



PRESS RELEASE



Collect, organise, preserve and share the Software Heritage of mankind

Inria announces today the Software Heritage project, an ambitious initiative to collect, organise, preserve, and make easily accessible any already publicly available source code.

Sending messages to our family and friends, paying bills, purchasing goods, accessing entertainment, interacting with the public administration, finding information, booking travels: practically *every act of our daily life* relies on computers and software to be performed.

That is just the tip of the iceberg: software controls the electronic equipment embedded in the machines we use to travel, communicate, trade and exchange. Software lies at the heart of medical equipment and devices; software ensures proper operation of the power, transportation, and telecommunication networks; software powers banks and financial institutions; software is just crucial for the proper operation of large organizations, be them public or private, in mobile devices or in the cloud.

In a word, software is today a key enabler for all aspects of our modern world: *our industry, our science, our lifestyle, and all of our society depend on software.*

A societal, technical and scientific challenge

Inria unveils today the **Software Heritage** project, whose goal is to build at the same time a modern « library of Alexandria » of software, a unique reference data base of all source code, a tool for new software projects and a research instrument for Computer Science: a societal, technical and scientific challenge.

Software Heritage is an essential building block for preserving, enhancing and sharing the scientific and technical knowledge that is increasingly embedded in software; it also contributes to our ability to access all the information stored in digital form.

Software Heritage will adopt a distributed infrastructure in order to ensure long term availability and reliability of its archive.

Software Heritage provides a reference knowledge base for all open source software used in industry, enabling better lifecycle management and long term preservation of industrial software. Once enriched with live update capabilities, Software Heritage is bound to become the reference archive for all industrial users, helping software developers of new software projects find, re-use and archive new source code.

Software Heritage is the foundation on which we can build a unique research instrument for studying all the software source code, enabling significant advances in all domains of computer science, and leading to better quality, security and safety in the software we depend on in our daily lives.

Software Heritage today

Inria, a French national science and technology institution dedicated to computational sciences, has decided to tackle these challenges, and set up a dedicated team and the initial resources to start the project.

As of today, **Software Heritage** already collected more than **20 million software projects**, archiving more than **two billions and a half unique source files**. Together with all their development history, this is *the richest collection of source code on the planet*.

Antoine Petit, INRIA's CEO, says "We decided to start working on Software Heritage more than a year ago, and we have now shown its feasibility. In order to scale up worldwide, the time has now come to open it up to the widest, national and international contributions".

Two early partners have already committed their support to Software Heritage, and will help it grow: Microsoft, one of the largest software industries in the world, and [DANS](#), an institution of the Royal Academy of the Arts and Sciences and the Netherlands Organisation for Scientific Research, dedicated to preserving and promoting sustained access to digital research data.

"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," says **Jean Paoli, General Manager at Microsoft Corp.** "We applaud 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. We are proud to be one of the first industry partners for this initiative and to provide the Azure infrastructure to ensure the data is highly available."

"Software Sustainability is one of the core elements to accelerate discovery. Reproducibility of scientific results, but also re-usability and findability of software are core principles in science DANS seeks to support" **says Peter Doorn, Director of DANS.** "We are committed to long-time preservation of scientific information and are delighted to be the first international academic partner to support the mission of Software Heritage".

We are all concerned, everybody can contribute

Software Heritage has already been [endorsed](#) by scientists, industry players, learned societies, foundations, as well as a variety of organisations, both public and private.

After launching the project, having shown its feasibility, and established the first partnerships, Inria is now calling all stakeholders worldwide to join. Inria opens the project's website:

<https://www.softwareheritage.org/>

Software Heritage is an ambitious project that requires contributions from multiple players:

- Collecting all the software: help us identify the thousand different sites where the world's software heritage is now spread around.
- Contribute to developing the infrastructure: the Software Heritage team has a long tradition of collaboration, and is well known in the free and open source arena; in the coming days, we are going to open up our own source code to the world, and we will welcome the developers that share our vision and want to help in this mission.
- Solving the scientific challenges coming with the development of a universal source code archive from disparate information will require new insights; the participation of researchers from all disciplines will be instrumental to succeed.

Preserving in the long term and sharing with the world the contents of the archive requires significant resources, in terms of manpower, infrastructure and funding, as well as partners all over the world.

Getting involved in **Software Heritage** is a unique chance to tackle a major challenge for society, to support an initiative with a global reach and to participate in the preservation of our global software heritage.

About Inria

Inria, the French National Institute for computer science and applied mathematics, promotes “scientific excellence for technology transfer and society”. Graduates from the world’s top universities, Inria’s 2,700 employees rise to the challenges of digital sciences. With this open, agile model, Inria is able to explore original approaches with its partners in industry and academia and provide an efficient response to the multidisciplinary and application challenges of the digital transformation. Inria transfers expertise and research results to companies (startups, SMEs and major groups) in fields as diverse as healthcare, transport, energy, communications, security and privacy protection, smart cities and the factory of the future.

Press Contacts :

Laurence Goussu – laurence.goussu@inria.fr
+ 33 (0)1 39 63 57 29 – + 33 (0)6 81 44 17 33

Hélène Latour – helene.latour@meanings.fr
+ 33 (0)7 60 46 66 24



PRESS KIT



**Collect, organise, preserve and share
the Software Heritage of mankind**

<https://www.softwareheritage.org/>

Software Heritage aims to collect, organise, preserve, and make easily accessible any already publicly available source code. By building a universal and sustainable software archive, the goal of Software Heritage is to create an essential infrastructure for society, science and industry.

Software at the heart of our society

Sending messages to our family and friends, paying bills, purchasing goods, accessing entertainment, interacting with the public administration, finding information, booking travels: practically *every act of our daily life* relies on computers and software to be performed.

That is just the tip of the iceberg: software controls the electronic equipment embedded in the machines we use to travel, communicate, trade and exchange. Software lies at the heart of medical equipment and devices; software ensures proper operation of the power, transportation, and telecommunication networks; software powers banks and financial institutions; software is just crucial for the proper operation of large organizations, be them public or private, in mobile devices or in the cloud.

Software is at the crux of all technological development and has become essential for all areas of scientific research. Software in the modern world plays a pivotal, even critical, role in our daily lives, our industries and our society.

Digital transformation entails the conversion of all aspects of our knowledge and cultural heritage into digital representations, but as **Vinton Cerf, co-creator of the Internet**, points out "without the software capable of interpreting these representations, all this knowledge would be a worthless and unintelligible series of bits".

To preserve this heritage and prepare for the technological and scientific challenges of tomorrow, it's critically important to start building a **universal and sustainable software archive**.

Software Heritage will benefit society, science and industry

The central role of software in modern societies means that Software Heritage is an infrastructure with a universal reach across society.



Preserve software, make easily accessible digital knowledge

The software that we use is an important part of our modern culture, and Open Source software plays a major role. Up to now, a large number of software projects have been developed and distributed using a variety of disparate infrastructures, which change often and do not make any long term commitment. That's a risky gamble, and it became evident just a few months ago, when two important hosts that contained hundreds of thousands

of software projects shut down: Gitorious and Google Code.

Today the need to build an infrastructure that aims at indexing all publicly available source code, with a long-term commitment is clear: it is not just a matter of preserving old source code, but to help developers of new software projects find, re-use, reference and archive new source code. **Software Heritage has set itself this goal.**

Providing a single and universal archive making software source code readily available will facilitate access to the knowledge contained therein, support programming education, and create a reference catalogue with all knowledge about this software, a kind of software Wikipedia.

Software heritage will adopt a distributed infrastructure, with international partners, in order to ensure long term availability and reliability of its archive.

Preserving and sharing software for Science

A highly significant share of scientific knowledge is produced by, or contained in, software¹. However, scientific software archiving is rare and poorly managed, with ineffective or nonexistent references, to the extent that the reproducibility of scientific experiments that use this software is not guaranteed. If we are to preserve technological and scientific knowledge, it is now imperative to preserve not only the published articles, but also the source code of the software described in the scientific articles and used for data processing.



Research in computer science has made great progress: it is now becoming possible to automatically identify programming errors, design faults or cyber attacks. However, unlike other disciplines such as physics, there is no large scale instrument to support systematic exploration of all software source codes.

Software Heritage provides a fundamental research infrastructure, with the objective of becoming the reference for access to software used in all scientific research, supplementing the Open Access approach, which aims to make scientific articles freely accessible.

Software Heritage is the foundation on which we can build a unique research instrument for studying all the software source code, enabling significant advances in all domains of computer science, and leading to better quality security and safety in the software we depend on in our daily lives.

¹ For example <http://www.nature.com/news/the-top-100-papers-1.16224> shows how a large share of the most cited Nature articles describe or use scientific software.



Improve software quality for industry

Software security and quality are crucial to many industrial sectors, such as the aerospace sector, healthcare, energy, transport and security, all of which lie at the heart of our society. Today, industrial software is built by assembling a wide range of software components, for which **traceability** and **quality assurance** are essential.

Software Heritage is building a reference database for all free and open source software, enabling improved life cycle management and long-term accessibility for present and future industrial software development. All software artefacts in **Software Heritage** will obtain a unique and long-lasting reference that can be used as a base for services related to security, certification, traceability and code qualification. When equipped with real-time updates, **Software Heritage** is bound to become the reference source code catalogue for all industrial users.

Software Heritage today

The project has been initiated by **Inria**, a research institute for computer science that has a well-established role in the software arena for a long time, owing to its conviction regarding the essential role of software in the development of a digital society. Inria has decided to tackle these challenges, and set-up a dedicated team and the initial resources to start the project.



As reiterated by **Antoine Petit, Inria Chairman and CEO**, "Software is now at the heart of the majority of human activities: from healthcare to entertainment, planes to agriculture... It puts the "smart" into transports, towns, homes and even factories. Software therefore has a central and even critical role in our daily lives. For almost 50 years, researchers from Inria, a digital science research institution, have produced several thousand software packages, including some that have led to major scientific breakthroughs and others that are the basis of commercial products, namely the 130 Inria spin-offs. It is therefore legitimate for Inria to address the issue of preserving all software-related technological and scientific knowledge, to help build and conserve world software heritage and to provide access to the latter for industry, science, education and the society at large".

For **Roberto Di Cosmo, Software Heritage Director**, "Mankind's technological knowledge is inseparable from software: our entire society and our entire culture depends on it. The Source Code of software is special: it is executable knowledge, which is human readable by design, an all time new in human history. We have a duty to preserve this knowledge. Supporting **Software Heritage** means taking part in an essential project that preserves the past to prepare for the future".

As a global initiative, **Software Heritage** faces important challenges.

Organising, and indexing in an efficient and effective way the source code of all software that is publicly available in order to make it easily accessible, and extracting useful information to improve its quality, pose a lot of questions.

From a technical point of view, it entails ensuring data can be maintained over the long term, through widespread replication of the infrastructure, providing stable references for each software component, taking into account the specific aspects of software source code, which is different from other digital objects.

From a strategic point of view, an institutional legitimacy is of course required, as is a real capacity for openness to build a wide consensus. This was already the case for the World Wide Web consortium, for which Inria was the first European host.

For Jean-François Abramatic, Inria Senior Researcher, and former Chairman of W3C, "More than 20 years ago W3C was hosted in Europe, with the vision of providing an independent, open and free Web, accessible to all. Today, the Software Heritage project has kicked off with a similar drive and vision for international contribution to preserve and nurture the software commons on the long term and at a global scale".



To date, **Software Heritage** has collected more than **20 million software projects**, archiving more than **two and a half billion unique source files**, as well as all related development history, already making it **the world's leading and most comprehensive source code archive** (<https://www.softwareheritage.org/archive/>).

Stefano Zacchiroli, Software Heritage CTO, "Our ambition is to collect, preserve, and share the entire software commons, the vast majority of which consists of Free and Open Source Software. Openness and transparency will be key to our success. That's why our own software will be entirely open source and we invite all Free Software actors to join us in the pursuit of this ambitious goal."

This is just the beginning: by visiting www.softwareheritage.org, you can monitor in the growth of the database containing the collected codes in real time and discover how to contribute to it.

The first partners

We're pleased to announce today that two early international partners have already committed their support to Software Heritage, and will help it to grow: Microsoft, one of the largest software industries in the world, and DANS (<https://dans.knaw.nl/en>), an institution of the Royal Academy of the Arts and Sciences and the Netherlands Organisation for Scientific Research, dedicated to preserving and promoting sustained access to digital research data.

"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," says **Jean Paoli, General Manager at Microsoft Corp.** "We applaud 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. We are proud to be one of the first industry partners for this initiative and to provide the Azure infrastructure to ensure the data is highly available.”

“Software Sustainability is one of the core elements to accelerate discovery. Reproducibility of scientific results, but also re-usability and findability of software are core principles in science DANS seeks to support” says Peter Doorn, Director of DANS. “We are committed to long-time preservation of scientific information and are delighted to be the first international academic partner to support the mission of Software Heritage”.

Software Heritage: we are all concerned, and everyone can contribute !

Software Heritage has already been endorsed by scientists, industrial players, learned-societies, foundations, as well as a variety of organisations, both public and private.

All testimonials are available www.softwareheritage.org/support/testimonials/

After launching the project, shown its feasibility, and established the first partnerships, Inria is now calling all stakeholders worldwide to join, and opens the project's website:

<https://www.softwareheritage.org/>

The objective is to rapidly create an international community, to be managed by an autonomous consortium, as was the case for W3C. The first to be concerned are key players from the digital sector, public interest foundations, scientific organisations and French and International public players.

But all manner of contribution is expected:

- Collecting all the software: help us identify the thousand different sites where the world's software heritage is now spread around.
- Contribute to developing the infrastructure: the Software Heritage team has a long tradition of collaboration, and is well known in the free and open source arena ; in the coming days, we are going to open up our own source code to the world, and we will welcome the developers that share our vision and want to help in this mission.
- Solving the scientific challenges coming with the development of a universal source code archive from disparate information will require new insights, and researchers, from all disciplines, will be instrumental to succeed.

Preserving in the long term and sharing with the world the contents of the archive requires significant resources, in terms of manpower, infrastructure and funding, as well as partners all over the world.

Over the coming days, **Software Heritage** will open its own source code, so that external contributors, particularly from the Open Source Software sector, can participate in the drive to develop this common infrastructure, and help the archive to grow.

Getting involved in **Software Heritage** is a unique chance to tackle a major challenge for society, to support an initiative with a global reach and to participate in the preservation of our global software heritage.