



→ LA SCIENCE OUVERTE POUR LES DIRECTRICES ET DIRECTEURS D'UNITÉ
ET LES RESPONSABLES D'ÉQUIPES DE RECHERCHE

ALGORITHMES, CODES SOURCES ET LOGICIELS LIBRES

Les algorithmes, codes et logiciels constituent des résultats de la recherche, au même titre que les publications et les données. Leur ouverture contribue à leur amélioration et à l'émergence de nouvelles idées.

L'**algorithme** décrit le déroulé pour la résolution d'un problème posé qui précède l'écriture d'un code.

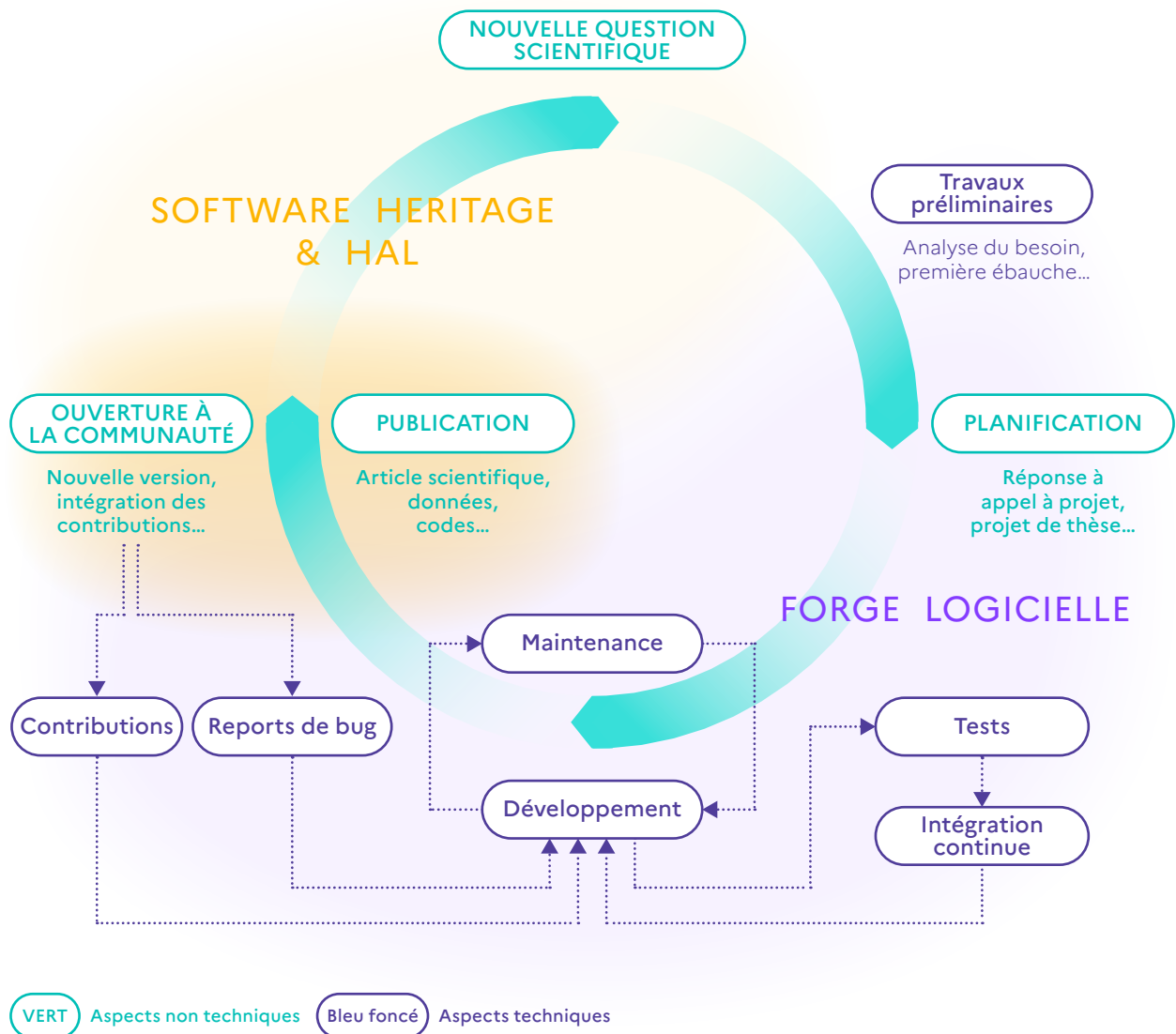
Le **code source** est la mise en œuvre et la formalisation d'un algorithme dans un langage informatique. Il s'agit du texte écrit par un ou plusieurs auteurs humains. Il existe de nombreux langages qui correspondent à des spécificités qui leurs sont propres (C, Java, Python, R, Ocaml, Scilab, etc.)

Le **logiciel** est la description, dans un ou plusieurs langages informatiques, d'un processus de traitement de données ou de modélisation que l'on souhaite faire réaliser par un ordinateur. C'est un ensemble comprenant le code source, l'exécutable (compréhensible par l'ordinateur), la documentation, la licence associée, etc.

En France, le législateur a perçu très tôt l'intérêt de l'ouverture des données et des codes sources produits par les organisations publiques et a promulgué en 2016, la [Loi pour une république numérique](#) afin de favoriser l'ouverture des codes sources et la création de communs du numérique. Les logiciels et codes constituent, aux côtés des publications et des données de la recherche, un des trois piliers du deuxième Plan national pour la science ouverte publié en 2021 par le MESR : axe 3 "Ouvrir et promouvoir les codes sources produits par la recherche". L'ouverture de l'ossature de la recherche qui sous-tend la production de connaissance implique d'ouvrir les codes sources et logiciels. Le code logiciel est souvent à tort considéré comme une donnée.

Or le code constitue l'élément de base de tout logiciel. Il s'agit du texte écrit dans des langages informatiques par un développeur ou une communauté de développeurs et évolue au fil du temps. Il est compréhensible et utilisable par l'homme mais aussi exécutable par un ordinateur lorsqu'il est compilé ou interprétable grâce à des environnements informatiques dédiés.

— Outils dans le cycle de vie d'un code de recherche



Quelles sont les bonnes pratiques pour le développement d'un logiciel ?



SCIKIT-LEARN

En 2019, [Gaël Varoquaux](#) souhaitait créer un outil statistique propre, beau et reconnu par les pairs et qui puisse être utilisé par tous, monde académique et monde industriel : [Scikit-learn](#).

Il constitue une équipe qui se transforme petit à petit en une communauté basée sur des concepts d'ouverture et de respect : elle compte en 2022 plus de 10 000 contributeurs ! [Scikit-learn](#) est développé sous une licence BSD qui permet de le réutiliser sans restriction, qu'il soit intégré dans un logiciel libre ou propriétaire. Cette bibliothèque est devenue une référence mondiale pour l'intelligence artificielle et est utilisée pour les modèles prédictifs.

Gaël estime que cette aventure en science ouverte a été un atout gigantesque pour tous les membres de l'équipe. Sa devise : créer un monde où l'excellence technique est importante, pour que tous en bénéficient !

[Scikit-learn](#).

Phase de développement

Il est recommandé d'utiliser une [forge logicielle](#) pour le développement. Elle permet la rédaction, le partage et la maintenance collaborative. Elle intègre un système de gestion des versions tel que GIT et des outils permettant l'organisation du travail collaboratif (forums, documentation, gestion des tâches, suivi des bugs...). Elle favorise différentes contributions, ainsi que l'intégration continue et la reproductibilité de la recherche.

Quelle protection, quelle licence choisir ?

La propriété intellectuelle est le premier outil pour tracer les contributions et maîtriser les conditions de réutilisation des productions de la recherche. Les contributeurs aux codes doivent donc être crédités pour leurs rôles afin de permettre la réutilisation de leurs créations par des tiers dans des conditions adaptées au contexte. L'attribution de ces droits doit être anticipée dès le démarrage du projet.

Le choix de la licence est une étape importante de l'ouverture, qui peut avoir des conséquences importantes sur les évolutions du logiciel. Il existe différents types de licences, chacune adaptée à des pratiques de réutilisation différentes (cf. p. 7 du [Passport Codes et logiciels](#)). Il est à noter que le niveau d'ouverture des codes peut varier dans le temps.

Si on souhaite partager le logiciel et le rendre utilisable par d'autres, il faut en autoriser l'accès, aussi bien juridiquement que techniquement. Il vous faut également réfléchir au devenir du code source du logiciel sur le long terme, soit par le maintien d'une activité de recherche à son sujet, soit par la mise en place d'une démarche active de valorisation.

Le logiciel est considéré comme une œuvre de l'esprit, protégée par le droit d'auteur. Le monopole de l'exploitation (les droits patrimoniaux) revient à l'employeur.

— Diffusion et valorisation

Les bonnes pratiques pour la diffusion :

- Utilisation d'une forge logicielle (suivi du développement) ;
- Signalement dans HAL (description) et dans un entrepôt de données de confiance (ex. Recherche Data Gov) quand le code est adossé à un jeu ou une base de données ;
- Archivage du code dans Software Heritage (pérennisation et reproductibilité).

— Définir une politique de laboratoire pour accompagner les chercheurs

- Sensibiliser les équipes aux bonnes pratiques qui contribuent à la construction des communs de la science (licences préconisées, forges disponibles...)
- Expliquer comment capitaliser pour ne pas recommencer ;
- Inciter à fédérer des communautés de développeurs et d'utilisateurs, voire à les animer ;
- Rappeler aux encadrants la nécessité de former les doctorants sur ces sujets ;
- Informer sur les ressources et personnels d'appui à la recherche (services de valorisation ,ateliers de la donnée, ambassadeurs science ouverte...).

POUR ALLER PLUS LOIN :

- Science Ouverte - Codes et logiciels :
https://www.ouvrirlascience.fr/wp-content/uploads/2022/10/Passeport_Codes-et-logiciels_WEB.pdf
- Rapport sur les Forges de l'ESR – Définition, usages, limitations rencontrées et analyse des besoins :
<https://www.ouvrirlascience.fr/forges-de-lesr-definition-usages-limitations-rencontrees-et-analyse-des-besoins/>