



# Open Science et données de la recherche

Odile Hologne, Déléguée à l'information scientifique et technique @Holo\_08



# Agenda

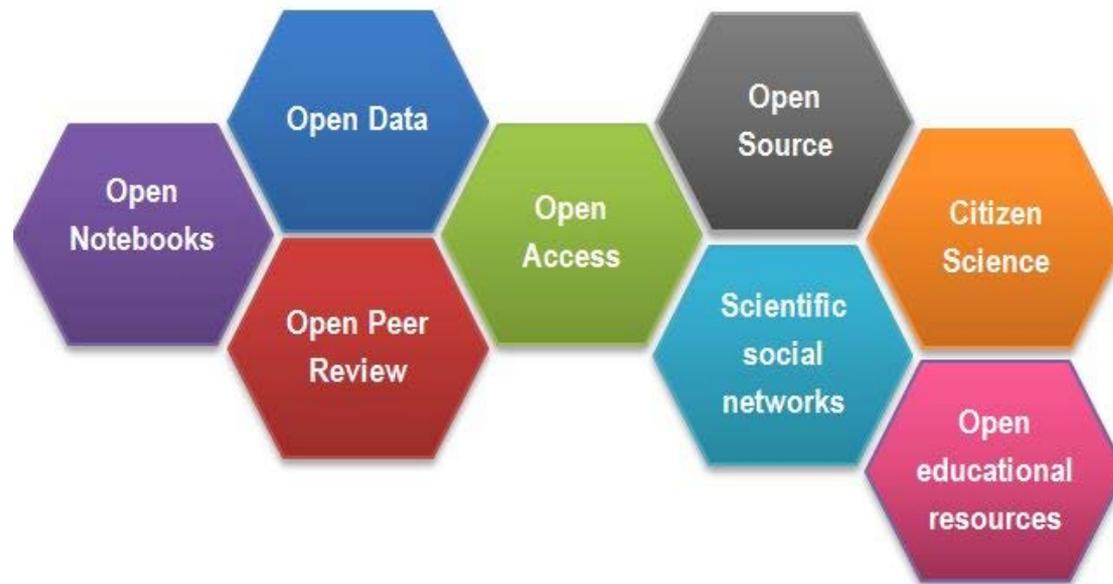
- ❖ Les facettes de l'Open Science
- ❖ Intégrité, reproductibilité
- ❖ FAIR Data, Open data
- ❖ Plan de gestion de données
- ❖ Conclusion

# Les facettes de l'open science



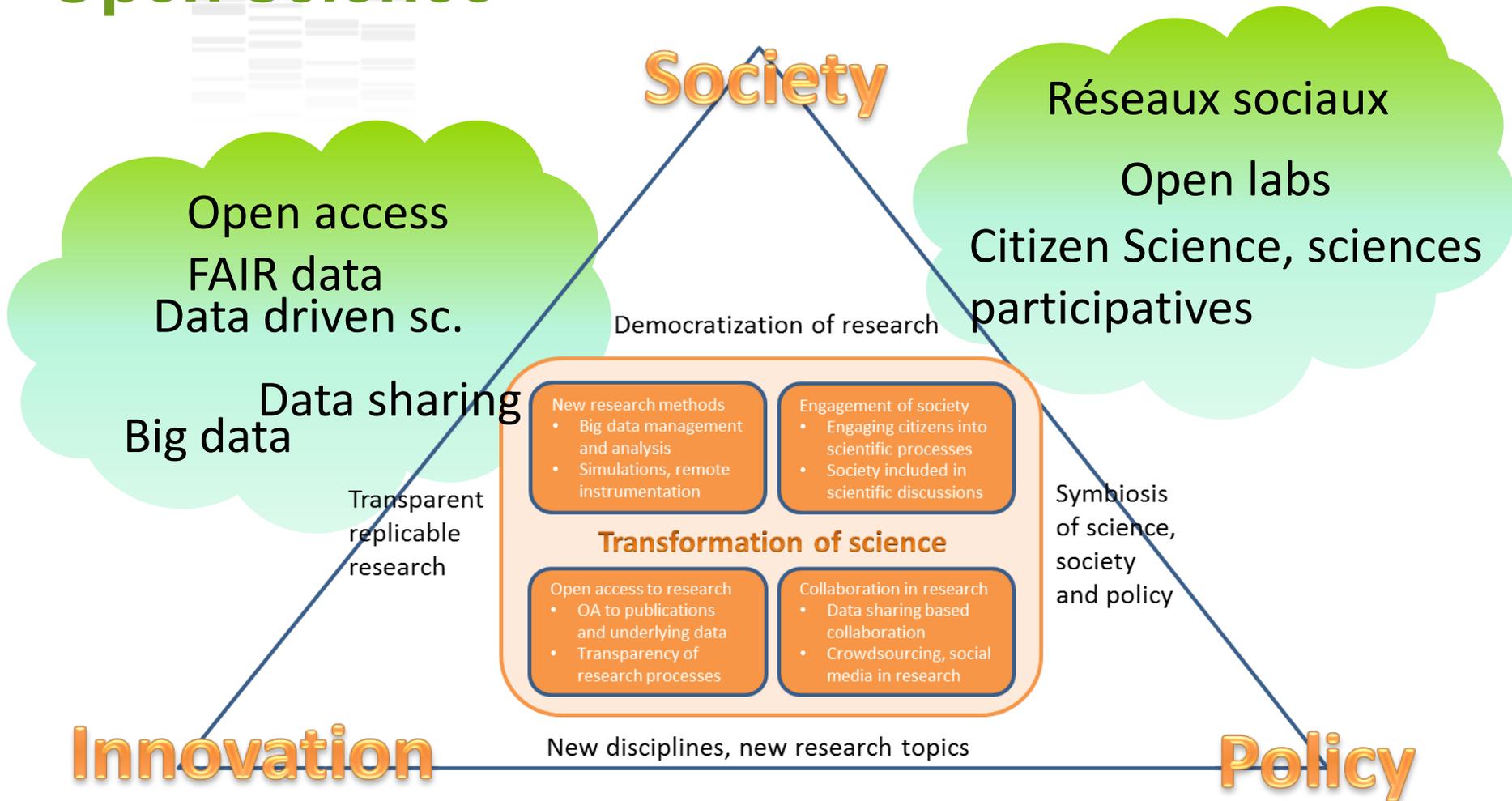
# Définition

Open Science is about **extending the principles of openness to the whole research cycle** (see figure 1), fostering sharing and collaboration as early as possible thus entailing a systemic change to the way science and research is done.



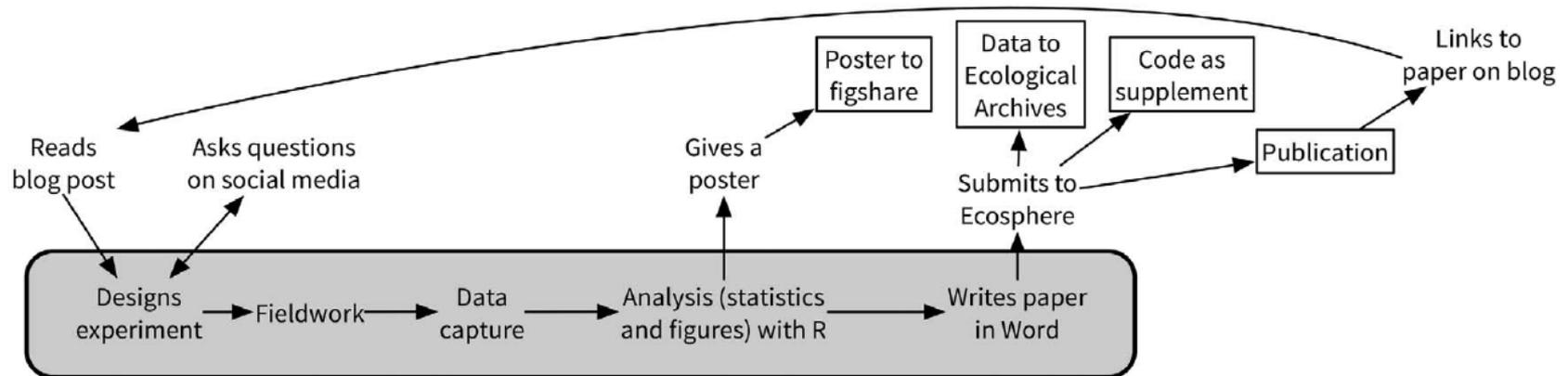
<https://www.fosteropenscience.eu/content/what-open-science-introduction>

# Open science



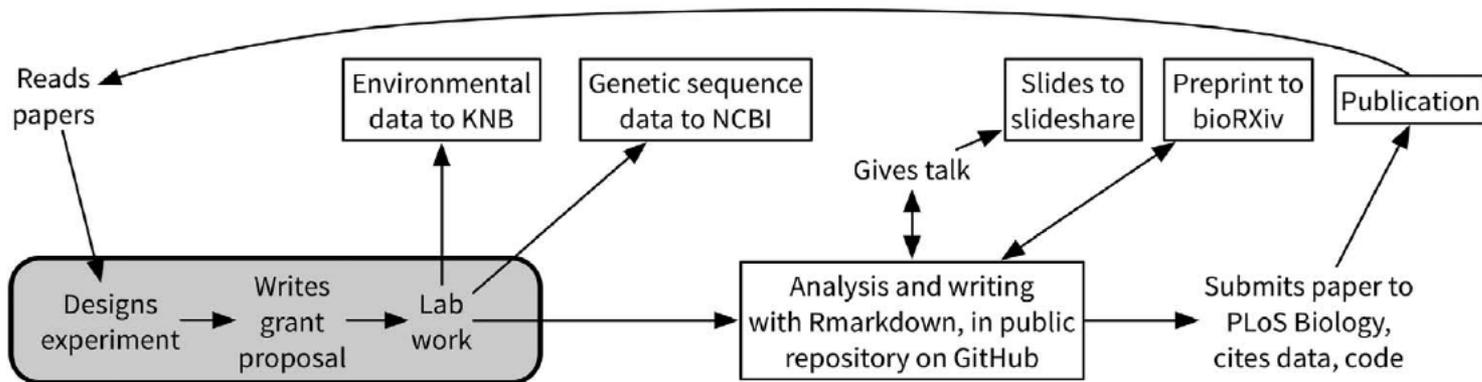
<https://ec.europa.eu/digital-agenda/en/open-science>

# 3 exemples de workflow « open science » (1/3)



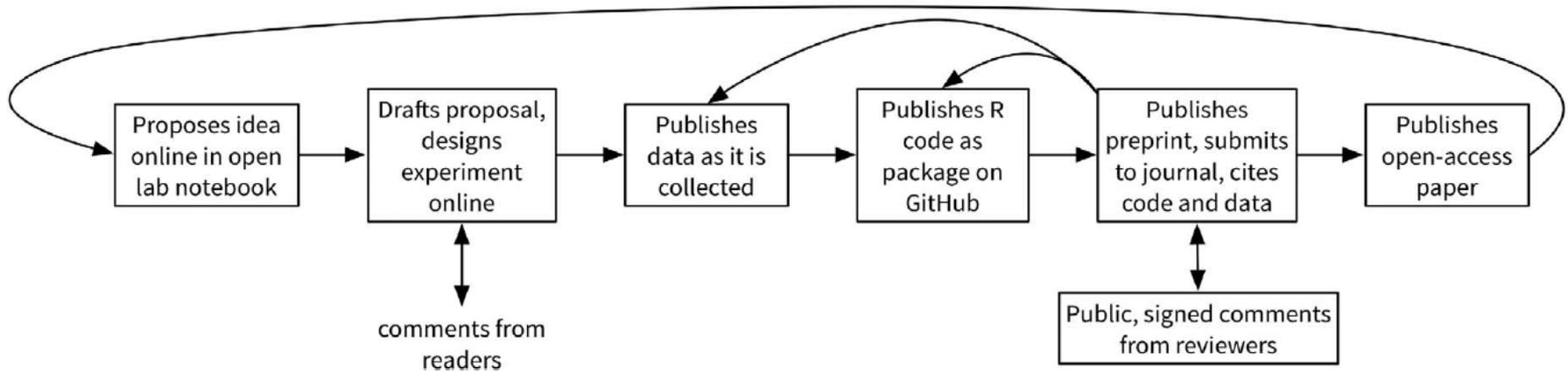
Hampton, S. E., S. S. Anderson, S. C. Bagby, C. Gries, X. Han, E. M. Hart, M. B. Jones, W. C. Lenhardt, A. MacDonald, W. K. Michener, J. Mudge, A. Pourmokhtarian, M. P. Schildhauer, K. H. Woo, and N. Zimmerman. 2015. The Tao of open science for ecology. *Ecosphere* 6(7):120. <http://dx.doi.org/10.1890/ES14-00402.1>

# 3 exemples de workflow « open science » (2/3)



Hampton, S. E., S. S. Anderson, S. C. Bagby, C. Gries, X. Han, E. M. Hart, M. B. Jones, W. C. Lenhardt, A. MacDonald, W. K. Michener, J. Mudge, A. Pourmokhtarian, M. P. Schildhauer, K. H. Woo, and N. Zimmerman. 2015. The Tao of open science for ecology. *Ecosphere* 6(7):120. <http://dx.doi.org/10.1890/ES14-00402.1>

# 3 exemples de workflow « open science » (3/3)



Hampton, S. E., S. S. Anderson, S. C. Bagby, C. Gries, X. Han, E. M. Hart, M. B. Jones, W. C. Lenhardt, A. MacDonald, W. K. Michener, J. Mudge, A. Pourmokhtarian, M. P. Schildhauer, K. H. Woo, and N. Zimmerman. 2015. The Tao of open science for ecology. *Ecosphere* 6(7):120. <http://dx.doi.org/10.1890/ES14-00402.1>

# Open science ... open access, open data



"The days of keeping our research results to ourselves are over. There is far more to gain from sharing data and letting others access and analyse that data."

For example, if sharing big data reveals that a certain kind of cancer activates a particular molecular pathway in most cases and it turns out that there is already a drug approved and available to block the activation of that molecular pathway, clinical trials can begin almost immediately. Saving time, money and lives.

Or if scientists want to monitor the effects of climate change on local ecosystems, they can use Open Science to engage citizen reporting, and rapidly multiply the data at their disposal.

To make the most of Open Science opportunities for Europe, I plan to focus on open data, open access and research integrity over the course of my mandate."

*Commissioner Carlos Moedas,  
"European research and innovation for global challenges", Lund, 4 December 2015*

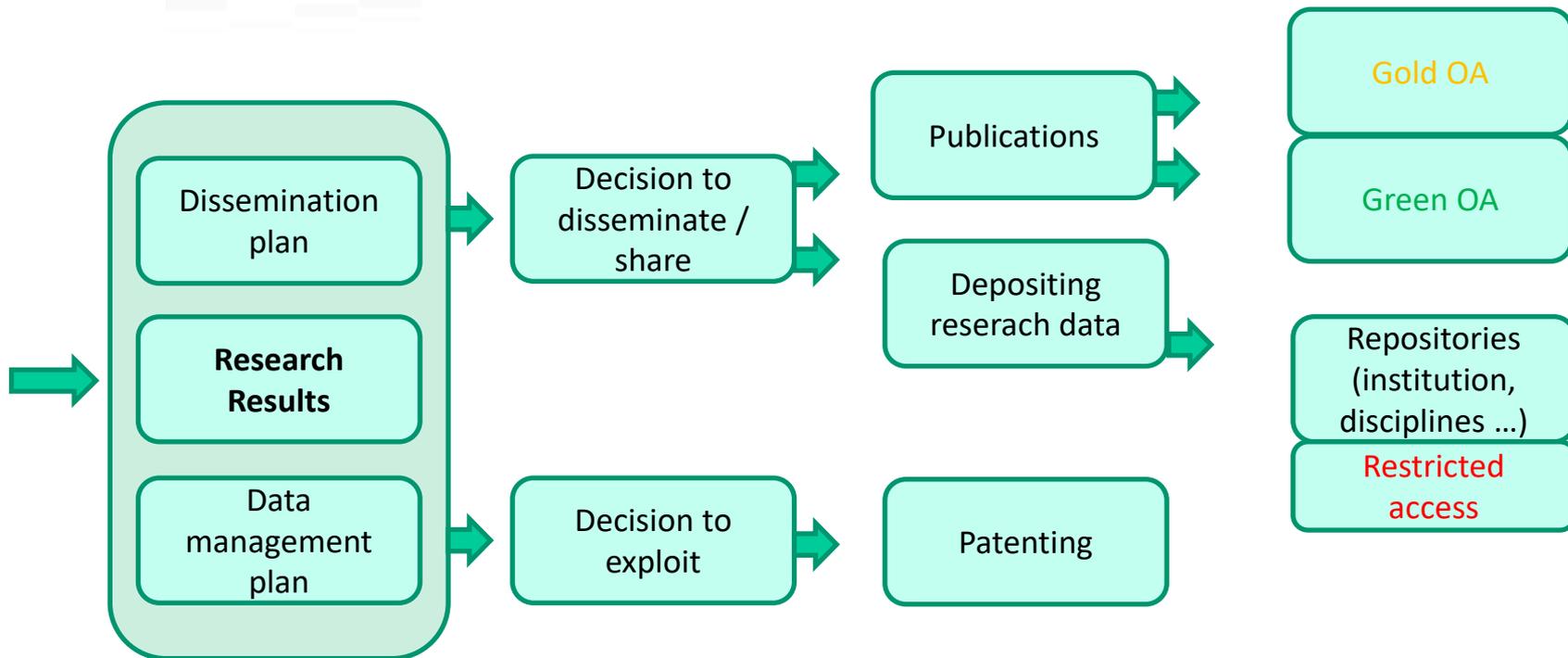


<http://bookshop.europa.eu/en/open-innovation-open-science-open-to-the-world-pbKI0416263/>

# Dissemination dans



R  
e  
s  
e  
a  
r  
c  
h



# Horizon Europe

**More openness:** The principle of 'open science' will become the modus operandi of Horizon Europe, requiring open access to publications and data. This will assist market uptake and increase the innovation potential of results generated by EU funding.

[http://europa.eu/rapid/press-release\\_IP-18-4041\\_en.htm](http://europa.eu/rapid/press-release_IP-18-4041_en.htm)





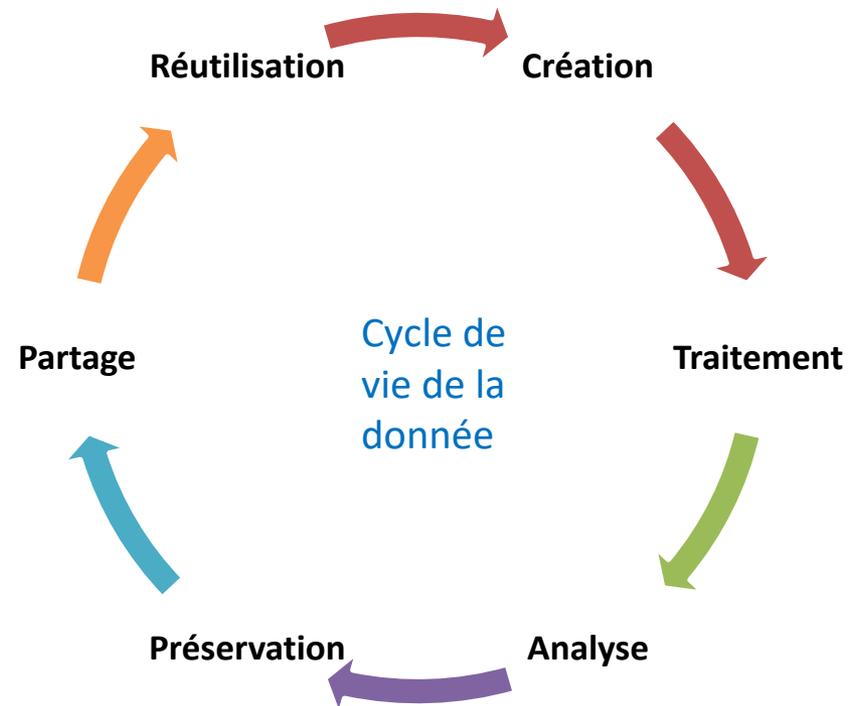
# Intégrité, reproductibilité et Open science

# Intégrité et cycle de vie de la donnée

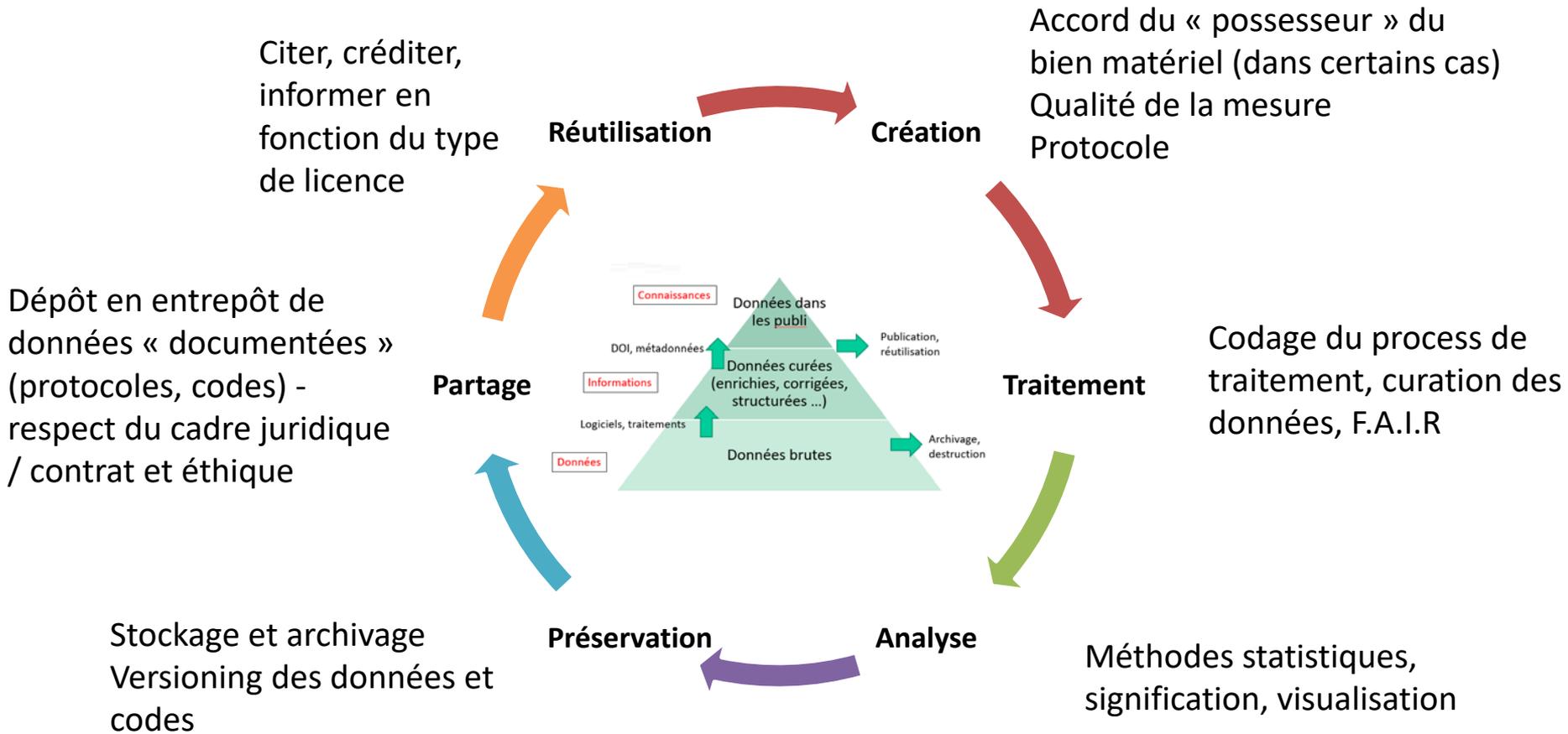
## Les valeurs de l'intégrité

- ❖ Fiabilité
- ❖ Honnêteté
- ❖ Respect
- ❖ Responsabilité

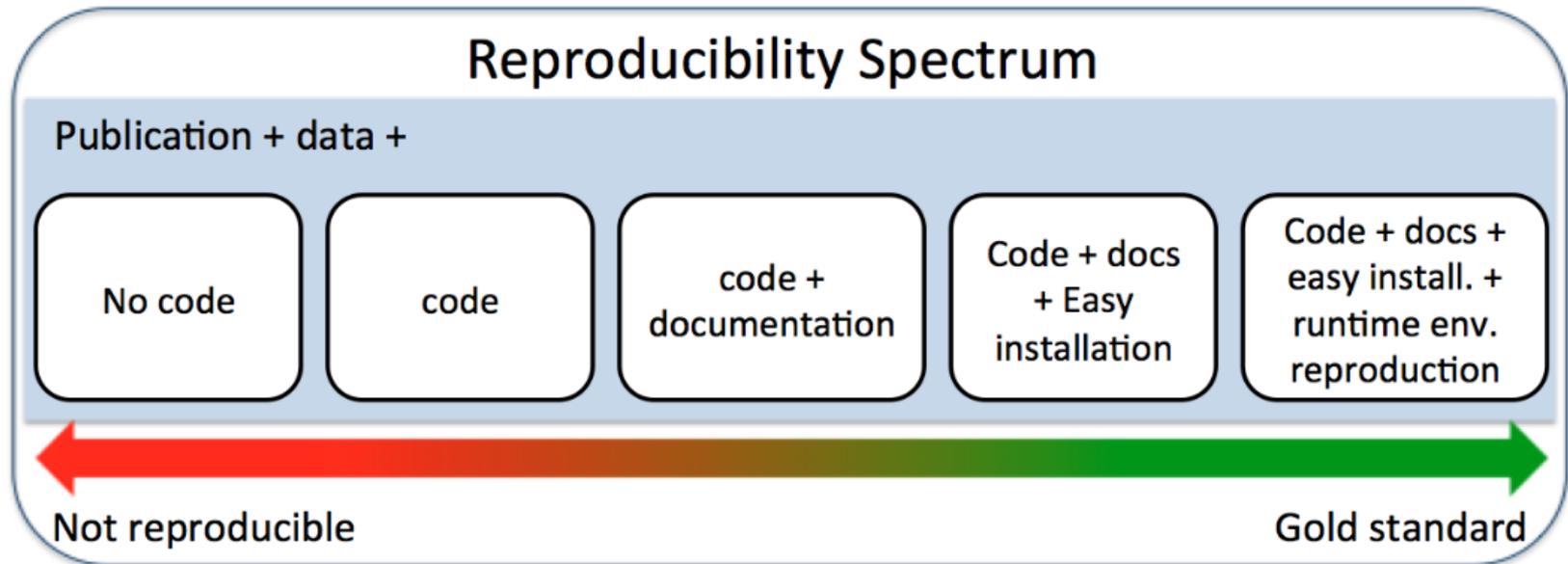
Ouverture : transparence + répliquabilité  
Science ouverte : levier pour l'intégrité  
Confiance des citoyens dans la science



# Intégrité et reproductibilité : même combat ?



# Le spectre de la reproductibilité



Reproducibility spectrum observed in publications. Sharing data and code is seen as enough to reproduce the data analysis by many. However, this is not enough.

<https://medium.com/@aakalin/scientific-data-analysis-pipelines-and-reproducibility-75ff9df5b4c5>

# Conclusion intermédiaire ...

**Jon Tennant**   
@Protohedgehog

Abonné

Open Science: Just science done right! Sticker from [@Mellmimming](#) now available on [@ZENODO\\_ORG](#) [zenodo.org/record/1285575...](https://zenodo.org/record/1285575...) [@BioMickWatson](#) #openscience

Traduire le Tweet

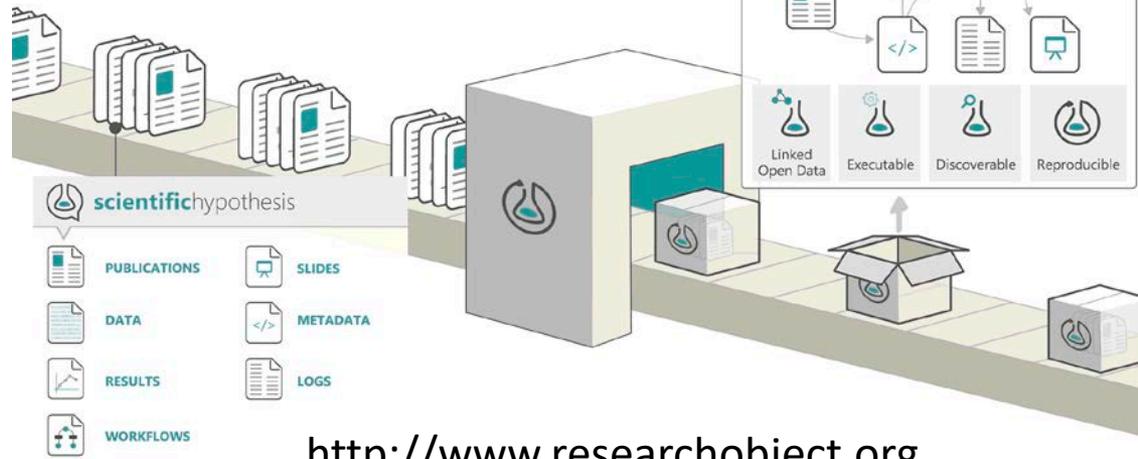


02:37 - 8 juin 2018

14 Retweets 31 J'aime



 Enabling **reproducible**, transparent research.



<http://www.researchobject.org>



# FAIR Data, Open Data

# FAIR Principles

<https://www.force11.org/group/fairgroup/fairprinciples>

## F

discoverable with machine readable metadata, identifiable and locatable by means of a standard identification mechanism

## A

available and obtainable to both human and machine

## I

both syntactically parseable and semantically understandable, allowing data exchange and reuse among scientific disciplines, researchers, institutions, organisations and countries

## R

sufficiently described and shared with the least restrictive licences, allowing the widest reuse possible across scientific disciplines and borders, and the least cumbersome integration with other data sources

### To be Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

### To be Accessible:

- A1 (meta)data 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 Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

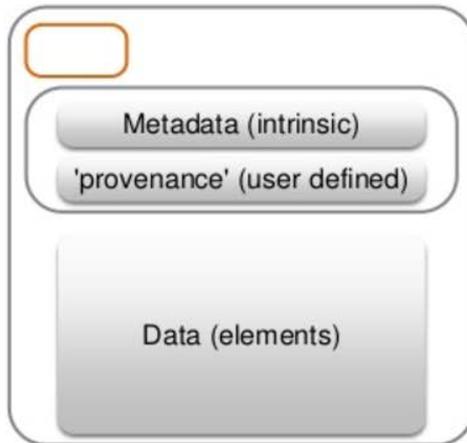
### To be Re-usable:

- R1. meta(data) have a plurality of accurate and relevant attributes.
- R1.1. (meta)data are released with a clear and accessible data usage license.
- R1.2. (meta)data are associated with their provenance.
- R1.3. (meta)data meet domain-relevant community standards.

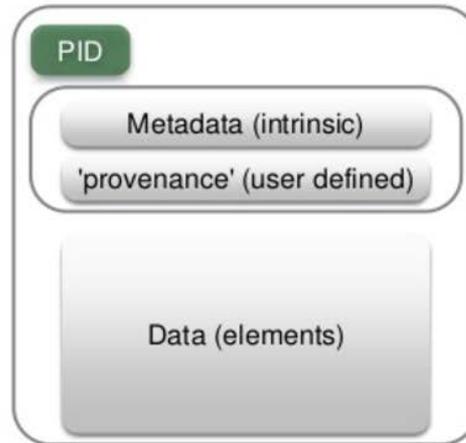
# Data as increasingly FAIR Digital Objects

Source : Barend Mons

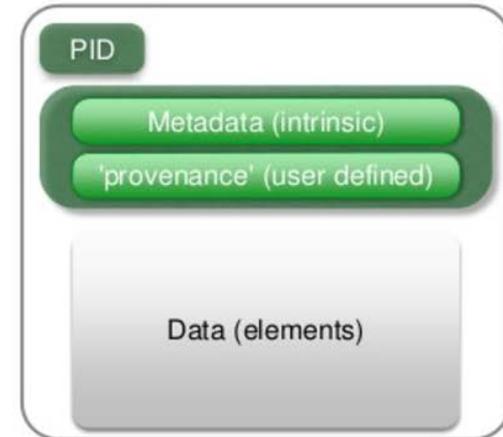
## Totally UNFAIR



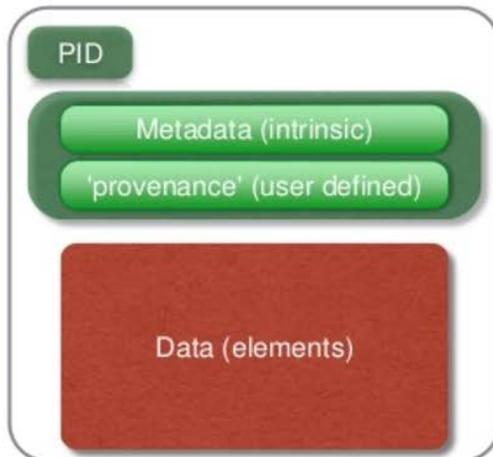
## Findable Usable for Humans



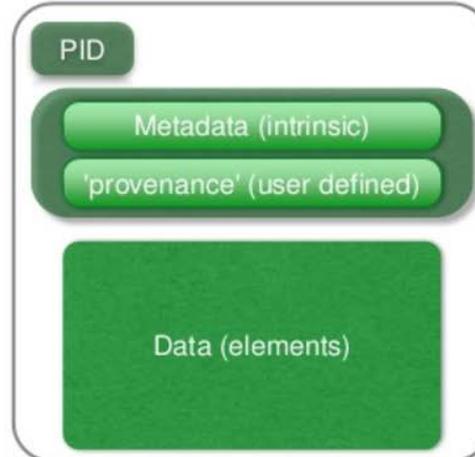
## FAIR metadata



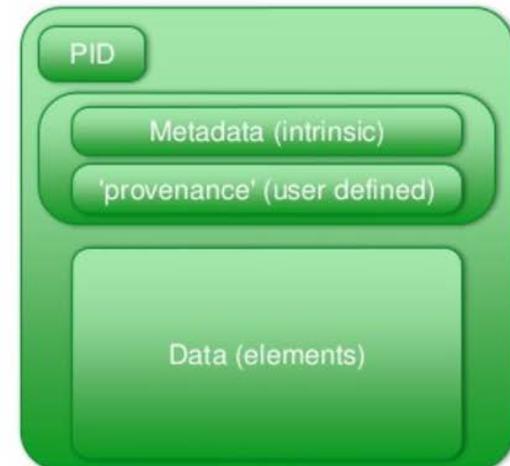
## FAIR data- restricted access



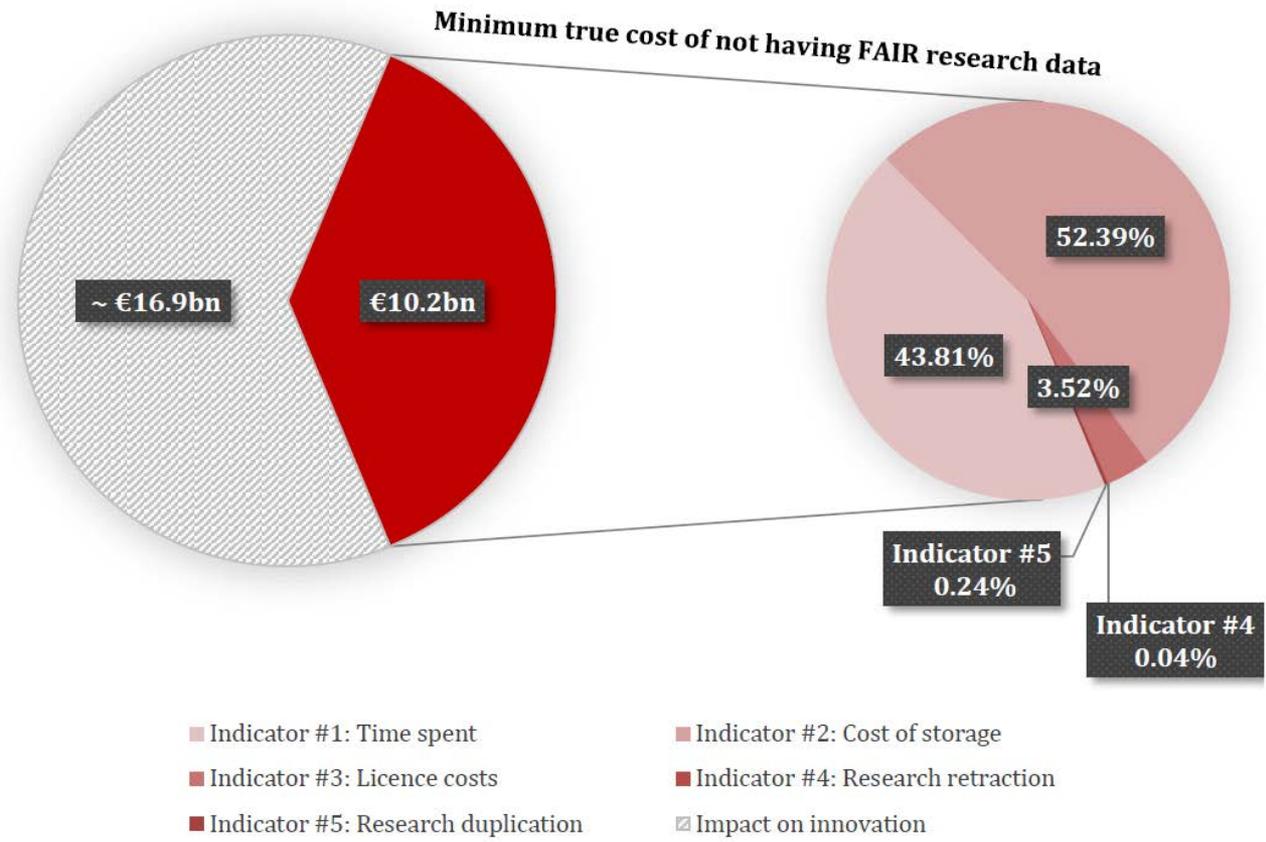
## FAIR data- Open Access



## FAIR data- Open Access/Functionally Linked



# Not FAIR Data : Un coût de 10 à 26 Milliards



<https://publications.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1/language-en>

# Code de la recherche modifié par la loi ESR de juillet 2013

## Chapitre II : Objectifs et moyens institutionnels de la recherche publique.

### Article L112-1 [En savoir plus sur cet article...](#)

Modifié par [LOI n°2013-660 du 22 juillet 2013 - art. 16](#)

La recherche publique a pour objectifs :

- a) Le développement et le progrès de la recherche dans tous les domaines de la connaissance ;
- b) La valorisation des résultats de la recherche au service de la société, qui s'appuie sur l'innovation et le transfert de technologie ;
- c) Le partage et la diffusion des connaissances scientifiques en donnant priorité aux formats libres d'accès ;
- c bis) Le développement d'une capacité d'expertise et d'appui aux associations et fondations, reconnues d'utilité publique, et aux politiques publiques menées pour répondre aux défis sociétaux, aux besoins sociaux, économiques et du développement durable ;
- d) La formation à la recherche et par la recherche ;
- e) L'organisation de l'accès libre aux données scientifiques.

Les établissements publics de recherche et les établissements d'enseignement supérieur favorisent le développement des travaux de coopération avec les associations et fondations, reconnues d'utilité publique. Ils participent à la promotion de la recherche participative et au développement des capacités d'innovation technologique et sociale de la Nation. Ces coopérations s'exercent dans le respect de l'indépendance des chercheurs et, en l'absence de clauses contraires, dans un but non lucratif. Les travaux de recherche menés dans le cadre de ces coopérations sont, en l'absence de clauses contraires, rendus publics et accessibles.

# Les lois

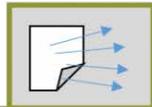


## Loi dite CADA

Relative à la liberté d'accès aux documents administratifs (dont indirectement aux données de la recherche) et à la réutilisation des informations publiques

1978

## Loi dite Valter



Relative à la gratuité et aux modalités de réutilisation des informations du secteur public

2015

## Loi dite Lemaire

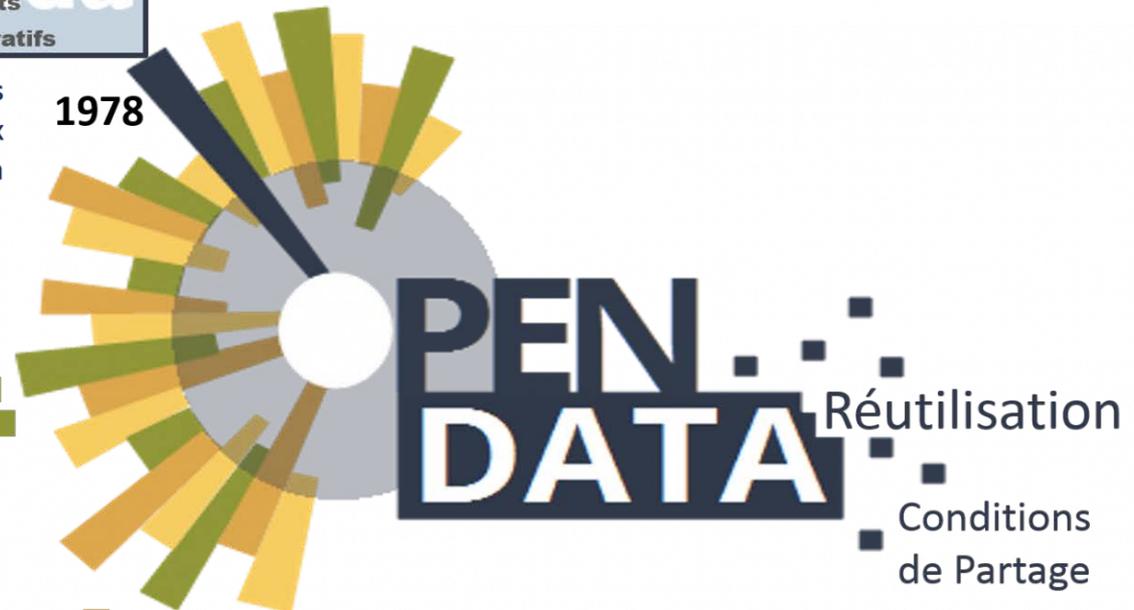


Pour une République Numérique

Relative à la circulation des données et du savoir, à la protection des droits dans la société numérique et à l'accès au numérique

2016

<http://dorandum.fr>



L'Open Data trouve ses bases dans la loi dite CADA et c'est la loi dite Valter qui introduit le principe d'ouverture des données de la recherche.

La loi dite Lemaire crée un principe d'ouverture des données publiques par défaut. Les articles 30 et 38 posent un principe de libre utilisation des données de la recherche.

# Les lois.... Ce qu'il faut retenir (simplifié)

- ❖ Pour les publi : le chercheur est titulaire des droits (moral et patrimonial)
- ❖ Ce n'est pas le cas pour les données - plutôt des personnes morales
- ❖ Un organisme de recherche doit :
  - ✓ porter à connaissance (catalogue)
  - ✓ rendre réutilisable gratuitement (données et logiciels)



# Check list « ouvrir/fermer »

1- Le jeu de données contient-il des données personnelles (qui identifient directement ou indirectement des individus) ?

- Oui
- Oui, mais les formalités CNIL sont respectées, ET les individus concernés ont donné leur accord écrit à la diffusion OU les données ont été anonymisées (non conservation des noms, des identifiants et des pseudonymes, cf [CNIL](#) et [CNRS](#)).
- Non
- Vous n'avez pas encore répondu

2- Avez-vous utilisé d'autres données que celles que vous avez produites vous-mêmes ?

- Oui
- Oui, mais j'ai bien vérifié la licence d'utilisation de la base de données et j'ai le droit de les extraire et de les réutiliser.
- Non
- Vous n'avez pas encore répondu

3- Est-ce que certaines de ces données, produites dans le cadre de l'établissement public, contiennent des données protégées par le droit d'auteur (textes, images, schémas, ...) ?

- Oui
- Oui, mais j'ai obtenu l'accord écrit de l'auteur.
- Non
- Vous n'avez pas encore répondu

4- Est-ce que ces données ont été produites dans le cadre d'un consortium avec le secteur privé ?

- Oui
- Oui, mais elles peuvent être diffusées librement
- Non
- Vous n'avez pas encore répondu

5- Votre Unité est-elle protégée ou en Zone à Régime Restrictif (ZRR) ?

- Oui
- Oui, mais ma Direction d'Unité m'a donné son accord pour diffuser mes données
- Non
- Vous n'avez pas encore répondu

6- Avez-vous produit des données contenant des informations relatives à l'environnement ou pouvant avoir un impact sur la conservation de la biodiversité ?

- Oui
- Non
- Vous n'avez pas encore répondu

7- Avez-vous produit des données pouvant impacter la sécurité publique (pathogènes, ressources en eau, etc) ?

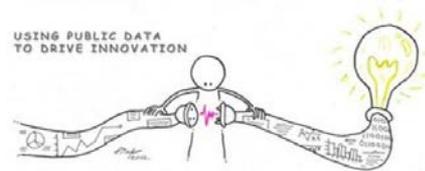
- Oui
- Non
- Vous n'avez pas encore répondu

8- Est-ce que parmi ces données, certaines proviennent de services statistiques publics comme l'INSEE ?

- Oui
- Oui, mais une autorisation a été obtenue auprès du Comité du Secret Statistique.
- Non
- Vous n'avez pas encore répondu

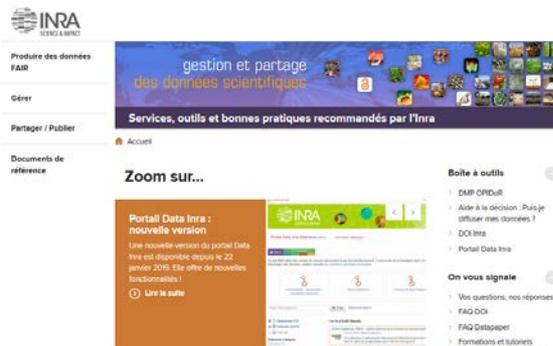
# « Open data, open access, open code » penser à d'autres façons de créer de la valeur

- ❖ Des peurs : qui va cueillir les fruits (de l'ouverture)
  - ✓ Crainte d'une science à 2 vitesses entre les producteurs de données et les consommateurs
  - ✓ Perte de revenus liée à la LRN
- ❖ D'autres façons d'innover
- ❖ Aller chercher les données des autres, investir dans les nouveaux outils et méthodes d'analyse



# Le site web Data Partage

Gestion et partage des données scientifiques  
Actualités, information, accès aux services



Attribution de DOI

Plans de gestion de données

Entrepôt de données

(aide au choix, annuaire, entrepôt et forge logicielle)

Appui à la publication de données

<http://www.datapartage.inra.fr>

Site web public

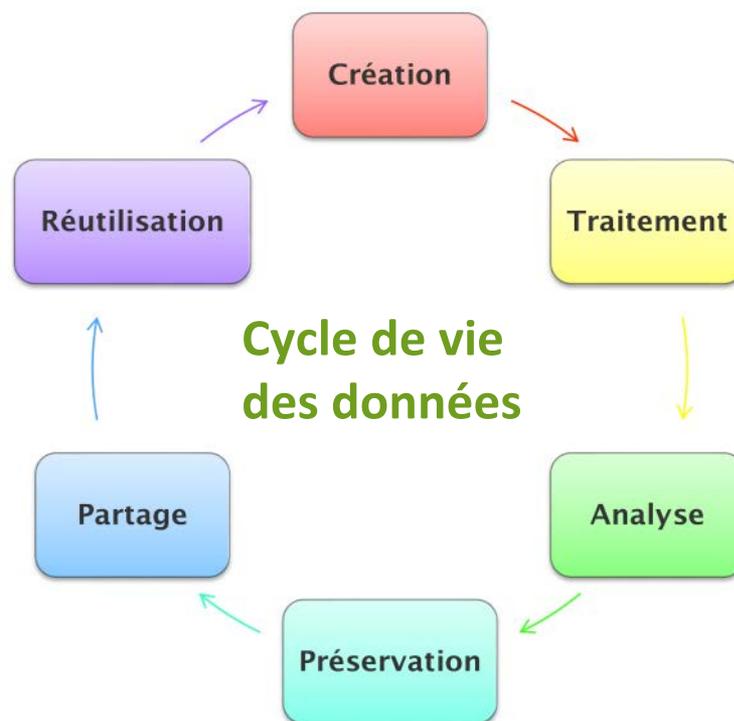


# Plan de gestion de données

Un outil pour réfléchir à tout ça ....

# Le PGD : une définition

- Document qui décrit la façon dont les données seront obtenues, traitées, organisées, stockées, sécurisées, préservées, partagées au cours et à l'issue d'un projet
- Aide à la mise en place de bonnes pratiques de gestion à toutes les étapes du cycle de vie des données



# A quelles questions répond un PGD ?



- Comment la gestion et le partage des données sont-ils financés, en particulier à long terme ?

## Ressources

- En quoi consiste le projet ?
- Qui sont les partenaires ?
- Quelle est la politique de gestion des données ?
- Qui est responsable de la gestion des données ?

## Responsabilités dans le projet

- Quelles données seront produites/utilisées au cours du projet ? (type, format, volume et accroissement...).
- Comment seront-elles produites ou transformées ?

## Collecte des données

- Comment les données seront-elles identifiées, décrites ?
- Quels standards de métadonnées utilisera-t-on ?
- Comment seront générées les métadonnées ?

## Documentation des données

- Qui pourra accéder aux données ?
- Les données seront-elles publiées ?
- Comment ?
- Dans quel délai ?
- Sous quelle licence ?

## Accès et partage des données

What's the plan?

Dans le cadre d'un projet de recherche ou non...

- Comment, où, par qui, seront stockées, sauvegardées et sécurisées les données ?

## Sauvegarde des données

- Qui sera propriétaire des données produites ?
- Des données externes seront-elles utilisées ?

## Propriété intellectuelle

- Des données sensibles seront-elles produites ou utilisées ?
- Comment sera assurée leur anonymisation ?

## Ethique

- Quel est le plan d'archivage et de préservation à long terme ?

## Archivage et préservation des données



# Core requirements for DMPs

Science Europe. (2018). [Practical guide to the international alignment of research data management.](#)



## 1. Data description and collection or re-use of existing data

**F2, R1.2.** How will new data be collected or produced and/or how will existing data be re-used?

**I1, R1.2, R1.3.** What data (for example the kinds, formats, and volumes) will be collected or produced?

## 2. Documentation and data quality

**I1, I2, I3, R1, R1.3.** What metadata and documentation (for example the methodology of data collection and way of organizing data) will accompany data?

**I2, R1, R1.2.** What data quality control measures will be used?

## 3. Storage and backup during the research process

a. How will data and metadata be stored and backed up during the research process?

b. How will data security and protection of sensitive data be taken care of during the research?

## 4. Legal and ethical requirements, codes of conduct

a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

**A1.2, R1.1.** How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

**A2.** How will possible ethical issues be taken into account, and codes of conduct followed?

## 5. Data sharing and long-term preservation

**A1.2, A2, I3, R1.1.** How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?

**A1.1, A1.2, I3.** What methods or software tools will be needed to access and use the data?

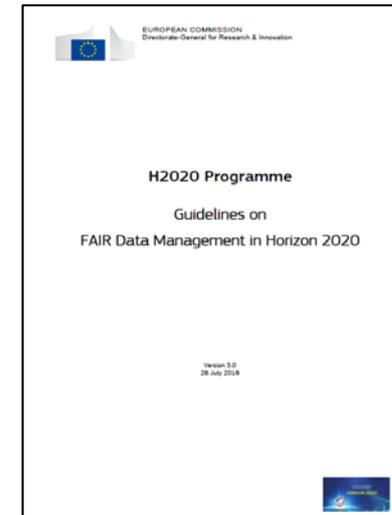
**F1, F3, A2.** How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

## 6. Data management responsibilities and resources

a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

# Data sharing plan



[https://erc.europa.eu/sites/default/files/document/file/ERC\\_Data ManagementPlan\\_template.docx](https://erc.europa.eu/sites/default/files/document/file/ERC_Data ManagementPlan_template.docx)

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

# Des outils – https://dmp.opidor.fr

Bienvenue sur **DMP OPIDoR V2.1.5**, découvrez la [liste des changements](#) apportés par cette nouvelle version.

## Titre du projet

 Plan de test, d'entrainement ou créé en vue d'une formation

## Choisissez un modèle

Vous pouvez choisir soit un modèle fourni par votre organisme soit par un autre organisme, ou un modèle financier. Le modèle par défaut est **Horizon 2020 FAIR DMP (anglais)**.  
[Retrouvez la liste des modèles disponibles](#)

INRA - Institut National de la Recherche Agronomique (Votre organisme)   Autre organisme   Financier

Plusieurs modèles sont disponibles, lequel souhaitez-vous utiliser ?

- INRA - General project template
- INRA - General project template
- INRA - Trame Structure
- INRA - Trame Structure (5 modes de gestion)
- INRA - Trame générique projet**

Créer un plan   Suivant   Utiliser le modèle par défaut

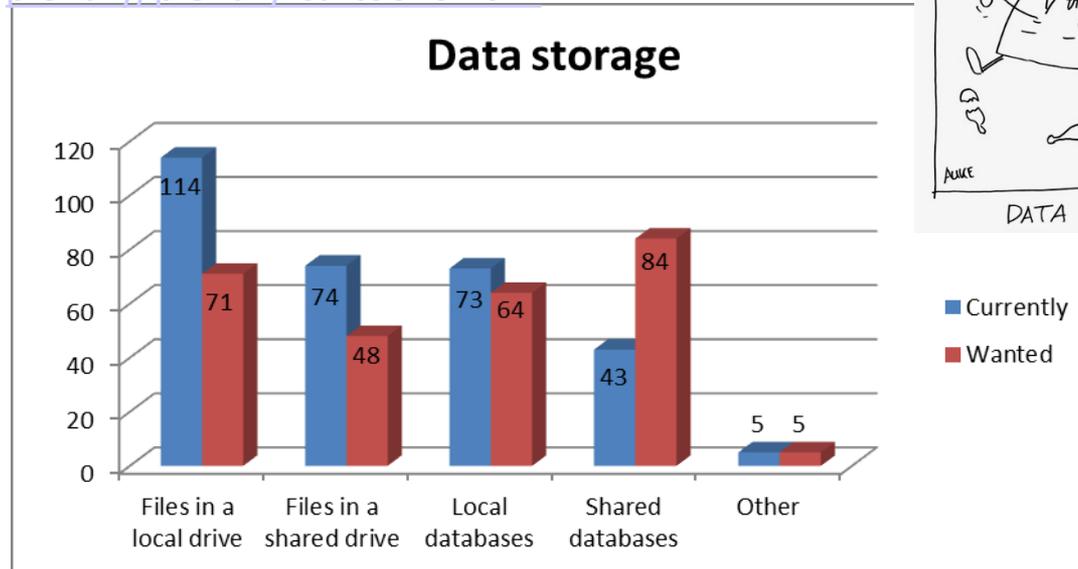


# Conclusion

# Gestion des données

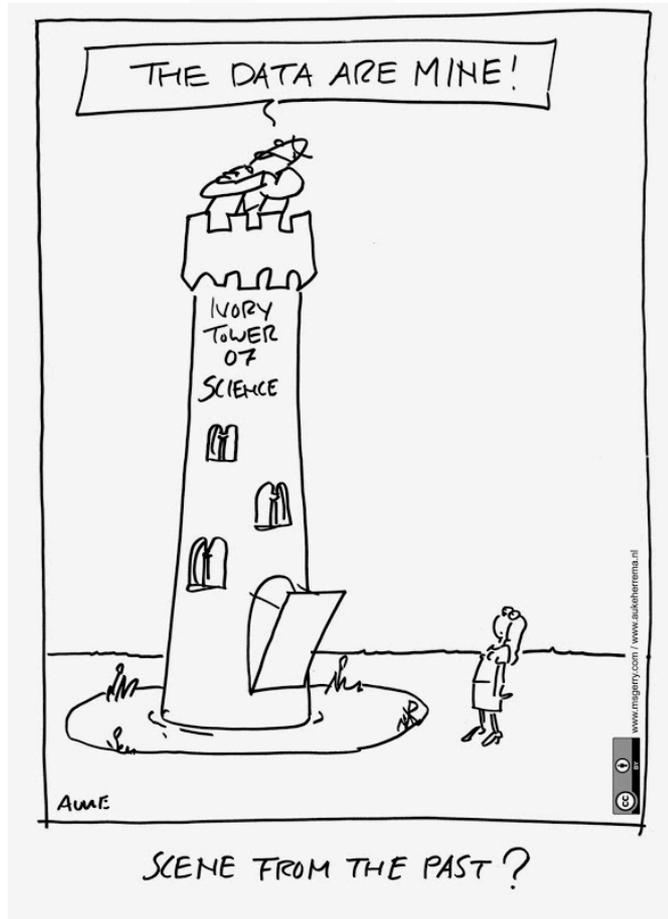
Source des dessins

<https://rd-alliance.org/plenary-meetings/fourth-plenary/plenary-cartoons.html>



- ❖ Faire évoluer les pratiques : formation
- ❖ Infrastructures collectives : du poste de travail au cloud
- ❖ Développer l'usage d'outil pour une recherche reproductible : Lab notebook électroniques, Codage des traitements, gestion de workflow, Container (Docker...)

# Un changement de culture





# Le plan de gestion, pour conclure

- ❖ Un outil pour la gouvernance des données

*Gouvernance : un processus pour mieux gérer et valoriser les données qui définit des rôles et des responsabilités et qui renvoie à des questions :*

- *Techniques*
- *Juridiques*
- *Scientifiques*
- *Economiques*
- *Politiques*

- ❖ Evolution vers des « Machine Actionable DMP » ([Active DMP](#))

- ❖ Il existe aussi des « Software management plan » voir <https://hal.archives-ouvertes.fr/hal-01802565/document>

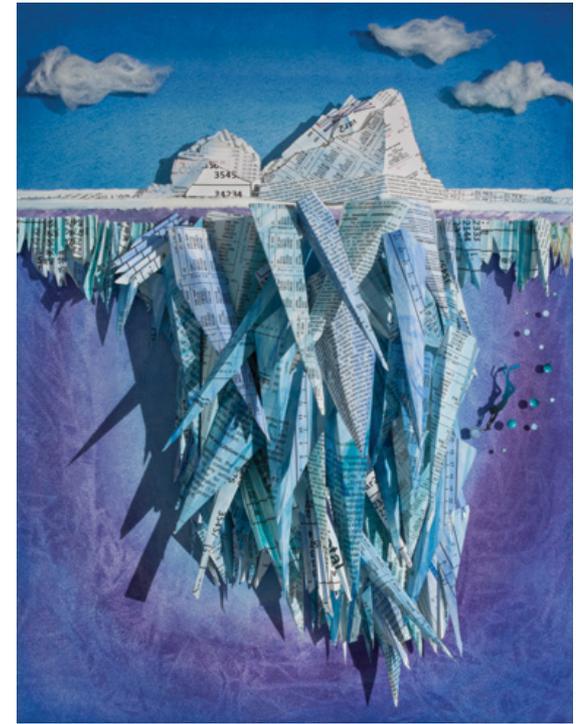
# Des supports et des outils

❖ <https://ist.inra.fr>

❖ <https://datapartage.inra.fr>

❖ Numéro spécial cahier des techniques

[https://www6.inra.fr/cahier\\_des\\_techniques/Les-Cahiers-parus/Les-n-Speciaux-et-les-n-Thematiques/Donnees-de-la-recherche-N-Thematique2](https://www6.inra.fr/cahier_des_techniques/Les-Cahiers-parus/Les-n-Speciaux-et-les-n-Thematiques/Donnees-de-la-recherche-N-Thematique2) -



Merci pour votre attention...

[odile.hologne@inra.fr](mailto:odile.hologne@inra.fr)

@Holo\_08



twitter

