

Intégrité Scientifique : Logiciels & Données

Table des matières

I.	Le logiciel, objet juridique.....	1
1)	Les droits patrimoniaux :.....	1
2)	Les droits moraux :	1
II.	Choix des Licences	2
1)	Licences libre ou propriétaires ?.....	2
2)	Quelles sont les catégories de licences libres ?	2
3)	Licence sur les données (data).....	3
III.	Manquements à l'intégrité scientifique.....	4
1)	Le non-respect de la licence ou des règles	4
2)	La non-reconnaissance des droits moraux.....	4
3)	La fraude sur les données	4
IV.	Grille d'identification	6
1)	Logiciels.....	6
2)	Données	6

I. Le logiciel, objet juridique

Il a deux types de droit concernant un logiciel : les **droits patrimoniaux** concernant l'exploitation de l'œuvre, ce sont des droits monnayables, cessibles, temporaires (détenteurs ou propriétaires, donc les employeurs) ; les **droits moraux** qui se rapportent à l'intérêt moral de l'auteur, et ils sont inaliénables.

1) Les droits patrimoniaux :

Les logiciels relèvent du droit d'auteur (Art. L112-2 du CPI) « sauf dispositions statutaires ou stipulations contraires, les droits patrimoniaux sur les logiciels et leur documentation créés par un ou plusieurs employés dans l'exercice de leurs fonctions ou d'après les instructions de leur employeur sont dévolus à l'employeur qui est seul habilité à les exercer » (Art.L.113-9 du CPI). Les Personnes « accueillies dans le cadre d'une convention », « réalisant de la recherche »... « leurs droits patrimoniaux sur ces logiciels ...sont dévolus à cette structure d'accueil » (Art.L.113-9-1 du CPI).

2) Les droits moraux :

Les droits moraux permettent aux auteurs de décider de l'utilisation, de l'étude de l'œuvre, de la modification, de la diffusion, de la paternité du logiciel ou la **citation, qui est la référence du travail accompli**.

- Droit à la paternité, relatif à la mention de l'auteur.
- Droit de divulgation, relatif au moment et aux conditions de livraison.

- Droit de repentir, permet de retirer une œuvre.
- Droit au respect de l'œuvre, permet de s'opposer aux modifications.

Pour faire respecter ces droits, les auteurs doivent choisir ou construire une **licence**, qui est un contrat qui lie les utilisateurs avec le logiciel, les droits moraux des auteurs et leur choix sur les divers aspects de diffusion (study, copy, modify, and redistribute).

AUTEUR	DROIT MORAL	DROIT PATRIMONIAL
Auteur non salarié / ni agent public (stagiaire, émérite)	Exercé par l'auteur	L'auteur détient les droits patrimoniaux
Créateur salarié ou agent public dans l'exercice de ses fonctions (chercheur, ingénieur, doctorant, post-doc, CDD)	Exercé par l'auteur	Droit patrimoniaux à l'employeur
Créateur salarié hors de son lieu de travail ou de ses horaires de travail	Exercé par l'auteur	Dévolution à l'employeur admise si la création a été faite dans le cadre de ses missions ou sur instruction de l'employeur
Doctorant CIFRE	Exercé par l'auteur	Droit patrimoniaux à l'entreprise qui embauche le doctorant
Professeur invité	Exercé par l'auteur	Droit patrimoniaux à l'université qui embauche le professeur

Figure 1: Droit Moral et droit patrimonial dans le laboratoire. (Sauf contrats spécifiques détaillant les droits, ex. doctorant CIFRE, les droits doivent être détaillés dans le contrat entre l'Université et l'entreprise).

II. Choix des Licences

Les licences sont des contrats qui protègent les auteurs, les utilisateurs, les éventuels collaborateurs au développement. Elles octroient des droits (et des libertés) et peuvent contenir des clauses de réciprocité ou imposer des obligations qui sont à respecter. Dans 99% des cas, les problèmes arrivent par manque de préparations et de précisions : pas de licence, pas de contrat, aucunes protections !

1) Licences libre ou propriétaires ?

Une licence propriétaire est une licence ad-hoc dont chaque clause est définie par rapport aux objectifs. Par exemple, une licence de test et d'évaluation est une licence propriétaire. Une licence payante vers un industriel est également appelée propriétaire.

Mais comme la recherche scientifique académique préconise une open science, une licence libre est préférable pour la vérification et la reproductibilité des résultats.

2) Quelles sont les catégories de licences libres ?

Il existe trois niveaux de catégorie de licences libres qui concernent la diffusion et la modification du logiciel :

- Copyleft fort (GNU GPL, CeCILL-2.1, EUPL 1.2 , etc...)

C'est le niveau classique : toute modification d'un logiciel dont la licence contient une telle clause doit être redistribuée sous les mêmes termes de licence. De même, si le logiciel est intégré dans un projet logiciel plus important, ce projet devra également être rendu public sous les mêmes termes de licence.

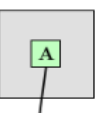
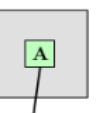
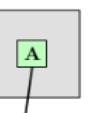
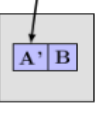
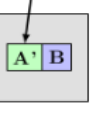
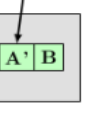
- Copyleft faible (GNU LGPL, Mozilla MPL, CeCILL-C, etc...)

Ces licences permettent de combiner le code qu'elles couvrent avec des modules tiers placés sous tous types de licences, y compris privatives (encore appelées licence propriétaires), à condition qu'il n'existe pas d'incompatibilité entre les licences. Cela permet de s'assurer que les évolutions du logiciel initialement couvert resteront toujours sous licence libre. Certaines licences à copyleft faible imposent en outre que le logiciel dérivé puisse être reconstruit à partir d'une nouvelle version du logiciel libre, afin que les améliorations apportées au logiciel couvert puissent bénéficier au logiciel combiné.

- Non-copyleft (MIT, Apache, BSD, CeCILL-B, etc...)

Ces licences, dites « permissives », n'ont pas d'effet copyleft. Elles permettent de modifier le code ou de l'intégrer dans un logiciel distribué sous tout autre type de licence, sans que l'utilisateur puisse obtenir le code source de ce logiciel et exercer les droits initialement attachés à la licence libre. Ce type de licence est à l'origine du fait que l'immense majorité des logiciels actuels – libres ou non – repose en partie sur du code initialement libre.

- Copyleft fort : GNU GPL, CeCILL v2, EUPL
- Copyleft faible : MPL, GNU LGPL, CeCILL-C
- Sans Copyleft : Apache, BSD, MIT, CeCILL-B

Licence		Non Copyleft	Weak Copyleft	Strong Copyleft
Software	Original			
	Modified			

3) Licence sur les données (data)

Pourquoi accorder des licences sur les données de recherche ?

Bien que la pratique varie d'une discipline à l'autre, la tendance est de plus en plus à la publication planifiée des données de recherche. Le besoin de licences de données découle directement de ces publications, et la première question à se poser est donc de savoir pourquoi les données de recherche doivent être publiées. Un nombre important de bailleurs de fonds de la recherche exigent désormais que les données produites dans le cadre de la recherche qu'ils financent soient mises à la disposition d'autres chercheurs pour qu'ils les découvrent, les examinent et s'en inspirent.

L'origine des données va généralement induire les règles d'utilisation des données. Des données provenant de simulations numériques vont être associées à la licence du code de calcul. Les données d'observation au sol ou spatiales seront régies par les règles du consortium (missions spatiales, consortia, collaborations ... ect). On peut donc parler de licence sur les données pour réguler l'utilisation, la copie, la diffusion et les citations des données comme les logiciels.

Voir l'article : <https://www.dcc.ac.uk/guidance/how-guides/license-research-data>

III. Manquements à l'intégrité scientifique

1) Le non-respect de la licence ou des règles

Le premier manquement, le plus évident, concernant les logiciels de recherche et les données, est le non-respect de la licence ou des règles qui concernent l'utilisation, la copie, la modification et la redistribution, et des citations du travail des auteurs du logiciel ou de l'origine des données.

Pour les logiciels, c'est l'article Art. L. 335-2 du CPI du code de la propriété intellectuelle qui devrait s'appliquer : toute personne utilisant, copiant, modifiant ou diffusant le logiciel sans autorisation explicite des détenteurs des droits patrimoniaux est coupable de contrefaçon et passible de trois ans d'emprisonnement et de 300 000 euros d'amende.

Ce manquement serait bien sûr aggravé s'il est fait sciemment. Une copie d'un logiciel ou d'une partie sans citation ou respect de sa licence s'apparente aussi à du plagiat.

2) La non-reconnaissance des droits moraux.

Dans la construction du logiciel, l'auteur principal pourrait omettre la contribution d'auteurs ayant travaillé sur le logiciel en codant ou en apportant des techniques innovantes pour la création du logiciel. On peut attribuer une part des droits moraux même à des personnes ne contribuant pas au codage, mais qui sont à l'origine de concepts fondamentaux.

L'utilisation de suiveurs de version (gitlab, github) permet d'avoir un historique des contributions de chacun (thésards, post-doc et stagiaires). Leurs contributions respectives devraient être notées dans l'en-tête de la licence, mais comme le PI du projet ou du logiciel, ils n'auront pas des droits patrimoniaux qui reviendront à l'employeur (Université, Institut...).

Avoir choisi et décidé des modalités, d'une licence libre avant le début d'un nouveau projet, protège l'auteur principal (chercheur ou IR), des contestations sur les droits moraux et patrimoniaux des différents intervenants sur le projet.

3) La fraude sur les données

Les fraudes scientifiques génériques, qualifiées de graves et intentionnelles car elles faussent les résultats de l'enquête scientifique : il s'agit de ce qui est communément désigné, dans l'acronyme FFP, comme **Fabrication de données, Falsification de données**.

Les pratiques questionnables de recherche (QRP en anglais pour *Questionable Research Practices*) : dans cette « zone grise », les pratiques inappropriées, qui nuisent à la fiabilité des résultats de l'enquête scientifique et/ou au bon fonctionnement des communautés de recherche, peuvent concerner les données (**archivage ou gestion déficient, rétention, omission ou sélection, traitements statistiques problématiques, embellissement des images, ...**).

Voir : <https://www.ofis-france.fr/quest-ce-que-lintegrite-scientifique/>

IV. Grille d'identification

1) Logiciels

Nom du logiciel :

Liste des auteurs :

Type de licence :

Site web/Forge :

	Pratiques souhaitées	Manquement intentionnel	Manquement accidentel
Utilisation/études			
Copies totales ou partielles			
Modification			
Diffusion			
Respect des citations			
Respect de la liste des auteurs/contributeurs			

2) Données

Type de donnée :

Origine des données :

licences/règles :

Site web/database :

	Pratiques souhaitées	Manquement intentionnel	Manquement accidentel
Utilisations/études			
Diffusion			
Respect des citations			
Respect de la liste des auteurs/contributeurs			
Falsification de données			
Rétention, omission ou sélection			
Analyses déficientes			

Références:

- Les logiciels de la recherche et leurs licences : trois visions sur un objet
Teresa Gomez-Diaz

<https://hal.science/hal-02434287v2>

Trois plaquettes CNRS avec le réseau DevLog disponibles :

- Je code : Les bonnes pratiques de développement

<https://hal.science/hal-02083801v1>

- Je code : Les bonnes pratiques de diffusion

<https://hal.science/hal-02400300v1>

- Je code : Quels sont mes droits ? Quelles sont mes obligations ?

<https://hal.science/hal-02399517v1>

Cahier de laboratoire CNRS 2024 : elabFTW

https://qualite-en-recherche.cnrs.fr/wp-content/uploads/2022/11/Guide_bonnes_pratiques_CLE_V1.pdf

Dans quelles revues dois-je publier mon logiciel ?

<https://www.software.ac.uk/top-tip/which-journals-should-i-publish-my-software>

How to License Research Data

<https://www.dcc.ac.uk/guidance/how-guides/license-research-data>