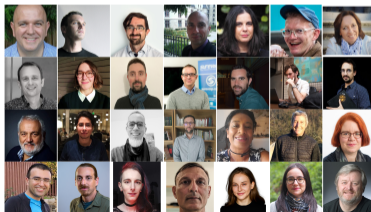
Advisory Board



Members and Sponsors, 2025 Symposium



Ambassadors



Partnerships (History and Heritage)



- Computer History Museum (Mountain View, US)
- Deutsches Museum (Munich, DE)
- Cité des Sciences (Paris, FR)

Infrastructure (Consolidation ongoing)



- GENCI
- CINES
- AI2F

Policy (International engagement)



- DPG Registry
- UN Open Source Principles



Standards (Software Hash ID, SWHID)



- Reference implementation
- CRA standards



New diamond sponsor



Technology Innovation Institute
United Arab Emirates

R. Di Cosmo roberto@dicosmo.org @rdicosmo



2025 annual
report



(CC-BY 4.0) Software Heritage softwareheritage.org

28/01/2026

9 / 13

Strategic Metrics and Insights



Global analysis
of contributions
to public software

CyberSecurity



Make open-source
more secure

SWHSec at [PTCC](#)



Support Open Source Program Offices



Track, find and reuse
research software



Data for Transparent AI



Traceable source code
for LLMs

[CodeCommons](#)



A global undertaking needs global support

Recognition is coming



En matière de climat, de connaissances de la planète, en matière de sciences du vivant, nous avons besoin de poursuivre ce travail, de consolider les plateformes et d'avancer. Réfléchissons collectivement à la lourde tâche d'archivage de certains savoirs spécifiques. Je pense aux travaux aussi de *Software Heritage*, qui, en France, archive l'ensemble des logiciels créés dans le monde, avec une initiative européenne à construire sur le sujet. Je souhaite, au-delà de la question de notre capacité, à attirer chercheurs, enseignants-chercheurs, que nous ayons aussi une capacité à préserver, consolider, attirer plateformes de collaboration et données au service de l'intérêt général.

May 5th 2025 : Choose Europe for Science – Resilience infrastructure



Worldwide outreach

Open Science and AI
CEPAL, Chile



High level meeting at CEPAL in Chile

Open Science
Cyber and AI
Delhi, India



Open Science,
Compliance, AI
Tokyo, Japan

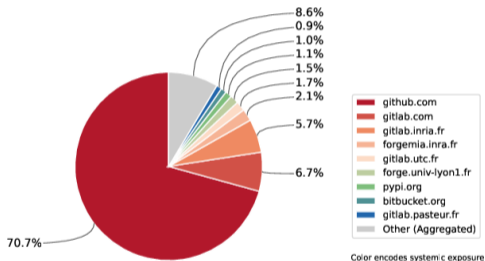


OSPO for Good
United Nations
New York, USA



The new frontier: ecumenical resilience

Global Concentration = Global Risk



Collaborative development depends **massively**
on **very few** single points of failures

Software Heritage is the systemic response

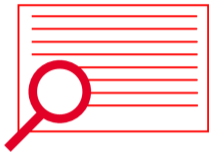
Today

- All the code, also niche and legacy code
- SWHID, ISO 18670 for robust traceability
- Enables long term reproducibility

Tomorrow (needs 10x resources!)

- ecumenical resilience
(strategic autonomy for all)
- automatic recovery (registry failure, network partition)
- global fault tolerance (mirrors)

Supporting transparent and inclusive AI



Source Code as Digital Public Good



Sustaining the Commons



The way forward

