On the sharing and dissemination of Research Software and Research Data in the Open Science context

Teresa Gomez-Diaz

CNRS - Laboratoire d'informatique Gaspard-Monge

Includes work in collaboration with Prof. T. Recio (Univ. Nebrija, Madrid)

Open Science Days,

Max Planck Digital Library, Berlin, 29th January 2024

Goal

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

The goal of this talk is to present our work on Research Data, where we propose:

- Definition
- Dissemination procedure
- Evaluation protocols: CDUR

We will start by setting the Open Science context (preamble I),
following with our work on Research Software (preamble II),
and we will show part of the pathway (preamble III),
that drives us to build these propositions for Research Data.

Work in collaboration with Prof. Tomas Recio, University Antonio de Nebrija, Madrid (http://www.recio.tk/).

Plan

Open Science: proposed definition - 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

Open Science is...

> [EN] *the political and legal framework where research outputs are shared and disseminated in order to be rendered visible, accessible and reusable*
>
> [FR] *le cadre politique et juridique dans lequel les productions scientifiques sont partagées et diffusées afin d'être rendues visibles, accessibles et réutilisables.*

Version	**Title**	**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 With ref.: 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
V1	A policy and legal Open Science...	09/2020	https://zenodo.org/record/4075106

Goal: to understand what Open Science is, the motivation for the proposition of this definition, but also to structure information in a complex landscape.

Open Science: structuring the landscape - poster

The future of Open Science asks for a common understanding

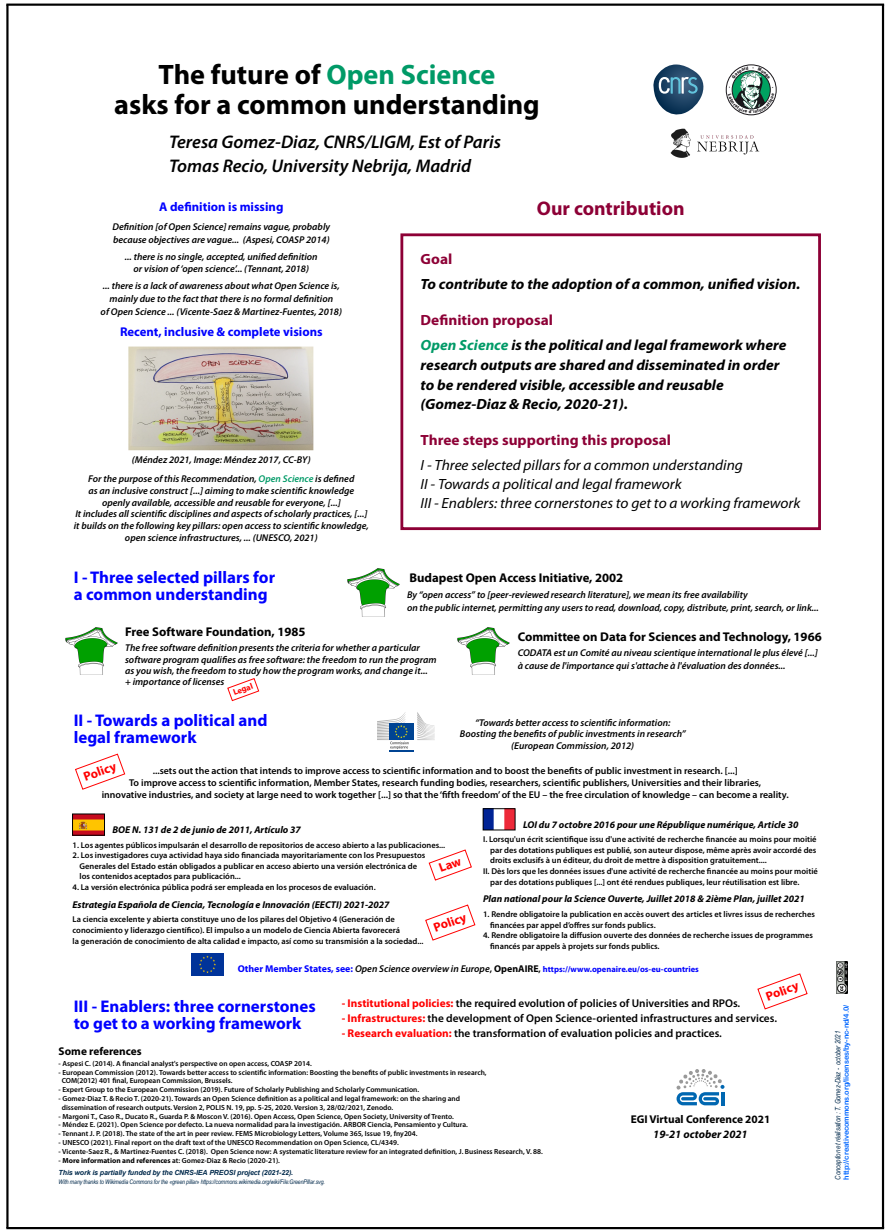
Open Science is
the political and legal framework
where research outputs are shared
and disseminated in order to be rendered
visible, accessible and reusable.

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



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... (Aspesi, COASP 2014)

... there is no single, accepted, unified definition or vision of 'open science'... (Tennant, 2018)

... there is a lack of awareness about what Open Science is, mainly due to the fact that there is no formal definition of Open Science ... (Vicente-Saez & Martinez-Fuentes, 2018)

Recent, inclusive & complete visions

(Méndez 2021, Image: Méndez 2017, CC-BY)

For the purpose of this Recommendation, Open Science is defined as an inclusive construct [...] aiming to make scientific knowledge openly available, accessible and reusable for everyone, [...] It includes all scientific disciplines and aspects of scholarly practices, [...] it builds on the following key pillars: open access to scientific knowledge, open science infrastructures, ... (UNESCO, 2021)

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

Budapest Open Access Initiative, 2002
By "open access" to [peer-reviewed research literature], we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link...

Free Software Foundation, 1985
The free software definition presents the criteria for whether a particular software program qualifies as free software: the freedom to run the program as you wish, the freedom to study how the program works, and change it... + importance of licenses
Legal

Committee on Data for Sciences and Technology, 1966
CODATA est un Comité au niveau scientique international le plus élevé [...] à cause de l'importance qui s'attache à l'évaluation des données...

II - Towards a political and legal framework

"Towards better access to scientific information: Boosting the benefits of public investments in research" (European Commission, 2012)

Policy

...sets out the action that intends to improve access to scientific information and to boost the benefits of public investment in research. [...] To improve access to scientific information, Member States, research funding bodies, researchers, scientific publishers, Universities and their libraries, innovative industries, and society at large need to work together [...] so that the 'fifth freedom' of the EU – the free circulation of knowledge – can become a reality.

BOE N. 131 de 2 de junio de 2011, Artículo 37

1. Los agentes públicos impulsarán el desarrollo de repositorios de acceso abierto a las publicaciones...
2. Los investigadores cuya actividad haya sido financiada mayoritariamente con los Presupuestos Generales del Estado están obligados a publicar en acceso abierto una versión electrónica de los contenidos aceptados para publicación...
4. La versión electrónica pública podrá ser empleada en los procesos de evaluación.

Law

LOI du 7 octobre 2016 pour une République numérique, Article 30

I. Lorsqu'un écrit scientifique issu d'une activité de recherche financée au moins pour moitié par des dotations publiques est publié, son auteur dispose, même après avoir accordé des droits exclusifs à un éditeur, du droit de mettre à disposition gratuitement....
II. Dès lors que les données issues d'une activité de recherche financée au moins pour moitié par des dotations publiques [...] ont été rendues publiques, leur réutilisation est libre.

Estrategia Española de Ciencia, Tecnología e Innovación (EECTI) 2021-2027

La ciencia excelente y abierta constituye uno de los pilares del Objetivo 4 (Generación de conocimiento y liderazgo científico). El impulso a un modelo de Ciencia Abierta favorecerá la generación de conocimiento de alta calidad e impacto, así como su transmisión a la sociedad...

Policy

Plan national pour la Science Ouverte, Juillet 2018 & 2ième Plan, juillet 2021

1. Rendre obligatoire la publication en accès ouvert des articles et livres issus de recherches financées par appel d'offres sur fonds publics.
4. Rendre obligatoire la diffusion ouverte des données de recherche issues de programmes financés par appels à projets sur fonds publics.

Other Member States, see: *Open Science overview in Europe*, **OpenAIRE**, https://www.openaire.eu/os-eu-countries

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

- **Institutional policies:** the required evolution 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.

Policy

Some references

- Aspesi C. (2014). A financial analyst's perspective on open access, COASP 2014.
- European Commission (2012). Towards better access to scientific information: Boosting the benefits of public investments in research, COM(2012) 401 final, European Commission, Brussels.
- Expert Group to the European Commission (2019). Future of Scholarly Publishing and Scholarly Communication.
- Gomez-Diaz T. & Recio T. (2020-21). Towards an Open Science definition as a political and legal framework: on the sharing and dissemination of research outputs. Version 2, POLIS N. 19, pp. 5-25, 2020. Version 3, 28/02/2021, Zenodo.
- Mangoni T., Caso R., Ducato R., Guarda P. & Moscon V. (2016). Open Access, Open Science, Open Society, University of Trento.
- Méndez E. (2021). Open Science por defecto. La nueva normalidad para la investigación. ARBOR Ciencia, Pensamiento y Cultura.
- Tennant J. P. (2018). The state of the art in peer review. FEMS Microbiology Letters, Volume 365, Issue 19, fny204.
- UNESCO (2021). Final report on the draft text of the UNESCO Recommendation on Open Science, CL/4349.
- Vicente-Saez R., & Martinez-Fuentes C. (2018). Open Science now: A systematic literature review for an integrated definition, J. Business Research, V. 88.

- **More information and references at:** Gomez-Diaz & Recio (2020-21).

This work is partially funded by the CNRS-IEA PRECOS project (2021-22).

With many thanks to Wikimedia Commons for the «green pillar» https://commons.wikimedia.org/wiki/File:GreenPillar.svg

EGI Virtual Conference 2021
19-21 october 2021

UNESCO Recommendation on Open Science

- 40th session, 11/2019
- Members States asked for an international standard-setting instrument
- 2020: awareness, consequence of the COVID-19 pandemic
- Large consultative process, global and regional consultations, preliminary report
  https://unesdoc.unesco.org/ark:/48223/pf0000374409
- 41st session, 23/11/2021 - final UNESCO Recommendation on Open Science
  https://unesdoc.unesco.org/ark:/48223/pf0000379949
- Unanimously adopted by the Member States (193 countries)

For the purpose of this Recommendation, **open science** is defined as an **inclusive construct** that combines various movements and practices **aiming to make** multilingual **scientific knowledge openly available, accessible and reusable for everyone**, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community.

It comprises **all scientific disciplines and aspects of scholarly practices, including basic and applied sciences, natural and social sciences and the humanities**, and it builds on the following key pillars: open scientific knowledge, open science infrastructures, science communication, open engagement of societal actors and open dialogue with other knowledge systems.

Followed by Open Science working groups,
https://www.unesco.org/en/open-science/implementation#open-science-working-groups

Plan

Research Software definition

(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.

- goal: to do research
- a member of the lab participates to the code writing (similar to publications)
- the main production is the publication, software is an associated object

(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...

Research Software dissemination procedure

Easy to adapt to many situations, also valid for (research) data.

- Choose a name, avoid trademarks and proprietary names, associate date, version...
  (2018) DoRANum or Harvard. File Naming Conventions
- (*) (research team step) Establish list of authors/contributors (% participation, affiliations). Consider a Software Management Plan (2018, TGD, G. Romier, PGLR V3.2, PRESOFT)
- (*) Establish the list of main functionalities.
- (*) Establish the list of included software & data components, their licenses
- **Choose a license**, have an agreement (signed) with rightholders and authors, consider FLOSS licenses. Beware of license compatibility and inheritance issues.
- Choose a website, forge, deposit for dissemination, indicate licenses and how to cite the work. Use PIDs if possible.
- (*) Archive a tar.gz or similar regularly to keep track of added functionalities.
- Inform your laboratories and head institutions (if not done in the license step).
- Set and indicate clearly a contact address.
- **Distribute** your (research) software or data component.
- Inform the target scientific community. Consider Software or Data papers...

(*) To be reviewed with each new version.

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

Protocol(s) **CDUR**: research evaluation and 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)

Designed to help evaluated researchers, evaluation committees, decision makers...
also valid for (research) data.

(C) Citation measure if RS is well identified as a research output:
good citation form, but also metadata, best citation practices...
legal point: authors, affiliations, participation %

(D) Dissemination best dissemination practices, in agreement with
the scientific policy of the evaluation context
policy point: Open Science, **legal point:** licenses

(U) Use "software" aspects **of RS**: correct results, facilitate reuse, good softw.
practices: doc, test, install, up to read the code, launch RS...
point reproducibility: validation of scientific results

(R) Research "research aspects": quality of the scientific work, proposed and coded
algorithms & data structures, related publications, collaborations...
point research: impact

Flexibility of application: each decision maker or evaluation committee **sets its own CDUR protocol** adapted to the evaluation context and goals.

Plan

Article vs. Logiciel : questions juridiques et de politique scientifique

Article vs. Software: legal and scientific policy issues in software production

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

To understand and to explain legal and scientific policy issues related to Research Software, by comparison with the main production of Research Articles.

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

All references given at the end address issues related to legal and scientific policy issues.

Articles, software, data: study of the dissemination of the scientific production

Scientific policy issues in the software and data production

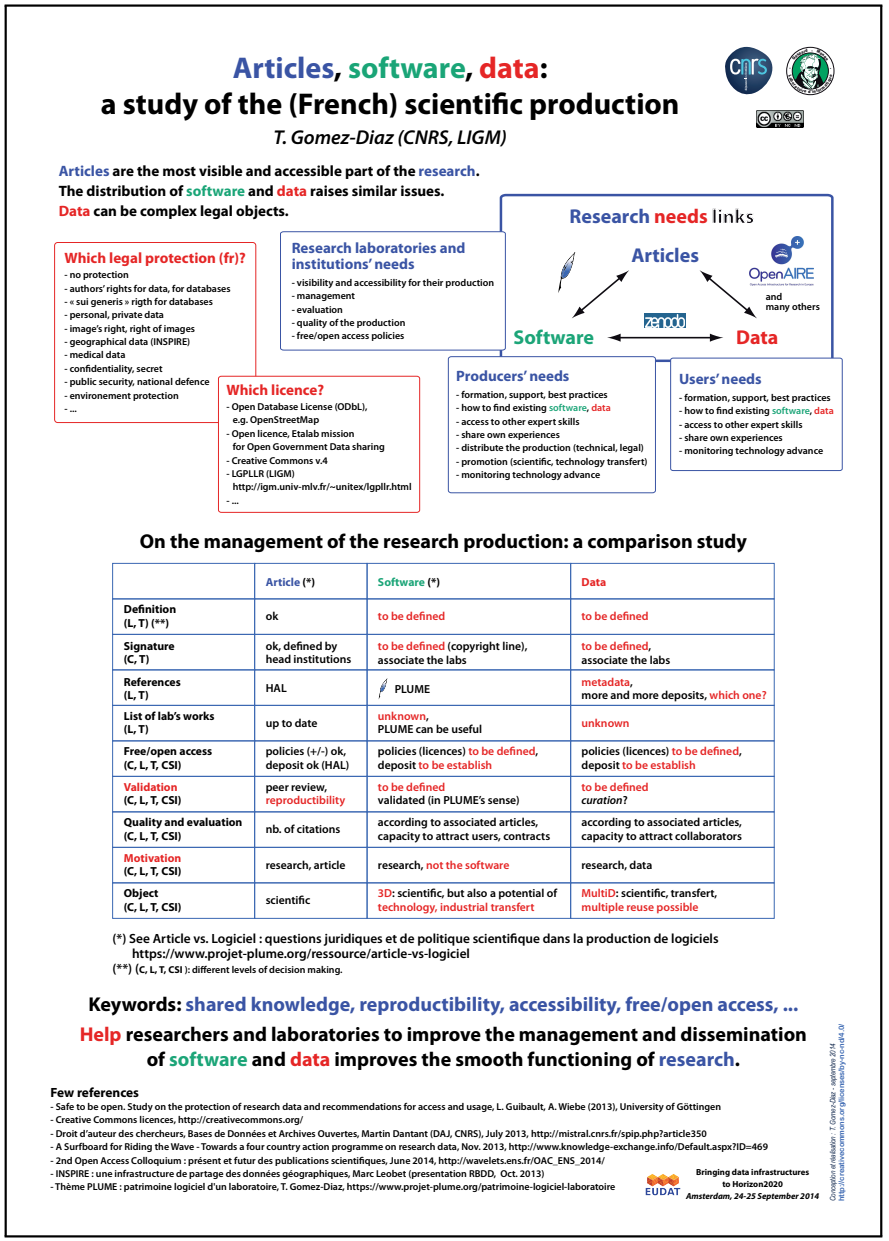
Three pillars in the scientific production:

- Articles
- Software
- Data

- Articles are the most visible and accessible part of research.
- Data can be complex legal objects.
- The disemination of software and data raises similar decision-making questions (and at the same time).

(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



Articles, software, data:
a study of the (French) scientific production

T. Gomez-Diaz (CNRS, LIGM)

Articles are the most visible and accessible part of the **research**.
The distribution of **software** and **data** raises similar issues.
Data can be complex legal objects.

Which legal protection (fr)?
- no protection
- authors' rights for data, for databases
- « sui generis » rigth for databases
- personal, private data
- image's right, right of images
- geographical data (INSPIRE)
- medical data
- confidentiality, secret
- public security, national defence
- environement protection
- ...

Research laboratories and institutions' needs
- visibility and accessibility for their production
- management
- evaluation
- quality of the production
- free/open access policies

Research needs links

Articles — Software — Data

OpenAIRE and many others

Which licence?
- Open Database License (ODbL), e.g. OpenStreetMap
- Open licence, Etalab mission for Open Government Data sharing
- Creative Commons v.4
- LGPLLR (LIGM) http://igm.univ-mlv.fr/~unitex/lgpllr.html
- ...

Producers' needs
- formation, support, best practices
- how to find existing software, data
- access to other expert skills
- share own experiences
- distribute the production (technical, legal)
- promotion (scientific, technology transfert)
- monitoring technology advance

Users' needs
- formation, support, best practices
- how to find existing software, data
- access to other expert skills
- share own experiences
- monitoring technology advance

On the management of the research production: a comparison study

	Article (*)	Software (*)	Data
Definition (L, T) (**)	ok	to be defined	to be defined
Signature (C, T)	ok, defined by head institutions	to be defined (copyright line), associate the labs	to be defined, associate the labs
References (L, T)	HAL	PLUME	metadata, more and more deposits, which one?
List of lab's works (L, T)	up to date	unknown, PLUME can be useful	unknown
Free/open access (C, L, T, CSI)	policies (+/-) ok, deposit ok (HAL)	policies (licences) to be defined, deposit to be establish	policies (licences) to be defined, deposit to be establish
Validation (C, L, T, CSI)	peer review, reproductibility	to be defined validated (in PLUME's sense)	to be defined *curation*?
Quality and evaluation (C, L, T, CSI)	nb. of citations	according to associated articles, capacity to attract users, contracts	according to associated articles, capacity to attract collaborators
Motivation (C, L, T, CSI)	research, article	research, not the software	research, data
Object (C, L, T, CSI)	scientific	3D: scientific, but also a potential of technology, industrial transfert	MultiD: scientific, transfert, multiple reuse possible

(*) See Article vs. Logiciel : questions juridiques et de politique scientifique dans la production de logiciels https://www.projet-plume.org/ressource/article-vs-logiciel
(**) (C, L, T, CSI): different levels of decision making.

Keywords: shared knowledge, reproductibility, accessibility, free/open access, ...

Help researchers and laboratories to improve the management and dissemination of software and data improves the smooth functioning of research.

Few references
- Safe to be open. Study on the protection of research data and recommendations for access and usage, L. Guibault, A. Wiebe (2013), University of Göttingen
- Creative Commons licences, http://creativecommons.org/
- Droit d'auteur des chercheurs, Bases de Données et Archives Ouvertes, Martin Dantant (DAJ, CNRS), July 2013, http://mistral.cnrs.fr/spip.php?article350
- A Surfboard for Riding the Wave - Towards a four country action programme on research data, Nov. 2013, http://www.knowledge-exchange.info/Default.aspx?ID=469
- 2nd Open Access Colloquium : présent et futur des publications scientifiques, June 2014, http://wavelets.ens.fr/OAC_ENS_2014/
- INSPIRE : une infrastructure de partage des données géographiques, Marc Leobet (presentation RBDD, Oct. 2013)
- Thème PLUME : patrimoine logiciel d'un laboratoire, T. Gomez-Diaz, https://www.projet-plume.org/patrimoine-logiciel-laboratoire

EUDAT — Bringing data infrastructures to Horizon2020 — *Amsterdam, 24-25 September 2014*

Software is a legal object

(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)

- [FR] Code de la propriété intellectuelle (CPI), Article L. 112-2 :
  un *logiciel* est une œuvre de l'esprit protégée par le droit d'auteur.
- [FR] 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.
- [EC] Directive 2009/24/EC, 23/04/2009, on the legal protection of computer programs:

  For the purpose of this Directive, the term *computer program* shall include programs in any form, including those which are incorporated into hardware.

  This term also includes *preparatory design work* leading to the development of a computer program provided that the nature of the preparatory work is such that a computer program can result from it at a later stage.

(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...

Data as a legal object?

(2011) The Knowledge Exchange report. The Legal Status of Research Data in the Knowledge Exchange Partner Countries

(2018) A-L Stérin. Diffuser des données de la recherche dans le respect du droit et de l'éthique

(Stérin) *En droit, on ne sait pas ce qu'est une donnée :*

une donnée, en soi, ne relève pas d'un régime juridique spécifique.

data, in itself, does not fall under any specific legal regime.

(K. Exchange) *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.

Analyze on a case-by-case basis, in a team (roles), experts are maybe necessary:

- legal point of view, see all applicable legal rules: copyright, sui generis, image rights, personal data, respect for privacy, reuse of public information...
- scientific point of view, examples: temperature, STRENDA guidelines (for describing investigations of enzyme activities), Lexicon-Grammar tables at LIGM, environmental information, geographical information...
- technical point of view, data collected, processed, analyzed, shared & disseminated

See: A. Robin, (2022) Droit des données de la recherche, (2023) Journée d'ouverture du Printemps de la donnée

Plan

Goals for Research Data

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

We propose for Research Data:

- Definition
- Answers to "Borgman's conundrum challenges"
- Dissemination procedure
- Protocols CDUR for Research Data evaluation

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

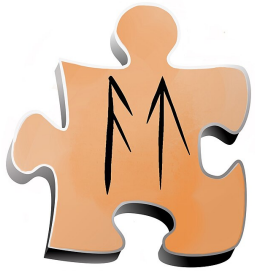
"Borgman's conundrum challenges":

> *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.*

Definition for Research Data

(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).

Four main characteristics:

- the purpose of collection and analysis is to do research, to answer scientific questions
- it has been produced by a research team
- the produced research results are presented in scientific articles (or other contributions)
- data has associated software for its manipulation, which can also be Research Software

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 + how? + where?

Answers inform policy makers and research funders

(2012) Christine L. Borgman. The conundrum of sharing research data
(2022) TGD, T. Recio. Research Software vs Research Data I (Definition)

Which data to be shared? Team decision on the RD to be shared, in which form, and when

By whom? The RD production team who has collected, processed, analyzed, and decides to share & disseminate the RD

How? Following a dissemination procedure (see next page)

Where? Many repositories available, see for example
Registry of Research Data Repository, https://www.re3data.org/

With whom? Each scientific communication act has its own target public, initially it can be the one of the associated publications, but maybe there is some interdisciplinary value, so...

intended users may vary from researchers within a narrow specialty to the general public

Under what conditions? the license gives the RD sharing conditions

Why, and to what effects? to answer funding demands (Institution, project funding...), to follow Open Sciences policies and/or best practices, for validation and reproduction of published results...

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

Research Data dissemination procedure

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

- Choose a name for the RD, avoid trademarks and proprietary names, associate date, version... (2018) DoRANum or Harvard. File Naming Conventions
- (*) (research team step) Establish list of contributors (maybe with % of participation) + affiliations. Deal with author rights if any. Consider a Data Management Plan (DMP).
- ★ Deal with other legal (or ethical) issues, if any, (with experts?).
- (*) Establish the list of software and data components, their licenses.
  Best citation practices.
- **Choose a license**, have an agreement (signed) with rightholders and authors. Beware of license compatibility and inheritance issues.
- Choose a website, forge, deposit for dissemination, indicate licenses and how to cite the work. Use PIDs if possible.
- ★ (*) (research work step) Facilitate reuse: how the RD was used for your research.
  Traceability is important, archive in .tar.gz regularly, keep track of related publications...
  Review documentation, PGD, website...
- Inform your laboratories and head institutions (if not done in the license step).
- Set and indicate clearly a contact address.
- **Distribute** your Research Data.
- Inform the target scientific community. Consider Data papers...

(*) To be reviewed with each new version.

Protocol(s) **CDUR**: research evaluation and 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)

Designed to help evaluated researchers, evaluation committees, decision makers...

(C) Citation measure if the RD is well identified as a research output:
good citation form, but also metadata, best citation practices...
legal point: authors (if $\exists$ copyright) and/or contributors, affiliations, participation %

(D) Dissemination best dissemination practices, in agreement with the scientific policy of the evaluation context, DMPs, legal (and ethical) issues
policy point: Open Science, $\star$ **legal point:** other legal issues, licenses

$\star$ **(U) Use** "data" aspects **of the RD**: quality, facilitate reuse, documentation, use examples... and information on needed software (maybe RS)
point reproducibility: validation of scientific results

(R) Research "research aspects": quality of the scientific work, proposed and coded algorithms & data structures, related publications, collaborations...
point research: impact

Flexibility of application: each decision maker or evaluation committee **sets its own CDUR protocol** adapted to the evaluation context and goals.

Conclusions

We have built a framework to study Research Software, to understand and to explain them, and to promote their contribution to Open Science.

It is constructed in three stages: definition, dissemination, evaluation.

We have proposed a similar framework for Research Data:

- Definition
- Answers to "Borgman's conundrum challenges"
- Dissemination procedure
- CDUR evaluation protocols

We have studied the complexity of Research Data, these objects can be complex from a legal, scientific and/or technical point of view.

Our objective is to bring elements of discussion to contribute to this complex and under construction building of Open Science.

As a consequence, one of the strengths of our work on Research Software is **its adaptability to other contexts** (RS $\rightarrow$ RD, nanomaterials (2023 TRAAC))

With my sincere thanks...

- to Tomas Recio, http://www.recio.tk/
- to Martin Boosen and the Open Science Days Team,

  https://osd.mpdl.mpg.de/
- to my lab, the LIGM, https://siteigm.univ-mlv.fr/
- to the MATESCO lab, for kind hospitality,

  https://web.unican.es/departamentos/matesco
- to Peter Schmidt, for Code for Thought podcasts

  (2021, 2023) P. Schmidt, (2023) TGD, T. Recio, https://codeforthought.buzzsprout.com/
- to You, listening here...

References (1/3)

2011 de Cock BM, van Dinther B, Jeppersen 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/

2012 Christine L. Borgman. The conundrum of sharing research data, Journal of the American Society for Information Science and Technology, vol. 63, no 6, p. 1059-1078. https://zenodo.org/record/7063154

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, Sécretariat Général N. 2013-01, https://www.enssib.fr/bibliotheque-numerique/documents/60533-guide-data-culture.pdf

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/

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/
See also "File Naming Conventions", Harvard, https://datamanagement.hms.harvard.edu/plan-design/file-naming-conventions

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 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, NanoImpact, 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

References (2/3)

2007 TGD. Autour de la valorisation de logiciels développés dans un laboratoire de recherche, LIGM.

2009 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/2014septeudat_70x100.pdf

2014 TGD. Free software, Open source software, licenses. A short presentation including a procedure for research software and data dissemination, Zenodo preprint, 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, document bilingue (FR/EN), https://zenodo.org/record/1405614

2019 TGD. Le Projet PLUME et le paysage actuel des logiciels de la recherche dans la science ouverte, Zenodo preprint, https://zenodo.org/record/2591474

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 & dissemination of research outputs, **POLIS N. 19, 2020**, https://doi.org/10.58944/yuro5734, **V3 du 28/02/2021**, https://zenodo.org/record/4577066

References (3/3)

2021 TGD. Free/Open source Research Software production at the Gaspard-Monge Computer Science laboratory. Lessons learnt, FOSDEM21, https://archive.fosdem.org/2021/schedule/event/open_research_gaspard_monge/

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 P. Schmidt. Podcast Code for Thought: Open Science and Research Software, https://codeforthought.buzzsprout.com/1326658/10822132-open-science-and-research-software

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. Les logiciels de la recherche et leurs licences : trois visions sur un objet, Training support, Gustave Eiffel University, https://hal.science/hal-02434287v2

2023 P. Schmidt. Podcast Code for Thought: Open Data, Open Software - with Teresa Gomez-Diaz, https://codeforthought.buzzsprout.com/1326658/13216530-en-open-data-open-software-with-teresa-gomez-diaz

2023 TGD, T. Recio. Podcast Code for Thought: Research Software and Research Data in Open Science, Zenodo preprint, https://zenodo.org/record/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.v3i4.17132