

# Collect, organise, preserve and share the Software Heritage of mankind » Inria@SiliconValley



Software Heritage: Universal software archive

On June 30th 2016, Inria announced the launching of [“Software Heritage”](#), an ambitious initiative to collect, organise, preserve, and make easily accessible the source code of all software that is publicly available.

By building a universal and sustainable software archive, the goal of Software heritage is to create an essential infrastructure for society, science and industry.

Find out more about the project in this article featuring an [Interview of Roberto DiCosmo, Founder and CEO of Software Heritage](#). A societal, technical and scientific challenge

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 depends on software.

The Software Heritage project wants 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. Software Heritage is an essential building block for preserving 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 the open source code 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 developers of new software projects find, re-use and archive new source code.

Software is now at the heart of the majority of human activities: from healthcare to entertainment, planes

to agriculture... It is therefore legitimate for Inria to address the issue of preserving all software-related 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.

**Antoine Petit, CEO of Inria**

### **Software Heritage today : from Inria to Microsoft**

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 make it scale it up worldwide, the time has now come to open it up to the widest, national and international contribution "*.

Software Heritage has already been [endorsed](#) by scientists, industry players, learned societies, foundations, as well as a variety of organisations, both public and private. Besides, two international partners have endorsed the project : Microsoft and DANS, a public institute of the Royal Academy of Netherlands.

### **We are all concerned, everybody can contribute**

After launching the project, shown its feasibility, and established the first partnerships, Inria is now calling all stakeholders worldwide to join, and opening the [project's website](#).

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.

Source: <https://www.inria.fr/en/news/news-from-inria/launching-of-software-heritage> Interview: Roberto Di Cosmo (Founder, CEO)



**About Roberto:** After teaching for almost a decade at [Ecole Normale Supérieure](#) in Paris, Roberto Di Cosmo became full professor in Computer Science at [University Paris Diderot](#). He is currently on leave at [Inria](#)

to lead the [Software Heritage](#) project.

His research interests span a wide spectrum from foundational aspects of logical systems to functional programming, parallel and distributed programming. He created and directed the european research project [Mancoosi](#) to improve the quality of large collections of software quality, and is investigating now the scientific problems posed by the general adoption of Free Software, with a particular focus on static analysis of large software collections.

A long term Free Software advocate, contributing to its adoption since 1998, he has created the [Free Software thematic group](#) of [Systematic](#) in October 2007, which has helped fund over 40 research and development projects, and he is now director of [IRILL](#), a research structure dedicated to Free and Open Source Software quality.

- Roberto, what compelled you to engage in this project?

– Software lies at the heart of our society and is driving all aspects of the digital transformation. Software `_source code_` is a unique form of knowledge that is executable by a machine and yet meant to be read by humans: as Harold Abelson famously said, “Programs must be written for humans to read, and only accessorially for machines to execute”.

Software Source Code is a precious form of knowledge: as Len Shustek puts it, it “provides a view in the mind of the designer”, and it is of high value even when the machines on which it could run are no longer there (see for example the interest spawned by the release of the code of the Apollo 11, that Margareth Hamilton coordinated). When I realised that nobody was taking care of this precious form of knowledge that is the new literature of the digital age, I decided that it was our responsibility, as computer scientists and technologists, to take action. It was fantastic to see Inria support the project, with a clear will to make it grow into an international, non profit organisation.

It is a great mission: building the Library of Alexandria of Software, and the Very Large Telescope of Source Code at the same time.

- Can you tell us about your cooperations with the US? With California partners?

– We have of course established connections with many organisations and industries in the US, explaining what we plan to do, and how we plan to do it. Many of them share our vision (GitHub, GitLab, FSF, OSI, Eclipse, and dozens more, see [www.softwareheritage.org/support/testimonials](http://www.softwareheritage.org/support/testimonials)).

Microsoft has been the first US company to join forces with us, and is providing computing resources on Azure for establishing a first mirror: we are very happy of their engagement.

More recently, Nokia Bell Labs and Huawei become sponsors of the project. We would be delighted to see all other major IT companies (Google, Facebook, IBM,...) follow soon: this is a mission that is clearly for all of them.

Contact has been made with the Internet Archive project, as well as with the Computer History Museum, and we do hope to see concrete collaborations emerge from this.

More generally, all our own software is open source ([forge.softwareheritage.org](http://forge.softwareheritage.org)) and we welcome all contributors that want to try their hand at building this unique infrastructure that preserves and shares the

“software commons”

- What is next for Software Heritage?

– Software Heritage is a long term, worldwide undertaking and we are here for the long run: we have focused all our energies to collect the source code (in particular the one on discontinued platforms like Gitorious and Google Code). In a sense, we are filling the shelves of the library.

The very next step is to open the doors of the library to the world, to allow people to actually look at its content. Contributions to this effort are welcome: from individuals to corporations, from public to private institutions, everybody can help.

**More information:** <https://www.softwareheritage.org/>