

Ouverture des codes, algorithmes, et logiciels

Ce document a pour objectif de rappeler et préciser les préconisations en matière d'ouverture des codes sources, algorithmes et logiciels issus des travaux de recherche d'INRAE.

Table des matières

Quelques rappels sur la Science Ouverte	2
Introduction au mouvement de la Science Ouverte	2
INRAE et la Science Ouverte.....	2
La notion de valeur dans le contexte de la politique Science Ouverte de l'Institut	3
Le principe par défaut d'ouverture des codes sources, algorithmes et logiciels, issus des travaux de recherche	3
La propriété intellectuelle des codes sources, algorithmes et logiciels.....	4
L'intérêt de déposer sur une archive et d'obtenir un identifiant pérenne.....	5
Le choix de la licence libre	5
Valorisation.....	6
Valorisation auprès de la communauté scientifique	6
Valorisation vers/avec les acteurs socio-économiques.....	7

Quelques rappels sur la Science Ouverte

Introduction

La science ouverte est la diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique. Ce mouvement n'est pas récent en France, et la loi pour une République numérique, publiée au Journal officiel du 8 octobre 2016, vise à favoriser l'ouverture et la circulation des données et du savoir, à garantir un environnement numérique ouvert et respectueux de la vie privée des internautes et à faciliter l'accès des citoyens au numérique. Les codes sources, comme les données de recherche, rentrent dans le cadre des documents administratifs de l'État et sont soumis aux mêmes statuts d'ouverture par défaut.

Chaque ministère décline sa propre feuille de route pour l'ouverture des données, des algorithmes et des codes et parmi ceux qui nous concernent, les trois ministères en charge de la recherche, de l'agriculture, des territoires et de l'environnement¹.

En 2021, le MESR avec son deuxième Plan National pour la Science Ouverte (2021-2024) (PNSO2) élargit son domaine vers les codes sources, les algorithmes et les logiciels à travers l'intégration d'un nouvel axe. Ce Plan traduit une politique cohérente et dynamique dans le domaine de la science ouverte, coordonnée par le Comité pour la Science Ouverte (COSO) piloté par le MESR et associe les établissements de recherche et d'enseignement supérieur et la communauté scientifique.

En complément, dans sa recommandation de novembre 2021 pour la Science Ouverte², l'Unesco désigne un ensemble de connaissances scientifiques ouvertes dont la principale manifestation est le **libre accès** aux **publications scientifiques**, mais également le libre accès, notamment, aux **données de recherche**, aux **métadonnées**, aux **logiciels**, et aux **codes sources** relevant du domaine public. Ces ressources doivent pouvoir bénéficier à tous les acteurs de manière immédiate ou aussi rapidement que possible et ce gratuitement.

INRAE et la Science Ouverte

INRAE s'est engagé depuis 2012 à ouvrir et partager ses données, ses résultats afin d'induire une démocratisation de l'accès aux savoirs, utiles à l'enseignement, à la formation, à l'économie, aux politiques publiques, aux citoyens et à la société dans son ensemble. Il s'appuie désormais sur « La Politique de Science Ouverte d'INRAE³ » et se saisit ainsi de l'opportunité offerte par le mouvement de la Science Ouverte pour en faire un accélérateur d'innovation, en encourageant la réutilisation des productions dans des approches partenariales nouvelles qui s'appuient sur un panel d'acteurs élargi. Cette position n'empêche pas pour autant de préserver découvertes et innovations chaque fois que nécessaire selon le principe : « la science doit être aussi ouverte que possible et aussi fermée que nécessaire ». **L'institut a posé un certain nombre d'éléments de principes généraux sur la valorisation et l'innovation ouverte dans plusieurs documents de références.**

- [Politique Science ouverte](#)
- [Gouvernance des données](#)
- [Note sur le choix d'une licence logicielle](#)

¹<https://www.numerique.gouv.fr/actualites/donnees-algorithmes-codes-sources-mobilisation-generale-sans-precedent-15-feuilles-de-route-ministerielles/>

² Cf. le texte de la déclaration : https://unesdoc.unesco.org/ark:/48223/pf0000379949_fre

³ https://www.inrae.fr/sites/default/files/pdf/Politique_INRAE_Science_Ouverte.pdf

La notion de valeur dans le contexte de la politique Science Ouverte de l'Institut

La valorisation dans le milieu de la recherche ne se limite pas à la seule valorisation économique directe. Il s'agit aussi et surtout de **faciliter la diffusion et l'enrichissement des connaissances** au sein de la communauté internationale de recherche et dans les sphères politiques et socio-économiques. Ainsi, la science ouverte est pensée comme un levier permettant une nouvelle forme de valorisation économique et sociale des résultats des institutions de recherches publiques, en permettant leur mise à disposition rapide et ainsi favoriser l'innovation et l'émergence de nouveaux services, au bénéfice de la société dans son ensemble, en premier lieu les citoyens.

Le code de la recherche, qui régit notamment les activités de valorisation de la recherche publique, concilie les principes d'ouverture et de protection. Il affirme, d'une part, **un principe de mise à disposition en accès libre et ouvert des productions de la recherche** et, d'autre part, encourage les personnels de la recherche à **publier leurs résultats pour en garantir la paternité et un accès pérenne**. En conséquence, la valorisation économique par le biais de licences payantes pour des acteurs privés n'est pas le modèle privilégié. De ce fait il est particulièrement important pour l'Institut de :

1. Montrer sa capacité à générer des innovations, réalisées par un panel large de tiers, à travers des accès libres et gratuits à ses productions afin de démontrer sa contribution à la création de valeur en direction de la sphère socio-économique (et à tous les autres acteurs de la société) ;
2. Engager des dynamiques ouvertes par la mise en place de licences compatibles avec des pratiques vertueuses pour rendre les écosystèmes socio-économiques plus réactifs.

Le principe par défaut d'ouverture des codes sources, algorithmes et logiciels, issus des travaux de recherche

Comme le rappelle le second Plan National pour la Science Ouverte, « le logiciel joue un rôle clé dans la recherche scientifique, dont il est à la fois un outil, un résultat et un objet d'étude. La mise à disposition des codes sources des logiciels, avec la possibilité de les modifier, les réutiliser et les diffuser, est un enjeu majeur pour permettre la reproductibilité des résultats scientifiques et soutenir le partage et la création de connaissances, dans une logique de science ouverte ».

Ainsi, sans entraver le principe de liberté du chercheur dans ses choix de production et de diffusion, la Science Ouverte doit devenir la pratique par défaut des chercheurs d'INRAE et les codes sources, algorithmes et logiciels doivent être diffusés sous l'une des licences reconnues libres par la Free Software Foundation et l'Open Source Initiative, pour autoriser leur réutilisation, leur modification et leur partage. Il est recommandé de les déposer sur l'archive universelle des logiciels Software Heritage⁴ et sur l'archive ouverte Hal⁵ afin de garantir la disponibilité et la traçabilité du code source et assurer leur archivage pérenne et leur accessibilité.

Cependant, la loi prévoit de respecter les contraintes légales en vigueur qui fournissent un certain nombre d'exceptions, dans lesquelles les données, codes sources, algorithmes et logiciels ne doivent pas être publiés et rendus réutilisables. Il s'agit des situations intégrant la protection de la vie privée et des données à caractère personnel, l'articulation avec les droits de propriété intellectuelle et celles de la protection de la confidentialité et des secrets.

Cependant, afin de soutenir leur accessibilité et accroître leur visibilité auprès d'acteurs socio-économiques, de décideurs de politiques publiques ou de nouveaux partenaires, certaines productions logiciels libres ont besoin d'accompagnement pour construire des stratégies de valorisation et de nouveaux modèles économiques, pour augmenter l'impact qu'ils peuvent avoir sur la société. Un appui à la valorisation peut donc être activé dans ce cas.

⁴ <https://www.softwareheritage.org/?lang=fr>

⁵ <https://www.archives-ouvertes.fr/> ou <https://hal.inrae.fr/>

La propriété intellectuelle des codes sources, algorithmes et logiciels

Le logiciel libre est, comme tout logiciel, une œuvre de l'esprit, qui peut, sous certaines conditions, être protégé par le droit d'auteur. Le régime juridique applicable aux logiciels déroge par ailleurs sur certains points au droit d'auteur dit « classique ».

En droit de la propriété intellectuelle, toute œuvre de l'esprit, si elle est originale, est protégée par le droit d'auteur⁶, quels qu'en soient le genre, la forme d'expression, le mérite ou la destination, et les logiciels et le matériel de conception préparatoire ont été ajoutés à la liste des œuvres de l'esprit « protégeables ». Ainsi, les éléments du logiciel protégés au titre du droit d'auteur sont l'architecture des programmes, le code source, le code objet, les différentes versions, les écrans et les modalités d'interactivité s'ils sont originaux et le matériel de conception préparatoire. La protection par le droit d'auteur est automatique dès la création de l'œuvre, et aucun dépôt n'est nécessaire.

Cependant, l'originalité est la condition centrale de la protection par le droit d'auteur : l'œuvre doit ainsi porter l'empreinte de la personnalité de son auteur. Ainsi, les algorithmes, les interfaces et les langages de programmation, davantage assimilés à des principes mathématiques qui relèvent du domaine de l'idée, ne sont pas protégeables au titre du droit d'auteur.

Le droit d'auteur est composé de deux types de droits :

- le droit moral qui constitue le prolongement de la personnalité de l'auteur à travers son œuvre et est donc perpétuel, non cessible et imprescriptible. Il est constitué de 4 attributs : le droit à la paternité, le droit à la divulgation, le droit au respect de l'œuvre et le droit de retrait et de repentir ;
- le droit patrimonial qui regroupe le droit de reproduction et de représentation de l'œuvre et se distingue en droit de reproduction, de distribution et d'adaptation.

A la différence du droit d'auteurs sur les autres œuvres de l'esprit pour lesquelles l'auteur dispose des droits moraux et patrimoniaux, les droits patrimoniaux sur les logiciels et leur documentation, créés dans l'exercice de leurs fonctions ou d'après les instructions de leur employeur, sont dévolus à l'employeur qui est seule habilité à les exercer⁷.

- Dans le cas d'un agent de l'état, le droit patrimonial portant sur le logiciel créé est automatiquement dévolu à l'organisme public ([L.113-9](#) et [L.113-9-1](#) du code de la propriété intellectuelle). Mais l'auteur conservant le droit moral sur son logiciel, il conserve le respect de son nom : ce droit de paternité permet à l'auteur de revendiquer que son nom apparaisse aux côtés de son logiciel.
- Les personnels ni salariés, ni agents publics (ex : les stagiaires, doctorants étrangers et professeurs ou directeurs émérites) s'alignent sur le régime des agents de l'état s'ils sont accueillis dans le cadre d'une convention réalisant de la recherche et qu'ils créent des logiciels dans l'exercice de leur mission ([Ordonnance 2021-1658 du 15 décembre 2021](#)). Sauf stipulations contraires, les droits sont dévolus à la structure d'accueil. Cette condition nécessite qu'elles perçoivent une contrepartie (financière ou matérielle) et qu'elles soient placées sous l'autorité d'un responsable de la personne morale.
- Notons néanmoins que la représentation visuelle de l'interface graphique reste protégée par le droit d'auteur commun et qu'elle reste donc la propriété des auteurs ;

Par ailleurs, il est à noter que, dans le cas des logiciels libres, ce n'est pas parce que l'auteur originel d'un logiciel libre fournit le code source et permet le libre usage, la reproduction, la modification, la distribution du logiciel modifié ou non, qu'il renonce à ses droits patrimoniaux et moraux.

⁶ Article L112-1 du code de la propriété intellectuelle

⁷ Article L113-9 du code de la propriété intellectuelle

Enfin, un logiciel libre n'a pas vocation à rester le travail d'une seule personne. Le code source étant libre d'accès, des personnes, des « contributeurs » ont la possibilité de modifier le logiciel. Elles sont obligées par la suite de faire connaître leur contribution aux autres membres de la communauté. Le logiciel libre est donc par essence le travail de plusieurs personnes.

- Le logiciel libre peut donc être une œuvre collective. Selon l'article L113-2, alinéa 3 du Code de la propriété Intellectuelle, l'œuvre collective est « l'œuvre créée sur l'initiative d'une personne physique ou morale qui l'édite, la publie, la divulgue sous sa direction et son nom et dans laquelle la contribution personnelle des divers auteurs participant à son élaboration se fond dans l'ensemble en vue duquel elle est conçue, sans qu'il soit possible d'attribuer à chacun d'eux un droit distinct sur l'ensemble réalisé ». Mais ce cas reste rare.
- En effet, le logiciel libre est le plus souvent considéré comme une œuvre de collaboration, c'est-à-dire une œuvre créée par plusieurs personnes ensemble. L'article L113-2 du Code de la propriété Intellectuelle énonce « est dite de collaboration l'œuvre à la création de laquelle ont concouru plusieurs personnes physiques ».
- Enfin, le logiciel libre pourrait être considéré comme une œuvre composite, c'est-à-dire une œuvre réalisée à partir d'une autre œuvre. L'œuvre nouvelle sera alors incorporée dans une œuvre préexistante sans la collaboration de l'auteur de l'œuvre originelle (article L113-2, alinéa 2 du Code de la Propriété Intellectuelle).

L'intérêt de déposer sur une archive et d'obtenir un identifiant pérenne

Bien que le droit d'auteur sur les logiciels s'acquière sans formalité, déposer ses codes dans une archive ouverte et obtenir un identifiant pérenne permet d'apporter une protection juridique reconnaissant son droit d'auteur avec une datation certifiée. Cela permet ainsi de garantir la pérennité de ses fichiers et des accès tout en assurant un accès libre et gratuit à la recherche.

Pour les codes sources, l'utilisation de Software Heritage est recommandée pour l'archivage et le référencement. La standardisation du SWHID (SoftWare Heritage IDentifier), système d'identifiant pérenne sous forme de code alphanumérique qui identifie de manière unique les codes sources des logiciels, complètera les DOI pour les logiciels. Ainsi, Software Heritage conserve les lignes de code mais aussi tout l'historique des développements des logiciels, chaque SWHID comportant le numéro de version.

Le choix de la licence libre

Une licence est un consentement du propriétaire à l'utilisation de la propriété intellectuelle en échange d'une compensation (en argent ou autre). Afin de soutenir la diffusion des productions logicielles, l'utilisation de licences libres, c'est-à-dire publiées sous l'une des licences reconnues libres par la Free Software Foundation et l'Open Source Initiative, sera privilégiée. Ces licences doivent garantir le respect des quatre libertés fondamentales : la liberté d'exécution du logiciel (utilisation à l'infini), la copie sans restriction, l'accès direct au code source, et le droit de modification et de redistribution du code source. Afin d'éviter la prolifération des licences, la [loi pour une République numérique](#) a prévu la création d'une liste, fixée par décret, de licences qui peuvent être utilisées par les administrations pour la réutilisation à titre gratuit de leurs informations publiques, qu'il s'agisse de données ou de code source d'un logiciel ([article D.323-2-1 du code des relations entre le public et l'administration \(CRPA\)](#)).

Le choix de la bonne licence open source dépend des contraintes liées à la composition et à ses objectifs. Le premier critère à prendre en considération est l'analyse des dépendances de son logiciel. Il est rare qu'un logiciel soit créé entièrement *de novo* et nombre d'entre eux intègrent des bibliothèques logicielles. Les bibliothèques logicielles sont des collections de routines, qui peuvent être déjà compilées et prêtes à être utilisées par des programmes. Elles sont fréquemment utilisées pour réaliser des interfaces de programmation, des framework, des plugins ainsi que des langages de programmation. Bien que la plupart des développeurs fondent leurs choix sur des mérites purement techniques, le choix de ces bibliothèques peut avoir un impact important sur le choix de la licence libre du logiciel et devrait donc être réfléchi en amont de la conception en fonction de ses objectifs.

- Si chacune des licences est "permissive" (donne au public l'autorisation d'utiliser, de modifier et de partager, sans condition pour les licences en aval), vous pouvez utiliser la licence que vous voulez. Les licences permissives les plus connues sont Apache 2.0, BSD2, BSD3, CeCILL-B et MIT.
- Si l'une des licences de dépendances est "strong copyleft" (donne également les mêmes permissions publiques, sous condition d'utiliser la même licence en aval), alors votre projet devra utiliser la même licence. GPLv2, GPLv3, AGPLv3, et CeCILL-C sont les principales licences communes de copyleft.

En second lieu, il faut considérer les communautés de contributeurs et d'utilisateurs qui contribueront au projet, mais l'utilisation d'une licence connue sera de toute façon favorable au développement d'une communauté. Ainsi, l'utilisation d'une licence permissive permettra à son développement d'être facilement intégré à différents projets, quitte à ce que les modifications soient placés sous une nouvelle licence libre, voire sous licence propriétaire alors que l'utilisation d'une licence copyleft imposera l'utilisation d'une licence particulière.

Une difficulté importante des *librairies* qu'il convient de garder à l'esprit est leur mise à jour. Cela peut impliquer des problèmes de portabilité du code et quelques fois de licence, puisque certaines librairies peuvent changer leur licence au gré des mises à jour. Cela implique donc que les logiciels construits sur leur base doivent non seulement être mis à jour, mais également potentiellement leurs licences. Ainsi, la façon dont un code informatique est construit peut interférer avec les possibilités de choix d'une licence. **Il faut absolument être en mesure de tracer tous les auteurs et leurs statuts ainsi que tous les composants tiers mis dans le code.** Pour ce faire, il est préconisé d'utiliser une forge logicielle.

Valorisation⁸

Valorisation auprès de la communauté scientifique

Le partage des codes sources d'un logiciel permet la réutilisation de celui-ci, mais également sa modification. Celui-ci peut alors être amélioré ou adapté à de nouveaux usages. Au-delà de cette diffusion, il y a aussi un fort bénéfice à animer une communauté de développeur autour de celui-ci. Les gains sont multiples :

- Le logiciel devient plus robuste et plus évolutif grâce à la correction de bug et la mise en place de pratiques de codage partagées et explicites.
- Le logiciel continue d'exister et d'évoluer même avec la disparition de certains développeurs ce qui accroît sa pérennité.
- Le logiciel bénéficie d'une force de développement accrue avec des compétences variés. Le logiciel va s'enrichir plus rapidement de fonctionnalités.
- L'architecture du logiciel est pensée collectivement, rassemblant les compétences et points de vue d'usagers.
- Les tests des nouvelles versions se font avec un panel de développeurs et d'utilisateur plus large

L'utilisation d'une forge logiciel devient alors incontournable. Le suivi des contributeurs est alors à considérer avec attention, car ils acquièrent un droit d'auteur qu'ils peuvent opposer à certaines décisions sur l'évolution des licences d'utilisation. Le choix des licences devient alors primordiale pour la bonne entente de la communauté et ainsi la pérennité du logiciel.

⁸ On entend par valorisation l'action de rendre utilisable ou commercialisable le savoir et les résultats de la recherche, c'est-à-dire leur fournir une valeur ajoutée. Toutefois, cela ne signifie pas qu'elle est centrée sur la valeur marchande de la recherche. On distingue ainsi deux types de valorisation : une valorisation *financière* de type commerciale et une valorisation *sociale* de type non marchande. Dans le second cas, les retombées économiques peuvent exister mais n'est pas la finalité première de la valorisation.

Valorisation vers/avec les acteurs socio-économiques

Dès lors que ceux-ci présenteront un intérêt pour une ou plusieurs applications socio-économiques, INRAE accompagnera le développement des stratégies de valorisation permettant la diffusion et l'utilisation la plus large de ces outils par le monde socio-économique. Pour déclencher cet accompagnement, une Déclaration d'Invention et de Résultats Valorisables (DI-RV) sera alors rédigée par le scientifique ou l'équipe de recherche à l'origine de l'invention et la stratégie de valorisation sera décidée en Groupe Sectoriel en concertation avec le responsable scientifique.

Différents modèles de valorisation seront envisagés tels que :

1/ **Des services autour du logiciel libre** : La mise en place de formations sur le logiciel libre ou sur des concepts associés à son utilisation, des contrats d'assistance technique ou d'expertise permettant une meilleure prise en main du logiciel, mais également des contrats de transfert de savoir-faire associé à cet outil. La mise en place de clubs utilisateurs, avec une animation adaptée, pourra également être envisagée.

2/ **La mise en place de collaborations de recherche avec un ou plusieurs partenaire(s) socio-économique(s)** afin d'adapter le logiciel libre à son/leurs besoin(s) spécifique(s). Les évolutions, adaptations ou nouveaux modules, co-développés avec le partenaire, pourront être détenus en copropriété et faire l'objet d'un contrat de transfert sous licence propriétaire à l'issue du partenariat.

3/ Le développement d'une version améliorée du logiciel en vue d'obtenir une version industrialisable de l'outil d'intérêt socio-économique pour une filière. Cette nouvelle version pourra rester propriétaire et être transférée à un ou plusieurs partenaires socio-économiques. Elle pourra notamment avoir été développée au cours de programmes de pré-maturation ou maturation, éventuellement cofinancée ou co-construite par des partenaires.

Dans les modèles 2 et 3, des versions libres et propriétaires d'un même logiciel coexisteront. Les versions libres seront identifiées grâce à leur identifiant SWHID⁹ et les versions propriétaires déposées à l'Agence de Protection des Programmes et se verront attribuer un numéro IDDN.

La limitation de la licence exclusive à un certain domaine d'application offre à l'institut la possibilité d'exploiter les mêmes résultats dans d'autres secteurs, augmentant l'impact potentiel de sa recherche. De plus, pour défendre l'intérêt général, les licences exclusives par domaine doivent être limitées dans le temps et assorties de seuils d'obligations d'exploiter commercialement afin que l'entreprise privée ne bloque pas l'innovation : (1) en deçà d'un seuil d'exploitation (qui se mesure au niveau des redevances), il perd son exclusivité ; et (2) en cas de non exploitation, la licence est résiliée. Ceci est souvent mis en œuvre au moyen de clauses anti-rayonnage dans les accords de licence.

Afin d'appuyer la mise en place de ces stratégies de valorisation, INRAE s'appuiera sur une vitrine des outils numériques valorisés et/ou valorisable permettant de faire connaître et diffuser les outils pouvant avoir un impact socio-économique.

⁹ <https://www.softwareheritage.org>