



**HAL**  
open science

## Les principes FAIR

Sébastien Guyader, Frédéric Bray

► **To cite this version:**

Sébastien Guyader, Frédéric Bray. Les principes FAIR. Ateliers de la FAIRisation, Jun 2024, Petit Bourg, Guadeloupe, France. hal-04642345

**HAL Id: hal-04642345**

**<https://hal.inrae.fr/hal-04642345v1>**

Submitted on 9 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open licence - etalab

# Ateliers de la FAIRisation

## 1. Les principes FAIR

Sébastien Guyader 

sebastien.guyader@inrae.fr

Frédéric Bray 

frederic.bray@inrae.fr

13 juin 2024



LICENCE OUVERTE  
OPEN LICENCE

Ce travail est sous [Licence Ouverte](#).

# Plan

## Introduction

- Répliquer / Reproduire
- Historique : la “crise de la reproductibilité”
- Sortir de la crise

## Les principes FAIR

- F...
- A...
- I...
- R...

# La “crise de la reproductibilité”

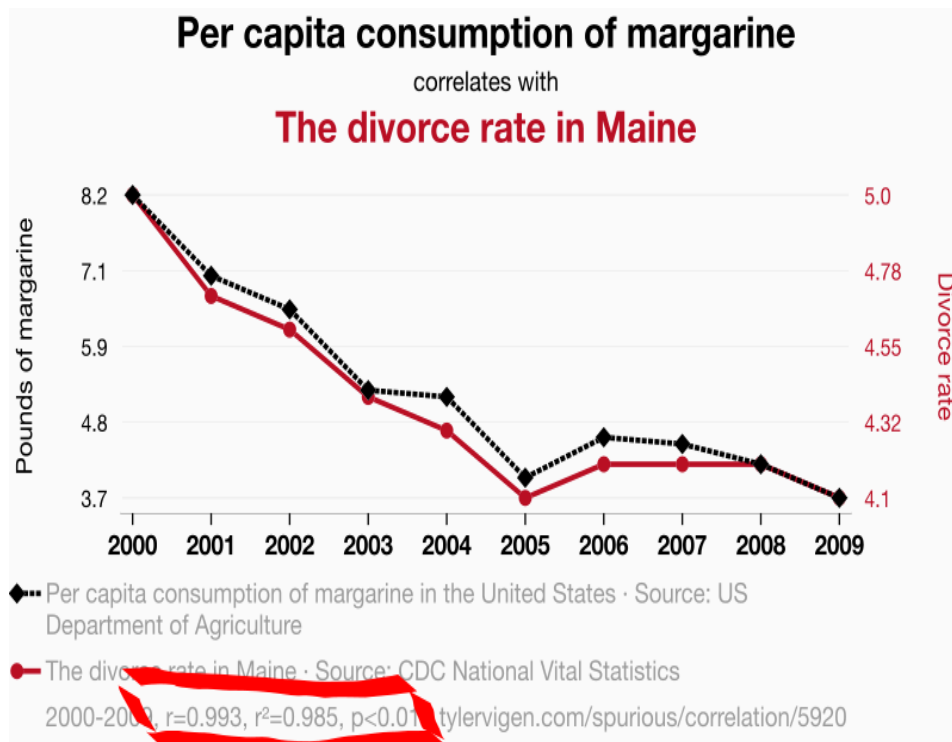
# Reproductibilité : pierre angulaire de la science ?

- **Répliquer** un résultat : reproduire ce résultat dans les mêmes conditions
- On parle de **réplication** :
  - *directe* ou *exacte* : reproduire les procédures au plus proche des conditions de l'étude initiale
  - *systematique* : quelques paramètres accessoires varient
  - *conceptuelle* : des différences plus notables, afin de tester la genericité des résultats
- La **reproductibilité** consiste à **reproduire** des résultats à partir du même jeu de données

# Reproductibilité : stupeur et tremblements...

- Dans les années 2010, des études publiées ont doublement marqué les esprits :
  - de par des **conclusions marquantes...**
  - ... puis de par l'**impossibilité de reproduire ou répliquer** les résultats
- Constats :
  - des études ont estimé que seulement **11-20%** des résultats précliniques en oncologie pouvaient être répliqués
  - mise en évidence du phénomène de **p-hacking**

# Reproductibilité : ils sont gonflés !



19th century  
scientist

I must find the explanation for this phenomenon in order to truly understand Nature...



21st century  
scientist

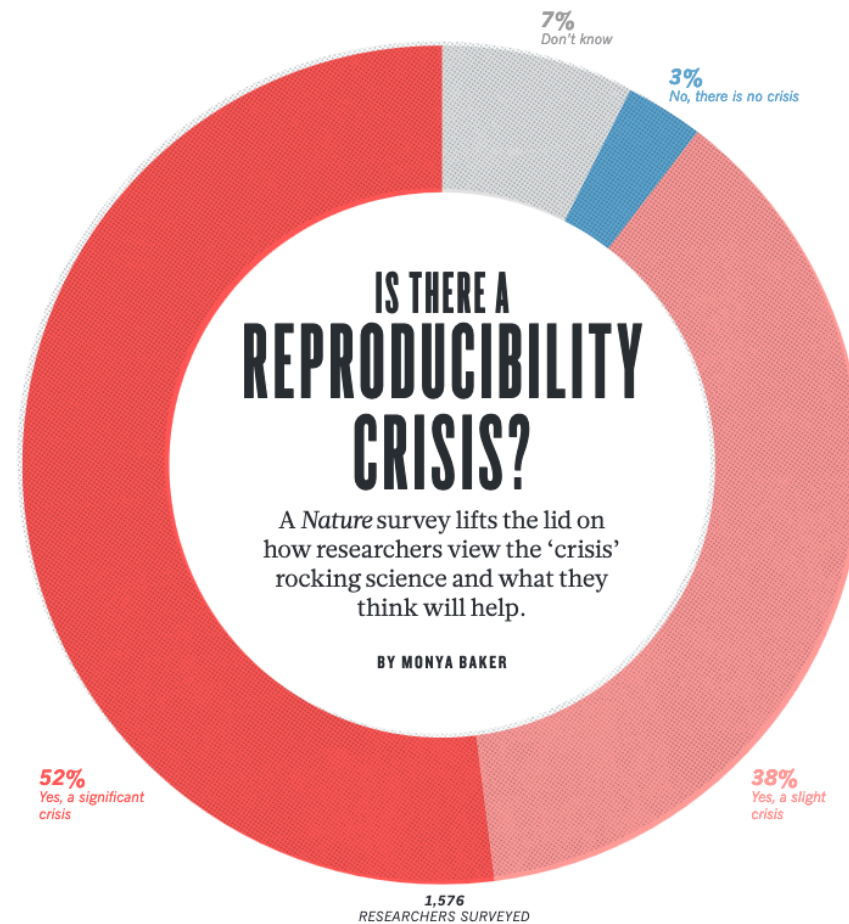
I must get the result that fits my narrative so I can get my paper into Nature..



facebook.com/pedromics

facebook.com/pedromics

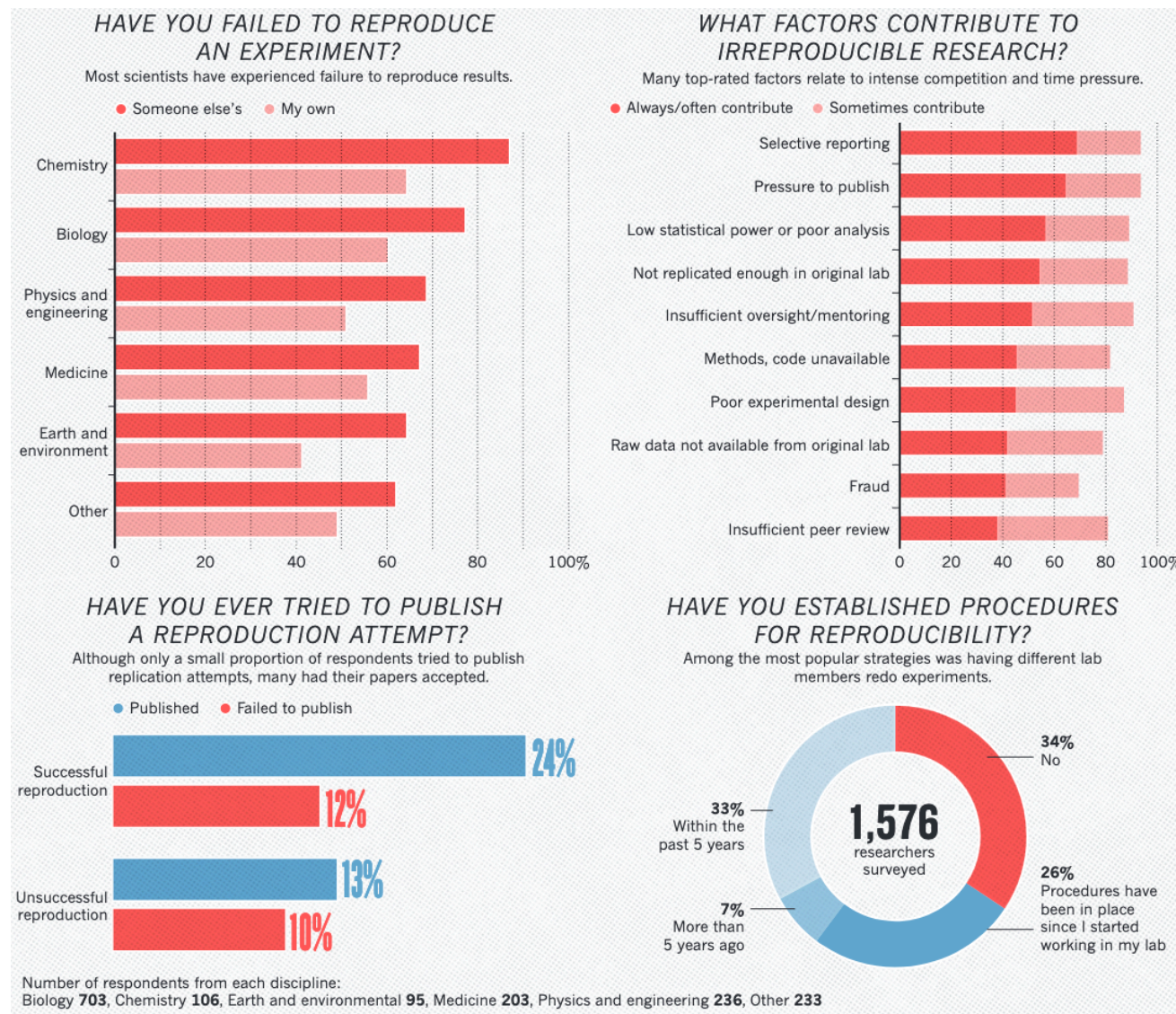
# Reproductibilité : la crise ?



Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454 (2016)



# Reproductibilité : la crise ?



# Reproductibilité : la crise ?

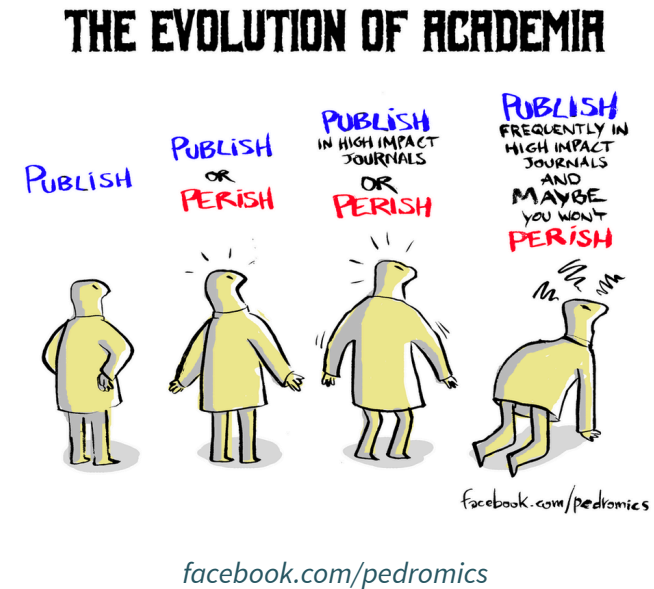
- Il est apparu que la faible reproductibilité est partagée au travers des disciplines : médecine, SHS, sciences environnementales...
- De plus en plus de scientifiques deviennent suspicieux : trituration ➡ fraude



E. Bik vs D. Raoult (Mark Harris pour BuzzFeed News, Getty Images)

# Reproductibilité : causes de la crise

- Les causes sont multiples :
  - **biais de publication :**  
*les résultats négatifs sont rarement publiés*
  - **“publish or perish” :**  
*publier toujours plus, augmenter son h-index...*
  - **standards de publication faibles :**  
*les revues ne sont pas toujours très regardantes*
  - on ne s’adresse pas toujours à **statisticien...**



facebook.com/pedromics

# Reproductibilité : sortir de la crise

- Changer nos pratiques en corrigeant les faiblesses de nos travaux :
  - avoir une approche réflexive
  - demander de l'aide si besoin
- Laisser à nos pairs et à tout un chacun la possibilité de scruter nos travaux
- Texte fondateur en 2016 :

SCIENTIFIC DATA 110110  
0111101  
11011110  
011101101

Amended: Addendum

**OPEN**

SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

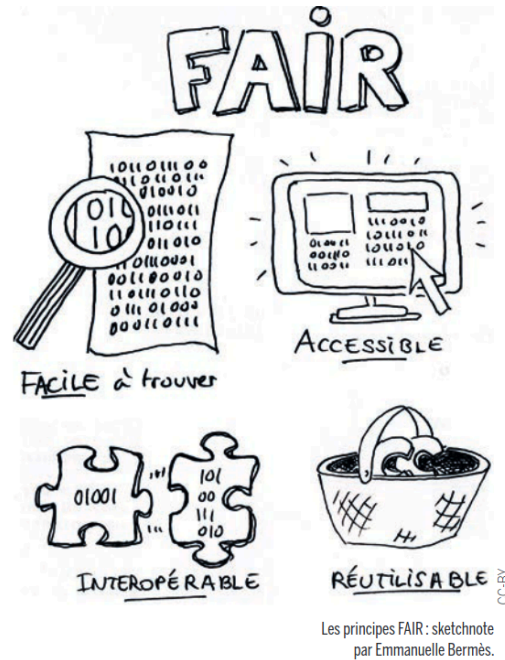
**Comment: The FAIR Guiding Principles for scientific data management and stewardship**

Mark D. Wilkinson *et al.*<sup>#</sup>

# Les principes FAIR

# Mais alors, que FAIRe ?


- Proposer une ligne directrice pour améliorer la **réutilisabilité** des données produites par la science
- Les scientifiques doivent faire en sorte que leurs données soient :

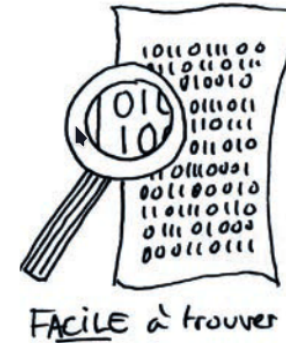


- 🔍 **F**aciles à trouver
- 👉 **A**ccessibles
- ⚙️ **I**nteropérables
- ♻️ **R**éutilisables

# Comment FAIRe ?

## 1. Facile à trouver

- Métadonnées
  - riches : décrire les données (et les méthodes)
  - donner des **identifiants uniques permanents** (PIDs) et les inclure dans les métadonnées
  - utiliser des entrepôts de données **indexés**, ex.  Data INRAE



Common Metadata	
Persistent Identifier	doi:10.57745/GNWXVA
Publication Date	2023-06-07
Title	Replication Data for: Diouf MB, et al. (2023). Molecular diversity of yam virus Y and identification of banana mild mosaic virus isolates infecting yam ( <i>Dioscorea</i> spp.)
Point of Contact	Use email button above to contact. Guyader, Sébastien (INRAE)
Author	Diouf, Mame Boucar (INRAE - CIRAD) - ORCID: 0000-0001-6314-611X Guyader, Sébastien (INRAE) - ORCID: 0000-0001-5438-8790 Umber, Marie (INRAE) - ORCID: 0000-0002-5759-0748 Teycheney, Pierre-Yves (CIRAD) - ORCID: 0000-0002-9754-0745

# Comment FAIRe ?

## 2. Accessible

- **Rendre l'accès possible**
  - accès par des **protocoles ouverts, gratuits et standardisés**
  - **informer des modalités d'accès** (autorisations, authentification...)
  - a minima, les **métadonnées** doivent être **accessibles librement**



The screenshot shows a file management interface with the following elements:

- Navigation tabs: Fichiers, **Métadonnées**, Conditions, Versions
- Search bar: Chercher dans ce jeu de données...
- Filter options: Type de fichier: Tout, Accès: Tout
- Grouping and Sorting: Group by Folder, Group by Tag, Trier
- File list: 1 à 4 de 4 Fichiers
- Files listed:
  - 0\_Dataset\_notice.pdf** (Adobe PDF - 33.1 Ko, 2 téléchargements)
  - Supplementary\_Table\_SII.4.tab** (Données tabulaires - 49.3 Ko, 0 téléchargement) - This file name and its download icon are circled in red.
  - Supplementary\_Table\_SII.5.tab** (Données tabulaires - 21.8 Ko, 0 téléchargement)
- Actions: Télécharger, Demander l'accès (circled in red)





# Comment FAIRe ?

## 4. Réutilisable

- But ultime du FAIR, les (méta)données doivent :
  - être décrites de la manière la plus complète
  - être accompagnées d'une licence claire concernant l'usage des données
  - se conformer aux standards du domaine



recherche Data Gouv > Data INRAE > Experimental - Observation - Simulation Dataverse >

### Replication Data for: Epidemiology of yam viruses in Guadeloupe: role of cropping practices and seed-tuber supply

Version 1.1

Diouf, Mame Boucar; Guyader, Sébastien; Gaspard, Olyvia; Francius, Eric; Teycheney, Pierre-Yves; Umer, Marie, 2022, "Replication Data for: Epidemiology of yam viruses in Guadeloupe: role of cropping practices and seed-tuber supply", <https://doi.org/10.57745/KD1GEB>, Recherche Data Gouv, V1, UNF:6:34HgP31WUFIMsliouqJtQ== [fileUNF]

[Cite Dataset](#) [Learn about Data Citation Standards](#)

**Description** This dataset reports the full data on virus indexing and field plots characterisation used in the article "Epidemiology of yam viruses in Guadeloupe: role of cropping practices and seed-tuber supply" English (2022-10-07)

**Subject** Agricultural Sciences

**Related Publication** Diouf MB, Guyader S, Gaspard O, Francius E, Teycheney P-Y, Umer M. Epidemiology of Yam Viruses in Guadeloupe: Role of Cropping Practices and Seed-Tuber Supply. Viruses. 2022; 14(11): 2366. <https://doi.org/10.3390/v14112366> doi: 10.3390/v14112366

**License/Data Use Agreement** etalab 2.0

**Access Dataset**

Contact Owner Share

Make Data Count (MDC) Metrics [?](#)  
since 2020-07-01

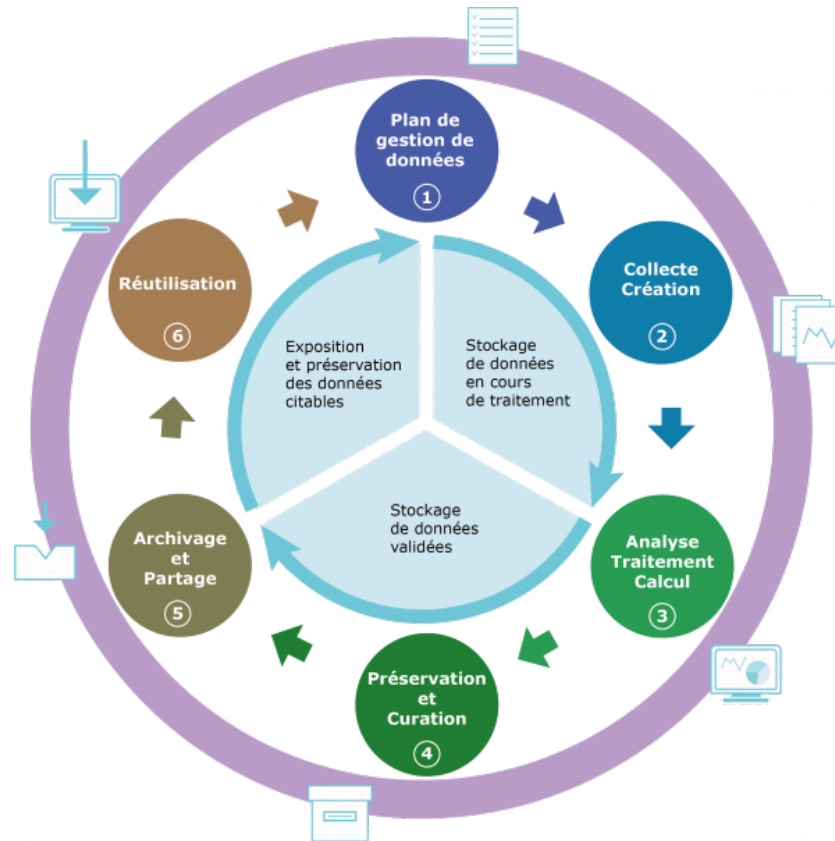
273 Views [?](#)

10 Downloads [?](#)

0 Citations [?](#)

# A nous de FAIRe !

## Intégrer les principes FAIR au cycle de vie des données



CC-BY Nicole Lambert / GRICAD / CNRS

- Importance de rédiger des PGD
- Prévoir comment et où **stocker les (méta)données**
- Choisir des outils et infrastructures de **stockage** qui répondent aux principes FAIR
- **Ouvrir autant que possible**, ne pas fermer plus que nécessaire
- Un exemple : **SYGADE**