

Inria pose la première pierre d'un "Wikipedia" du logiciel



© Capture d'écran Software Heritage - Inria

L'Institut de recherche dédié au numérique travaille depuis 18 mois sur le projet Software Heritage. Sorte de « Bibliothèque d'Alexandrie » du logiciel, il vise à collecter, organiser, préserver et rendre accessibles les codes sources de tous les logiciels disponibles publiquement. 2,5 milliards de fichiers sources uniques ont déjà été archivés. Pour aller plus loin, [Inria](#) entend créer une galaxie de partenaires.

Pendant un an et demi, les équipes de l'[Institut national de recherche en informatique et en automatique \(Inria\)](#) ont travaillé en catimini à créer les fondations d'un projet particulièrement ambitieux baptisé **Software Heritage**. Dévoilée ce jeudi 30 juin, l'initiative d'envergure mondiale consiste, ni plus ni moins, à développer le patrimoine logiciel de l'humanité. Réseaux de transport, électricité, finance, appareils médicaux, automobile, contrats... « *Le logiciel est partout, dans toutes les activités économiques, industrielles, individuelles et associatives. Ils (les logiciels, ndlr) forment un patrimoine extraordinaire qu'il faut préserver* » a souligné **Antoine Petit**, le PDG d'[Inria](#).

Un catalogue unique qui recense l'ensemble du code source disponible

Le projet est actuellement mené par une poignée de chercheurs, pilotée par **Roberto Di Cosmo**, professeur à l'université **Paris Diderot**. Il a expliqué que le cœur de cette initiative consistait à prendre soin du code source des logiciels libres, « *qui contient la vraie connaissance* ». « *Désormais, toute l'industrie adopte l'approche open source afin d'accélérer les développements. La plupart des start-up qui se créent, se développent en rajoutant seulement une petite touche au patrimoine logiciel déjà existant* » a-t-il expliqué.

Le hic ? Aujourd'hui, le logiciel est à la fois partout et nul part. « *Il n'existe pas de catalogue qui recense tous les logiciels. Il faudrait qu'il y ait un seul endroit pour trouver tout le code source disponible et être sûr qu'il sera là dans deux mois, dans 2 ans ou dans 20 ans pour faciliter la vie des développeurs* » a-t-il déclaré.

2,5 milliards de fichiers sources uniques déjà archivés

Software Heritage a donc pour mission de collecter tous les codes sources publiquement disponibles, d'organiser et de structurer ces données, de les préserver et de les partager. « *L'objectif est de préserver le code ancien, mais aussi de construire une architecture qui trace tous les développements fait par les développeurs de la planète pour améliorer les codes aujourd'hui et avoir une informatique meilleure demain* » a précisé Roberto Di Cosmo.

En 18 mois, l'équipe est déjà parvenue à récupérer et intégrer dans ses archives 21 millions de projets logiciels, 500 millions de comits (modifications effectuées par un développeur dans un logiciel) et 2,5 milliards de fichiers sources uniques. Selon le chercheur, cela représente 20% de ce qui existe. Reste donc à collecter les 80% qui subsistent.



Développer une galaxie de partenaires

L'[Inria](#) ne compte pas combler ce gap tout seul. L'institut a profité de la présentation du projet pour lancer un appel à contributions auprès des chercheurs, développeurs, archivistes et industriels pour retrouver le reste du code source disponible et développer l'infrastructure nécessaire à l'hébergement de ces données.

Microsoft est le premier partenaire industriel à rejoindre cette initiative. « *Nous allons héberger gratuitement le premier miroir sur notre plate-forme **Azure*** » a indiqué **Frédéric Aatz**, responsable de la stratégie open source en France de Microsoft. Software Heritage compte, en effet, développer une architecture distribuée avec plusieurs miroirs de manière à garantir la robustesse des données et leur disponibilité. La firme de Redmond participera aussi à l'alimentation de cette forge informatique. « *Nous serons également les premiers consommateurs de cette plate-forme* » a ajouté Frédéric Aatz.

En créant les fondations de ce "**Wikipédia**" du logiciel, [Inria](#) espère qu'une galaxie de partenaires participeront au développement de différentes applications dans l'industrie (traçabilité, conformité, qualification du code, détection de faille), la recherche (reproductibilité scientifique, études logicielles avancées sur des données massives) et l'éducation (compilation structurée d'exemples de codes sources). « *L'objectif est de faire de Software Heritage un instrument mondial en développant des partenariats avec des organisations à travers le monde. L'idée est qu'[Inria](#) disparaisse dans ce projet ; que le projet devienne si important, que ce ne soit plus à l'[Inria](#) de le porter* » a assuré, souriant, Antoine Petit. « *Nous souhaiterions garder une toute petite place dans la cabine de pilotage* » a conclu Roberto Di Cosmo.