Software Heritage

Membership
Program

Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collecting, preserving and
sharing software source code
since 2015

About Us

Software Heritage is a non profit multi-stakeholder initiative launched by Inria in partnership with UNESCO, hosted by the Inria Foundation, and with a growing number of partners. It is building the **universal archive and knowledge base of software source code**, at the service of society as a whole.

The Software Heritage Symposium and summit 2024 took place at UNESCO's headquarters on February 1st bringing together advisors, sponsors, ambassadors, and the entire community.
© Inria | Photo M. Magnin.

The **Software Heritage archive** is the largest collection of publicly available source code ever built, containing, as of October 2024, more than 20 billion unique source files from over 320 million software origins.

Hosted by

In collaboration with

RÉPUBLIQUE FRANÇAISE
*Liberté
Égalité
Fraternité*

Inría

Inría La Fondation

unesco

Software Heritage has been launched by Inria in 2015.

Our mission

We are building an essential infrastructure, that is meant to ensure three main properties for the source code we collect:

○ **Availability**

The code will be stored, preserved and made accessible on the long term.

○ **Traceability**

Eeach software component will get a unique identifier, called **SWHID**, that can be relied upon in the long term.

○ **Uniformity**

Despite the great variety of origins, all of the source code collected in our archive will be accessed through the same uniform Application Programmer's Interface (**API**)

A catalog to find them all

Software is spread all around: it is developed on many collaborative platforms and distributed through a variety of different channels. Software Heritage is building a **universal catalog** to let you **find** all software projects, no matter where they are developed, or how they are distributed.

An archive to preserve them

Modern software development relies on collaborative platforms, and many of them can be used free of charge. One can **create**, but also **modify** or **delete** projects: *they are not archives*. In recent years, we have seen several platforms come and go, sometimes suddenly, endangering hundreds of thousands of software projects all at once. Software Heritage is building the **universal archive** that is needed to ensure we will not loose source code any more.

© iStock | kynny

An instrument to explore and study them

Software underlies all aspects of our modern societies, and in a few decades we have built software systems of incredible complexity: some are huge programs, with tens of millions of lines of code, some are smaller programs, but mostl rely on hundreds or thousands of other components. We need to master this complexity, in order to build better, safer systems, and protect against malware.

Humankind has been able to build marvelous instruments to explore the universe, now it's time to build a common, shared infrastructure to explore and study the galaxy of software development. With enough support, Software Heritage can evolve into such an infrastructure.

HUMANKIND HAS BEEN ABLE TO BUILD MARVELOUS INSTRUMENTS TO EXPLORE THE UNIVERSE, NOW IT'S TIME TO BUILD A COMMON, SHARED INFRACTURE TO EXPLORE AND STUDY THE GALAXY OF SOFTWARE DEVELOPMENT.

## All the public code history in a giant graph

Merkle graphs and SWHID

A massive crawler harvests source code from different sources and converts it, with all its development history, into a single giant Merkle directed acyclic graph, using SWHID cryptographic identifiers for all its nodes.

50 | 700
billion nodes | billion edges

THE SOFTWARE HERITAGE DATA STRUCTURE IS A NATURAL EXTENSION OF MERKLE TREES, A CLASSICAL CRYPTOGRAPHIC CONSTRUCTION, COMBINING A TREE AND A HASH FUNCTION. [MERKLE, 1987]
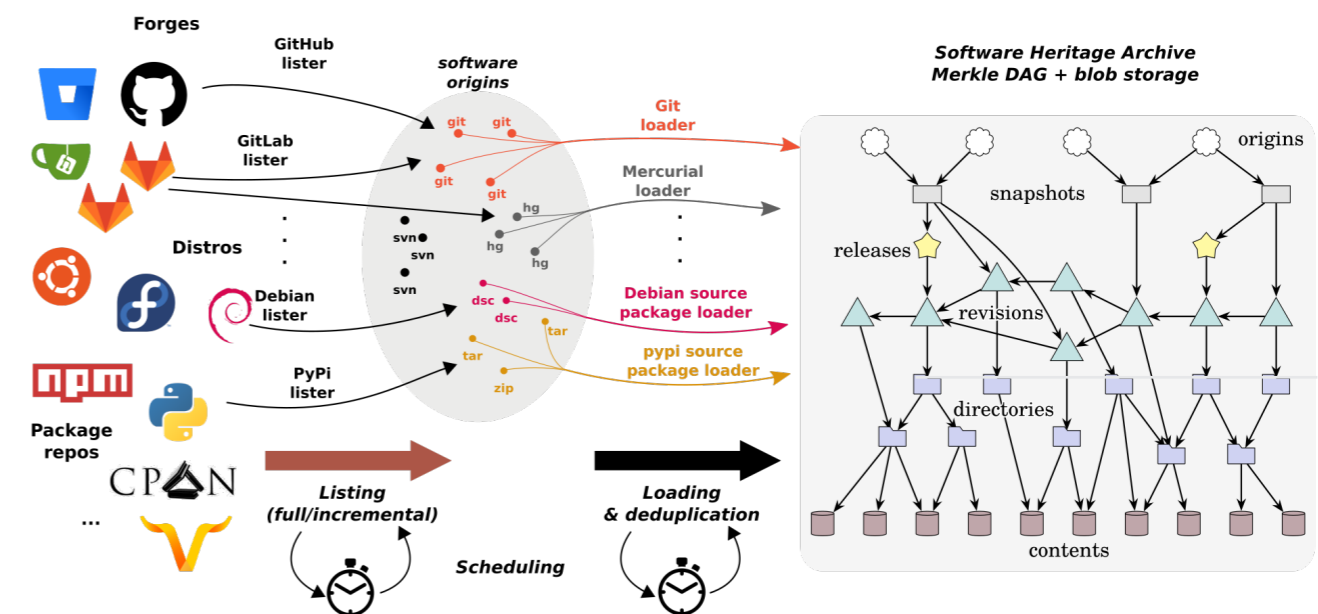
The process is separated into three phases: *listing software sources*, *scheduling updates* and *collecting the software artifacts* into the archive.

Industry

Securing the Open Source software supply chain

The *uniform, technology-neutral global Merkle graph* provides, with the growing mirror network, **a transparent source of trust.** Source code artifacts, at all levels, are tracked using the **S**oft**w**are **H**ash **ID**entifier (SWHID) that provides uniform, technology-independent, and cryptographically strong identification. SWHID normalization is ongoing.

Complete and corresponding source code

Industry players can **delegate** to the Software Heritage archive the task to preserve and make available to third parties the complete and corresponding source code of any open source component they use. They can use the Software Heritage as a trusted reference in their agreements.

Enhancing Cybersecurity through Software Heritage

The Software Heritage for Security project (SWHSec) brings together eight expert research teams specializing in security and software engineering to harness the power of Software Heritage's robust infrastructure and create cutting-edge tools for cybersecurity.

The SWHID intrinsic persistent identifier

All artefacts in the Software Heritage archive get a **S**oft**w**are **H**ash **ID**entifier, or **SWHID** for short, that is guaranteed to remain stable (persistent) over time.
A SWHID consists of two parts, a mandatory *core identifier*, and an optional list of *qualifiers* that specify the context and can pinpoint a subpart.


```
schema_version          object_id

swh:1:cnt:41ddb23118f92d7218099a5e7a990cf58f1d07fa

prefix      object_type
```

▭	"snp" - snapshot
☆	"rel" - release
△	"rev" - revision
▭	"dir" - directory
⬓	"cnt" - content

origin_ctxt → `;origin=https://github.com/chrislgarry/Apollo-11`

visit_ctxt → `;visit=swh:1:snp:206c27c0c031c6aac6b5fedddba8fe082dea9836`

anchor_ctxt → `;anchor=swh:1:rev:3913f198f4383d4d638c0485d6aa902ff2f35828`

path_ctxt → `;path=/Luminary099/BURN_BABY_BURN--MASTER_IGNITION_ROUTINE.agc`

lines_ctxt → `;lines=64-72`

Contribution to implementation of security regulations

The Cyber Resilience Act (CRA) is changing the way industry will handle security for open source components. Software Heritage is a founding member of the Open Regulatory Compliance Working Group (ORCWG), dedicated to helping open-source projects navigate emerging regulatory requirements while ensuring that innovation and collaboration continue to thrive.

Compliance for source code distribution

Software Heritage is a long term archive for source code
Software Heritage is a non profit multi-stakeholder organisation whose stated mission is to collect, preserve and the source code of all available software. Forever.

Software Heritage simplifies source code distribution
Companies adopting Software Heritage can effectively *outsource their source code distribution obligations*. Given its mission of long term software preservation, Software Heritage is the **perfect steward** for the source code it receives.

Software Heritage provides perpetual identifiers
Software Heritage provides **SWHIDs**, uniform, cryptographic identifiers for all of the tens of billions of software artefacts it archives. These *intrinsic* identifiers allow to *independently* verify the integrity of the software artefacts they denote. **Without having to trust any third party**.

Software Heritage can be integrated in your compliance process, saving time and resources
Source code deposit in Software Heritage can be integrated in any compliance process. As soon as code is deposited, you get the corresponding intrinsic identifier: that's the only piece of information you will ever need to be compliant. Software Heritage **takes over** all other distribution obligations from there. By contrast, typical in-house industry approaches require to maintain dedicated infrastructure and resources **over a very long term**.

Software Heritage allows you to rely on a mutualised infrastructure
When you join Software Heritage, you adopt an **open infrastructure** that is backed by a growing worldwide community that brings together stakeholders in **industry, open science,** and **digital preservation**.

How it works

1. You prepare
the complete and corresponding source code (CCS) archive, *as usual*.

2. You deposit the archive
into the Software Heritage platform, using an API that you can integrate with your continuous delivery process.

3. You get back the SWHID,
a perpetual identifier that anybody can use to retrieve and browse the source code you deposited.

4. You are happy to know
that Software Heritage preserves the archive and will host it forever.

And yes, this seamless process is fully compliant with all major copyleft licenses!

SUPPORTING SOFTWARE HERITAGE

Becoming a sponsor

Cultural Heritage	Industry	Research	Public Administration

Software Heritage

One common infrastructure shared

Software Heritage builds a key digital shared infrastructure:

- it needs your support, because it is a non-profit operation;
- as a digital commons, you mutualize the cost with other sponsors.

Your contribution supports

- **Infrastructure Development:** storage, compute, and network resources to support the ever-growing archive.
- **Software and Standards Development:** new key features to enhance archive functionality and usability.
- **Innovation and Alignment:** stay at the forefront of innovation and aligned with the latest advancements (like AI and Cybersecurity).
- **Community Engagement:** foster adoption, grow the community and connecting with partners worldwide.
- **Open Source Preservation:** a safe haven for Open Source software, safeguarding it against loss and making it perpetually retrievable.
- **Operational Sustainability:** infrastructure expansion, software development, and normal operations to keep the archive and the organization running smoothly.
- **Long-Term Mission Support:** bring the infrastructure to operational and financial maturity, ensuring the long-term sustainability of operations.

Benefits

1. Supporting Software Heritage's long term mission contributes to your **social responsibility agenda**
2. You have an opportunity to **contribute your views on the technical roadmap**
3. You get **advanced information on new developments**
4. The annual general meeting is a great occasion to **meet the supporter network** and learn about the latest progress and perspectives
5. Depending on local legislation, your contribution **may be tax deductible**

Deposit Interest Group

The **Deposit Interest Group** brings together stakeholders that are interested in using, supporting and expanding the *Software Heritage infrastructure as the archive of choice* to deposit open-source software artifacts that they produce or use, with appropriate metadata.

Sponsorships tiers

Sponsors can choose to join at different tiers, depending on the level of support they want to provide for the stability, long term availability and evolution of the Software Heritage archive infrastructure.

Diamond sponsors

See Software Heritage as a strategic infrastructure and commit significant resources that contribute to the *stability of the platform* in the medium to long term, as well as to its necessary *technological development*. They contribute to the *strategic roadmap* of Software Heritage.

Platinum sponsors

Consider Software Heritage as a core infrastructure and provide funding for its *regular operation* as well as advice for its short term development. They contribute at different levels, according to tier, to the *technical roadmap* of Software Heritage.

Gold sponsors

See Software Heritage as an important infrastructure for preserving and sharing software source code, and want to *get involved in the community* that sustains it.

Silver sponsors

Use Software Heritage for their open-source software archival needs.

Fees and Benefits

Software Heritage is a Foundation, hosted by the Inria Foundation.
The sponsorship fees and associated benefits, valid in the period 2023-2024, are summarised below:

Annual sponsorship fees and benefits*	Diamond	Platinum			Gold	Silver
		Tier 3	Tier 2	Tier 1		
	250K€	200K€	150K€	100K€	50K€	25K€
Strategic Advisory Board	*					
Contribute to Platform Stability	*	*				
Technical Advisory Board	*	*	*			
Compliance Working Group	*	*	*	*		
General Assembly	*	*	*	*	*	
Deposit Code	*	*	*	*	*	*

(*) a reduced academic fee may be available for qualifying institutions, inquiry to learn more.

Join us! Becoming a sponsor is very easy: contact us at **sponsor@softwareheritage.org** or fill the online form

You are in great company!
Thrive together!

Software is key in **CEA**'s commitment to transferring knowledge from research to industry.
With the Software Heritage Foundation, we stand behind the preservation and sharing of this knowledge.

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

The National Open Science Plan was launched on 4 July 2018 by the Minister of Higher Education, Research and Innovation. This plan includes a provision to support Software Heritage, an initiative that we consider a major pillar of open science. In addition to enabling open access to publications and research data, making research software source code openly available is critical to success of the open science program that we are collectively building.

CNRS's support to Software Heritage, a universal, open and sustainable software archive, is a natural part of our proactive approach in favour of open science, a necessary revolution in which everyone must play a part.

Microsoft has been involved in open source initiatives by enabling, integrating, releasing and contributing to many open source projects and communities for well over a decade. We applaud the Software Heritage as an open project that will help curate and conserve human knowledge in the form of code for future generations as well as help today's generations of developers find and re-use code worldwide.

Intel has been at the forefront of open source development for nearly two decades and today is a top contributor to the Linux kernel, as well as dozens of leading projects across technology markets and industries.
Intel is committed to support Software Heritage in its mission to collect, preserve and share code, as we believe open source is critical in transforming our world through innovation in enterprise, consumer technology, the Internet of Things and beyond.

Huawei has been working with the open source communities for decades: we are active contributors in projects ranging from the Linux kernel to cloud native computing and machine learning, and we will keep increasing our participation and investment in this open innovation world. We share Software Heritage's vision that publicly available source code, including open source software, is a precious heritage of mankind, and should be collected, preserved and shared for the benefit of all.

Google

Google is proud to support Software Heritage in its mission to collect, preserve, and share software for future generations. We look forward to the variety of services that can be built atop this unique collection of software.

Hugging Face

Partnering with Software Heritage was a great journey for BigCode and **Hugging Face**. The foundation's focus on preservation, reproducibility, availability and traceability mirrors many of the values and mission of Hugging Face as a central platform for sharing and collaborating in the ML community.

openinvention network

Open source software has been one of the instrumental, driving forces of innovation this century. Software Heritage is an important organization for software, (...). Archiving of code in a curated form maintains the technical and scientific knowledge that goes along with the code, preserving the innovation while also providing a means for determining prior art.

servicenow

At **ServiceNow** we recognize the value and importance of preserving open-source software (...). We firmly believe in the capacity of Software Heritage to cultivate goodwill and collaboration within the technology ecosystem, while promoting a more sustainable and open software industry.

SOCIETE GENERALE

We are aware of the code's value for our digital transformation, it has become a major asset for the bank and we firmly believe that we must preserve it in the long term. Open Source lies at the heart of our strategy, as it is in line with our needs and our values: team spirit, innovation, responsibility and commitment to better serve our clients.

SORBONNE UNIVERSITÉ

Firmly committed to open science, which is at the heart of its project, **Sorbonne University** supports Software Heritage. By helping to collect and to share software, Software Heritage contributes to one of the key missions of the university: the preservation and transmission of knowledge and of our scientific heritage.

Université Paris Cité

By supporting the Software Heritage initiative, Université de Paris continues its commitment to the free and responsible sharing of knowledge and research software.

AdaCore

RÉPUBLIQUE FRANÇAISE
Liberté · Égalité · Fraternité
La DINUM

MINISTÈRE DES ARMÉES
Liberté Égalité Fraternité

DGA
DIRECTION GÉNÉRALE DE L'ARMEMENT

GitHub

UNIVERSITÀ DI PISA

SCANOSS

SCUOLA NORMALE SUPERIORE

Alma Mater Studiorum

Ready to make a difference?
Join us!
Together, let's ensure the
legacy of technology thrives for
generations to come.

Software Heritage will provide solid, common foundations to serve the different needs of heritage preservation, science, and industry.

Becoming a sponsor is very easy: fill the online form →

🌐 softwareheritage.org
in @swheritage
𝕏 @SwHeritage
🐘 @swheritage@mstdn.social
▶ @softwareheritage4978

Contact us at sponsor@softwareheritage.org