Software as Critical Infrastructure

addressing Europe's Strategic Blind Spot

Roberto Di Cosmo

Director, Software Heritage
Inria and Université Paris Cité

March 18 2026
DG RTD / DG CONNECT, Brussels

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Outline

Software is all around us

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 memo
        */

        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;
};
```

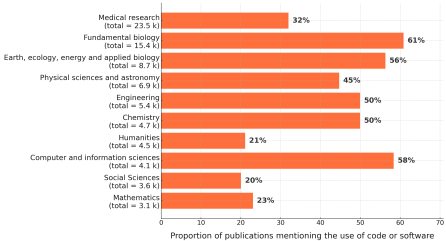
Covid Sim (excerpt)

Len Shustek, Computer History Museum 2006

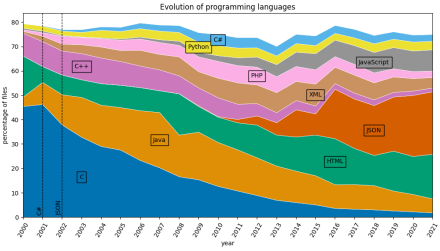
"Source code provides a view into the mind of the designer."

A Global Undertaking…

Pillar of Science across all research areas



Many programming languages



From all continents

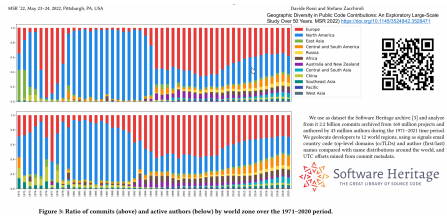
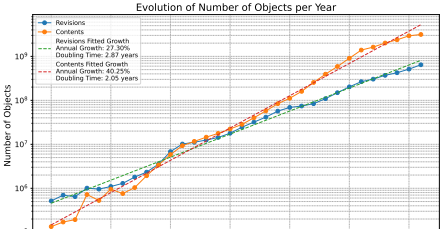


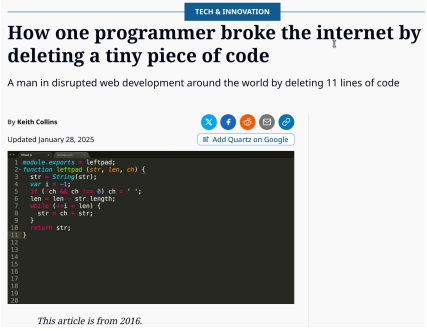
Figure 3: Ratio of commits (above) and active authors (below) by world zone over the 1971–2020 period.

An exponential growth



… and a global dependency network

An early warning: left-pad (March 2016)



TECH & INNOVATION

How one programmer broke the internet by deleting a tiny piece of code

A man in disrupted web development around the world by deleting 11 lines of code

By **Keith Collins**

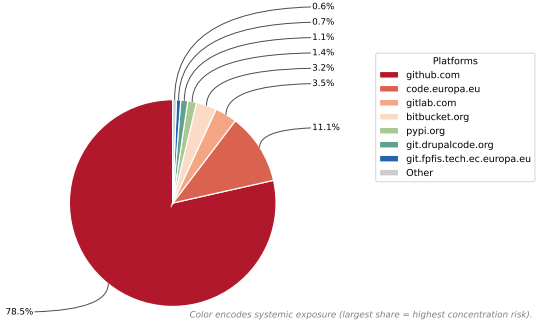
Updated January 28, 2025

Add Quartz on Google

```
1  module.exports = leftpad;
2  function leftpad (str, len, ch) {
3    str = String(str);
4    var i = -1;
5    if (!ch && ch !== 0) ch = ' ';
6    len = len - str.length;
7    while (++i < len) {
8      str = ch + str;
9    }
10   return str;
11 }
12
13
14
15
16
17
18
19
20
```

This article is from 2016.

A clear and present danger



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

See the Wikipedia article:

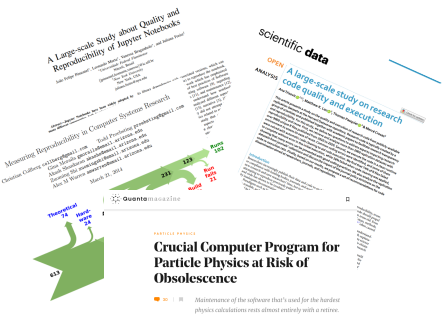
- Facebook, PayPal, Netflix, Spotify affected
- web broken worldwide ~2.5 hours

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

Outline

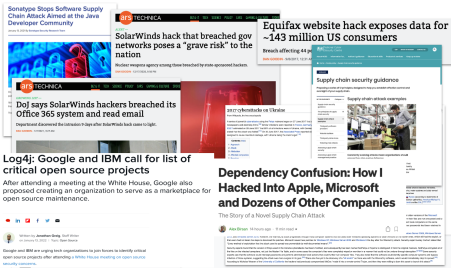
Different angles, same bad news

Reproducibility, maintenance in Academia



(articles: here, here, here and here)

Security, integrity, traceability in Industry



Can they track the software that they

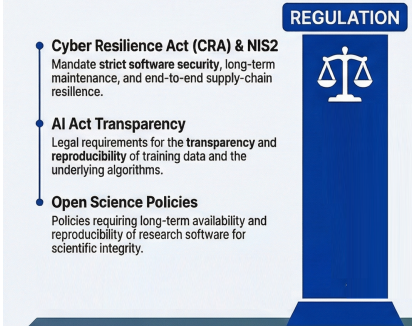
- ship, use, acquire
- has that bug or vulnerability

Policy is evolving in response to this challenge

Outline

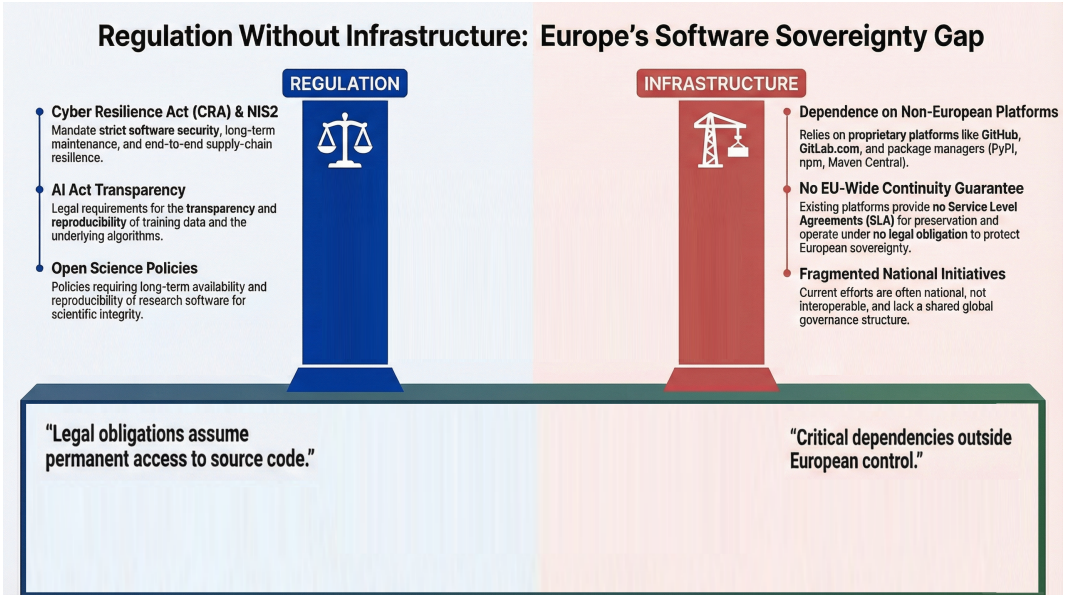
Regulation

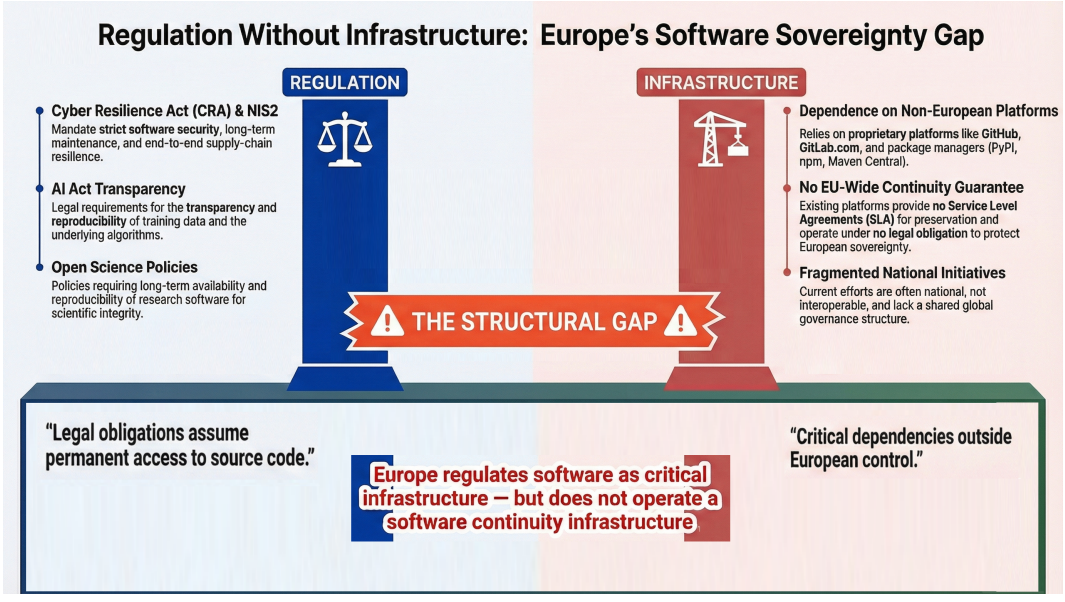
Europe's Software Sovereignty



- **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."
"Critical dependencies outside European control."
Europe regulates software as critical infrastructure — but does not operate a software continuity infrastructure

Outline

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Inria with unesco

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

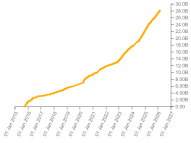
Inria with unesco

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

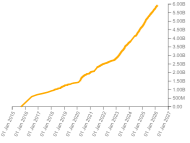
Unique digital common good *built in France since 2015*

Cultural Heritage | **Industry** | **Research** | **Public Administration**

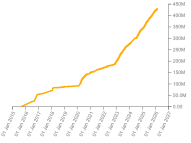
Source files
28,152,651,759



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

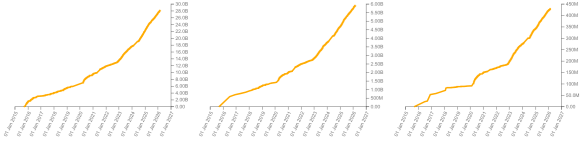
Unique digital common good *built in France since 2015*

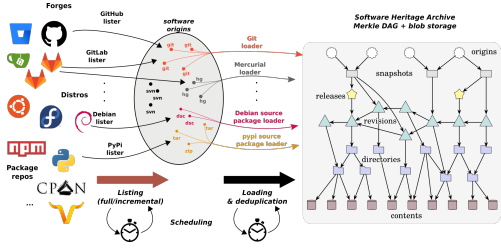
Cultural Heritage | **Industry** | **Research** | **Public Administration**

Source files
28,152,651,759

Commits
5,920,787,012

Projects
429,963,770





5000+ platforms

All versions, all history development in a single graph

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

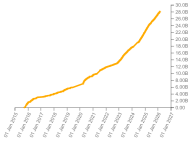
Inria with unesco

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

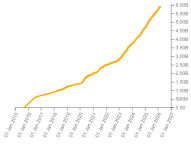
Unique digital common good *built in France since 2015*

Cultural Heritage | **Industry** | **Research** | **Public Administration**

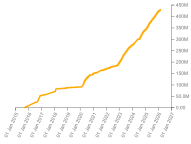
Source files
28,152,651,759

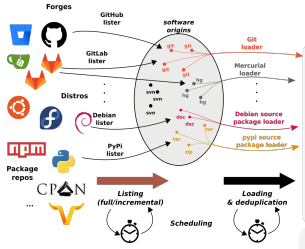


Commits
5,920,787,012



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

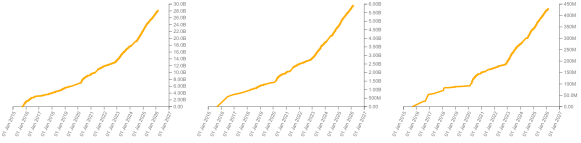
Unique digital common good *built in France since 2015*

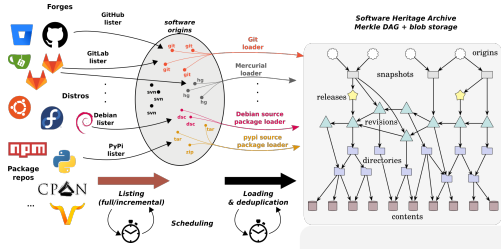
Cultural Heritage | **Industry** | **Research** | **Public Administration**

Source files
28,152,651,759

Commits
5,920,787,012

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**

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

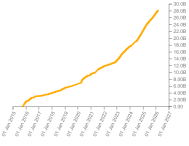
Inria with unesco

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

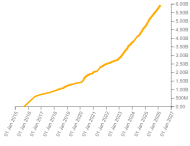
Unique digital common good *built in France since 2015*

Cultural Heritage | **Industry** | **Research** | **Public Administration**

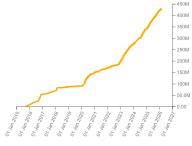
Source files
28,152,651,759

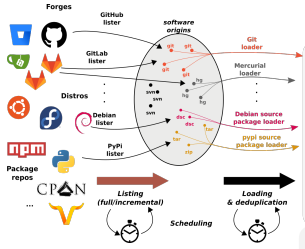


Commits
5,920,787,012



Projects
429,963,770





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**

IBM Microsoft TII Technology Innovation Institute HUAWEI cnrs Infosys cea Ministère de l'Enseignement Supérieur et de la Recherche aws Google Université Paris Cité openinventionnetwork

AdaCore GitHub Université di Pisa Sorbonne Université Université de Lorraine Sant'Anna NWO université PARIS-SACLAY

Outline

Selected hilights

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?

Proof of capability: Guix integration

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

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

Guix demonstrates automated recovery using the Software Heritage archive

When a platform or registry fails… Guix fallbacks to Software Heritage and continues to build!

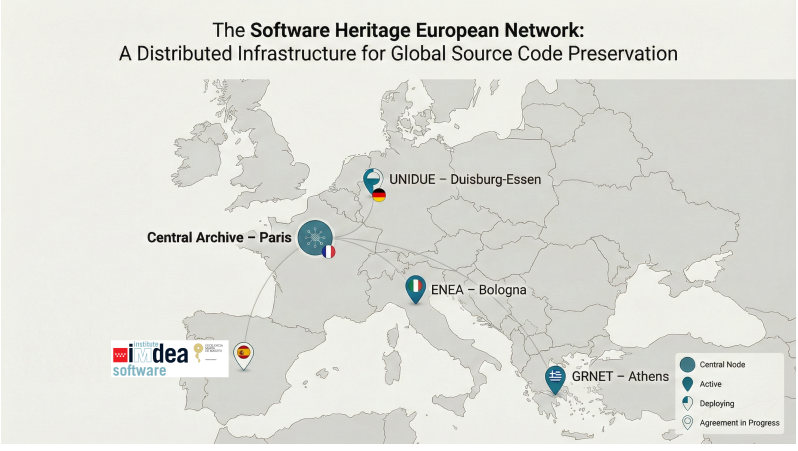
Since 2019

Next step: Sovereign fallback for Europe's build chains

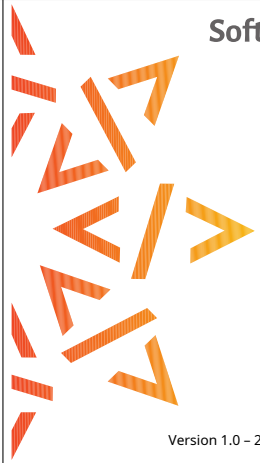
Extend to key package managers

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

Realize ecumenical resilience



Technology preview: the Strategic Software Insights report



Contents

A peek inside the French Academic Software report

3531626

Contributions

46533

Software Projects

Active last 12 months: 3425

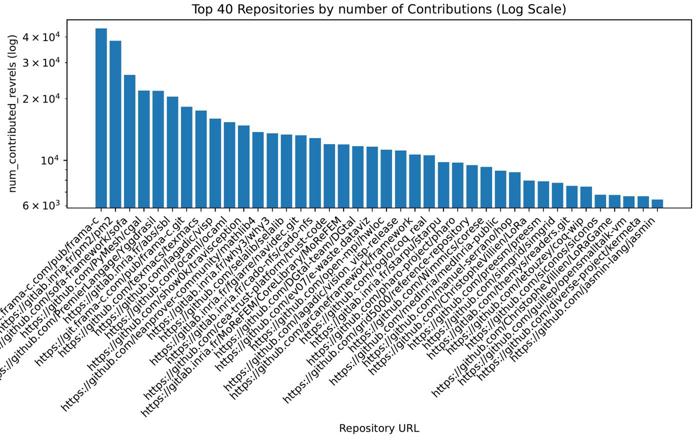
Still crawled: 44874

30101

Contributors

1992-07-08
2025-05-17

Time Span



domain	count
inria.fr	3833
univ-lyon1.fr	2806
univ-orleans.fr	1555
uca.fr	1371
jussieu.fr	1247
utc.fr	1216
cea.fr	1187
insa-toulouse.fr	747
ens-lyon.fr	706
univ-artois.fr	696
inrae.fr	690
univ-lemans.fr	688
cnrs.fr	682
ens.fr	675
inra.fr	644
univ-fcomte.fr	634
unilim.fr	541
insa-rennes.fr	491
univ-avignon.fr	489
univ-pau.fr	483
utt.fr	472
polytechnique.fr	437
uha.fr	422
pasteur.fr	420
telecom-paris.fr	412
univ-cotedazur.fr	363
univ-smb.fr	352
univ-tours.fr	350
u-picardie.fr	349
enpc.fr	328
sciencespo.fr	327
universite-paris-sa	323
univ-poitiers.fr	312
u-paris.fr	289
univ-angers.fr	267
univ-paris8.fr	264
inserm.fr	211
centralesupelec.fr	192
ird.fr	171
imt-atlantique.fr	159
onera.fr	158
univ-tlse2.fr	131
irstea.fr	130
enac.fr	126
univ-eiffel.fr	125
agroparistech.fr	105
cnam.fr	104
mnhn.fr	95
uphf.fr	94
univ-tln.fr	89

Note: Showing top 50 of 371 total domains

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

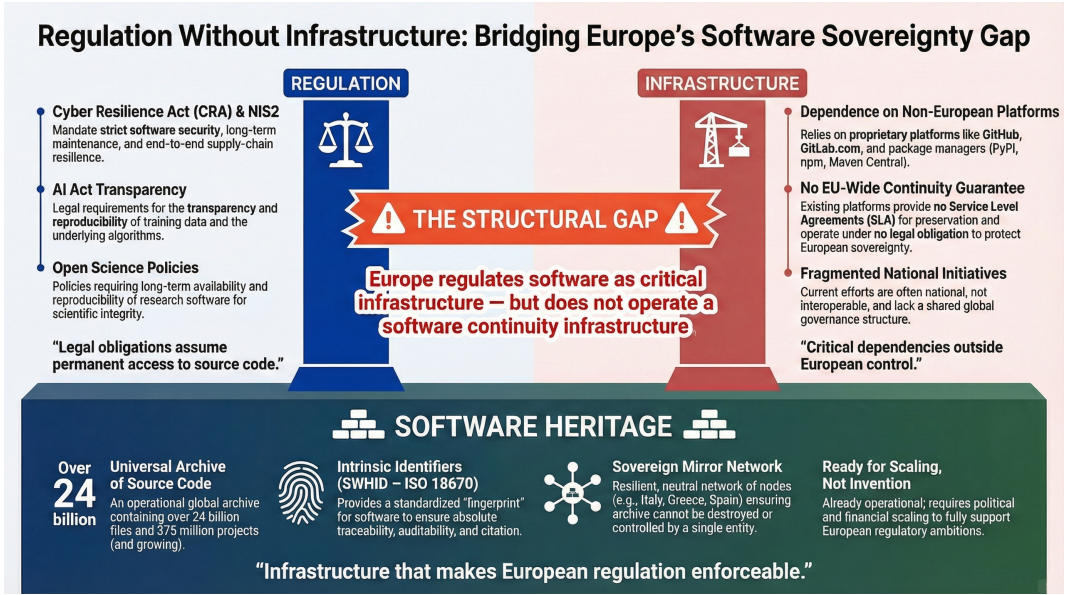
Transparent AI

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

Metrics we can count on

- Objective Strategic Insights
- Support policy implementation

The missing foundation to fill the gap is here



Outline

A plan for Europe

Europe does not need to invent this infrastructure…

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

Next steps

All recognise software continuity as critical enabler for

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

All avoid balkanisation, support mutualisation

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

DG RTD / DG CONNECT coordinate a joint effort

- grow Software Heritage into the EU strategic software backbone
- unlock funding *now* to close the gap (~20 to 40Me table stakes)
- build the long term resilience and continuity strategy