Software Heritage's contribution to the EU Open Source strategy

Software Heritage – January 2026

This document is Software Heritage's contribution to the European Commission's consultation on the European Strategy for Open Digital Ecosystems. Here is an excerpt from the consultation:

> *"The initiative seeks to support the EU's tech sovereignty and competitiveness agenda. It will identify some of the barriers that are currently hampering the EU's open-source potential and propose a path forward to eliminate them. In particular, the initiative will focus on improving the EU open-source sector's ability to grow and develop further. A strong and developed open-source sector can effectively contribute to further EU innovation and accelerate standardisation, strengthening the EU's international competitiveness, preserving its sovereignty, and ensuring its continuous economic prosperity, security, resilience and global influence."*

This note proposes targeted EU actions to strengthen Free Libre Open Source Software (hereafter FLOSS[1]) as digital public infrastructure, with Software Heritage - the universal, non-profit archive of software source code - as a core enabling service.

Support FLOSS for accountable governments

- As the "Rules as Code" movement increasingly transforms legislation into executable software, FLOSS becomes essential for ensuring government accountability and public transparency.
- European regulations are most easily and cost-effectively achieved through interoperable software. Since the use of Free Software inherently promotes interoperability, FLOSS should become the default choice whenever interoperability is important.
- As software becomes more interdependent, interoperability also requires persistent identifiers such as the SWHID[2].

The EU should:

- recognize the importance of FLOSS for accountable government;
- recognize the importance of FLOSS interoperability for cost-effective implementation of European legislation;
- promote SWHID wherever interoperability between government infrastructures is important, e.g. in cybersecurity databases.

1 FLOSS software is software released under a license that gives users the freedom to use the software, study its source code, modify it, and redistribute modified versions of it.

2 A SoftWare Hash IDentifier (SWHID) is an ISO-standardized, intrinsic, and persistent identifier that uniquely references software source code artifacts by computing a cryptographic hash directly from their content.

Support SWHID in SBOM[3] for CRA-ready industry actors

- The Cyber Resilience Act (CRA) requires that products with digital elements sold in the EU meet security requirements across the lifecycle.
- A SBOM item isn't just "we use OpenSSL", it must identify the *exact* artifact included - in a machine-verifiable way, independent of Git host, forge migration, or URL rot.
- Vendors will increasingly refuse to integrate components that cannot be reliably identified and archived.

In this context, Software Heritage persistent identifiers (SWHIDs) provide exactly this kind of stable, content-based reference for source code and other software artifacts.

The EU should:

- recognise content-based, persistent identifiers such as SWHIDs as best practice for SBOMs;
- encourage or require their support in EU-promoted SBOM formats and tooling (e.g. SPDX, CycloneDX); and
- ensure that FLOSS components used in regulated products can be reliably identified and archived via Software Heritage.

Support FLOSS for reproducible research

Software is now ubiquitous in scientific research. In theory, this is good news for reproducibility, but in practice, too little source code is published with papers, and too little of it is easily reproducible today. The EU should:

- treat publication of source code as first-class research outputs and require in research funding (Horizon Europe and successors);
- require that funded software be released under FLOSS licenses, with adequate support for long-term archiving and citation, for instance via dedicated calls and evaluation criteria in Horizon Europe work programmes.

Support practical data portability of FLOSS projects

- Most FLOSS projects are hosted on github.com, which is owned by a US company.
- Source code can be easily moved from one forge to another, but all other data cannot.
- Data portability for software forges should be treated as part of the EU's broader data and platform regulation agenda (building on the Data Act and DMA principles, and future updates to the EU Open Source strategy).
- Initiatives like ForgeFriends (https://f3.forgefriends.org/) and CodeCommons (https://codecommons.org) address this gap and deserve EU support.

The EU should:

- support open, documented APIs and formats for full project export/import (issues, merge requests, CI configs, metadata, etc.);
- fund and politically back initiatives that implement practical cross-forge portability and archiving.

3 A Software Bill of Materials (SBOM) is a formal, machine-readable inventory that lists all components, libraries, and dependencies within a software application to ensure supply chain transparency and security.

Support resilient digital services for FLOSS

- Critical build and distribution services (e.g. Maven Central, PyPI, npm, container registries) are now part of Europe's digital critical infrastructure, but often depend on non-EU or single-provider services.
- The EU should develop contingency plans to remain independent in the event of a shutdown of these services.
- Software Heritage can help by providing useful snapshots for rebuilding packages.

The EU should:

- map these services as part of its digital critical infrastructure;
- support EU-based mirrors and fallback services;
- use Software Heritage snapshots and archives as a recovery layer for rebuilding packages when upstream services are unavailable or compromised.

Support responsible use of source code for training LLM[4]s

- Almost all LLMs are trained on source code, not just those specializing in code generation.
- Ensuring that output code is traceable isn't just important for copyright reasons, it's also critical for security.

The EU should promote responsible use of source code in AI training by:

- supporting mechanisms to trace generated code back to archived origins (e.g. via SWHIDs);
- aligning AI-related regulation with the Software Heritage statement on LLMs for code.

The goal is not to prevent the use of open source in AI, but to ensure that such use remains compatible with security, legal certainty, and the sustainability of the upstream FLOSS ecosystem.

In summary

This note builds on the experience of Software Heritage in archiving, identifying and preserving the global body of software, and on its collaborations with EU institutions and research projects.

Policy recommendations: Use FLOSS to promote interoperability and government accountability; treat FLOSS as infrastructure for reproducible research; enforce data portability for software forges; promote responsible use of source code in LLM training.

Operational recommendations: Recognise SWHIDs as best practice for SBOMs under the CRA; treat major software registries as critical infrastructure and plan EU-based resilient mirrors, backed by Software Heritage archiving.

4 A Large Language Model (LLM) is an AI system trained on massive datasets to understand, generate, and manipulate human-like text based on probabilistic patterns.