

Partager les logiciels de la recherche, questions juridiques et projets en cours : PRESOFT et FG-SOL

Teresa Gomez-Diaz

CNRS - Laboratoire d'informatique Gaspard-Monge

Cette œuvre est mise à disposition selon les termes de la licence
Creative Commons Attribution - Pas d'utilisation commerciale - Pas de modification 4.0 International

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Partager les données de la recherche : pour qui, pourquoi, comment ?

LAAS, Toulouse, 14 novembre 2017

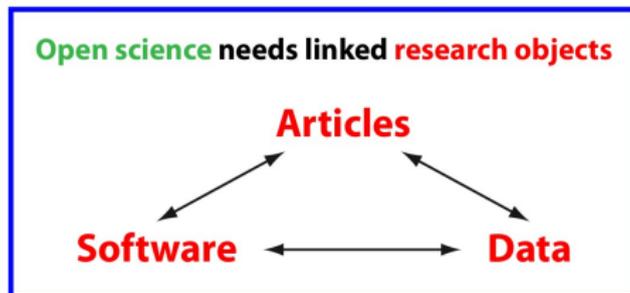


Motivation

Depuis 2006 : travail centré sur la diffusion des logiciels avec mission logiciels LIGM, PLUME, et nouveaux projets : **PRESOFT**, **FG-SOL**

Cadre général - Open Science : diffusion et partage des productions scientifiques

- Les articles sont encore la partie la plus visible et accessible
- Beaucoup de logiciels et données sont produits par les chercheurs pour obtenir des résultats scientifiques publiés
- Mais pas de procédure claire pour leur publication ou distribution
- Conséquence : la majorité sont en *unknown status*
- Améliorer la diffusion et le partage demande une connaissance du **cadre légal**
- Nous avons besoin des objets de la recherche et des liens entre eux



Plan

- 1 Cadre légal : licences
 - Motivation
 - Les types de licences
 - Les licences les plus utilisées
- 2 Cadre légal : le droit d'auteur des œuvres et le droit sui generis
 - Le droit d'auteur des œuvres
 - Les logiciels
 - Les données et bases de données
 - Le droit sui generis
- 3 PRESOFT : pérennisation des logiciels de la recherche
 - Le projet
 - Plan de gestion du logiciel PRESOFT ou SMP
 - Bénéfices du SMP
- 4 FG-SOL : ensemble de services pour les logiciels de la recherche
 - Vision et objectifs
 - Les services
- 5 Conclusions
- 6 Annexes

Motivation : pourquoi une licence ? (1/2)

La diffusion des logiciels libres (2005)

Dominique Dalmas, Directrice juridique (CNRS), Lyasid Hammoud, Juriste (CNRS)

Il est regrettable de constater que les aspects juridiques sont encore trop souvent méconnus et négligés par les créateurs lors de la diffusion de leurs logiciels.

Les surprises peuvent être douloureuses notamment en cas de litiges car les auteurs de bonne foi risquent de voir leur licence invalidée ou leur responsabilité mise en cause.

Il est pour eux essentiel de veiller à accompagner leur logiciel d'une licence bien construite, seule garantie que celui-ci vivra selon les principes qu'ils auront choisis.

Ceci est aussi valable pour les données.

Motivation : pourquoi une licence ? (2/2)

Qui peut utiliser un logiciel ?

Art. L. 335-2 du CPI

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 300000 euros d'amende.

Source : T. Aimé, Guide pratique d'usage des logiciels libres dans les administrations

<https://www.projet-plume.org/ressource/guide-logiciels-libres-administrations>

Voir aussi : Directive 2009/24 EC du Parlement européen et du Conseil du 23 avril 2009

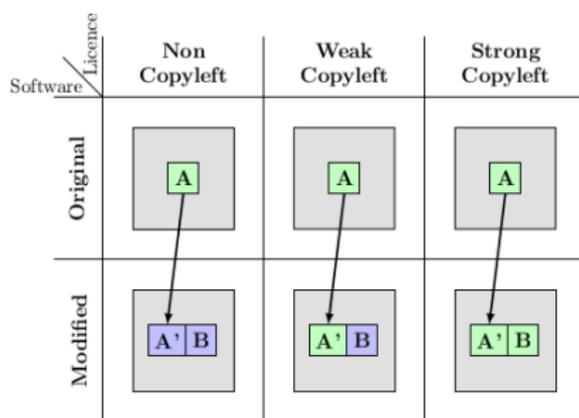
Les licences complètent le cadre juridique établi par la loi : s'il n'y a pas de droit explicitement donné, utiliser un logiciel relève de la contrefaçon.

Les licences [articles, logiciels, données] clarifient le cadre légal du partage : utiliser, copier, modifier, (re)distribuer, explorer...

Elles octroient des droits (et libertés) et peuvent contenir des clauses de réciprocité ou imposer des obligations qui sont à respecter.

Logiciels : les types de licences libres/open source

- Copyleft fort (*diffusives*)
 - Licence initiale s'impose sur tout.
 - Obligation de réciprocité, évite de fermer un code libre.
- Copyleft faible (*persistantes*)
 - Licence initiale reste.
 - Ajouts peuvent avoir autre licence.
- Sans Copyleft (*évanescentes*)
 - Licence initiale ne s'impose pas.
 - Les dérivés peuvent avoir n'importe quelle licence.



GPLv2 : « *You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.* »

Image : T. Aimé, Guide pratique d'usage des logiciels libres dans les administrations
<https://www.projet-plume.org/ressource/guide-logiciels-libres-administrations>
Voir aussi : <https://www.projet-plume.org/ressource/faq-licence-copyright>

Les licences les plus utilisées

Loi pour une République numérique : liste de licences, fixée par

Décret n°2017-638 du 27 avril 2017 <http://www.data.gouv.fr/fr/licences>

- Licence ouverte, v2 en cours, compatible CC-BY
- ODC Open Database License (ODbL) version 1.0
- Licences des logiciels : Apache-2.0, BSD, CECILL, CECILL-B, CECILL-C, MIT, GPL-3.0, LGPL-3.0, AGPL-3.0, MPL 2.0

Autres : demande homologation

- CC - Creative Commons <https://creativecommons.org/choose/?lang=fr>
 - ▶  CC-0 : renoncer aux droits de propriété ≠ 
 - ▶  CC-BY : libre non copyleft
 - ▶  CC-BY-SA : libre copyleft
 - ▶  CC-BY-ND
 - ▶  CC-BY-NC
 - ▶  CC-BY-NC-SA
 - ▶  CC-BY-NC-ND
 - ▶ version International ou France, v4.0 adapté à sui generis pour BD
- EUPL v1.2 (copyleft) <https://joinup.ec.europa.eu/page/eupl-guidelines-faq-infographics>
- autres : LGPL (LIGM, ressources linguistiques), GNU-FDL (documentation), LAL...

Proposition : ajouter la licence à la signature.

- 1 Cadre légal : licences
- 2 Cadre légal : le droit d'auteur des œuvres et le droit sui generis
 - Le droit d'auteur des œuvres
 - Les logiciels
 - Les données et bases de données
 - Le droit sui generis
- 3 PRESOFT : pérennisation des logiciels de la recherche
- 4 FG-SOL : ensemble de services pour les logiciels de la recherche
- 5 Conclusions

Le droit d'auteur des œuvres (1/2)

Les droits protégés par le Code de la propriété intellectuelle (CPI) sont automatiquement associés à l'auteur lors de la création de l'œuvre, sous condition de son **originalité** (ceci dépend de la date).

L'œuvre doit être **mise en forme** : les idées, les concepts ne sont pas protégeables.

Deux types de droits associés : droits moraux et droits patrimoniaux.

Droits moraux : ce sont des droits imprescriptibles, inaliénables, incessibles, ils sont en général associés à des personnes physiques (les auteurs ou leurs héritiers). Il y en quatre :

- 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.

Le droit d'auteur des œuvres (2/2)

Droits patrimoniaux : concernent l'exploitation de l'œuvre, ce sont des droits monnayables, cessibles, temporaires.

On considère qu'il y a deux types d'exploitation :

- la représentation (par exemple d'une œuvre de théâtre) et
- la reproduction (musique sur CD par exemple).

Ce sont des droits associés souvent à des personnes morales (suite à des cessions effectuées par les auteurs), on parle alors des **détenteurs** des droits patrimoniaux, ou des **propriétaires**.

Terminologie :

- Œuvres orphelines :
il n'y a plus de personne physique associée aux droits moraux.
- Œuvres de domaine public :
fin des droits patrimoniaux, 70 ans après le décès de l'auteur.
 Ce terme est parfois (mal) utilisé dans le cadre de LL.

Le droit d'auteur du logiciel : traitement spécial

Pour les logiciels, il y a des **différences** :

- Droits moraux réduits : paternité.
- L'auteur ne peut (sauf stipulations contraires) s'opposer à la modification de l'œuvre ou exercer son droit de retrait.
- Les droits patrimoniaux (sauf stipulations contraires) sont dévolus à l'employeur. Cela s'applique aussi à leur documentation.
- Originalité : effort personnalisé allant au-delà de la simple mise en œuvre d'une logique automatique et contraignante.

Les détenteurs des droits patrimoniaux (propriétaires) d'un logiciel sont établis en fonction de :

- les auteurs
- leur statut et/ou le mode de collaboration
- les contrats : employeurs, collaboration, commande, conventions...
- ⚠ laboratoires : les accords entre tutelles (quadriennaux...)

Les données et bases de données

Les données prennent des formes différentes : chiffres, statistiques, textes, thésaurus, images, sons, vidéos et être au format papier, numérique, ou échantillons... Peuvent être organisées (ou pas) en bases de données.

Art. L.112 - 3 du CPI

On entend par base de données (BD) un recueil d'oeuvres, de données ou d'autres éléments indépendants, **disposés** de manière systématique ou méthodique, et **individuellement** accessibles par des moyens électroniques ou par tout autre moyen.

Protection double : droit d'auteur, droit sui generis du producteur de BD.

Le droit d'auteur peut s'appliquer

- aux données,
- à la structure de la base (l'architecture), la disposition des matières...
- aux éléments qui permettent de la faire fonctionner.

Le droit sui generis du producteur de BD

Le producteur :

- prend l'initiative et est responsable de la base,
- assume le risque financier, matériel et/ou humain (invest. **substantiel**).

Protection du contenu, le producteur a le droit d'interdire :

- L'extraction (par transfert permanent ou temporaire sur un autre support) de la totalité ou d'une partie qualitativement ou quantitativement substantielle du contenu d'une BD, par tout moyen et sous toute forme que ce soit.
- La réutilisation (par la mise à la disposition du public) de la totalité ou d'une partie qualitativement ou quantitativement substantielle du contenu de la BD, quelle qu'en soit la forme.
- Le producteur peut également interdire l'extraction ou la réutilisation répétée et systématique de parties qualitativement ou quantitativement non substantielles du contenu de la BD lorsque ces opérations excèdent manifestement les conditions d'utilisation normale de la base de données.

Durée de 15 ans à partir du 1er janvier de l'année qui suit :

- l'achèvement de la fabrication,
- la première mise à disposition,
- ou un nouvel investissement substantiel.

Référence : M. Dantan (CNRS/DAJ).

- 1 Cadre légal : licences
- 2 Cadre légal : le droit d'auteur des œuvres et le droit sui generis
- 3 PRESOFT : pérennisation des logiciels de la recherche**
 - Le projet
 - Plan de gestion du logiciel PRESOFT ou SMP
 - Bénéfices du SMP
- 4 FG-SOL : ensemble de services pour les logiciels de la recherche
- 5 Conclusions

PRESOFT : pérennisation des logiciels de la recherche

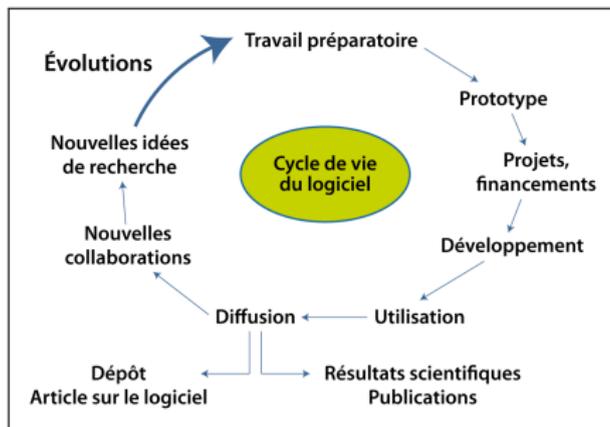
Projet IN2P3 (2017-18), partenaires : CC-IN2P3, IdGC, LIGM



Objectif :

étudier l'implémentation de plans de gestion de logiciels dans les unités de recherche afin d'améliorer leur pérennisation.

Développer des procédures et modèles **adaptés au cycle de vie** du logiciel.



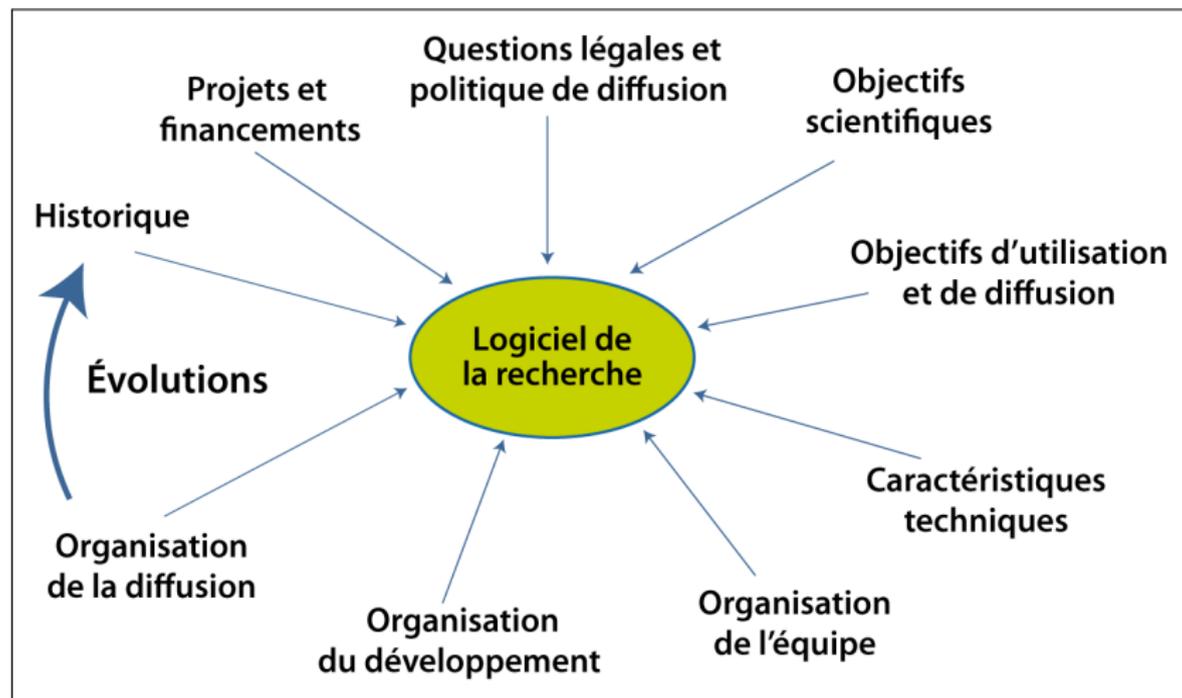
Plan de gestion du logiciel PRESOFT ou SMP (1/2)

Un SMP ...

- Nouvel outil pour la pérennisation du logiciel
- Feuille de route pour gérer le cycle de vie du logiciel
- Doit être adapté au logiciel
- Doit être adapté aux besoins de l'équipe et des chercheurs
- Outil de réflexion, de planification
- Permet d'identifier les étapes, les acteurs, les responsabilités
- Permet de mieux planifier la diffusion
- Outil pour mieux gérer la production scientifique
- **À ne pas confondre avec un plan de développement**

En cours version bilingue du modèle de SMP, en test au LIGM
(4 plans en cours de rédaction).

Plan de gestion du logiciel ou SMP (2/2)



Bénéfices du SMP

- Au niveau équipe de développement
 - ▶ Une vue d'ensemble de chaque logiciel
 - ▶ Un suivi facile dans une équipe
 - ▶ L'amélioration des méthodes de gestion (gain de temps)
 - ▶ L'identification de ce qui marche bien
 - ▶ L'identification des risques
 - ▶ La mise en œuvre d'une stratégie
- Au niveau laboratoire ou institution
 - ▶ La connaissance des productions scientifiques
 - ▶ Une meilleure visibilité des productions importantes
 - ▶ La mise en œuvre d'une politique

- 1 Cadre légal : licences
- 2 Cadre légal : le droit d'auteur des œuvres et le droit sui generis
- 3 PRESOFT : pérennisation des logiciels de la recherche
- 4 **FG-SOL : ensemble de services pour les logiciels de la recherche**
 - Vision et objectifs
 - Les services
- 5 Conclusions

Créer l'environnement nécessaire où les logiciels de la recherche peuvent être découverts, étudiés et réutilisés par d'autres chercheurs afin de valider et de reproduire les résultats publiés, ce qui produira de nouvelles connaissances.

Objectifs :

- Initialement conçu dans le cadre d'une proposition de projet H2020 (SOL2020, mars 2016) : CNRS, EGI.eu, IASA, Univ. d'Édimbourg
- Tenant compte des contextes européen et français
- Pour s'intégrer aux infrastructures existantes
- Proposer un ensemble cohérent de services
- Placer les logiciels à leur place dans le paysage scientifique
- Vitrine pour la production logicielle de nos laboratoires

Les services (1/2)

● Accès

- ▶ PLUME-RELIER (logiciels des laboratoires), EGI AppDB
- ▶ publication de fiches descriptives, interface de recherche
- ▶ liens vers les publications et les données associées
- ▶ FOSS licence recommandée

● Conseil & Support

- ▶ Patrimoine logiciel d'un laboratoire et Software Sustainability Institute
- ▶ améliorer les conditions de diffusion des logiciels de la recherche
- ▶ documentation, formation

● Préservation

- ▶ premier niveau de préservation, copie figée
- ▶ FOSS licence

● Utilisation

- ▶ copie re-utilisable du logiciel, accès au code source
- ▶ faciliter la reproductibilité des résultats publiés
- ▶ FOSS licence

Les services (2/2)



FG-SOL :

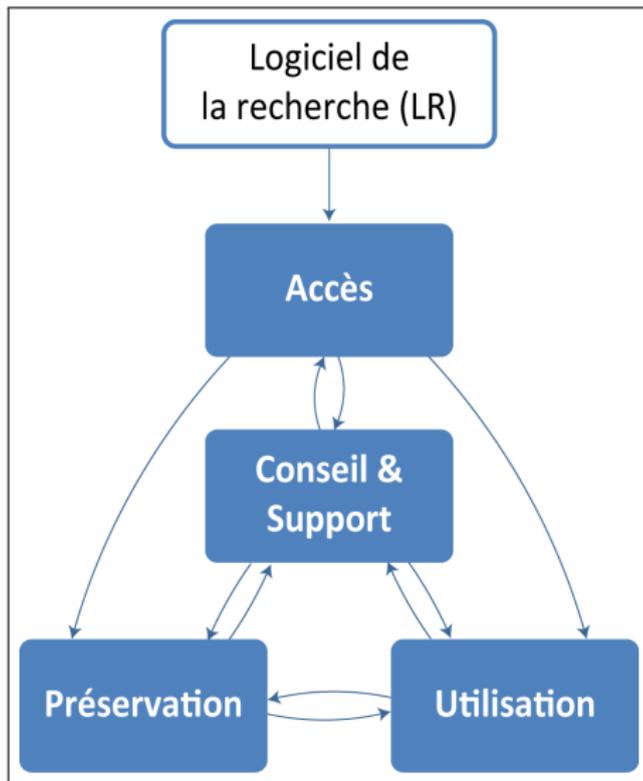
- nouveau service France Grilles
- <http://www.france-grilles.fr/fg-sol/>
- services en France,
- prototype en construction,
- CC-IN2P3, CREATIS, IPHC, LIGM,



- adoption de **VIP** (CREATIS) :
 - utilisation des grilles FG et EGI
 - interaction avec :
FG-Cloud, FG-DIRAC, FG-iRODS

Voir :

- <http://www.france-grilles.fr/fg-sol-acces/>
- Journées SUCCES 2017 :
[poster](#), Utilisation : [intro](#), [dém](#)
- [JRES 2017](#) (cet apm !) :
article, présentation



Conclusions

- Le paysage Open Science est en construction
- Beaucoup est fait et continue à se faire pour les articles
- En cours : les données, mais que veut dire “research data” ?
- L'élément mystérieux : les logiciels, mais comment manipuler les données de la recherche sans les logiciels de la recherche ?
Comment reproduire les résultats publiés ?
De plus en plus de voix, dont Dr Ben Goldacre, Univ. Oxford
Data access and transparency, 10/2016 <https://www.youtube.com/watch?v=XqJ7QMTSKME>
sans oublier D. Donoho, <http://reproducibleresearch.net/>
- Le libre accès à la science se construit sur un contexte legal, qui implique l'utilisation de licences et des évolutions des lois (EC, pays).
- Par ailleurs...

Pas de science ouverte possible sans les logiciels de la recherche

Références

- Droit d'auteur des chercheurs, bases de données et archives ouvertes, M. Dantant, CNRS/DAJ (2013), http://mistral.cnrs.fr/IMG/pdf/Droit_d_auteur...
- Framabook Option Libre. Du bon usage des licences libres, B. Jean (2011) - <http://framabook.org/option-libre-du-bon-usage-des-licences-libres>
- Thème PLUME : patrimoine logiciel d'un laboratoire - <https://www.projet-plume.org/patrimoine-logiciel-laboratoire>
- L'open data : quelle protection ? Quelles obligations pour le chercheur ?, H. Skrzyniak, U. Bordeaux/Irdap (2017), <http://doranum.fr/enjeux-juridiques...>
- La directive Inspire pour les néophytes, F. Merrien, M. Léobet, M. Francès, Min. Écologie (2015) <https://geo.valdemarne.fr/portail/...>
- Ouverture et partage des données publiques culturelles pour une (r)évolution numérique dans le secteur culturel, C. Domange, Min. Culture et Communication (2013), <http://www.culturecommunication.gouv.fr/...>
- Ouverture des données de la recherche ? Guide d'analyse du cadre juridique en France (2016) <http://www.bibliothequescientifiquenumerique.fr/ouverture...>

Comprendre les logiciels de recherche : Article vs. Logiciel

Article vs. Logiciel : questions juridiques et de politique scientifique dans la production de logiciels

<http://www.societe-informatique-de-france.fr/wp-content/uploads/2015/04/1024-5-gomez-diaz.pdf>

Aspects légaux		
	Article	Logiciel
Droit auteur	droits moraux, droits patrimoniaux	droits moraux réduits droits pat. dévolus à l'employeur
Œuvre	article	code source, code objet, doc., ...
Auteurs	signataires, même %	notion complexe, pb. légal , établir % de participation
Propriétaires	auteurs, même % cession des droits	tutelles en général, mais dépend du régime salarié , des contrats , ...
Dates	soumission, publication	matériel de conception, versions
Évolution	œuvre indépendante	œuvre indépendante ? il faut revoir auteurs, dates, lic., ...
Travaux préc.	références, citations	briques : compatibilité , héritage lic.
Diffusion	éditeur, web	web, forges, besoin de licence
Droits	lire, citer, ne pas copier	lire, ne pas utiliser , ..., besoin lic.
Licences	droits et obligations, CC (web)	droits et obligations, libres, propriétaires

C'est clair pour les articles. Il faut prendre des précautions pour les logiciels.

Procédure de distribution des logiciels

S'adapte à chaque situation, valable pour les données.

- Choisir un nom, éviter les noms déjà utilisés, les marques.
- (*) Établir la liste des auteurs (avec % de participation), leurs affiliations.
- (*) Établir la liste des fonctionnalités principales.
- (*) Établir la liste des briques logicielles ou les données utilisées, avec licences.
- Choisir une licence, avec l'accord des auteurs et propriétaires des droits. Si possible : un accord signé. Attention à la compatibilité et héritage des licences.
- Choisir un site web, forge, dépôt pour la distribution. Indiquer les licences et les conditions d'utilisation, copie, etc., et comment citer l'œuvre.
Utiliser métadonnées, identificateurs persistants, standards ouverts...
Donner des licences à la documentation, au site web...
- Créer et indiquer une adresse courriel de contact.
- (*) La traçabilité est importante, archiver en .tar.gz régulièrement.
- Informer la direction des laboratoires et les tutelles (si pas fait au point licence).
- Diffuser le logiciel et/ou les données.
- Informer la communauté cible, considérer les data ou les software papers.

(*) À revoir à chaque nouvelle version du logiciel.

Mise en place d'une licence

La licence doit être mise en place **avant la diffusion** du logiciel.
Attention aux cahiers des charges et aux contrats (clauses PI, licences).

En-tête pour tous les fichiers :

- Nom du fichier, nom du logiciel
- Copyright (©, Droits patrimoniaux), année(s), p. morale ou physique aussi : tous droits réservés, quelques droits réservés
- Auteur(s), une adresse de contact
- Licence(s)
- Utiles : date de création, date de la dernière version
- Utile : format SPDX ou Open source cartouche

Et ajouter un fichier de licence (COPYING, LICENCE, README, ...) à l'ensemble des fichiers, avec le texte complet ou une URL.

En plus :

- Indiquer les briques logicielles utilisées et leurs licences.
- Indiquer clairement la licence (**et les auteurs**) dans la documentation, sur le site Web.