



L'Inria veut archiver tous les codes sources

L'[Institut national de recherche en informatique](#) et en automatique vient de dévoiler son projet [Software Heritage](#) , qui vise à archiver l'intégralité des codes sources de logiciels accessibles au public.

À l'instar de ce que fait [Archive.org](#) , l'[Inria](#) va pour se faire indexer automatiquement tous les dépôts de code connus, y compris les historiques de modification, avec le soutien de nombreux acteurs majeurs du secteur, comme [GitHub](#), la [Fondation Linux](#) ou encore [Microsoft](#).

Le but du projet est de centraliser tout le code dans une interface unique, pour le rendre plus accessible, et ainsi aider au partage de la connaissance et à l'apprentissage de la programmation. L'archivage ne sera d'ailleurs pas limité aux logiciels libres, l'[Inria](#) souhaitant accueillir tous les codes sources publics, quelque soit leur licence (y compris par exemple des codes sources dont seule la lecture est autorisée).

Cette archive pourra aussi apporter une certaine robustesse, en évitant que des grandes quantités de code disparaissent suite à la fermeture ou la panne d'un dépôt de code. La fermeture il y a peu de [Google Code](#) avait déjà causé pas mal de dégâts, et on peine à imaginer les conséquences désastreuses que pourraient avoir la disparition de [GitHub](#) ou [SourceForge](#).

Lancé il y a dix-huit mois, le projet a déjà collecté près de 3 milliards de fichiers sources, issus de près de 600 millions de commits dans 23 millions de projets. Le tout pèse pas moins de 200 To, hébergés notamment par [Microsoft](#) via son service [Azure](#). Les contenus archivés correspondent pour l'instant à l'ensemble des dépôts publics [GitHub](#) (hors forks), des snapshots des sources des packages [Debian](#) et des archives de sources publiées par le projet [GNU](#).

Outre l'augmentation de la couverture de l'archive (à terme, il sera même possible de soumettre des URL de dépôts à archiver), l'équipe en charge du projet travaille actuellement à rendre cette archive utilisable. Pour l'instant, les fonctions ouvertes au public se limitent à vérifier si un fichier source donné est présent dans l'archive (à partir de sa signature [SHA1](#)), mais des fonctions de recherche, visualisation et téléchargement vont arriver par la suite.

Source : [Silicon.fr](#) (cache : , [MAFF](#))