

Software Heritage : les logiciels ont leur bibliothèque d'Alexandrie



L'inauguration en présence des partenaires (Unesco, [INRIA](#), Open Source Initiative, Société Générale, Microsoft).

© Anaïs Marechal

La première bibliothèque universelle du code source du logiciel vient d'être mise en ligne, fruit d'un partenariat entre l'[INRIA](#) et l'Unesco. Avec l'objectif de réunir tous les logiciels (libres ou payants) sur une même plateforme, elle préservera ce patrimoine culturel et scientifique de façon uniforme.

Jeudi 7 juin était mise en ligne la première archive mondiale des logiciels , au siège de l'Unesco à Paris. Baptisée Software Heritage, l'initiative lancée par [Inria](#) a pour but de collecter, préserver à travers le temps et rendre disponibles à tous les codes sources de tous les logiciels. « *Une sorte de bibliothèque d'Alexandrie du code source du logiciel* », s'extasie Roberto Di Cosmo, le directeur du projet. A travers ce projet soutenu par l'Unesco, l'équipe espère asseoir la place précieuse des logiciels dans notre patrimoine culturel et scientifique : « *C'est une initiative très ambitieuse, un moment historique* », déclare Indrajit Banerjee, le directeur de la division des sociétés du savoir à l'Unesco.

Lancée en 2015, l'archive contient déjà plus de 4,5 milliards de fichiers source pour un total de 83 millions de projets. L'objectif est en effet de conserver une trace de l'historique des logiciels. On y trouve l'histoire du développement du système Unix, toute la plateforme Github, les archives GNU jusqu'en août 2015, etc.



[Visualiser l'article](#)

Et même le code source de l'ordinateur de navigation de la mission Apollo 11, qu'exhibe avec fierté Roberto Di Cosmo, en montrant l'archive juste mise en ligne. La stratégie de collecte dépend des logiciels. Pour les codes open-source, les développeurs comptent sur la production participative pour recenser ceux qui existent déjà. Les logiciels libres développés actuellement seront automatiquement collectés sur les plateformes de développement en ligne, sans aucun a priori de sélection. Pour les codes propriétaires, enfin, des historiens et informaticiens fouillent les archives, et l'équipe espère à l'avenir ouvrir une section spéciale pour ceux qui sont sous embargo, et seraient rendus visibles à l'expiration du copyright.

Disponibilité, traçabilité et uniformité

L'infrastructure repose sur la disponibilité, la traçabilité et l'uniformité. Des identifiants intrinsèques permettront de localiser précisément des versions spécifiques de chaque logiciel. L'archive est structurée par un arbre de Merkle, et sécurisée par un réseau de miroirs. « *Pour le moment, deux copies sont chez [Inria](#) et une chez Microsoft* », explique Roberto Di Cosmo. *A terme nous allons développer un réseau de 10 miroirs sur au moins 3 continents.* » Dans la liste des partenaires, on trouve aussi les français Qwant, qui développeront un moteur de recherche pour la plateforme. Le projet mobilise une dizaine de personnes à temps plein, l'appui d'une équipe de recherche à [Inria](#), et à terme un coût de fonctionnement de 6 millions d'euros par an.

Pour les auteurs du projet, l'industrie est de plus en plus dépendante des logiciels en open-source. « *Pour une entreprise de technologie comme la nôtre, le code source est au cœur de notre transformation* », détaille Alain Voiment, directeur adjoint des nouvelles technologies à la Société Générale, l'un des partenaires du projet. Grâce à cette base de données ouverte, universelle, le transfert technologique depuis le monde de la recherche vers l'industrie devrait être facilité, et l'innovation accélérée.