

Open science : la diffusion et le partage des productions

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

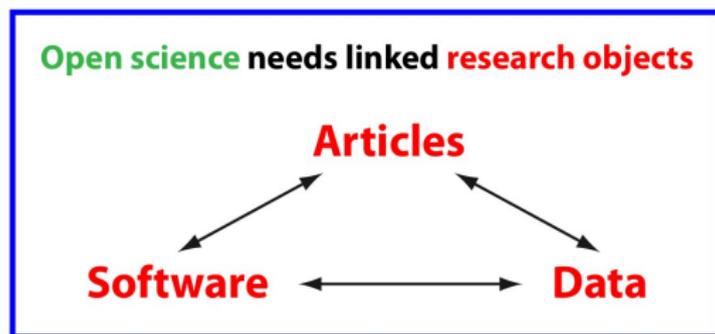
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

XLIM, Limoges, 19 juin 2017



Motivation

- Open Science : diffusion et partage des productions scientifiques
- Mouvements Open Access, Open Data : articles et données
- 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*
- Nous avons besoin des objets de la recherche et des liens entre eux



Plan

- 1 Les logiciels
 - Définitions : free et/ou open source software
 - Les licences des logiciels
- 2 Le droit d'auteur des œuvres
 - Les œuvres
 - Les logiciels
 - Article vs. Logiciel : aspects légaux
 - Procédure de distribution
- 3 Les articles et les données : le libre accès (open access)
 - La déclaration de Budapest
 - La déclaration de Berlin
 - Aujourd'hui
 - La politique de la Commission européenne
 - En France
- 4 Des aspects plus pratiques dans la science ouverte
 - Data Management Plan
 - PRESOFT : Software management Plan
 - Infrastructures
- 5 Conclusions

Définition de logiciel libre - free software

Selon la Free Software Fondation (FSF, 1985), fondée par R. M. Stallman, un logiciel est libre si ces quatre libertés sont garanties :

<http://www.gnu.org/philosophy/free-sw.fr.html>

- liberté d'exécuter le logiciel (tout usage),
- (*) liberté d'**étudier** et de modifier le fonctionnement,
- liberté de redistribuer des copies,
- (*) liberté de distribuer aux autres des copies de vos versions modifiées
⇒ (*) **condition nécessaire : accès au code source**

Exemples : T_EX by D. Knuth (1978), the Berkeley Software Distribution (BSD) by U. of California (1977-1995).

Garantie des quatre libertés : il faut une licence.

Un logiciel libre n'est pas *libre de droits* (protégé par CPI).

Un logiciel qui n'est pas libre se dit propriétaire (*privatif*).

Cette définition est née dans des milieux universitaires.

Définition de logiciel code ouvert - open source software

Selon l'Open Source Initiative (OSI), 1998, un logiciel est open source si sa licence respecte ces conditions :

<http://www.opensource.org/docs/osd>

1. Free Redistribution
The license shall not restrict any party from **selling** or giving away...
2. Source Code (⇒ **disponibilité du code**)
3. Derived Works
4. Integrity of The Author's Source Code
5. No Discrimination Against Persons or Groups
6. No Discrimination Against Fields of Endeavor
7. Distribution of License
8. License Must Not Be Specific to a Product
9. License Must Not Restrict Other Software
10. License Must Be Technology-Neutral

Cette définition est née dans des milieux proches des entreprises.
Besoin de licence pour garantir ces 10 conditions.

FOSS, FLOSS : free (libre) et/ou open source software (les mêmes licences).

Pourquoi une licence ?

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

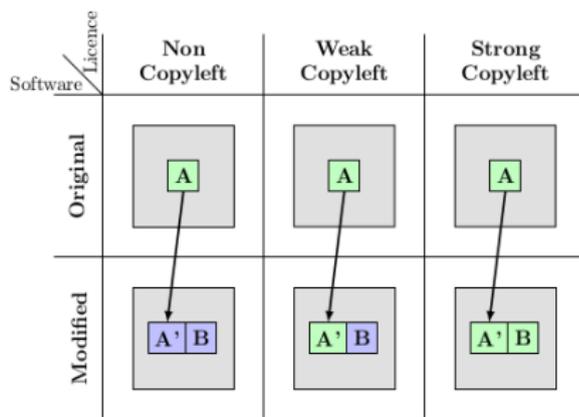
Les licences complètent le cadre juridique établi par la loi : s'il n'y a pas de droit explicitement donné, utiliser un logiciel relève de la contrefaçon.

Les licences sont des **contrats** et protègent les auteurs, les utilisateurs et les éventuels collaborateurs au développement.

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

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 (*persistentes*)
 - 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>

- 1 Les logiciels
- 2 Le droit d'auteur des œuvres
 - Les œuvres
 - Les logiciels
 - Article vs. Logiciel : aspects légaux
 - Procédure de distribution
- 3 Les articles et les données : le libre accès (open access)
- 4 Des aspects plus pratiques dans la science ouverte

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...)

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.

Mettre en place 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.

- 1 Les logiciels
- 2 Le droit d'auteur des œuvres
- 3 Les articles et les données : le libre accès (open access)**
 - La déclaration de Budapest
 - La déclaration de Berlin
 - Aujourd'hui
 - La politique de la Commission européenne
 - En France
- 4 Des aspects plus pratiques dans la science ouverte
- 5 Conclusions

Le libre accès : déclaration de Budapest (BOAI, 2002)

Convergence d'une tradition ancienne et d'une technologie nouvelle : les politiques d'accès libre (open access) sont de plus en plus importantes dans la communauté scientifique internationale. Elles sont rendues possibles par Internet et ses conséquences, en partie inspirées par les mouvements du FOSS, et motivées par les problèmes de reproductibilité.

<http://www.budapestopenaccessinitiative.org/>

Par "accès libre" à cette littérature, nous entendons sa mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces articles, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celles indissociables de l'accès et l'utilisation d'Internet. La seule contrainte sur la reproduction et la distribution, et le seul rôle du copyright dans ce domaine devrait être de garantir aux auteurs un contrôle sur l'intégrité de leurs travaux et le droit à être correctement reconnus et cités.

Le libre accès : déclaration de Berlin (2003)

Suite à BOAI, il y a la déclaration de Bethesda (2003) pour identifier :

les mesures concrètes nécessaires à la promotion d'un passage rapide et efficace à la publication en libre accès

Elle est suivie de la Déclaration de Berlin (2003), qui a été signée par le CNRS, la CPU, INRA, INRIA...

Les contributions au libre accès se composent de résultats originaux de recherches scientifiques, de données brutes et de métadonnées, de documents sources, de représentations numériques de documents picturaux et graphiques, de documents scientifiques multimédia. Elles doivent satisfaire deux conditions :

Droits et licence Les auteurs et les propriétaires des droits concèdent à tous les utilisateurs un droit gratuit, irrévocable et mondial d'accéder à l'**œuvre**, et une licence les autorisant à la copier, l'utiliser, la distribuer, la transmettre et la montrer en public, et de réaliser et de diffuser des œuvres dérivées.[...]

Dépôt Une version complète de cette œuvre, ainsi que de tous ses documents annexes, y compris une copie de la permission définie dans ce qui précède, est déposée (et, **de fait, publiée**) sous un format électronique approprié auprès d'au moins une archive en ligne.[...]

<http://legacy.earlham.edu/~peters/fos/bethesda.htm>

<http://openaccess.mpg.de/Berlin-Declaration>

La politique de la Commission européenne

- **Mandat** Open Access pour les publications en H2020 : assurer OA aux publications *peer-reviewed*, version finale dans un dépôt (par ex. HAL), y compris les métadonnées. Vise aussi les données nécessaires à la validation des résultats (dépôt simultané).
- Pilote en 2015, **mandat** en 2016 pour les données de la recherche : concerne les données sous-jacentes pour valider les résultats, ou autres données selon défini dans le DMP, obligation des DMPs pour les projets participants (opt-out possible).

Dépôt, accès, explorer (mine), exploiter, reproduire, distribuer, sans coût additionnel.
Donner (les informations sur) les outils et les instruments si possible.

- **Financement** de projets :

DARIAH, DataCite, EGI, **EOSC**, EUDAT, FOSTER, OpenAIRE, RDA (Europe)...

- **Policy framework** : Digital Single Market

user centric - service oriented - reliable - innovative eInfrastructure

- [From vision to action Next steps on the European Open Science Cloud and more](#), EC/DG RTD, 3/2017
- [European Digital Single Market Strategy : changing the game with a service oriented e-infrastructure](#), A. Burgueño-Arjona, EC/DG CONNECT, 9/2015

En France (1/2)

- HAL <https://hal.archives-ouvertes.fr/>

Convention de partenariat en faveur des archives ouvertes et de la plateforme mutualisée HAL, 2 avril 2013

engage tous les Partenaires à mettre en œuvre une politique d'archives ouvertes conforme aux modalités décrites

l'AMUE, l'ANDRA, l'ANR, la BNF, Le BRGM, la CDEFI, le CEA, la CEE, la CGE, le CIRAD, le CNRS, la CPU, le CSTB, l'IFPEN, l'IFREMER, l'IFFSTAR, l'INERIS, l'INRA, l'INRIA, l'INSERM, l'INVS, l'IRD, l'IRSN, l'IRSTEA, l'Institut Pasteur

http://cache.media.enseignementsup-recherche.gouv.fr/file/HAL/93/3/01_Convention_HAL_246933.pdf

- 36 “repositories OpenAIRE compliant” :

repository managers expose publications, datasets and CRIS metadata via the OAI-PMH protocol in order to integrate with OpenAIRE infrastructure

- Les enjeux éthiques du partage des données scientifiques, Avis du COMETS, mai 2015

http://www.cnrs.fr/comets/IMG/pdf/2015-05_avis-comets-partage-donnees-scientifiques.pdf

- Loi pour une République numérique, 7 oct. 2016

<https://www.economie.gouv.fr/republique-numerique>

En France (2/2)

- Bénéfices de certains éditeurs prédateurs
Perceptions de l'Open Access par les marchés financiers, Sami Kassab,
http://www.canal-u.tv/video/fmsh/les_6emes_journees_open_access_sami_kassab.19545
Journées open access Couperin, 10/2015, <https://jao2015.sciencesconf.org/>
- M. Farge, CNRS & ENS : Scholarly publishing and peer-reviewing in open access, 03/2017
http://openscience.ens.fr/MARIE_FARGE/2017_BOOK_CHAPTER_...
Reflections of the Research, Innovation and Science Policy Experts (RISE) High Level Group
- J.C. Guédon, U. Montréal : évaluation et Impact Factor
 - ▶ Journées à Toulouse Open Access et évaluation de la recherche, 10/2016
<https://openeval2016.sciencesconf.org/>
 - ▶ Open Access : Toward the Internet of the Mind, 02/2017
<http://www.budapestopenaccessinitiative.org/open-access-toward-the-internet-of-the-mind>
 - ▶ Carrefour de l'IST à Nancy, 03/2017
<https://carist.sciencesconf.org/>

- 1 Les logiciels
- 2 Le droit d'auteur des œuvres
- 3 Les articles et les données : le libre accès (open access)
- 4 Des aspects plus pratiques dans la science ouverte
 - Data Management Plan
 - PRESOFT : Software management Plan
 - Infrastructures
- 5 Conclusions

Data Management Plan

- DMP : livrable dans les projets H2020 (et autres) pour formaliser la gestion des données

« A DMP describes the data management life cycle for all datasets to be collected, processed or generated by a research project. »

Guidelines on Data Management in Horizon 2020, 02/2016, Commission européenne

- Objectif FAIR : findable, accessible, interoperable, re-usable
<https://www.force11.org/group/fairgroup/fairprinciples>
- Digital Curation Center (DCC), section DMP :
DMPonline + Checklist for a Data Management Plan, v4.0 (2013)
- FOSTER 2015 : Open Access et gestion des données dans H2020, 06/2015, <https://www.fosteropenscience.eu/event/open-access-et-gestion...>
- Plan de gestion de données, contraintes supplémentaires ou réelle opportunité ? Quels enjeux ? G. Romier, 10/2016,
<http://informatique.in2p3.fr/li/spip.php?article445>
- Outil CNRS/INIST (adaptation DCC) : <https://dmp.opidor.fr/>
DMP OPIDoR : Data Management Plan pour une Optimisation du Partage et de l'Interopérabilité des Données de la Recherche

Conclusions (1/2)

- 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 se construit sur un contexte legal, qui implique l'utilisation de licences et des évolutions des lois (EC, pays).
Comment avoir un Digital Single Market ?

Data and computing infrastructures for open scholarship - a legal perspective, T. Margoni, U. of Stirling

Conclusions (2/2)

- 2015 : impossible d'arrêter...
- 2017 : on y va à fond
- Mais encore un énorme décalage entre certains discours et
 - ▶ la politique scientifique de nos institutions
 - ▶ l'évaluation de la recherche :
facteur d'impact, production de données et de logiciels
 - ▶ le quotidien des chercheurs, des laboratoires
- Par ailleurs...

Pas de science ouverte possible sans les logiciels de la recherche

Références

- Le droit lié aux logiciels, par Valérie Hospital (DAJ, CNRS) -
https://www.projet-plume.org/files/20110615_rencontrescnrs_droitlogiciels_hospital.pdf
(et autres : [site DAJ CNRS sur le droit d'auteur](#), CPI, FSF, OSI, ...)
- 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>
- E-Infrastructures & RDA for data intensive science 9/2015 -
<https://rd-alliance.org/sixth-plenary/e-infrastructures-rda-data-intensive-science.html>
- Open Science Conference, mars 2017 <https://www.open-science-conference.eu/>