Web FRA

www.republik-it.fr

Famille du média : **Médias professionnels**
Audience : **6314**
Sujet du média : **Communication - Médias - Internet, Informatique-Télécommunications**

14 Janvier 2026
Journalistes : **Bertrand Lemaire**
Nombre de mots : **1242**

Visualiser l'article

Software Heritage : dix ans de préservation du patrimoine applicatif

Le 28 janvier 2026, Software Heritage fêtera ses dix ans à l'Unesco. Morane Gruenpeter, directrice de la verticale Open-Science, et Bastien Guerry, responsable des partenariats, présentent cette initiative qui vise à préserver le patrimoine applicatif quelque soit le sort des éditeurs ou les feuilles de route de ceux-ci.

L'équipe de Software Heritage. - © Inria / B. Fourrier

Pour commencer, pouvez-vous nous présenter Software Heritage ?

Morane Gruenpeter : Software Heritage est une infrastructure lancée par l'Inria en 2016. Ses deux co-fondateurs sont les chercheurs Roberto Di Cosmo (directeur) et Stefano Zacchiroli (directeur scientifique). Aujourd'hui, Software Heritage est une fondation abritée par la Fondation Inria avec le soutien de l'Unesco. La Fondation Inria offre un cadre pour les opérations de gestion courante de Software Heritage. L'Inria apporte de son côté des moyens matériels notamment en hébergeant les données de Software Heritage. Afin d'anticiper les évolutions de Software Heritage, un advisory board réunissant des experts de tous horizons géographiques et thématiques a été constitué. Roberto Di Cosmo, chercheur et enseignant, est une personnalité reconnue dans le logiciel libre, à l'échelle internationale. L'équipe de Software Heritage compte 25 collaborateurs, et est dirigée par un comité exécutif.

Les objectifs de Software Heritage sont de préserver tous les codes sources applicatifs publiés un jour et de les rendre accessibles au plus grand nombre.

Bastien Guerry : Si on pense évidemment aux logiciels libres issus de forges, il peut aussi s'agir de codes sources sans licence ou sous licence propriétaire, car nous collectons tout ce qui est en ligne. Et nous pouvons aussi récupérer des codes sources anciens jamais publiés, si les éditeurs nous les confient, notamment s'ils abandonnent un produit.

Quelle est l'utilité d'une telle initiative concernant les logiciels libres, par nature ouverts et disponibles ?

Morane Gruenpeter : Une forge va en effet héberger le code source de logiciels libres et sauvegarder automatiquement les historiques de modifications. Mais une forge peut fermer et ses données disparaître ! Parmi les grandes forges qui ont fermé, on peut citer Google Code, Code Plex, Gitorious... Et parfois des forges disparaissent simplement parce qu'elle n'ont aucun modèle économique ou qu'elle sont abandonnées.

Bastien Guerry : Github est très loin d'être la seule forge, même aujourd'hui. Rien que dans l'administration d'État française, on compte au moins 80 forges ! Parfois, des petites forges ne sont ouvertes que le temps d'un projet ou d'un programme.

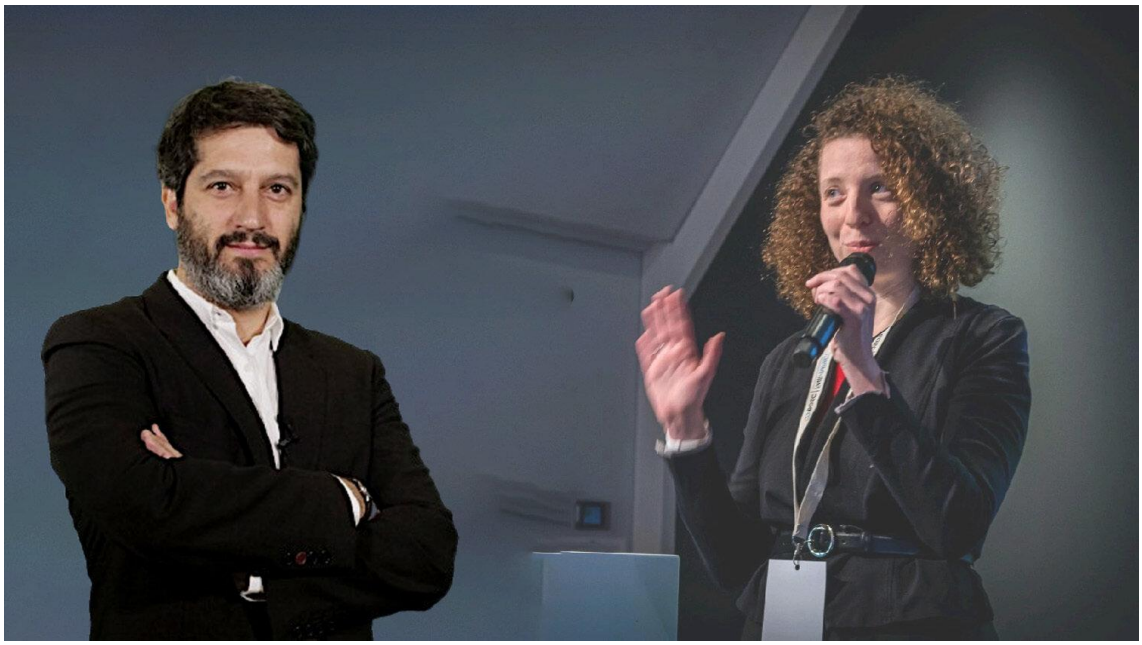
Pourquoi est-il utile d'archiver les sources des applicatifs ?

Morane Gruenpeter : archiver, c'est d'abord préserver sur le long terme et pouvoir retrouver les sources en cas de besoin parce que l'applicatif est toujours utilisé quelque part ou pourrait inspirer la création d'un nouveau logiciel. Mais il y a aussi un autre avantage à tout archiver au même endroit : étudier les sources. On peut ainsi, par exemple, réaliser des statistiques sur les licences choisies, les mises-à-jour opérées, les langages utilisés...

Concrètement, comment peut-on archiver tout ce foisonnement et le rendre disponible ?

Morane Gruenpeter : On peut mettre en place une aspiration automatique des forges, notamment les grandes. Un admin de forge peut aussi demander à ce que sa forge soit ajoutée à l'archivage. On peut aussi pousser manuellement une sauvegarde de code.

Bastien Guerry : Nous conservons tous les dépôts sur notre propre infrastructure en optimisant le stockage, optimisation facilitée par le fait que les logiciels utilisent très souvent Git pour gérer leurs versions. Avoir tous les codes au même endroit permet aussi de donner accès à des jeux de données utiles à l'entraînement des IA génératives de code, de réaliser des analyses sur les liens entre commits, sous réserve d'accepter nos principes LLM , ou encore de faire des statistiques sur l'usage des dépendances.



De gauche à droite : Bastien Guerry et Morane Gruenpeter. - © D.R.

Avoir des sources, c'est bien. Mais que peut-on en faire si elles sont dans un langage obsolète ou utilisent un framework qui n'est plus disponible ?

Bastien Guerry : Nous rendons disponibles les scripts, notamment ceux annexés à des articles de recherche. Le patrimoine de code est un patrimoine comme un autre. Mais, en effet, le script peut être dans une version obsolète d'un langage.

Morane Gruenpeter : Le code est créé par des humains (ou aujourd'hui des IA) et est lisible. La compilation de ces sources en exécutable est une autre question. En archivant les sources, nous archivons la connaissance de l'algorithme. Nous archivons aussi ses évolutions, ses mises-à-jour, son historique. Le code est un produit vivant.

Bastien Guerry : Si on reprend un code vieux d'une vingtaine d'années, on en connaît l'historique et l'environnement d'exécution. Quand on écrit un logiciel, son code peut très bien disposer de dépendances non-maîtrisées. Mais la source de la dépendance peut souvent être téléchargée sur Software Heritage. L'identification des différentes sources est rendue possible par le SWHID (SoftWare Hash IDentifiers) .

Comment est financée Software Heritage ?

Bastien Guerry : Nous sommes financés par l'Inria, la Fondation Inria et des sponsors publics et privés, en France et à l'international. Il peut s'agir de financement liés à la participation à des comités techniques ou associés à des projets de développement de fonctionnalités...

Pourquoi organisez-vous un colloque le 28 janvier 2026 ?

Morane Gruenpeter : A l'occasion de nos dix ans, nous organisons notre symposium annuel à l'Unesco. Nous en profitons pour réaliser une grande fête et ouvrir nos portes. Nous rassemblons nos soutiens et notre communauté mais il y a aussi une partie ouverte et gratuite.

Nous avons, jusqu'à présent, constitué l'initiative, amélioré et consolidé la gouvernance. Avec les dix ans, nous voulons lancer une nouvelle phase. Il s'agit notamment de développer une collaboration entre l'Unesco et Software Heritage.

Bastien Guerry : Nous tenons à une ouverture internationale. Nous traitons un sujet qui n'est pas seulement technique mais qui a aussi des implications, par exemple, sur la maîtrise des applicatifs et sur la souveraineté.

A propos de l'Inria et de la Fondation Inria

L'Institut national de recherche en informatique et en automatique (Inria) est un établissement public français à caractère scientifique et technologique spécialisé en mathématiques et informatique. Créé dans le cadre du Plan Calcul en 1967 sous le nom IRIA avant de devenir INRIA en 1980, l'institut est, depuis un décret de 2021, l'Institut national de recherche en sciences et technologies du numérique sans que son sigle n'ait changé. L'Inria mène plus de 300 projets de recherche grâce à ses 3 500 scientifiques, ingénieurs et personnels d'appui, en partenariat avec des universités et des acteurs divers de l'écosystème numérique.

L'INRIA a notamment développé, en 1971 sous le pilotage de Louis Pouzin, le réseau expérimental Cyclades, premier réseau à commutation par paquets et posant les bases théoriques du protocole IP, alors que son concurrent Transpac, à commutation de circuits, a été porté par le Centre national d'études des télécommunications (Cnet). Cyclades a été abandonné en 1979 et Transpac choisi pour supporter les services Télétel/Minitel.

Fondée en 2017, la Fondation Inria « a pour vocation de mobiliser de nouveaux moyens financiers pour permettre à l'institut de soutenir des projets audacieux et qui donnent du sens au numérique. » Succédant à Nelly Haudegand, Henri Verdier est devenu directeur général de la Fondation INRIA en octobre 2025 après avoir été, notamment, ambassadeur pour le numérique et directeur interministériel du numérique.