

# COLLOQUE ANNUEL DE LA SOCIÉTÉ ROYALE DES SCIENCES DE LIÈGE

NOMBRES, IMAGES ET AUTRES DONNÉES...  
À LA BASE DE LA SCIENCE EXPÉRIMENTALE

## Bonnes pratiques pour la gestion des données

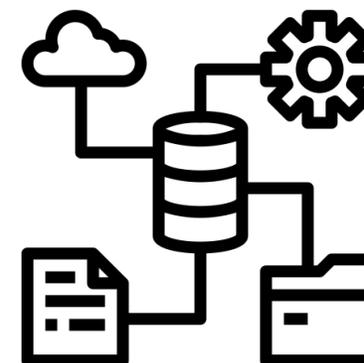
Judith BIERNAUX

ULiège RISE - Recherche, Innovation, Support et Entreprises

Research Data Officer

Head of Research Management Unit

[jbiernaux@uliege.be](mailto:jbiernaux@uliege.be)





# Research Data Officier ?

Le rôle principal de Research Data Officer (RDO) est de coordonner le **développement**, la **mise à disposition** et l'**amélioration** des **services** de gestion des données de recherche de l'Université.

*Définition par le GT Open Science du [CReF](#)*



# Research Data Officier ?

Le rôle principal de Research Data Officer (RDO) est de coordonner le **développement**, la **mise à disposition** et l'**amélioration** des **services** de gestion des données de recherche de l'Université.

*Définition par le GT Open Science du [CReF](#)*

*En pratique...*

Aider les chercheurs à être conformes à des critères de qualité en gestion des données



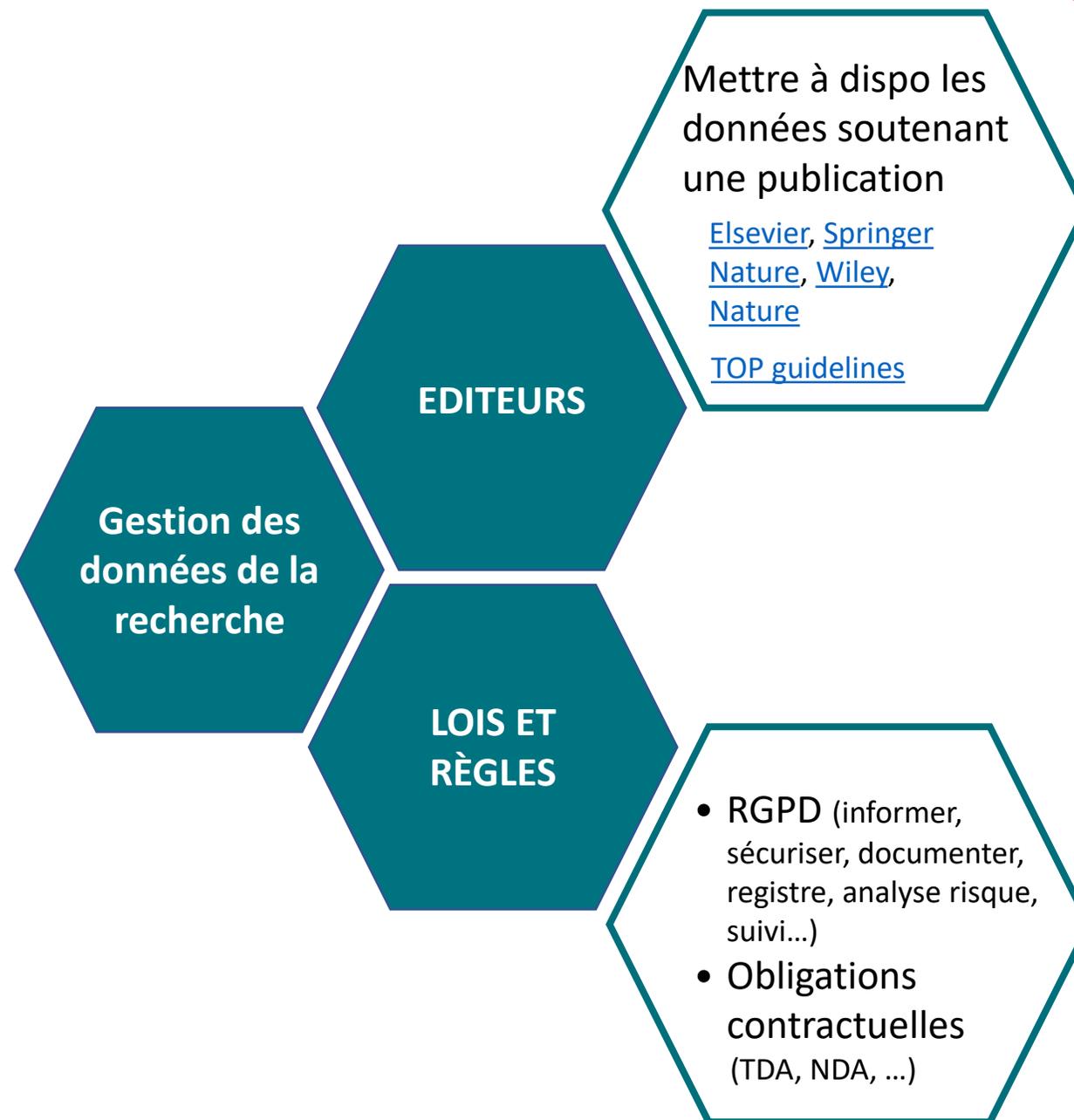
# Critères de qualité ?

**Gestion des  
données de la  
recherche**

# Critères de qualité ?

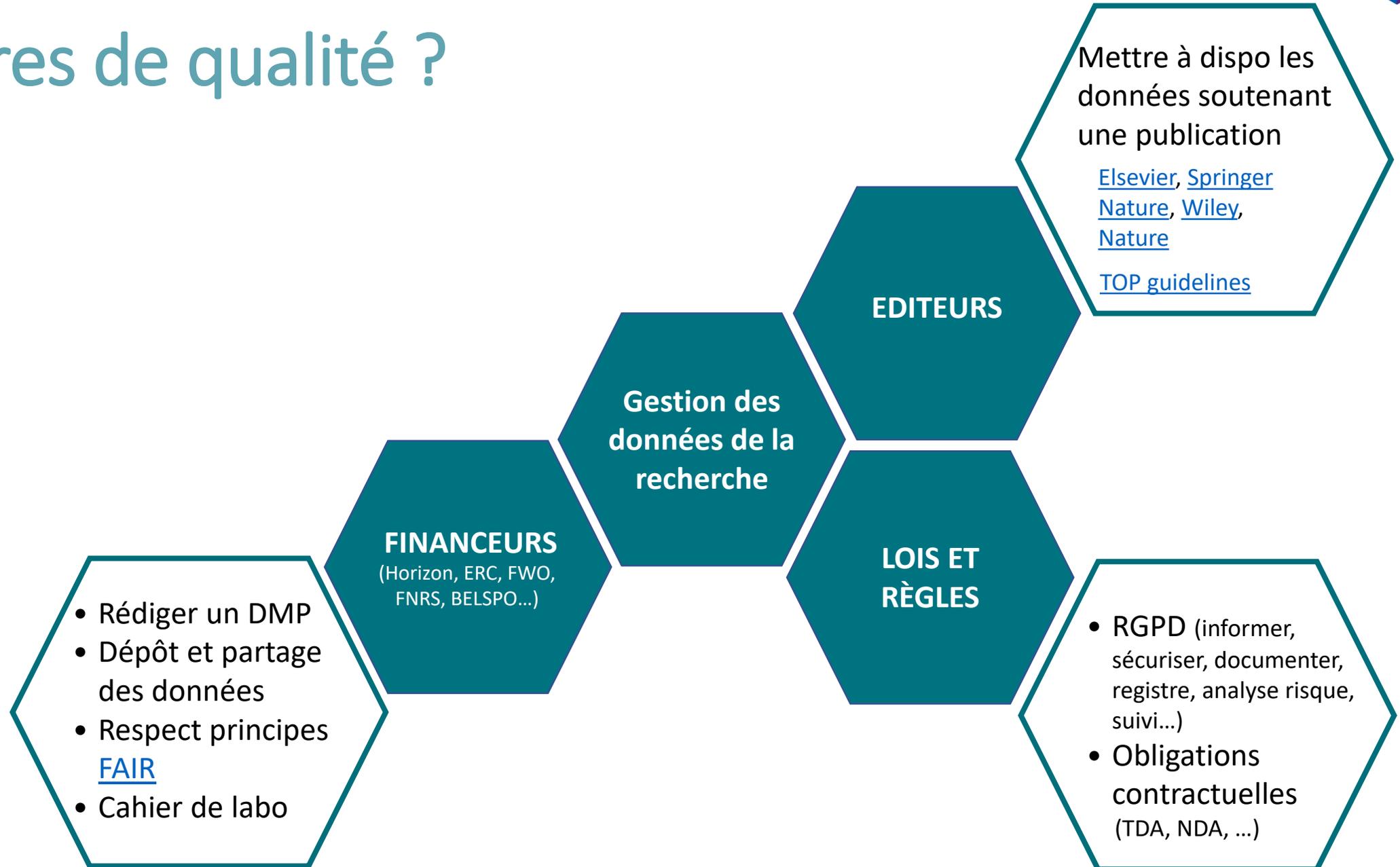


# Critères de qualité ?





# Critères de qualité ?





# Critères de qualité ?

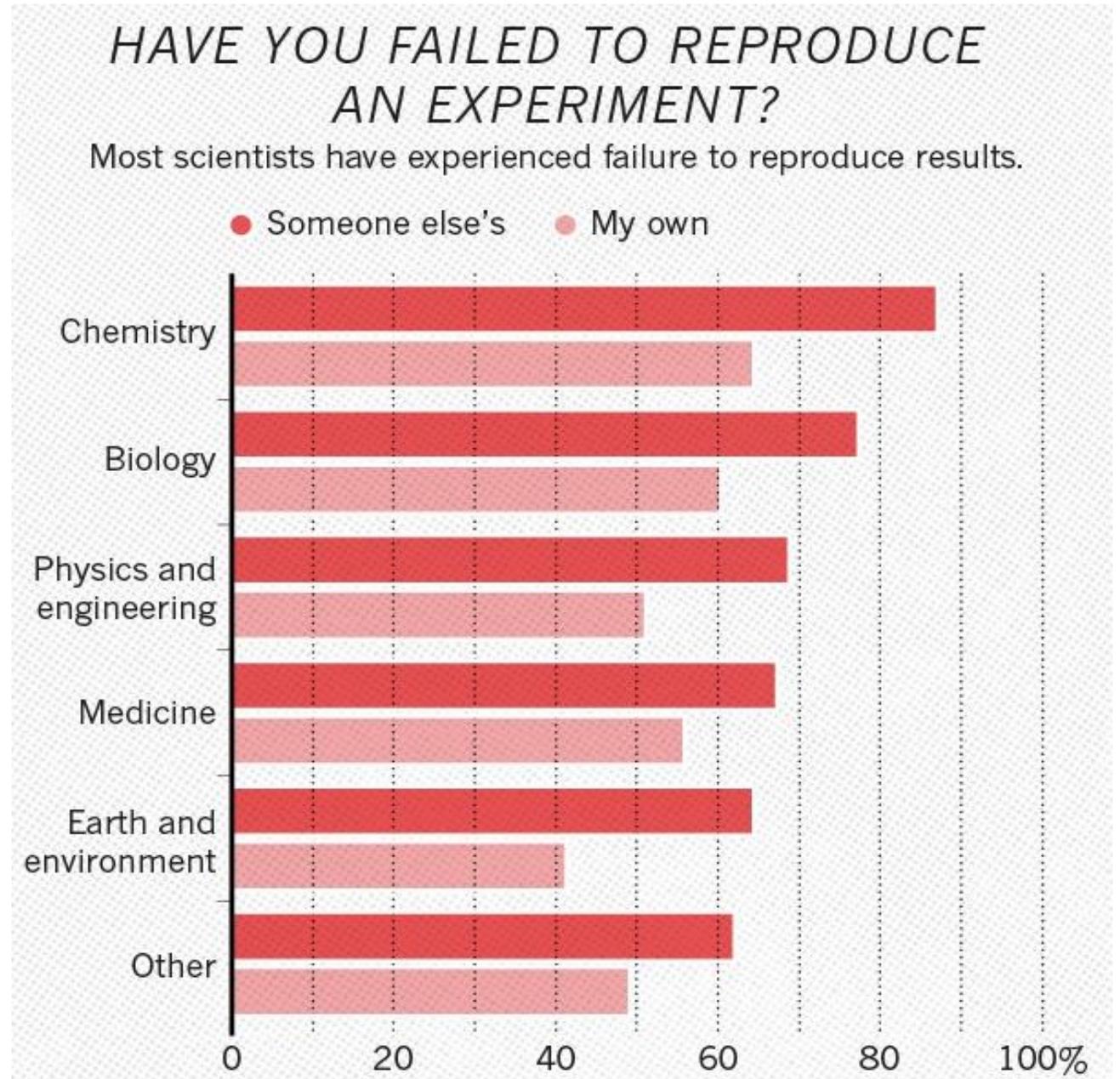




# Pourquoi ces critères ?

# Pourquoi ces critères ?

## Crise de la reproductibilité



Number of respondents from each discipline: Biology 703, Chemistry 106, Earth and environmental 95, Medicine 203, Physics and engineering 236, Other 233

# Pourquoi ces critères ?

## Crise de la reproductibilité

Méthodes statistiques, *publish or perish*, biais humains...

Une partie de la crise de la reproductibilité s'explique par **l'indisponibilité** de données originales.

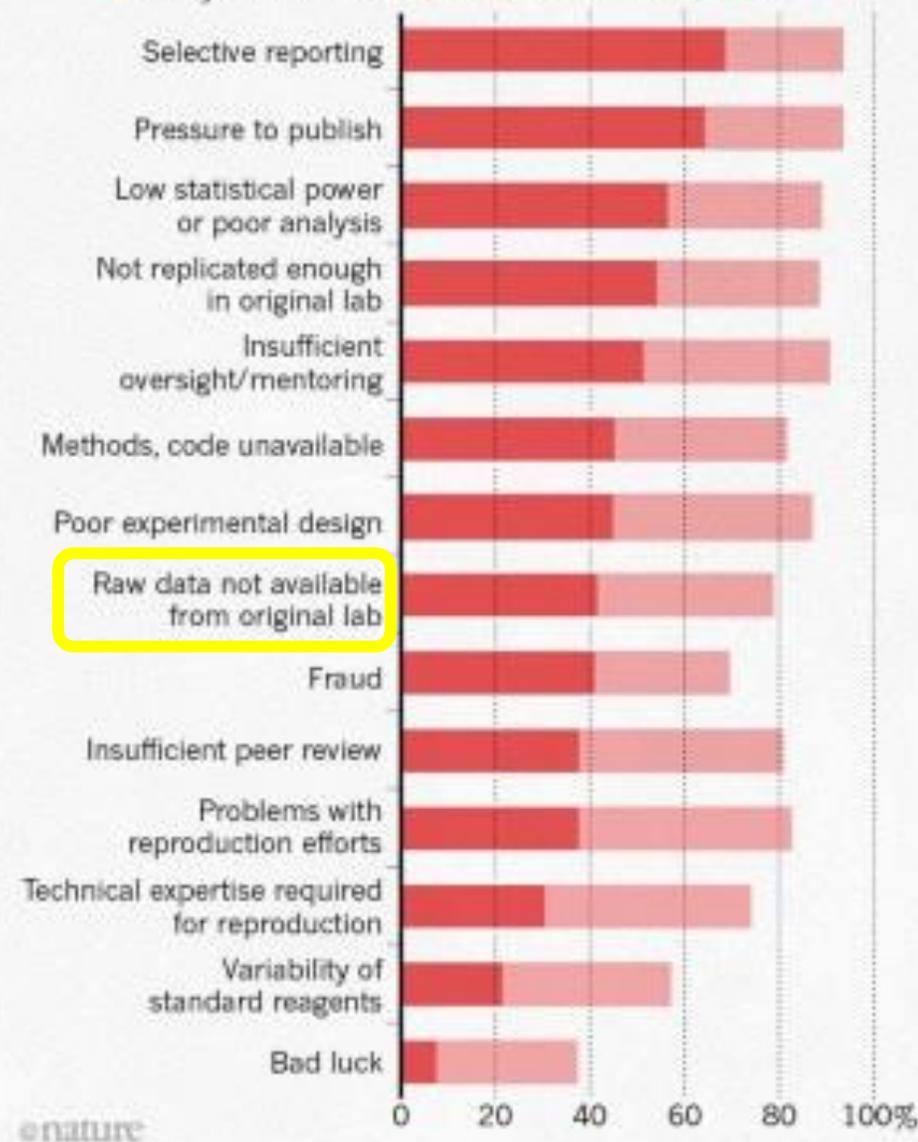
Il ne s'agit pas que de partage mais aussi de **traçabilité**, d'archivage, à l'échelle du labo ou individuelle.

**La reproductibilité scientifique passe aussi par des choix triviaux de format, de lieu de stockage, de partage, de documentation, de structure de métadonnées, ...**

## WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.

● Always/often contribute ● Sometimes contribute





# Gestion des données et reproductibilité

Le rôle principal de Research Data Officer (RDO) est de coordonner le **développement**, la **mise à disposition** et l'**amélioration** des **services** de gestion des données de recherche de l'Université.

*Définition par le GT Open Science du [CReF](#)*

*En pratique...*

Aider les chercheurs à être conformes à des critères de qualité en gestion des données



# Gestion des données et reproductibilité

Le rôle principal de Research Data Officer (RDO) est de coordonner le **développement**, la **mise à disposition** et l'**amélioration** des **services** de gestion des données de recherche de l'Université.

*Définition par le GT Open Science du [CReF](#)*

*En pratique...*

Aider les chercheurs à être conformes à des critères de qualité en gestion des données

**Soutenir la reproductibilité scientifique** en orientant les chercheurs dans leurs démarches de gestion des données



# Gestion des données et reproductibilité

**Soutenir la reproductibilité scientifique** en orientant les chercheurs dans leurs démarches de gestion des données

**Aider tout chercheur qui le souhaite à :**

Remplir, suivre et exécuter son plan de gestion de données

Connaître et respecter les réglementations applicables

Se repérer dans les considérations éthiques liées aux données de la

Sécuriser, organiser et rationaliser le stockage de données

Favoriser la réutilisabilité des données : format, stockage, documentation, métadonnées, accès

Organiser l'archivage des données à long terme

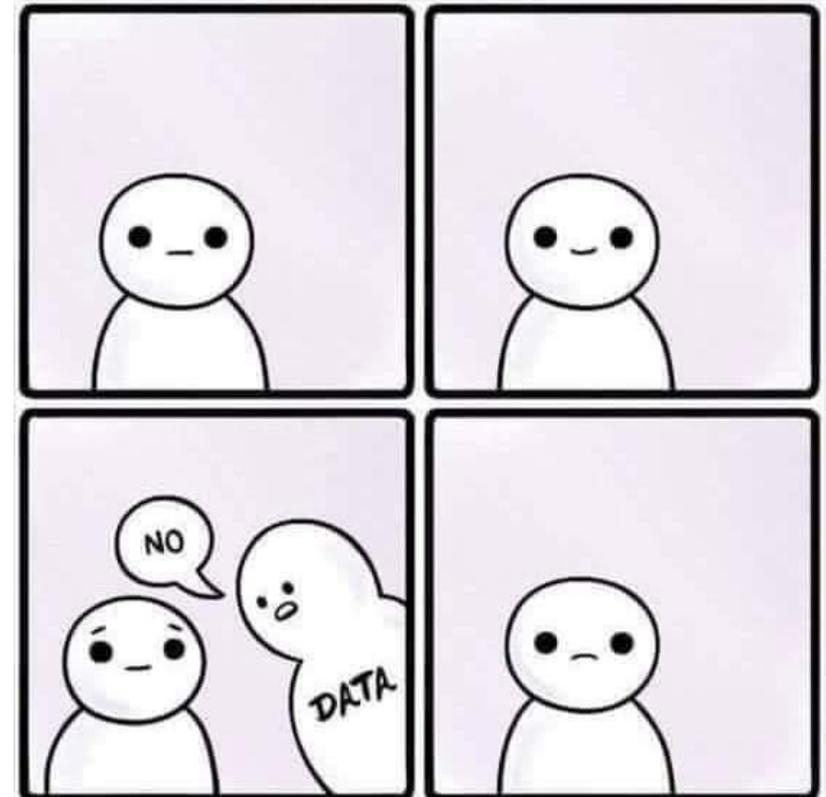
Repérer, développer, déployer et améliorer les outils nécessaires à toutes ces démarches

Former les chercheurs ULiège à la gestion responsable des données de la recherche

Choisir le meilleur mode de partage des données (répertoire)

[...]

The (real) scientific method.



# Gestion des données : quelques bonnes pratiques

**Soutenir la reproductibilité scientifique** en orientant les chercheurs dans leurs démarches de gestion des données

## Exemples réels du travail de RDO : le « de-bunk »

*« Pour appuyer mon article,  
mieux vaut publier les données  
qui le sous-tendent directement  
avec l'article, auprès de  
l'éditeur. »*

# Gestion des données : quelques bonnes pratiques

**Soutenir la reproductibilité scientifique** en orientant les chercheurs dans leurs démarches de gestion des données

## Exemples réels du travail de RDO : le « de-bunk »

*« Pour appuyer mon article, mieux vaut publier les données qui le sous-tendent directement avec l'article, auprès de l'éditeur. »*

Publication éditeur = outsourcing = perte de la mainmise sur le patrimoine que représente vos données

Bonne pratique = **garder une copie** « in-house » + choisir un bon **répertoire de données**

-> possibilité d'un **identifiant permanent** (DOI, handle ...)

-> quelques possibilités de **licences**

-> métadonnées **standardisées** (« galaxies » = « Galaxies » = « Galaxy » = « Galactic »...)

-> **reconnu** et **utilisé** dans la discipline

Génériques : [Zenodo](#), [OSF](#), [Harvard Dataverse](#), répertoires institutionnels, ...

Spécifiques : Astro: [CDS](#), Genomics: [NCBI](#), Particles: [HEPData](#), ...

Catalogue de répertoires : [Re3data](#), [FAIRsharing](#)...

# Gestion des données : quelques bonnes pratiques

**Soutenir la reproductibilité scientifique** en orientant les chercheurs dans leurs démarches de gestion des données

## Exemples réels du travail de RDO : le « de-bunk »

*« Si je ne veux pas / peux pas partager mes données ouvertement, le FAIR data ne s'applique pas du tout à ma recherche. »*

# Gestion des données : quelques bonnes pratiques

**Soutenir la reproductibilité scientifique** en orientant les chercheurs dans leurs démarches de gestion des données

## Exemples réels du travail de RDO : le « de-bunk »

« *Si je ne veux pas / peux pas partager mes données ouvertement, le FAIR data ne s'applique pas du tout à ma recherche.* »

Data FAIRness

Confidentiel

Open data

Le FAIR data est un ensemble de 12 principes qui incarnent la maxime « *as open as possible, as closed as necessary* »

Ils permettent de se **positionner entre** « confidentiel » et « open data »  
Ils ne concernent pas que le partage mais aussi le **stockage**, la **conservation**, la **traçabilité**, la **réutilisabilité** au sein du labo et par soi-même...

Exemples : **documenter** les données en anglais, suivant des **standards**, dans un **format flexible et ouvert**, **identifiées** clairement (PID, nom, code, ...), sur un **stockage pérenne** (serveur UR ou univ et pas disque dur portable)

**Reproductibilité ≠ transparence complète**

# Gestion des données : quelques bonnes pratiques

**Soutenir la reproductibilité scientifique** en orientant les chercheurs dans leurs démarches de gestion des données

## Exemples réels du travail de RDO : le « de-bunk »

*« Comme j'utilise des données personnelles ou sensibles, il m'est impossible de partager mes données. »*

# Gestion des données : quelques bonnes pratiques

**Soutenir la reproductibilité scientifique** en orientant les chercheurs dans leurs démarches de gestion des données

## Exemples réels du travail de RDO : le « de-bunk »

« *Comme j'utilise des données personnelles ou sensibles, il m'est impossible de partager mes données.* »

La plupart des répertoires de données permettent le partagent *a minima* de métadonnées, sans fichiers associés

**Publier les métadonnées permet de « planter un drapeau »** (à condition de bien conserver et protéger ses données)

Une fois **anonymisées**, des données personnelles peuvent être partagées  
MAIS une bonne anonymisation est **rare et difficile...**

*AOL search records (2006)...*

*Publication 20M requêtes par 500k utilisateurs mal anonymisés*

*[Journalistes](#) identifient Thelma Arnold, Liburn, GA*

*[Netflix Prize Dataset](#) (2007)...*

# Gestion des données : quelques bonnes pratiques

**Soutenir la reproductibilité scientifique** en orientant les chercheurs dans leurs démarches de gestion des données

## Exemples réels du travail de RDO : le « de-bunk »

*« En sciences fondamentales, contrairement aux sciences humaines ou de la santé, on est rarement soumis à des réglementations »*

# Gestion des données : quelques bonnes pratiques

**Soutenir la reproductibilité scientifique** en orientant les chercheurs dans leurs démarches de gestion des données

## Exemples réels du travail de RDO : le « de-bunk »

*« En sciences fondamentales, contrairement aux sciences humaines ou de la santé, on est rarement soumis à des réglementations »*

Sciences appliquées : données numériques, IoT, données de consommation, ...

⇒ Données sensibles, RGPD

Géographie, sciences de la Terre : images satellites, données de trafic, ...

⇒ Utilisation à mauvais escient (exemple de Gembloux)

Biologie, agronomie : ressources génétiques

⇒ Protocole de Nagoya

Ne pas oublier le Dual Use...

...et les autres obligations (éditeurs, financeurs, contractuelles, éthiques, stratégiques...)

# À vous de jouer!

## Que puis-je faire en tant que chercheur ?

Rester en contact pour toute question GDR

S'outiller (dmponline.be, more to come)

Se former

Formations périodiques ULiège, sur demande, ...

## Participer au [réseau FWB de Data Ambassadors](#)

- + Accès à des évènements, formations, espaces de réseautage
- + Visibilité, plus-value carrière
- + Co-crédation d'outils et de service avec les RDOs

**RDM @ Bicfb**  
*All you need to know about data management*

Home Plan Data collection Search for existing datasets Manage Save Protect your data Publish & share Get visible & promote **Contact us** **Train & Learn** Data Ambassadors

### MAIN MENU

- > Home
- > Plan
- > Data collection
- > Search for existing datasets
- > Manage
- > Save
- > Protect your data
- > Publish & share
- > Get visible & promote
- > Contact us
- > Train & Learn
- > **Data Ambassadors**

## Data Ambassadors

Réseau de Data Ambassadors de la Fédération Wallonie-Bruxelles

### Pourquoi un réseau de Data Ambassadors ?

- Les bonnes pratiques de gestion des données de la recherche, les dépôts de données et les prérequis éthiques **varient selon les disciplines scientifiques**.
- La littérature montre que les transmissions de savoirs et d'expériences sont plus efficaces **entre pairs**.

Depuis 2021, les Data Ambassadors de la Fédération Wallonie Bruxelles sont organisés en **réseau** ! Les Data Ambassadors des six universités de la FWB forment un réseau collaboratif au sein duquel chacun peut se **former, s'informer, s'inspirer et inspirer ses collègues** en termes de gestion ouverte et FAIR des données de la recherche.

Ce sont des chercheurs, doctorant·e·s, logisticien·ne·s, informaticien·ne·s, membres de la communauté universitaire.

### Quel est leur rôle ?

- **Promouvoir les bonnes pratiques** de gestion des données de la recherche dans leur discipline scientifique et contribuer à leur adoption au sein de la FWB.
- Organiser une ou plusieurs **activités de sensibilisation** aux questions de FAIR data, de collecte, stockage, partage, archivage, réutilisation, protection, traitement ou analyse de données destinée à leurs collègues.
- Dialoguer avec les Research Data Officers pour **relayer les problèmes** courants rencontrés et contribuent ainsi à l'optimisation des outils et services institutionnels mis à disposition des chercheurs en FWB.

### Pourquoi devenir Data Ambassadors ?

- Pour obtenir l'accès à des **formations et évènements** qui permettent d'accroître ses compétences et sa visibilité.
- Parce que la gestion des données de la recherche devient un pilier de l'évaluation de la carrière des chercheurs, et la mission de Data Ambassador s'aligne dans cette dynamique, ce qui représente **une plusvalue pour une carrière** de chercheur



Merci !

**Judith BIERNAUX**

**ULiège RISE - Recherche, Innovation, Support et Entreprises**  
Research Data Officer  
Head of Research Management Unit

**[jbiernaux@uliege.be](mailto:jbiernaux@uliege.be)**

# Sources

J. Ioannidis, 2005, Contradicted and Initially Stronger Effects in Highly Cited Clinical Research, JAMA. 2005;294(2):218-228. [doi:10.1001/jama.294.2.218](https://doi.org/10.1001/jama.294.2.218)

Ioannidis JPA (2005) Why Most Published Research Findings Are False. PLoS Med 2(8): e124. [doi.org/10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124)

Mark Otto Baerlocher et al., 2010, Data integrity, reliability and fraud in medical research, Elsevier European Journal of Internal Medicine 21 (2010) 40–45. [doi.org/10.1016/j.ejim.2009.11.002](https://doi.org/10.1016/j.ejim.2009.11.002)

Monya Baker, 1,500 scientists lift the lid on reproducibility, Nature 533, 452–454 (26 May 2016) [doi:10.1038/533452a](https://doi.org/10.1038/533452a)

Christian T-M, Gooch A, Vision T, Hull E (2020) Journal data policies: Exploring how the understanding of editors and authors corresponds to the policies themselves. PLoS ONE 15(3): e0230281. [doi.org/10.1371/journal.pone.0230281](https://doi.org/10.1371/journal.pone.0230281)

“Three Camps, One Destination: The Intersections of Research Data Management, FAIR and Open”. Higman, Rosie, Daniel Bangert, and Sarah Jones. 2019. *Insights* 32 (1): 18. DOI: [doi.org/10.1629/uksg.468](https://doi.org/10.1629/uksg.468)

AOL, Netflix and the end of open access to research data, Chris Soghoian, Dec. 1, 2007 12:14 p.m. PT, [consulted on Nov 22, 2022](#)

A. Narayanan and V. Shmatikov, "Robust De-anonymization of Large Sparse Datasets," 2008 IEEE Symposium on Security and Privacy (sp 2008), 2008, pp. 111-125, doi: 10.1109/SP.2008.33. (<https://arxiv.org/pdf/cs/0610105.pdf>)

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

B. A. Nosek, G. Alter et al., Promoting an open research culture, *Science* Volume 348, Issue 6242 Jun 2015, <https://doi.org/10.1126/science.aab2374>