

From Dependence to Resilience

Building Ecumenical Digital Sovereignty for Software

Roberto Di Cosmo

CEO, Software Heritage
Inria & Université Paris Cité

20 May 2026
GARR conference
Pisa



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

- 1 Why software source code matters
- 2 The regulation / infrastructure gap
- 3 Ecumenical digital sovereignty
- 4 Software Heritage: the ecumenical infrastructure
- 5 From archive to strategic backbone
- 6 A Very Large Telescope to explore the software development galaxy
- 7 We (happen to) have the answer at arm's reach
- 8 A path forward, together

Digital sovereignty

Often framed in terms of *cloud, data, or infrastructure*.

Beneath lies a more fundamental layer:
software source code

Invisible fabric that powers our economies, administrations, and research systems.

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=44 yrs.
     * Global search and replace of 'unsigned short int' with 'int' would remove this limit, but use more memory.
     */

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

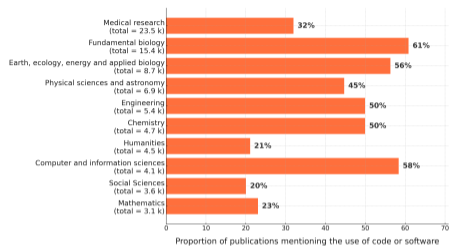
    int* places[MAX_NUM_PLACE_TYPES]; // list of places (of various place types) within microcell
    unsigned short int NumPlacesByType[MAX_NUM_PLACE_TYPES]; // number of places (of various place types) within microcell
    unsigned short int keyworkerproph, move_trig, place_trig, socdist_trig, keyworkerproph_trig;
    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](#))

Sovranità digitale: governare dati, infrastrutture e — innanzitutto — *codice sorgente*.

A global undertaking...

Pillar of Science across all research areas



From all continents

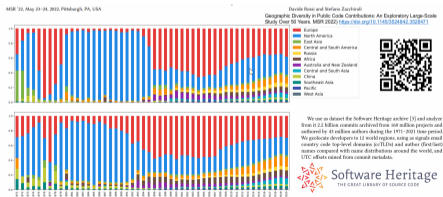
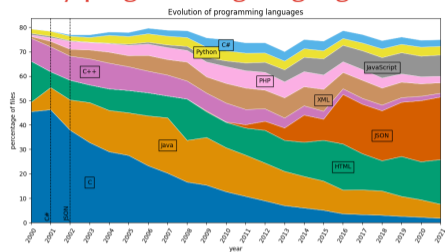
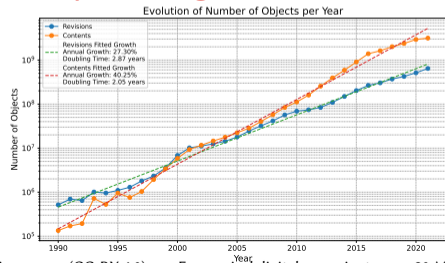


Figure 3: Ratio of commits (above) and active authors (below) by world zone over the 1971-2020 period.

Many programming languages

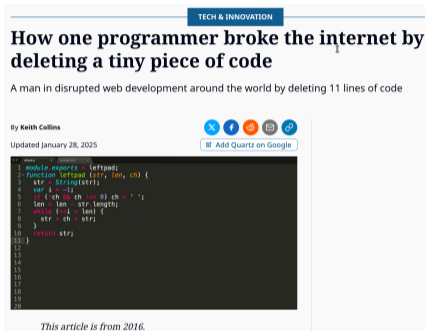


An exponential growth



... and a global dependency network

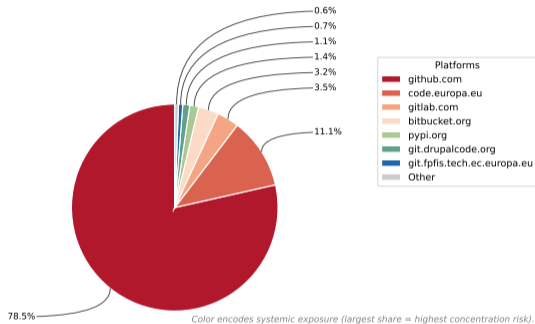
An early warning: left-pad (March 2016)



See the [Wikipedia](#) article:

- Facebook, PayPal, Netflix, Spotify affected
- web broken worldwide ~2.5 hours

A clear and present danger



Used by europa.eu devs (Source: Software Heritage)

"If GitHub or PyPI disappeared tomorrow, Europe would not just slow down — it would stop."

- 1 Why software source code matters
- 2 The regulation / infrastructure gap
- 3 Ecumenical digital sovereignty
- 4 Software Heritage: the ecumenical infrastructure
- 5 From archive to strategic backbone
- 6 A Very Large Telescope to explore the software development galaxy
- 7 We (happen to) have the answer at arm's reach
- 8 A path forward, together

Regulation

Europe's Software Sovereignty

REGULATION



- **Cyber Resilience Act (CRA) & NIS2**
Mandate strict software security, long-term maintenance, and end-to-end supply-chain resilience.
- **AI Act Transparency**
Legal requirements for the transparency and reproducibility of training data and the underlying algorithms.
- **Open Science Policies**
Policies requiring long-term availability and reproducibility of research software for scientific integrity.

“Legal obligations assume permanent access to source code.”

Regulation Without Infrastructure: Europe's Software Sovereignty Gap

REGULATION



- **Cyber Resilience Act (CRA) & NIS2**
Mandate strict **software security**, long-term maintenance, and end-to-end supply-chain resilience.
- **AI Act Transparency**
Legal requirements for the **transparency** and **reproducibility** of training data and the underlying algorithms.
- **Open Science Policies**
Policies requiring long-term availability and reproducibility of research software for scientific integrity.

INFRASTRUCTURE



- **Dependence on Non-European Platforms**
Relies on **proprietary platforms** like GitHub, GitLab.com, and package managers (PyPI, npm, Maven Central).
- **No EU-Wide Continuity Guarantee**
Existing platforms provide **no Service Level Agreements (SLA)** for preservation and operate under **no legal obligation** to protect European sovereignty.
- **Fragmented National Initiatives**
Current efforts are often national, not interoperable, and lack a shared global governance structure.

"Legal obligations assume permanent access to source code."

"Critical dependencies outside European control."

Regulation Without Infrastructure: Europe's Software Sovereignty Gap

REGULATION



- **Cyber Resilience Act (CRA) & NIS2**
Mandate strict **software security**, long-term maintenance, and end-to-end supply-chain resilience.
- **AI Act Transparency**
Legal requirements for the **transparency** and **reproducibility** of training data and the underlying algorithms.
- **Open Science Policies**
Policies requiring long-term availability and reproducibility of research software for scientific integrity.

INFRASTRUCTURE



- **Dependence on Non-European Platforms**
Relies on **proprietary platforms** like GitHub, GitLab.com, and package managers (PyPI, npm, Maven Central).
- **No EU-Wide Continuity Guarantee**
Existing platforms provide **no Service Level Agreements (SLA)** for preservation and operate under **no legal obligation** to protect European sovereignty.
- **Fragmented National Initiatives**
Current efforts are often national, not interoperable, and lack a shared global governance structure.

⚠ THE STRUCTURAL GAP ⚠

“Legal obligations assume permanent access to source code.”

Europe regulates software as critical infrastructure – but does not operate a software continuity infrastructure

“Critical dependencies outside European control.”

- 1 Why software source code matters
- 2 The regulation / infrastructure gap
- 3 Ecumenical digital sovereignty**
- 4 Software Heritage: the ecumenical infrastructure
- 5 From archive to strategic backbone
- 6 A Very Large Telescope to explore the software development galaxy
- 7 We (happen to) have the answer at arm's reach
- 8 A path forward, together

What it is *not*

- it is not *autarky*
- it is not *national silos*
- it is not *forking the world*
- it is not *owning everything*

What it is

- shared, **open**, **non-rival** infrastructure
- that benefits *all* while preserving autonomy for *each*
- resilient against disruption upstream
- *ecumenical* by design

“Sovereignty is not about owning everything
— it is about ensuring that nothing essential can disappear.”

- 1 Why software source code matters
- 2 The regulation / infrastructure gap
- 3 Ecumenical digital sovereignty
- 4 Software Heritage: the ecumenical infrastructure**
- 5 From archive to strategic backbone
- 6 A Very Large Telescope to explore the software development galaxy
- 7 We (happen to) have the answer at arm's reach
- 8 A path forward, together



Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Inria with  **unesco**





Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Inria with  unesco



The largest open source code archive: one infrastructure, open, shared, non profit

Unique digital common good *built in France since 2015*

Cultural Heritage



Industry



Research

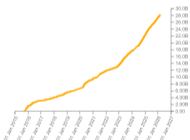


Public Administration



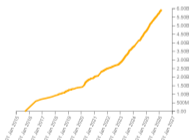
Source files

28,152,651,759



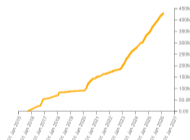
Commits

5,920,787,012



Projects [↗](#)

429,963,770





Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Inria with  unesco



The largest open source code archive: one infrastructure, open, shared, non profit
Unique digital common good *built in France since 2015*

Cultural Heritage



Source files

28,152,651,759

Industry



Commits

5,920,787,012

Research

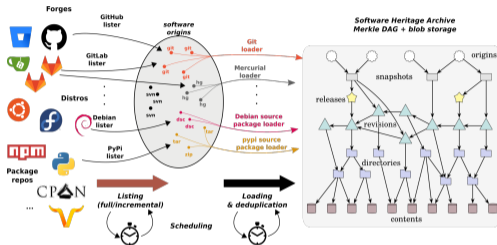
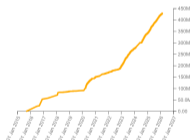
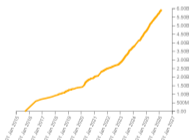
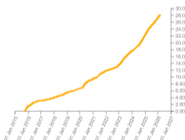


Public Administration



Projects

429,963,770



5000+ platforms

All versions, all history development in a single graph



Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Inria with  unesco



The largest open source code archive: one infrastructure, open, shared, non profit
Unique digital common good built in France since 2015

Cultural Heritage



Source files

28,152,651,759

Industry



Commits

5,920,787,012

Research

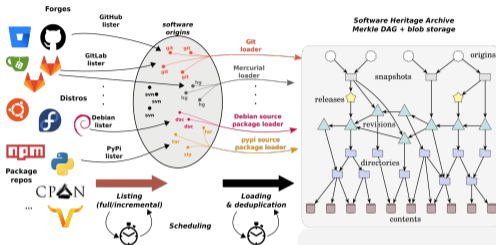


Public Administration



Projects

429,963,770



5000+ platforms

All versions, all history
development in a single graph

- 50 × 10⁹ nodes
- 1000 × 10⁹ edges
- ~ 3 PB of storage



Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Inria with  unesco



The largest open source code archive: one infrastructure, open, shared, non profit
Unique digital common good built in France since 2015

Cultural Heritage



Source files

28,152,651,759

Industry



Commits

5,920,787,012

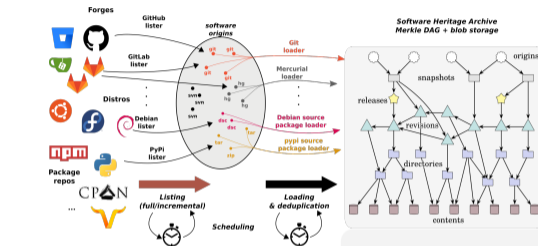
Research



Projects

429,963,770

Public Administration



5000+ platforms

All versions, all history
development in a single graph

- 50 × 10⁹ nodes
- 1000 × 10⁹ edges
~ 3 PB of storage

A revolutionary **infrastructure** ensures **availability** guarantees **integrity** enables **traceability**





Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Inria with unesco



The largest open source code archive: one infrastructure, open, shared, non profit
Unique digital common good built in France since 2015

Cultural Heritage



Source files

28,152,651,759

Industry



Commits

5,920,787,012

Research

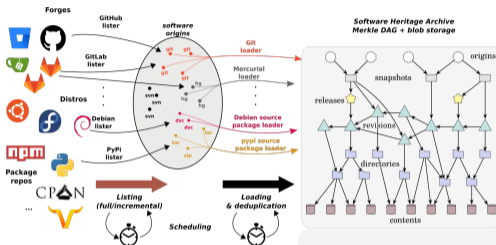


Public Administration



Projects

429,963,770



5000+ platforms

All versions, all history development in a single graph

- 50 × 10⁹ nodes
- 1000 × 10⁹ edges
~ 3 PB of storage

A revolutionary infrastructure ensures availability guarantees integrity enables traceability



- 1 Why software source code matters
- 2 The regulation / infrastructure gap
- 3 Ecumenical digital sovereignty
- 4 Software Heritage: the ecumenical infrastructure
- 5 From archive to strategic backbone**
- 6 A Very Large Telescope to explore the software development galaxy
- 7 We (happen to) have the answer at arm's reach
- 8 A path forward, together

Standards (Software Hash ID, SWHID)



Intrinsic identifier
integrity, traceability
50B+ SWHID in archive

Software Pillar of (Open) Science EOSC SIRS report (2020)



National Open Science Plan
(2021)
French National RI Roadmap
(2022)

Pillar of transparent AI



StarCoder2/The Stack v2
(Hugging Face)
CodeCommons (BPI, France)
AI Factory (AI2F, EU)

Next gen Cybersecurity



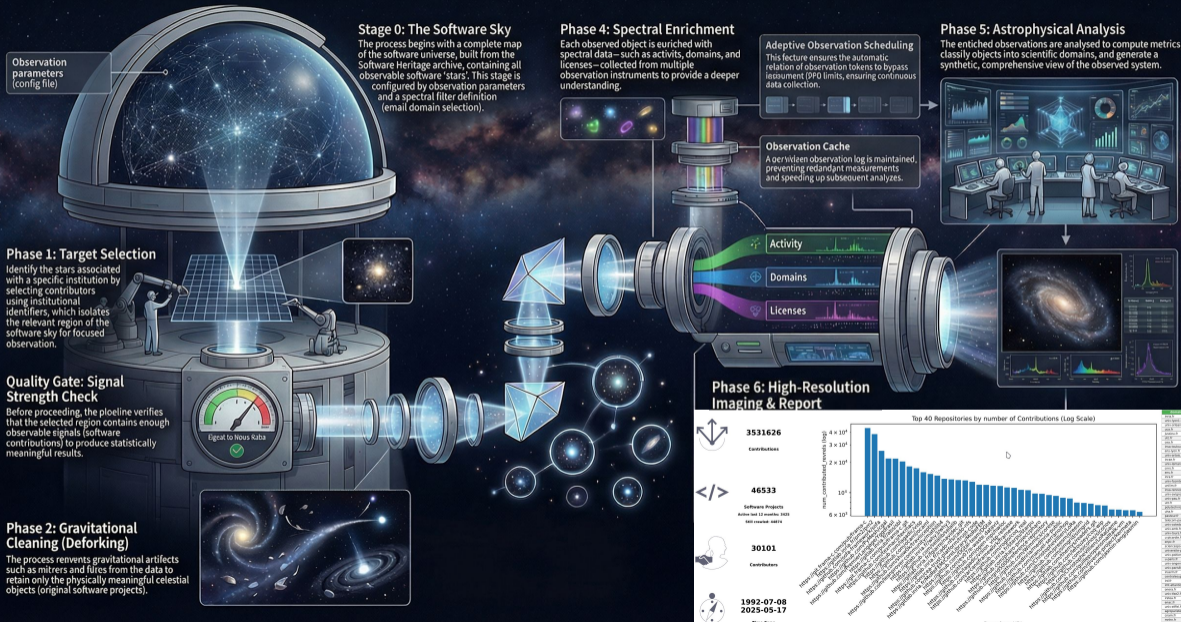
Global Vulnerabilities
(Sec4AI4Sec, SWHSec)
Software supply chain
(CRA, SBOM,...)

A (version of a) dependency of a critical software vanishes from GitHub, Gitlab, ...:

- how to **audit** it five years later?
- how to **reproduce** the results?
- how to **ensure** continuity of service?
- how to **retrieve** them?

- 1 Why software source code matters
- 2 The regulation / infrastructure gap
- 3 Ecumenical digital sovereignty
- 4 Software Heritage: the ecumenical infrastructure
- 5 From archive to strategic backbone
- 6 A Very Large Telescope to explore the software development galaxy**
- 7 We (happen to) have the answer at arm's reach
- 8 A path forward, together

From the Software Sky to Institutional Insight: Anatomy of an Observational Pipeline



Software Heritage

Report on the public software collected, preserved and referenced for Acad-It-Garr

Version 1.0 – 2026-05-11

Contents

1 Introduction	3
1.1 Bird's-Eye View: the key numbers	3
1.2 Technical Information	4
2 Analysis of Contributions	5
2.1 Density of contributions per project	5
2.2 Distribution of first contributions over time	7
2.3 Timespan of institutional contributions	9
2.4 Distribution of project lifetime	11
2.5 Distribution of Citation Files	13
2.5.1 Projects with Citation Files	13
2.5.2 Projects with Citation Files Only in Software Heritage	14
2.6 Top Projects by Repository Host	16
3 Analysis of the Projects to which Acad-It-Garr contributes	40
3.1 Distribution by number of stars	40
3.2 Main organizations	42
3.3 Top 40 projects by number of contributions	43
3.4 Top 40 projects by popularity	44
3.5 Top 40 projects by repository age	46
4 Programming Languages	66
5 License Analysis	68
5.1 License Distribution	68
5.2 License Status Overview	68
5.3 License Statistics	68
5.4 Top License Types	68
6 Platform Distribution Analysis	70
6.1 Platform Distribution Overview	70
6.2 Platform Distribution (Log Scale)	70
6.3 Platform Statistics	70
6.4 Top Platforms	70
6.5 Summary	72
7 Tentative Classification in Research Areas/Domains	73
7.1 Distribution of domains of contributors (all projects)	74
7.2 Distribution of domains of contributors (selected projects with more than 10 contributions)	75
8 Methodology: From Source Graph to Analysis-Ready Dataset	75
9 Conclusion	77

A peek inside: full analysis charts



2765958

Contributions



38309

Software Projects
Active last 12 months: 1993
Still crawled: 35505



19065

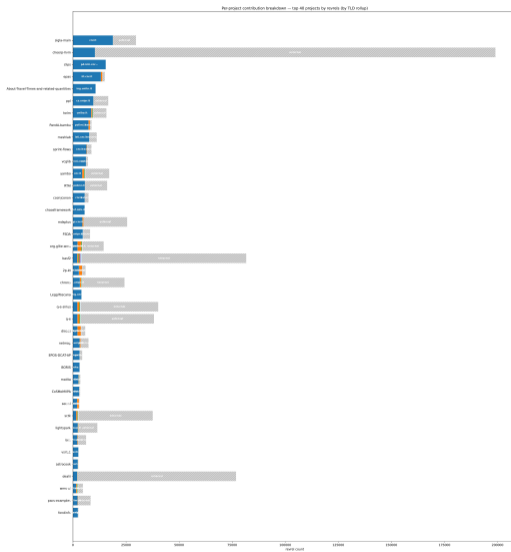
Contributors



1968-02-12

2026-05-06

Time Span



domain	count
polim.it	2901
uniba.it	2296
polito.it	1435
unipd.it	1039
univroma2.it	958
univr.it	949
univrb.it	732
unisi.it	661
unisa.it	619
unimi.it	549
unipi.it	529
uniba.it	522
unife.it	421
uniroma3.it	397
uniba.it	379
unina.it	336
unicam.it	287
unige.it	286
unimore.it	269
uniroma2.it	223
univpm.it	199
univa.it	183
uniba.it	175
unige.it	161
univ.it	161
unifi.it	154
unibg.it	151
unife.it	149
unisa.it	146
unige.it	127
uniroma3.it	117
unipa.it	109
univpm.it	103
unisa.it	92
unical.it	85
unipa.it	85
unisa.it	76
unipa.it	70
uninobilita.it	64
unige.it	63
unipi.it	56
polito.it	49
uniba.it	49
uniconnecta.it	48
univroma2.it	47
univalente.it	43
unipi.it	37
uniba.it	36
univpm.it	35

Note: Showing top 10 of 83 total domains

A peek inside: who contributes from Italian academia?

Major hubs

- **CNR** (7) — sigla-main, EPAS, sprint-flows, cool-jconon, MeshLab, VCGLib, ITTIG/LeggiToscana
- **INFN** (5) — chaosframework, Belle II, gLite WMS, wms-ui, INDIGO-IAM
- **Università di Bologna** (3) — HELM, Matita, tuProlog
- **Università di Parma** (4) — PPL, FSDA, Project Chrono, fondinfo
- **Politecnico di Milano** (3) — PandA-Bambu, RTAI, pacs-examples
- **Multi-institute Italian-led** (6) — Yambo, Q-E, Q-E SIRIUS, sctk, MDSplus, **deal.II**

Smaller hubs and single contributions

- **Università di Trento** (2) — CoSMoMVPA, About-Travel-Times (abouthydrology)
- **INAF** (2) — DISCOS, astrocook
- **Leaning Tech / Sant'Anna Pisa** (2) — Cheerp, lightspark
- **One each:**
 - UniPi — netmap
 - UniTo — BORIS
 - UniCa — SecML
 - INGV — EPOS-DCAT-AP
 - UniRoma1 — vsrlab
 - personal-research — CLIPS fork

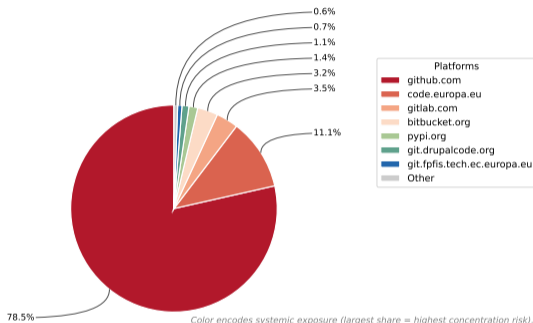
Source: Software Heritage — Strategic Software Insights, Italian academic cut

A peek inside: Italian flagships in the top-40

- **deal.II** — SISSA + UniPi + PoliMi:
world-class FEM library, massive external uptake
- **MeshLab + VCGLib** — CNR-ISTI:
biggest Italian 3D-graphics open-source impact
- **Quantum ESPRESSO** (q-e, q-e-sirius, sctk)
— SISSA/CNR-IOM/CINECA:
Italian DFT flagship
- **Matita / HELM** — UniBo:
Italian theorem-proving heritage
- **PandA-Bambu HLS** — PoliMi:
hardware/software co-design
- **netmap** — UniPi:
high-performance packet I/O
- **Parma Polyhedra Library** — UniPR:
foundational static-analysis library
- **Belle II / italiangrid / chaosframework / INDIGO-IAM** — INFN:
scientific-computing + grid pillars
- **DISCOS + astrocook** — INAF:
radio-telescope control + spectroscopy
- **Cheerp + lightspark** — Leaning Tech / Sant'Anna Pisa:
web-runtime engineering
- **BORIS** — UniTo:
behavioural-sciences research software

Back to our core digital sovereignty challenge, can we act?

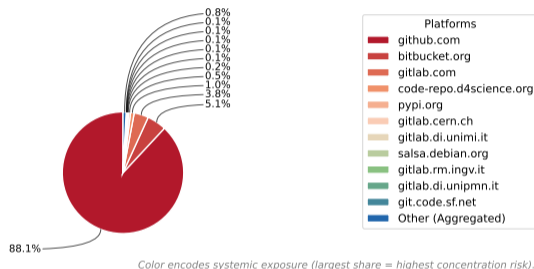
A clear and present danger



Used by europa.eu devs (Source: Software Heritage)

"If GitHub or PyPI disappeared tomorrow, Europe would not just slow down — it would stop."

A clear and present danger



Used by GARR devs (Source: Software Heritage)

"If GitHub or PyPI disappeared tomorrow, GARR members would not just slow down — they would stop."

- 1 Why software source code matters
- 2 The regulation / infrastructure gap
- 3 Ecumenical digital sovereignty
- 4 Software Heritage: the ecumenical infrastructure
- 5 From archive to strategic backbone
- 6 A Very Large Telescope to explore the software development galaxy
- 7 We (happen to) have the answer at arm's reach**
- 8 A path forward, together

Resilience, demonstrated: Guix integration

Software Heritage and GNU Guix join forces
to enable long term reproducibility



Guix demonstrates automated recovery using the Software Heritage archive

When a platform or registry fails...

Guix fallbacks to Software Heritage

Connecting reproducible deployment to a long-term source code archive



Ludovic Courtès — March 29, 2019

GNU Guix can be used as a “package manager” to install and upgrade software packages as is familiar to GNU/Linux users, or as an environment manager, but it can also provision containers or virtual machines, and manage the operating system running on your machine.

One foundation that sets it apart from other tools in these areas is reproducibility. From a high-level view, Guix allows users to declare complete software environments and instantiate them. They can share those environments with others, who can replicate them or adapt them to their needs. This aspect is key to reproducible computational experiments: scientists need to reproduce software environments before they can reproduce experimental results, and this is one of the things we are focusing on in the context of the Guix-HPC effort. At a lower level, the project, along with others in the [Reproducible Builds](#) community, is working to ensure that software build outputs are [reproducible](#), bit for bit.

Work on reproducibility at all levels has been making great progress. Guix, for instance, allows you to [travel back in time](#). That Guix can travel back in time and build software reproducibly is a great step forward. But there's still an important piece that's missing to make this viable: a stable source code archive. This is where [Software Heritage](#) (SWH for short) comes in.

When source code vanishes

and continues to build!



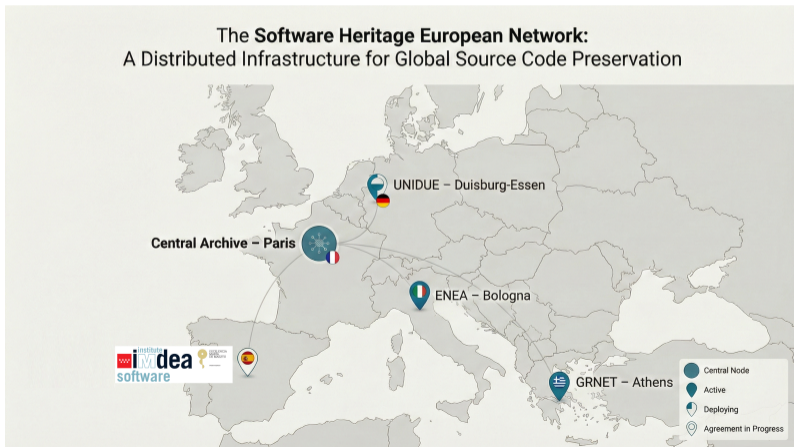
Since 2019

A sovereign fallback for Europe's build chains is possible

Extend to key package managers

npm, PyPI, Maven Central, Cargo, CPAN, RubyGems, Packagist, ...

Realize ecumenical resilience



Software Heritage: one infrastructure, six strategic levers

Research infrastructure

two facets:

- support Open Science
- enable Software Science

Traceability and compliance

- Software supply chain (SBOM)
- integrity (SWHID)
- availability
- support CRA/NIS2 mandates

Cybersecurity

Cyber forensics, vulnerability tracking at world scale

Resilience

when npm/PyPI/Maven **fail**,
Europe **continues to build**

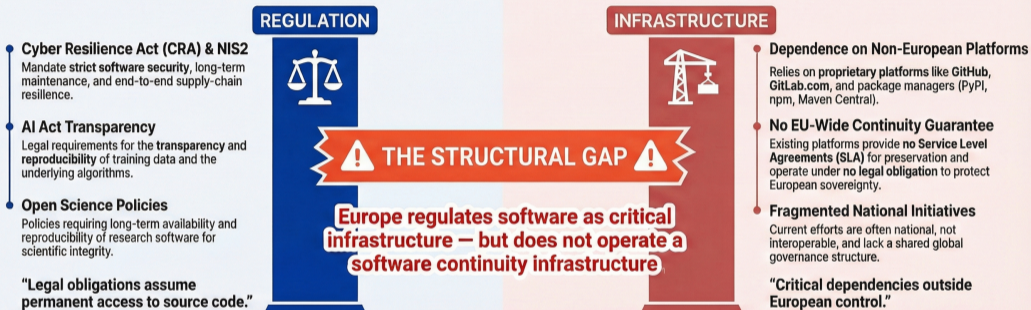
Transparent AI

- **CodeCommons** : provenance and attribution for LLMs
- AI Act, Art. 11, 12, 15:
 - technical documentation
 - record-keeping
 - reproducibility

Metrics we can count on

- Objective Strategic Insights
- Support policy implementation

Regulation Without Infrastructure: Bridging Europe's Software Sovereignty Gap



SOFTWARE HERITAGE

Over
24
billion

Universal Archive of Source Code

An operational global archive containing over 24 billion files and 375 million projects (and growing).



Intrinsic Identifiers (SWHID – ISO 18670)

Provides a standardized “fingerprint” for software to ensure absolute traceability, auditability, and citation.



Sovereign Mirror Network

Resilient, neutral network of nodes (e.g., Italy, Greece, Spain) ensuring archive cannot be destroyed or controlled by a single entity.

Ready for Scaling, Not Invention

Already operational; requires political and financial scaling to fully support European regulatory ambitions.

“Infrastructure that makes European regulation enforceable.”

- 1 Why software source code matters
- 2 The regulation / infrastructure gap
- 3 Ecumenical digital sovereignty
- 4 Software Heritage: the ecumenical infrastructure
- 5 From archive to strategic backbone
- 6 A Very Large Telescope to explore the software development galaxy
- 7 We (happen to) have the answer at arm's reach
- 8 A path forward, together

What this means for Italy and for Europe

Europe **does not need to invent** this infrastructure...

it can **adopt and scale** what took 10+ years to build.

Concrete next steps

All recognise **software continuity** as a critical enabler for

- digital sovereignty
- regulation enforcement (CRA, NIS2, AI Act)
- research, education and open science

All **avoid balkanisation**, support **mutualisation**

- build on common, shared, open, non-profit infrastructures

Public administrations and enterprises *join* as members, mirror operators, contributors

- turn *dependence* into *shared resilience*
- position Italy as a leader in this **ecumenical** approach

True digital sovereignty is not control — it is resilience.

<https://www.softwareheritage.org>

