

Sur le partage et la diffusion des logiciels et des données de la recherche dans le cadre de la Science ouverte. The Borgman's conundrum challenges.

Teresa Gomez-Diaz

CNRS - Laboratoire d'informatique Gaspard-Monge

Contient des travaux en collaboration avec Prof. T. Recio (Univ. Nebrija, Madrid)

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License
(CC-BY-NC-ND 4.0) – <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Séminaire à l'Institut de Systématique, Évolution, Biodiversité
Muséum national d'Histoire naturelle, Paris, 17 décembre 2024



Un peu de poésie

The Elephant's Child, Rudyard Kipling (1902)

I Keep Six Honest Serving Men

*I keep six honest serving-men
(They taught me all I knew);
Their names are What and Why and When
And How and Where and Who.
I send them over land and sea,
I send them east and west;
But after they have worked for me,
I give them all a rest.*

...



Josephine

The Kipling method, also known as the 5W1H Method:

What and Where and When, and How and Why and Who.

“Borgman’s conundrum challenges”: which data might be shared, by whom, with whom, under what conditions, why, and to what effects...

https://www.kiplingsociety.co.uk/poem/poems_serving.htm

Objectifs

(*) (2019) TGD, T. Recio. On the evaluation of research software : the CDUR procedure

(2022) TGD, T. Recio. Research Software vs Research Data I (Definition)

(2022) TGD, T. Recio. Research Software vs Research Data II (Dissemination, CDUR)

Notre objectif ici est de présenter ces travaux sur les données de la recherche, où nous proposons :

- Définition
- Procédure de diffusion
- Protocoles CDUR d'évaluation

Ces travaux sont basés sur nos travaux sur les logiciels de la recherche (préambule I) dans le cadre de la Science ouverte (préambule II). Nous montrerons une partie du chemin (préambule III) qui nous a menés à construire ces propositions.

Ces travaux sont réalisés en collaboration avec le Prof. Tomas Recio, Universidad Antonio de Nebrija, Madrid, (<http://www.recio.tk/>).

(*) Les références complètes sont à la fin de la présentation.

Plan

1 Préambule I : les logiciels de la recherche

- Le concept, la définition
- Procédure de diffusion
- Protocole(s) **CDUR** : évaluation de la recherche et RS

2 Préambule II : la Science ouverte

- Définition proposée de Science ouverte
- Recommandation de l'UNESCO sur une science ouverte

3 Préambule III : et pour les données ?

- Article vs. Logiciel : questions juridiques et de politique scientifique
- Articles, logiciels, **données** : étude de la diffusion de la production scientifique
- Le logiciel est un objet juridique
- Objectifs

4 Les données de la recherche

- Les données sont des objets juridiques ?
- Définition
- Borgman's conundrum challenges + Comment ? + Où ?
- Procédure de diffusion
- Protocole(s) **CDUR** d'évaluation de la recherche et RD
- Protocole(s) **CDUR** & FAIR Guiding Principles. CoARA.

5 Conclusions

Définition de logiciel de la recherche

(2007) TGD. Autour de la valorisation de logiciels développés dans un laboratoire de recherche

(2009) TGD. Guide laboratoire pour recenser ses développements logiciels

(2011, 2015) TGD. Article vs. Logiciel : questions juridiques et de politique scientifique dans la production de logiciels

Un logiciel du laboratoire est un programme utile pour faire avancer la recherche qui a été produit avec la participation d'un membre du laboratoire.

Il arrive souvent que des publications de recherche soient associées.

- finalité : recherche
- un membre du laboratoire participe à l'écriture du logiciel (idem aux publications)
- la production importante sont les articles, les logiciels sont des objets associés

(2019) TGD, T. Recio. On the evaluation of research software: the CDUR procedure

(2021) TGD, T. Recio. Open comments on the Task Force SIRS report: Scholarly Infrastructures for Research Software

Research software (RS) is a well identified set of code that has been written by a well identified research team. It is software that has been built and used to produce a result published or disseminated in some article or scientific contribution.

Each RS encloses a set of files containing the source code and the compiled code. It can also include other elements as the documentation, specifications, use cases...

Les définitions ne tiennent pas compte de **l'état du RS** :

"en projet", "fini", diffusé, qualité, portée, taille, documenté, maintenu, utilisé uniquement par une équipe pour réaliser une publication, ou utilisé dans plusieurs laboratoires

Procédure de diffusion des logiciels de la recherche

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

- Choisir un nom, éviter les noms déjà utilisés, les marques.
(2018) DoRANum or Harvard, Bien nommer ses fichiers/File Naming Conventions.
- (*) Établir la liste des auteurs (avec % de participation), leurs affiliations.
Plan de gestion de logiciels de la recherche (PGLR) (2018, TGD, G. Romier, Modèle de PGLR V3.2, PRESOFT)
- (*) Établir la liste des fonctionnalités principales.
- (*) Établir la liste des briques logicielles ou les données utilisées, avec licences.
- **Choisir et mettre en place une licence**, accord de tous les détenteurs 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..., comment citer l'œuvre.
- Créer et indiquer une adresse courriel de contact.
- (*) La traçabilité est importante, archiver en .tar.gz régulièrement. Réviser PGLR.
- 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.

(2010) TGD. Diffuser un logiciel de laboratoire : recommandations juridiques et administratives

(2014) TGD. Free software, Open source software, licenses. A short presentation including a procedure for RS and data diss...

(2022) TGD, T. Recio. Research Software vs Research Data II (Dissemination, CDUR)

Protocole(s) **CDUR** : évaluation de la recherche et RS

(2019) TGD, T. Recio. On the evaluation of research software: the CDUR procedure

(2022) TGD, T. Recio. Research Software vs Research Data II (Dissemination, CDUR)

Le protocole **CDUR** est proposé pour prendre en compte les logiciels de la recherche dans l'évaluation. Conçu pour aider les chercheurs évalués, les comités d'évaluation, les décideurs.

Il peut s'adapter aux **données de la recherche**.

- (C) Citation** mesure si RS bien identifié en tant que produit de la recherche : référence ou bien metadataset avec nom, auteurs, version, dates...
aussi bonnes pratiques pour citer les autres RS...
point légal : auteurs, affiliations, participation
- (D) Dissemination** les bonnes pratiques de diffusion suivies en fonction de la politique scientifique du contexte d'évaluation. Plan de Gestion de RS (PGLR).
point Science ouverte, légal : licences
- (U) Use** point dédié aux aspects purement "logiciel" **du RS** : produit des résultats corrects, facilite la réutilisation ; peut regarder aussi bonnes pratiques logiciel : exemples, docs., test, facilite install., voir le code, lancer le RS, normes de qualité, dépôt APP, valorisation, startup...
point reproductibilité, validation des résultats obtenus
- (R) Research** point dédié aux aspects purement "recherche" : qualité du travail scientifique, algorithmes et structures de données proposés et programmés, publications, utilisateurs, collaborations et projets...
point impact de la recherche

Flexibilité d'application : chaque ensemble décideurs/comité d'évaluation établit **sa propre procédure**, adaptée aux objectifs fixés et au contexte d'évaluation.

Plan

- 1 Préambule I : les logiciels de la recherche
 - Le concept, la définition
 - Procédure de diffusion
 - Protocole(s) **CDUR** : évaluation de la recherche et RS
- 2 Préambule II : la Science ouverte
 - Définition proposée de Science ouverte
 - Recommandation de l'UNESCO sur une science ouverte
- 3 Préambule III : et pour les données ?
 - Article vs. Logiciel : questions juridiques et de politique scientifique
 - Articles, logiciels, **données** : étude de la diffusion de la production scientifique
 - Le logiciel est un objet juridique
 - Objectifs
- 4 Les données de la recherche
 - Les données sont des objets juridiques ?
 - Définition
 - Borgman's conundrum challenges + Comment ? + Où ?
 - Procédure de diffusion
 - Protocole(s) **CDUR** d'évaluation de la recherche et RD
 - Protocole(s) **CDUR** & FAIR Guiding Principles. CoARA.

Définition proposée de Science ouverte - article

(2020-21) TGD, T. Recio. Towards an Open Science definition as a political and legal framework: on the sharing and dissemination of research outputs

La Science ouverte est :

[FR] *le cadre politique et juridique dans lequel les productions scientifiques sont partagées et diffusées afin d'être rendues visibles, accessibles et réutilisables.*

[EN] *the political and legal framework where research outputs are shared and disseminated in order to be rendered visible, accessible and reusable.*

Version	Titre	Date	Publication
V3	Towards an Open Science definition as a political and legal framework: on the sharing and dissemination of research outputs	02/2021	https://zenodo.org/record/4577066 Avec la réf. : Alma Swan, UNESCO, 2012
V2	Towards an Open Science definition as a political and legal framework...	12/2020	POLIS N. 19, pp. 36-56 - PDF https://doi.org/10.58944/yuro5734
V1	A policy and legal Open Science ...	09/2020	https://zenodo.org/record/4075106

Objectifs : comprendre ce qu'est la Science ouverte, la motivation qui nous mène à proposer cette définition et voir comment structurer l'information.

Définition proposée de Science ouverte - poster

The future of Open Science asks for a common understanding

Open Science is the political and legal framework where research outputs are shared and disseminated in order to be rendered visible, accessible and reusable.

I Three selected pillars

- BOAI (2002)
- Free Software Foundation (1985)
- CODATA (1966)

II Towards a political and legal framework

III Enablers:

- Institutional policies
- Infrastructures
- Research evaluation

EGI Virtual Conference 2021, Lisbon, 19-21 October 2021

<https://padlet.com/gwenfranck/EGI2021Posters>

(2020-21) TGD, T. Recio. Towards an Open Science definition as a political and legal framework: on the sharing and dissemination of research outputs

The future of Open Science asks for a common understanding
Teresa Gomez-Diaz, CNRS/LIGM, Est of Paris
Tomas Recio, University Nebrija, Madrid

A definition is missing
Definition (of Open Science) remains vague, probably because objectives are vague... (Borner, 2019)
...there is no single accepted, unified definition or set of agreed-upon... (TGD, 2020)
...Research output of researchers... (TGD, 2020)
...mostly share the fact that there is no shared definition of Open Science... (Gomez-Diaz & Recio, 2020-21)

Research, inclusive & complete visions

Our contribution

Goal
To contribute to the adoption of a common, unified vision.

Definition proposal
Open Science is the political and legal framework where research outputs are shared and disseminated in order to be rendered visible, accessible and reusable (Gomez-Diaz & Recio, 2020-21).

Three steps supporting this proposal
I- Three selected pillars for a common understanding
II- Towards a political and legal framework
III- Enablers: three cornerstones to get to a working framework

I Three selected pillars for a common understanding

- BOAI (2002)**
The BOAI (2002) is a report from the Board of the Open Access movement, published in 2002. It is a key document in the history of Open Access.
- Free Software Foundation, 1985**
The Free Software Foundation (FSF) was founded in 1985. It is a non-profit organization that promotes the development and use of free software.
- CODATA (1966)**
CODATA (1966) is a committee of data scientists from various countries, established in 1966. It focuses on the development of standards for data exchange and sharing.

II - Towards a political and legal framework

- European Open Access Initiative, 2002**
The European Open Access Initiative (EOAI) was established in 2002. It is a non-profit organization that promotes the development and use of open access.
- Committee on Data for Science and Technology, 1966**
The Committee on Data for Science and Technology (CODATA) was established in 1966. It is a committee of data scientists from various countries, established in 1966. It focuses on the development of standards for data exchange and sharing.

III - Enablers: three cornerstones to get to a working framework

- Institutional policies**
The required evaluation of policies of universities and RPOs.
- Infrastructures**
The development of Open Science-oriented infrastructures and services.
- Research evaluation**
The transformation of evaluation policies and practices.

References

EGI Virtual Conference 2021
19-21 October 2021

Recommandation de l'UNESCO sur une science ouverte

- À la demande des 193 pays participant à la Conférence générale de l'UNESCO, 40th session, 11/2019
- Prise de conscience, conséquence de la pandémie de COVID-19
- Une année de préparation, consultation, commentaires au rapport préliminaire
<https://unesdoc.unesco.org/ark:/48223/pf0000374409>
- Texte adopté par les 193 pays, 41st session, 11/2021
<https://unesdoc.unesco.org/ark:/48223/pf0000379949>

Aux fins de la présente Recommandation, la **science ouverte** s'entend comme un **concept inclusif qui englobe différents mouvements et pratiques** visant à **rendre les connaissances scientifiques** multilingues, **librement accessibles à tous et réutilisables par tous**, à renforcer la collaboration scientifique et le partage des informations au profit de la science et de la société, ainsi qu'à ouvrir les processus de création, d'évaluation et de diffusion des connaissances scientifiques aux acteurs de la société au-delà de la communauté scientifique traditionnelle.

Elle **inclut toutes les disciplines scientifiques et tous les aspects des pratiques savantes, y compris les sciences fondamentales et appliquées, les sciences naturelles et les sciences sociales et humaines**, et repose sur les piliers essentiels suivants : les connaissances scientifiques ouvertes ; les infrastructures de la science ouverte ; la communication scientifique ; la participation ouverte des acteurs de la société ; et le dialogue ouvert avec les autres systèmes de connaissances.

Suivi des groupes de travail :

<https://www.unesco.org/en/open-science/implementation#open-science-working-groups>

Plan

2 Préambule II : la Science ouverte

- Définition proposée de Science ouverte
- Recommandation de l'UNESCO sur une science ouverte

3 Préambule III : et pour les données ?

- Article vs. Logiciel : questions juridiques et de politique scientifique
- Articles, logiciels, **données** : étude de la diffusion de la production scientifique
- Le logiciel est un objet juridique
- Objectifs

4 Les données de la recherche

- Les données sont des objets juridiques ?
- Définition
- Borgman's conundrum challenges + Comment ? + Où ?
- Procédure de diffusion
- Protocole(s) **CDUR** d'évaluation de la recherche et RD
- Protocole(s) **CDUR** & FAIR Guiding Principles. CoARA.

5 Conclusions

Article vs. Logiciel : questions juridiques et de politique scientifique

(2011, 2015) TGD. Article vs. Logiciel : questions juridiques et de politique scientifique dans la production de logiciels

Pour comprendre et expliquer les questions juridiques et de politique scientifique sur les logiciels de la recherche par comparaison avec la production principale des articles de la recherche.

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

Aspects relatifs à la politique scientifique		
	Article	Logiciel
Définition (L, T)	ok	à définir
Signature (C, T)	ok, déf. par tutelles	à définir (copyright) associer les laboratoires
Références (L, T)	HAL	PLUME
Liste des œuvres (L, T)	document à jour	document inconnu , PLUME peut être utile
Libre accès (C, L, T, CSI)	politique (+/-) ok, dépôt ok (HAL)	politique (lic.) à définir , dépôt à établir
Validation (C, L, T, CSI)	procédure referee , reproductibilité	à définir , validé au sens PLUME
Qualité/évaluation (C, L, T, CSI)	nb. citations	articles associés, attirer utilisateurs, contrats
Motivation (C, L, T, CSI)	recherche, article	recherche, pas le logiciel
Objet (C, L, T, CSI)	scientifique	3D : scientifique, potentiel de transf. de tech., obj. industriel

(2011, PLUME) <https://zenodo.org/record/7063154>

(2015, SIF) <https://zenodo.org/record/18993>

Toutes les références données à la fin traitent des questions liées aux aspects juridiques et de politique scientifique.

Articles, logiciels, données : étude de la diffusion de la production scientifique

Aspects de politique scientifique dans la production de logiciels et des données

Trois piliers de la production scientifique :

-  Articles
-  Logiciels
-  Données

- Les articles sont la partie la plus visible et accessible de la recherche.
- Les données peuvent être des objets juridiques complexes.
- La diffusion de logiciels et des données pose des questions décisionnelles similaires (et au même temps).

Articles, logiciels, données : étude de la diffusion de la production scientifique

T. Gomez-Diaz (CNRS, LIGM)

Quelle protection légale ?

pour la production

droit de savoir pour les utilisateurs, pour le bien

- droit sur genome / proteoforme
- données personnelles
- droit à l'image, droit des images
- données géographiques (Géomatique)
- données médicales
- confidentialité, secret
- culture publique, diffusion publique
- protection de l'environnement

La recherche : besoin de savoir.

Articles

Logiciels ← ? → Données

Besoins des laboratoires et des institutions de recherche

- fluidifier et accélérer la production
- clarifier les missions
- évaluation
- qualité de l'enseignement
- fluidifier les échanges de savoirs

Quelle licence utiliser ?

Open Access (Creative Commons)

utilisation en open access

droits réservés, accès limité

pour l'enseignement des données publiques

données Commons (CC0)

accès aux données CC

CC-BY-NC-ND

https://creativecommons.org/licenses/by/4.0/

Besoins des producteurs

- formation, support, expertise des besoins particuliers
- connaître les logiciels, les données ouvertes
- être actifs sur les plateformes de savoirs experts
- partager les connaissances expertises
- diffuser leur production (technique, juridique)
- valider (scientifique, transfert de technologie)
- se tenir au courant des évolutions : veille technologique

Besoins des utilisateurs

- formation, support, expertise des besoins particuliers
- connaître les logiciels, les données ouvertes
- être actifs sur les plateformes de savoirs experts
- partager les données expertises
- se tenir au courant des évolutions : veille technologique

Aspects de politique scientifique dans la production de logiciels et des données

	Article (*)	Logiciel (*)	Données
Définition (L, T)	ok	à définir	à définir
Signature (C, T)	ok, définie par les tutelles	à définir (ligne de copyright), associer les laboratoires	à définir, associer les laboratoires
Références (L, T)	HAL	PLUME	mal définies, de a au z des dépôts, lequel ?
Liste des œuvres d'un laboratoire (L, T)	document à jour	document incomplet, PLUME peut être utile	document incomplet
Libre accès (C, L, T, CSD)	politique (+/-) ok, dépôt ok (HAL)	politique (licences) à définir, dépôt à établir	politique (licences) à définir, dépôt officiel à établir
Validations (C, L, T, CSD)	procédure de referee	à définir	à définir : expertise ?
Qualité et évaluation (C, L, T, CSD)	nombre de citations	en fonction des articles associés, capacité d'attirer d'utilisateurs, contrats	en fonction des articles associés, capacité d'attirer de collaborateurs
Motivation (C, L, T, CSD)	recherche, article	recherche, pas le logiciel	recherche, données
Objet (C, L, T, CSD)	scientifique	3D : scientifique, mais aussi potentiel de transfert de technologie, industriel	Multi : scientifique, transfert, réutilisations multiples

(*) Voir Article vs. Logiciel : questions juridiques et de politique scientifique dans la production de logiciels, <https://www.gutenberg.org/files/44444/44444-h/44444-h.html>

Mots clés : partage des connaissances, reproductibilité, accessibilité, libre accès, ...

Aider les producteurs et les laboratoires à mieux gérer et diffuser les logiciels et les données contribuera à un meilleur fonctionnement de la recherche.

Quelques références

- Droit d'Accès aux Données, Bureau de Données et Données Ouvertes, Martin Gavetti (DIA, CNRS), juillet 2015, <http://www.dia.cnr.fr/rapport/201507>
- Science Publica: Open Access, Les enjeux de l'Open Access juin 2013, <http://www.sciencedirect.com/science/article/pii/S0926641013000112>
- Le patrimoine culturel et la production scientifique : enjeux juridiques, A.-L. Stern, L. Moreau, juin 2015, http://www.culture.gouv.fr/IMG/pdf/le_patrimoine_culturel_et_la_production_scientifique.pdf
- OMPF : une infrastructure de partage des données géographiques, Sans L'adresse (preprint) 2022, 31/10/22
- Données de Recherche, <https://www.dataforpeople.com/>
- Réseau RE2D, <http://re2d.com/>
- Réseau PLUME : publication logiciel d'un laboratoire, T. Gomez-Diaz, <https://www.gutenberg.org/files/44444/44444-h/44444-h.html>

 **Formation des réseaux de la documentation**
Années 2014-2015
Gestion et valorisation des données de la recherche
Aussis, 7 - 10 octobre 2015

(2013) TGD. Articles, logiciels, données : étude de la diffusion de la production scientifique, Journées FRéDoc, RENATIS

(2014) TGD. Articles, software, data : a study of the (French) scientific production, EUDAT

T. Gomez-Diaz (CNRS/LIGM)

Logiciels et données de la recherche

Paris, décembre 2024

14 / 30

Le logiciel est un objet juridique

(2011, 2015) TGD. Article vs. Logiciel : questions juridiques et de politique scientifique dans la production de logiciels

(2022) TGD, T. Recio. Research Software vs Research Data I (Definition)

- Code de la propriété intellectuelle (CPI), Article L. 112-2 :
un *logiciel* est une œuvre de l'esprit protégée par le droit d'auteur.
- Arrêté du Ministère de l'Industrie du 22 décembre 1981 (vocabulaire de l'informatique) :
un *logiciel* est un ensemble des programmes, procédés et règles, et éventuellement de la documentation, relatifs au fonctionnement d'un ensemble de traitement de données.
- Directive 2009/24/CE du Parlement européen et du Conseil du 23 avril 2009 concernant la protection juridique des programmes d'ordinateur :
les termes *programme d'ordinateur* visent les programmes sous quelque forme que ce soit, y compris ceux qui sont incorporés au matériel. Ces termes comprennent également les travaux préparatoires de conception aboutissant au développement d'un programme...

(2019) TGD, T. Recio. On the evaluation of research software : the CDUR procedure

Research software (RS) *is a well identified set of code that has been written by a well identified research team. It is software that has been built and used to produce a result published or disseminated in some article or scientific contribution.*

Each RS encloses a set of files containing the source code and the compiled code. It can also include other elements as the documentation, specifications, use cases...

Objectifs : données de la recherche

(2022) TGD, T. Recio. Research Software vs Research Data I (Definition)

(2022) TGD, T. Recio. Research Software vs Research Data II (Dissemination, CDUR)

Nous proposons pour les données de la recherche :

- Définition
- Réponses au “Borgman’s conundrum challenges”
- Procédure de diffusion
- Protocoles CDUR d’évaluation

(2012) Christine L. Borgman. The conundrum of sharing research data. Journal of the American Society for Inf. Sci. Technol.

“Borgman’s conundrum challenges” :

Le partage des données est un énigme [...]

Les défis consistent à comprendre quelles données peuvent être partagées, par qui, avec qui, dans quelles conditions, pourquoi et à quels effets.

Les réponses éclaireront la politique et la pratique des données.

Data sharing is thus a conundrum. [...]

The challenges are to understand which data might be shared, by whom, with whom, under what conditions, why, and to what effects. Answers will inform data policy and practice.

Plan

2 Préambule II : la Science ouverte

- Définition proposée de Science ouverte
- Recommandation de l'UNESCO sur une science ouverte

3 Préambule III : et pour les données ?

- Article vs. Logiciel : questions juridiques et de politique scientifique
- Articles, logiciels, données : étude de la diffusion de la production scientifique
- Le logiciel est un objet juridique
- Objectifs

4 Les données de la recherche

- Les données sont des objets juridiques ?
- Définition
- Borgman's conundrum challenges + Comment ? + Où ?
- Procédure de diffusion
- Protocole(s) **CDUR** d'évaluation de la recherche et RD
- Protocole(s) **CDUR** & FAIR Guiding Principles. CoARA.

5 Conclusions

Les données sont des objets juridiques ? (1/2)

- Arrêté du Ministère de l'Industrie du 22 décembre 1981 (vocabulaire de l'informatique) :
Donnée : Représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement.
- (2018, Stérin) *En droit, on ne sait pas ce qu'est une donnée* :
une donnée, en soi, ne relève pas d'un régime juridique spécifique
- (2018, Maurel) *Il conviendra donc de vérifier si les corpus de données à ouvrir ne correspondent pas à des œuvres protégées par des droits de propriété intellectuelle*
- (2011, de Cock BM, et al. Knowledge Exchange report)
It is important to know the legal status of the data to be shared. [...] not all data are protected by law, and not every use of protected research data requires the author's consent. [...] Whether data are in fact protected must be determined on a case-by-case basis.

Côté Science : Mais qu'est-ce qu'une donnée ?

- Discussion et références : (2022) TGD, T. Recio. Research Software vs Research Data I (Definition)
- (2017-20, Borgman) *Malgré ses cinq siècles d'existence, le terme data ou "donnée" n'a toujours pas trouvé une définition consensuelle [...] les données sont des représentations d'observations, d'objets ou d'autres entités qui servent à mettre en évidence des phénomènes à des fins de recherche.*
- (2022, Gualandi et al.) What do we mean by "data" ? (arts and humanities)
 - *I know what [data] are until someone asks me to define them.*
 - *It is a complicated question [...] I don't know if I'm answering it or rather asking more questions.*

Les données sont des objets juridiques? (2/2)

Analyser au **cas par cas**, en équipe, avec différents rôles, des experts nécessaires? :

- **Juridique.** Voir toutes les règles de droit applicables : droit d'auteur, *sui generis*, droit à l'image, données personnelles (RGPD), respect de la vie privée, ouverture et réutilisation des informations publiques (open data)...

Exemples de définitions sectorielles (2023, A. Robin, Printemps de la donnée) :

- Informations concernant la santé : Article L1111-7, [Code de la santé publique](#)
- Donnée géographique, Article L127-1, [Code de l'environnement](#)
- Information relative à l'environnement, Article L124-2, [Code de l'environnement](#)
- Règles relatives à la conservation, ... du patrimoine archéologique, [Code du patrimoine](#)

- **Scientifique.** Exemples :

- Température
- STRENDA guidelines : recommandations pour fournir des données sur les enzymes
<https://www.beilstein-strenda-db.org/strenda/public/guidelines.xhtml>
- TDWG: Biodiversity Information Standards - occurrence data and species distributions
<https://www.tdwg.org/>
- Tables du Lexique-Grammaire au LIGM
<https://infolingu.univ-mlv.fr/DonneesLinguistiques/Lexiques-Grammaires/Telechargement.html>
- ...

- **Technique.** Données collectées, traitées, analysées, partagées & diffusées

Définition de données de la recherche



(2022) TGD, T. Recio. Research Software vs Research Data I (Definition)

(2022) TGD, T. Recio. Research Software vs Research Data II (Dissemination, CDUR)

Research Data (RD) *is a well identified set of data that has been produced (collected, processed, analyzed, shared & disseminated) by a research team.*

The data has been collected, processed and analyzed to produce a result published or disseminated in some article or scientific contribution.

Each research data encloses a set (of files) that contains the dataset maybe organized as a database, and it can also include other elements as the documentation, specifications, use cases, and any other useful material as provenance information, instrument information...

It can include the research software that has been developed to manipulate the dataset (from short scripts to research software of larger size) or give the references to the software that is necessary to manipulate the data (RS or other).

Quatre caractéristiques principales :

- le but de la collecte et l'analyse est de faire de la recherche, répondre question scientifique
- il a été produit par une équipe de recherche
- obtention des résultats présentés dans des articles (ou autre contribution) scientifique
- des logiciels associés pour manipuler les données, qui peuvent être RS

Image: Archibald Tuttle, <https://commons.wikimedia.org/wiki/File:Wiki-puzzle-piece-runes-AT.jpg>
<https://andreashelley.com/blog/futhark-runes-symbols-and-meanings/>

Borgman's conundrum challenges + Comment? + Où?

Les réponses éclaireront la politique et les organismes de financement

(1997) National Research Council. Bits of Power: Issues in Global Access to Scientific Data

(2012) Christine L. Borgman. The conundrum of sharing research data.

(2022) TGD, T. Recio. Research Software vs Research Data I (Definition) + (2024) The *Conundrum Challenges* for RS

RD = un ensemble de données de la recherche produites par une équipe.

Quelles données partager? Décision de l'équipe de recherche, quel est le RD à partager, de quelle forme, et quand

Par qui? Par l'équipe qui a produit le RD, ie., qui a collecté, traité, analysé le RD, et a décidé de le partager et de le diffuser

Comment? En suivant une procédure de diffusion comme celle proposée ici, voir aussi (2024) DoRANum. Comment publier ses données de recherche?

Où? Recherche Data Gouv, Institution, thématique de recherche...

<https://recherche.data.gouv.fr/fr/guide/ou-deposer-et-publier-ses-donnees>,
Registry of Research Data Repository : <https://www.re3data.org/>

Avec qui? Tout acte de communication académique a un public cible, initialement ce public est le même que celui des articles associés au RD, mais... y a-t-il de la valeur interdisciplinaire?

(2012) *intended users may vary from researchers within a narrow specialty to the general public*

Dans quelles conditions? la licence établit les conditions du partage des RD

Pourquoi et à quels effets? en réponse aux demandes des financeurs, de l'institution, de l'éditeur d'une revue, de la Science ouverte, pour la validation et la reproductibilité...

(1997+2012) *The value of data lies in their use.*

(2013) Michael Arentoft (EC, DG Research & Innovation, Open Science), OS FAIR 2023,

<http://opensciencefair.eu/keynotes/co-creating-enablers-for-practicing-open-science-the-ec-perspective>

Procédure de diffusion pour les données de la recherche

(2010) TGD. Diffuser un logiciel de laboratoire : recommandations juridiques et administratives

(2014) TGD. Free software, Open source software, licenses. A short presentation including a procedure for RS and data diss...

(2022) TGD, T. Recio. Research Software vs Research Data II (Dissemination, CDUR)

- Choisir un nom pour le RD, éviter les noms déjà utilisés, les marques.

(2018) DoRANum or Harvard, Bien nommer ses fichiers/File Naming Conventions.

- (*) (research team step) Établir la liste des personnes qui ont participé à la production, leurs affiliations. Traiter les droits d'auteur, s'il y en a. Plan de gestion de données (PGD).

★ Traiter les autres contextes juridiques (ou éthiques), s'il y en a, (avec experts juridiques ?).

- (*) Établir la liste des briques logicielles ou les données utilisées, avec licences. Bonnes pratiques de citation pour les données et les logiciels ?

- 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..., comment citer l'œuvre, PIDs...

★ (*) (research work step) Faciliter la réutilisation : comme les données ont été utilisées pour vos recherches. La traçabilité est importante, archiver en .tar.gz régulièrement, garder une trace des articles... Réviser la documentation, PGD, site web...

- Informer la direction des laboratoires et les tutelles (si pas fait au point licence).

- Créer et indiquer une adresse courriel de contact.

- **Diffuser** les données.

- Informer la communauté cible, considérer les data papers.

(*) À revoir à chaque nouvelle version.

Protocole(s) **CDUR** : évaluation de la recherche et RD

(2019) TGD, T. Recio. On the evaluation of research software: the CDUR procedure

(2022) TGD, T. Recio. Research Software vs Research Data II (Dissemination, CDUR)

Le protocole **CDUR** est proposé pour prendre en compte les RS et/ou les RD dans l'évaluation. Il est conçu pour aider les chercheurs évalués, les comités d'évaluation, les décideurs.

(C) Citation mesure si RS bien identifié en tant que produit de la recherche : référence ou bien metadataset avec nom, producteurs, version, dates, PID,...
aussi bonnes pratiques pour citer les autres RD...
point légal : auteurs (si PI), producteurs, affiliations, participation

(D) Dissemination les bonnes pratiques de diffusion en fonction de la politique scientifique du contexte d'évaluation, PGDs, attention aux questions légales (et éthiques)...
point Science ouverte, ★ légal : autres questions légales, licences

★ **(U) Use** point dédié aux aspects purement "données" **du RD** : qualité, faciliter la réutilisation avec exemples des d'utilisation, documentation, des informations sur le logiciel utilisé et/ou développé...
point reproductibilité, validation des résultats obtenus

(R) Research point dédié aux aspects purement "recherche" : qualité du travail scientifique, algorithmes et structures de données proposés et programmés, publications, utilisateurs, collaborations et projets...
point impact de la recherche

Flexibilité d'application : chaque ensemble décideurs/comité d'évaluation établit **sa propre procédure**, adaptée aux objectifs fixés et au contexte d'évaluation.

Open Science FAIR (Madrid, septembre 2023)

How to achieve FAIRER Research Data... ?

(2016) M. Wilkinson et al. The FAIR Guiding Principles for scientific data management and stewardship

(2023) TGD, T. Recio. How to achieve FAIRER research data by studying evaluation & assessment protocols

Comparaison CDUR et FAIR Guiding principes :

CDUR [2,7]	The FAIR Guiding Principles [1]
(C) Citation: The RD is well identified, involving issues concerning: <ul style="list-style-type: none">- citation form or reference- metadata (including PID)	To be Findable: F1. metadata are assigned a globally unique and persistent identifier F2. data are described with rich metadata (defined by R1 below) F3. metadata clearly and explicitly include the identifier of the data it describes To be Interoperable: I3. metadata include qualified references to other metadata To be Reusable: R1.2. metadata are associated with detailed provenance
(D) Dissemination: The RD is well disseminated, involving issues concerning: <ul style="list-style-type: none">- list of included components- RD license- RD deposit	To be Findable: F4. metadata are registered or indexed in a searchable resource To be Accessible: A1. metadata are retrievable by their identifier using a standardized communications protocol A1.1. the protocol is open, free, and universally implementable A1.2. the protocol allows for an authentication and authorization procedure, where necessary A2. metadata are accessible, even when the data are no longer available To be Reusable: R1.1. metadata are released with a clear and accessible data usage license
(U) Use: The RD facilitates its reuse, involving issues like: <ul style="list-style-type: none">- documentation, tutorials, examples...- reproducibility and replicability issues	To be Interoperable: I1. metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation I2. metadata use vocabularies that follow FAIR principles To be Reusable: R1. metadata are richly described with a plurality of accurate and relevant attributes R1.3. metadata meet domain-relevant community standards
(R) Research: Measures the impact of the RD related scientific work	Not applicable

Comment évaluer la production de données et/ou de logiciels ?

Nous proposons des protocoles précis qui peuvent aider à construire ce nouveau système.

CoARA : Coalition for Advancing Research Assessment, <https://coara.eu/>

OSMI : Open Science Monitoring Initiative <https://open-science-monitoring.org/>

Ces travaux auront, à terme, des conséquences dans nos laboratoires et institutions.

Conclusions

Nous avons construit un cadre d'étude pour les logiciels de la recherche, pour les comprendre, les expliquer et pour valoriser leur contribution à la Science ouverte.

Il se construit en trois étapes : définition, diffusion, évaluation.

Nous avons proposé un cadre similaire pour les données de la recherche :

- Définition
- Réponses au "Borgman's conundrum challenges"
- Procédure de diffusion
- Protocoles CDUR d'évaluation

Nous avons étudié la complexité des données de la recherche, ce sont des objets qui peuvent être complexes du point de vue juridique, scientifique et technique.

Notre objectif est de porter des éléments de discussion pour contribuer à cet édifice, complexe et en construction, de la Science ouverte.

Conséquence : cela montre qu'une des forces de nos travaux sur les logiciels est leur adaptabilité à d'autres contextes (RS → RD, nanomatériaux [TRAAC, 2023]).

Références - Questions juridiques (1/4)

- CPI Code de la propriété intellectuelle (CPI)
https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEXT000006069414/
- 1981 Arrêté du Ministère de l'Industrie du 22 décembre 1981 relatif à l'enrichissement du vocabulaire de l'informatique,
<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000842020>
- 2009 Directive 2009/24/EC of the European Parliament & Council 23/04/2009
<https://eur-lex.europa.eu/eli/dir/2009/24/oj>
- 2011 de Cock BM, van Dinther B, Jeppesen de Boer CG, et al. The Legal Status of Research Data in the Knowledge Exchange Partner Countries, Knowledge Exchange report. <https://repository.jisc.ac.uk/6280/>
- 2013 C. Domange (Rapporteur). Guide Data Culture. Pour une stratégie numérique de diffusion et de réutilisation des données publiques numériques du secteur culturel. Ministère de la Culture et de la Communication, Secrétariat Général N. 2013-01. <https://www.enssib.fr/bibliotheque-numerique/documents/60533-guide-data-culture.pdf>
- 2018 L Maurel. La réutilisation des données de la recherche après la loi pour une République numérique. La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques. Presses Universitaires de Provence <https://hal.science/hal-01908766/>
- 2018 A-L Stérin. Diffuser des données de la recherche dans le respect du droit et de l'éthique : Comment faire lorsqu'on n'est pas juriste? Guide de bonnes pratiques éthiques et juridiques. Presses Universitaires de Provence <https://hal.science/hal-02050510/>
- 2022 Agnes Robin. Droit des données de la recherche. Science ouverte, innovation, données publiques. Ed. Larcier, <https://hal.umontpellier.fr/hal-03630680>
- 2023 Agnes Robin. La disponibilité juridique des données de la recherche : pistes et schémas d'analyse, Journée d'ouverture du Printemps de la donnée, <https://www.canal-u.tv/chaines/printempsdonnee/edition-2023/la-disponibilite-juridique-des-donnees-de-la-recherche-pistes>
- 2023 A. Bassinet, F. Bouchet Moneret. Connaître le droit des données de la recherche, Love Data week, <https://www.enssib.fr/bibliotheque-numerique/documents/72124-connaître-le-droit-des-donnees-de-la-recherche.pdf>

Références (2/4)

- 1997 National Research Council. Bits of Power: Issues in Global Access to Scientific Data. Washington, DC: The National Academies Press. <https://doi.org/10.17226/5504>
- 2012 Alma Swan. Policy Guidelines for the Development and Promotion of Open Access, UNESCO, <http://unesdoc.unesco.org/images/0021/002158/215863e.pdf>
- 2012 Christine L. Borgman. The conundrum of sharing research data. Journal of the American Society for Information Science and Technology, <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.22634>
- 2016 M. Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship, <https://doi.org/10.1038/sdata.2016.18>
- 2017-20 Christine L. Borgman. Qu'est-ce que le travail scientifique des données ? Big data, little data, no data. OpenEdition Press, <https://books.openedition.org/oep/14692>.
English version : Big Data, Little Data, No Data : Scholarship in the Networked World. MIT Press, 2017.
- 2018 J.P. Tennant. "Open Science is just good science", DARIAH Annual Event, <https://www.youtube.com/watch?v=UEEcwRUgQu8>
- 2018 J.P. Tennant. The state of the art in peer review. FEMS Microbiology Letters, <https://doi.org/10.1093/femsle/fny204>
- 2018 DoRANum. Données de la recherche : apprentissage numérique. Comment bien nommer ses fichiers ? https://doranum.fr/stockage-archivage/comment-nommer-fichiers_10_13143_wgqw-aa59/. Or Harvard. "File Naming Conventions", <https://datamanagement.hms.harvard.edu/plan-design/file-naming-conventions>
- 2022 Gualandi, B., Pareschi, L., Peroni, S. What do we mean by "data" ? A proposed classification of data types in the arts and humanities. Journal of Documentation, <https://doi.org/10.1108/JD-07-2022-0146>
- 2023 Shandilya, N., Barreau, M. S., Suarez-Merino B., et al. TRAAC framework to improve regulatory acceptance and wider usability of tools and methods for safe innovation and sustainability of manufactured nanomaterials, Nanolmpact, Volume 30, April 2023, <https://www.sciencedirect.com/science/article/pii/S2452074823000125>
- 2023 Michael Arentoft, Keynote Speech, Co-creating enablers for practicing open science - the EC perspective, Open Science Fair 2023, (from 1:07:00), <https://www.youtube.com/watch?v=3JcX01iho8Y>
- 2024 DoRANum. Comment publier ses données de recherche ? https://doranum.fr/depot-entrepots/comment-publier-ses-donnees-de-recherche_10_13143_ajk3-ea67/

- 2007 TGD. Autour de la valorisation de logiciels développés dans un laboratoire de recherche, LIGM.
- 2009 JL Archimbaud, TGD. Licence & copyright pour les développements de logiciels libres de laboratoires de recherche, PLUME, <https://zenodo.org/record/7063146>
- 2010 TGD. Diffuser un logiciel de laboratoire : recommandations juridiques et administratives, PLUME, <https://zenodo.org/record/7096216>
- 2011 TGD. Article vs. Logiciel : questions juridiques et de politique scientifique dans la production de logiciels, PLUME, <https://zenodo.org/record/7063154>
- 2013 TGD. Articles, logiciels, données : étude de la diffusion de la production scientifique, Poster, Journées FRéDoc 2013, http://igm.univ-mlv.fr/~teresa/logicielsLIGM/documents/JourneesDR/2013octFREDOC_A0_Portrait.pdf
- 2014 TGD. Articles, software, data: a study of the (French) scientific production, Poster, EUDAT 2014, http://igm.univ-mlv.fr/~teresa/logicielsLIGM/documents/Internacional/2014septeutad_70x100.pdf
- 2014 TGD. Free software, Open source software, licenses. A short presentation including a procedure for research software and data dissemination, Zenodo, <https://zenodo.org/record/11709>
- 2015 TGD. Article vs. Logiciel : questions juridiques et de politique scientifique dans la production de logiciels, Société Informatique de France, <https://zenodo.org/record/18993>
- 2018 TGD, G. Romier. Modèle de Plan de Gestion de Logiciel de la Recherche (PGLR), V3.2, PRESOFT, <https://zenodo.org/record/1405614>
- 2019 TGD. Le Projet PLUME et le paysage actuel des logiciels de la recherche dans la science ouverte, Zenodo, <https://zenodo.org/record/2591474>
- 2019 TGD. Les logiciels de la recherche et leurs licences : trois visions sur un objet, Cours CERFACS, <https://hal.science/hal-02434287v1>
- 2019 TGD, T. Recio. On the evaluation of research software: the CDUR procedure, F1000Research, Research on Research, <https://doi.org/10.12688/f1000research.19994.2>
- 2020-1 TGD, T. Recio. Towards an Open Science definition as a political and legal framework: on the sharing and dissemination of research outputs, **POLIS N. 19, 2020**, <https://doi.org/10.58944/yuro5734>, **V3 du 28/02/2021**, <https://zenodo.org/record/4577066>
- 2021 TGD. Sur la production de logiciels libres au Laboratoire d'informatique Gaspard-Monge (LIGM) : ce que nous avons appris, Atelier BlueHats (DINUM), <https://zenodo.org/records/10970141>

- 2021 TGD, T. Recio. Open comments on the Task Force SIRS report: Scholarly Infrastructures for Research Software (EOSC Executive Board, EOSCArchitecture), RIO 7:e63872, <https://doi.org/10.3897/rio.7.e63872>
- 2021 TGD, T. Recio. The future of Open Science asks for a common understanding, Poster, EGI Virtual Conference 2021, <https://indico.egi.eu/event/5464/contributions/15729/>
- 2022 TGD, T. Recio. Research Software vs Research Data I: Towards a Research Data definition in the Open Science context, F1000Research, Research on Research, <https://f1000research.com/articles/11-118/v2>
- 2022 TGD, T. Recio. Research Software vs Research Data II: Protocols for Research Data dissemination and evaluation in the Open Science context, F1000Research, Research on Research, <https://f1000research.com/articles/11-117/v2>
- 2022 TGD, T. Recio. Research Software and Research Data : dissemination, evaluation and reusability in the Open Science context, Poster, IDCC22, <https://zenodo.org/record/6778872>
- 2023 TGD, T. Recio. Podcast Code for Thought : Research Software and Research Data in Open Science, Zenodo, <https://zenodo.org/records/8159906>
- 2023 TGD, T. Recio. How to achieve FAIRER research data by studying evaluation & assessment protocols, Poster, Open Science FAIR 2023, <https://zenodo.org/records/8398431>
- 2023 TGD, T. Recio. Articles, software, data: An Open Science ethological study, Maple Transactions, <https://doi.org/10.5206/mt.v3i14.17132>
- 2024 TGD, T. Recio. Open comments on the "Open Letter : Establishing a national research software award", Zenodo, <https://zenodo.org/records/10493797>
- 2024 TGD, T. Recio. Open comments for the NIH Request for Information (RFI): Best Practices for Sharing NIH Supported Research Software (NIH, 2023), Zenodo, <https://zenodo.org/records/10617585>
- 2024 TGD. On the sharing and dissemination of Research Software and Research Data in the Open Science context, Open Science Days 2024, Max Planck Digital Library, <https://zenodo.org/records/10617691>
- 2024 TGD. Les logiciels de la recherche et leurs licences : trois visions sur un objet, Cours Université Gustave Eiffel, <https://hal.science/hal-02434287v3>
- 2024 TGD, T. Recio. The *Conundrum Challenges* for Research Software in Open Science Computers 2024, 13(11), 302, <https://doi.org/10.3390/computers13110302>



Et voici une sélection de podcasts, réalisés par Peter Schmidt :

- 18/07/2022 Open Science and Research Software, avec Teresa Gomez-Diaz, <https://codeforthought.buzzsprout.com/1326658/episodes/10822132-open-science-and-research-software>
- 31/01/2023 Open Access, Open Knowledge, avec Jean-Claude Guédon, <https://codeforthought.buzzsprout.com/1326658/episodes/12129577-open-access-open-knowledge>
- 18/07/2023 [EN] Open Data, Open Software - with Teresa Gomez-Diaz, <https://codeforthought.buzzsprout.com/1326658/episodes/13216530-en-open-data-open-software-with-teresa-gomez-diaz>
- Et le document associé :
- TGD, T. Recio, Podcast Code for Thought : Research Software and Research Data in Open Science, Zenodo, 18/07/2023, <https://zenodo.org/records/8159906>
- 30/07/2024 [EN] Planning your research software - a workshop in Paris, <https://codeforthought.buzzsprout.com/1326658/episodes/15479406-en-planning-your-research-software-a-workshop-in-paris>
- 10/09/2024 [FR] L'apprentissage automatique et la reproductibilité - avec Pascal Monasse (LIGM), <https://codeforthought.buzzsprout.com/1326658/episodes/15649531-fr-l-apprentissage-automatique-et-la-reproductibilite-avec-pascal-monasse-ligm>

Voir : <https://codeforthought.buzzsprout.com/>