

Favoriser la reproductibilité de la recherche

Réseau de recherche reproductible



Sarah Cohen-Boulakia
Université Paris-Saclay

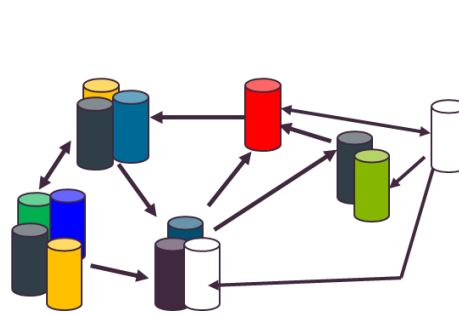


Analyse de données biologiques

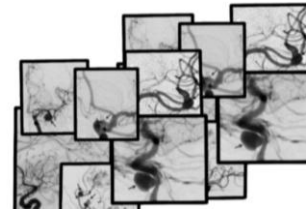
(1) Sources de données

Distribuées
Réseau hétérogène

> 1,000 (NAR)



```
TGACTTTTGTGACTTTTGGGC
GCCACTA TGACTTTTGGGC
TGGCGC CACTATGGCGC
```



(2) Outils – Scripts

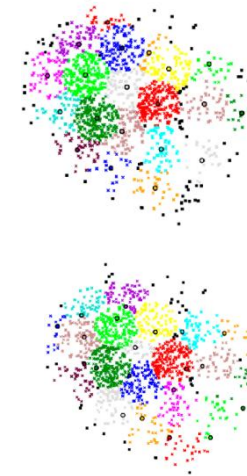
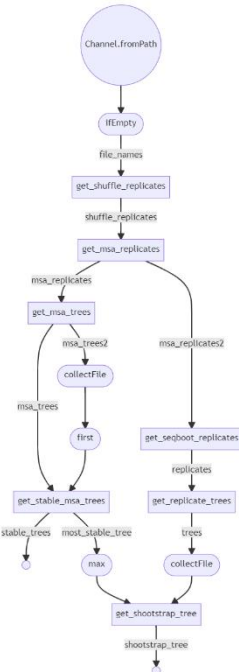
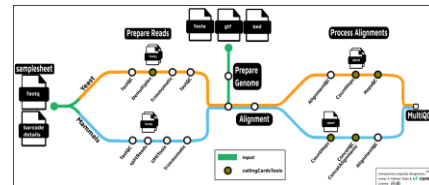
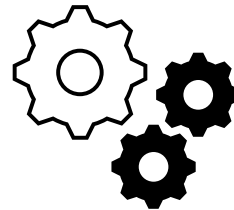
Distribués
Hétérogènes

> 30,000 (bio.tools)

(3) Pipelines d'analyse

Combinaison d'outils
Environnements &
plateformes variées

> 7,000 workflows
(GitHub, GitLab,)



*Quel jeu de données
d'entrée ai-je utilisé ?
Quel paramétrage ?
Quelles versions d'outils ?*

La crise de la reproductibilité

Chaque communauté a ses papiers de référence

Nekrutenko & Taylor - **Nature Genetics** (2012)

50 papers 2011 using the Burrows-Wheeler Aligner

31/50 (62%) provide no information

no version of the tool no parameters

no genomic reference sequence

7/50 (14%) provide all the necessary details

Alsheikh-Ali et al, **PLoS one** (2011)

10 papers in the top-50 IF journals → 500 papers

149 (30%) were not subject to any data availability policy (0% data available)

Of the remaining 351 papers

208 papers (59%) did not adhere to the data availability instructions

143 make a statement of willingness to share

47 papers (9%) deposited full primary raw data online



Must try harder
Too many sloppy mistakes are creeping into scientific papers, at the data — and at themselves.

Error prone
Biologists must realize the pitfalls massive amounts of data.

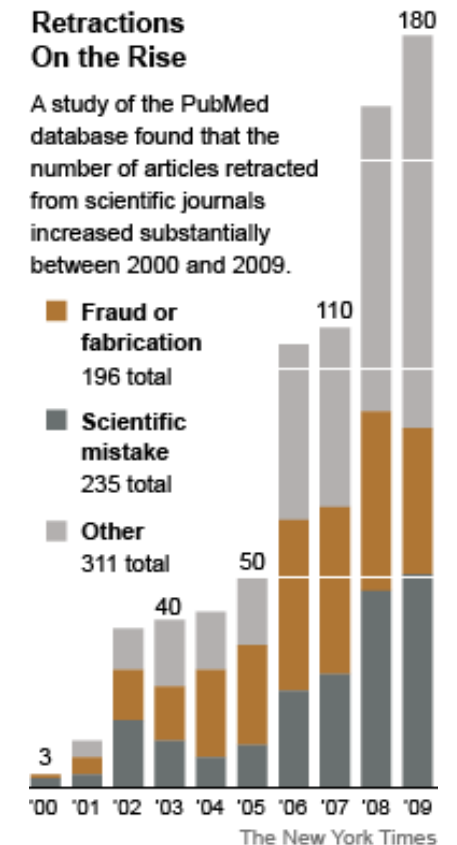
If a job is worth doing, it is worth doing twice
Researchers and funding agencies need to put a premium on ensuring that results are reproducible, argues Jonathan F. Russell.

Raise standards for preclinical cancer research
C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Six red flags for suspect work
C. Glenn Begley explains how to recognize the preclinical papers in which the data won't stand up.

Know when your numbers are significant

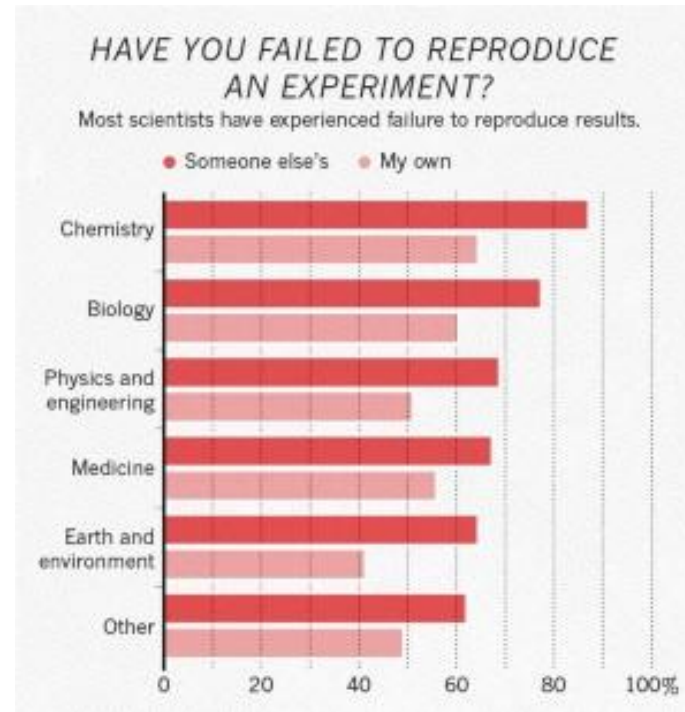
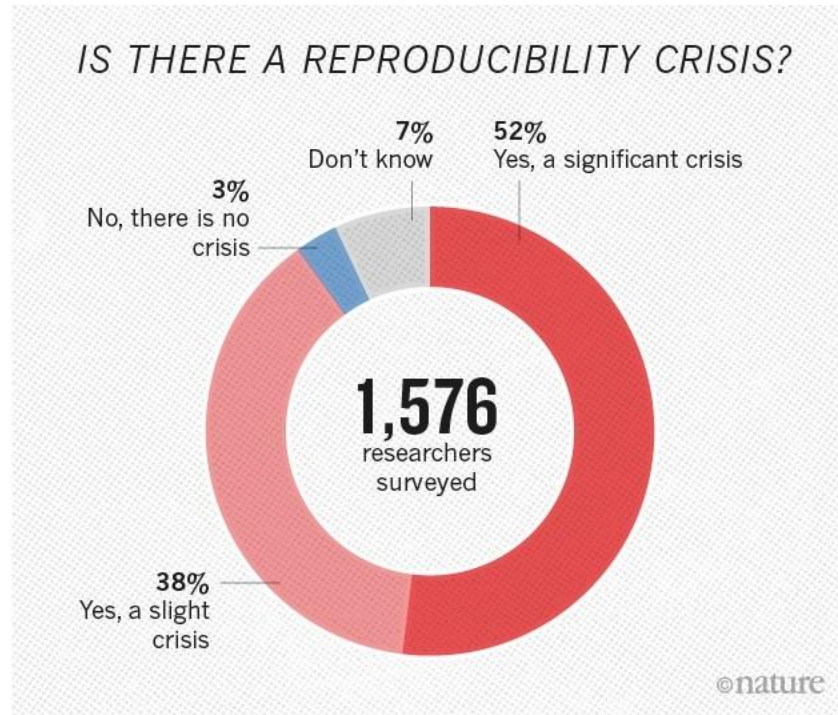
47/53 "landmark" publications could not be replicated
[Begley, Ellis Nature, 483, 2012]



New York Times 2011

Survey Nature

La crise a touché toutes les disciplines



nature

[Explore content](#) [About the journal](#) [Publish with us](#)

[nature](#) > [news feature](#) > [article](#)

News Feature | Published: 25 May 2016

1,500 scientists lift the lid on reproducibility

[Monya Baker](#)

[Nature](#) 533, 452–454 (2016) | [Cite this article](#)

209k Accesses | 2372 Citations | 5151 Altmetric | [Metrics](#)

Types de Reproductibilité

Reproductibilité Empirique

Informations détaillées sur les expériences

NB : Le chercheur contrôle le cadre de l'expérience

Son savoir faire peut entrer en jeu

Mise à disposition des données, méthodes de collecte des données

Reproductibilité Observationnelle

Informations détaillées sur les observations

NB : Le chercheur ne contrôle pas le cadre – il observe

Mise à disposition des données, méthodes de collecte des données

Le Monde

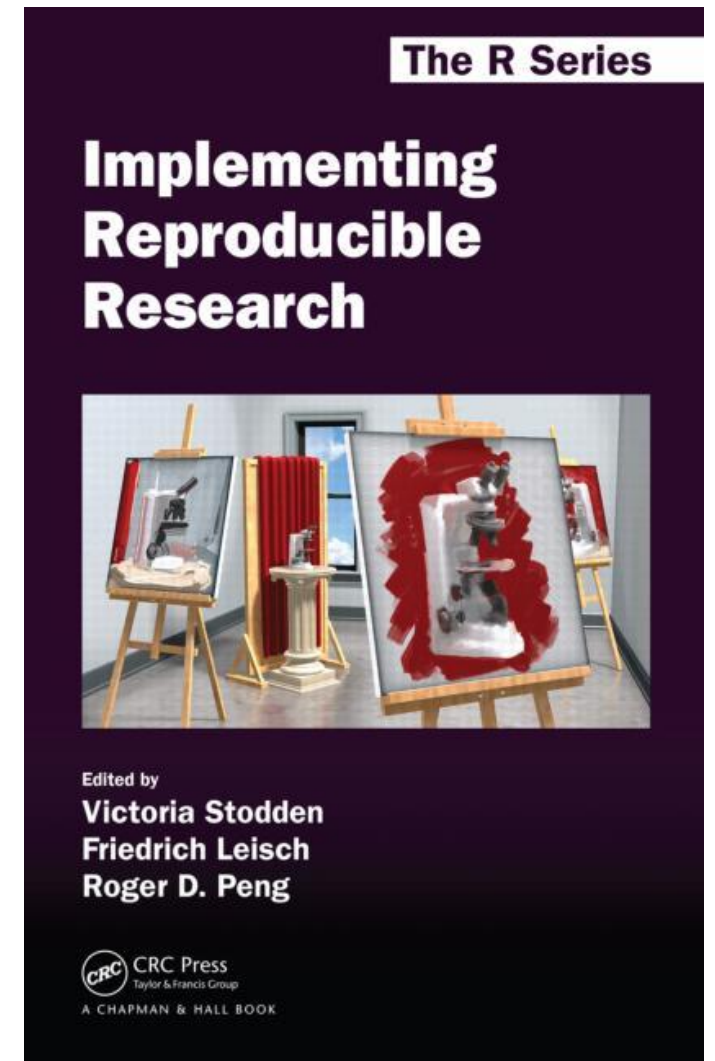
SCIENCES • PHYSIQUE

Des meringues pour faire goûter la démarche scientifique

Afin d'expliquer au grand public le défi que représente la reproductibilité des résultats en sciences, une équipe rennaise a choisi une approche originale : utiliser la confection de pâtisseries comme terrain d'expérimentation.

Par David Larousserie

Publié le 23 décembre 2024 à 06h00, modifié le 23 décembre 2024 à 15h17 • Lecture 3 min.



Types de Reproductibilité

Reproductibilité Empirique

Informations détaillées sur les expériences NB : Le chercheur contrôle le cadre de l'expérience Son savoir faire peut entrer en jeu Mise à disposition des données, méthodes de collecte des données

Reproductibilité Observationnelle

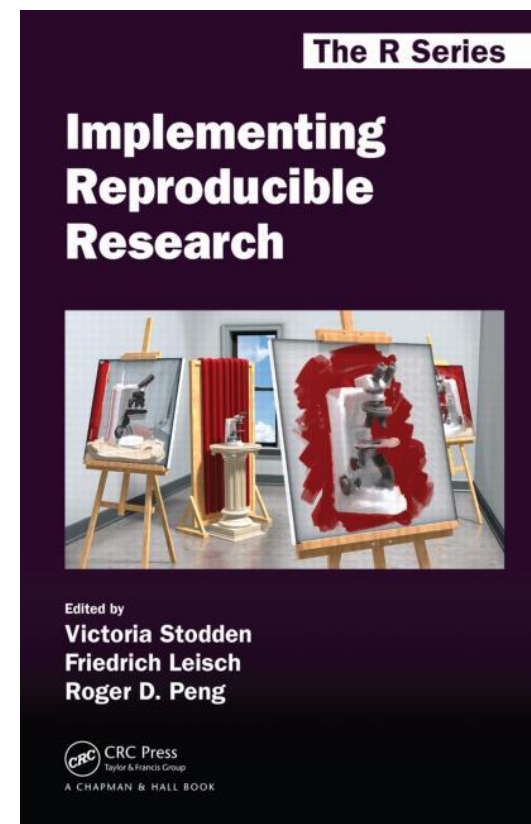
Informations détaillées sur les observations NB : Le chercheur ne contrôle pas le cadre – il observe Mise à disposition des données, méthodes de collecte des données

Reproductibilité Statistique

Informations détaillées sur le choix des tests statistiques, les paramètres des modèles, les seuils de décision... Pré-enregistrement du design de l'étude pour prévenir la manipulation des p-valeurs et autres manipulations

Reproductibilité Computationnelle

Informations détaillées sur le code, le logiciel, le matériel et les détails de l'implémentation. Documenter comment les données ont été produites



Remarques

Un même résultat scientifique peut avoir trait à **plusieurs types de reproductibilité**

E.g., Épidémiologie : reproductibilité observationnelle, statistique, computationnelle

La reproductibilité statistique ne concerne pas que les mathématiciens et la reproductibilité computationnelle pas que les informaticiens

Science des données touche un nombre toujours croissant de disciplines

La reproductibilité computationnelle a longtemps été considérée comme facile à obtenir (**à tort**)

Il existe des *niveaux de* reproductibilité : redo – replicate – reproduce – reuse...

Niveaux de reproductibilité : cadre (proposition)

Niveau 1

Repeat

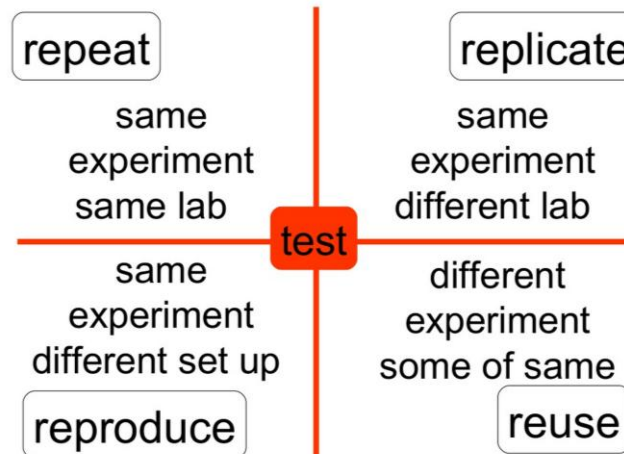
La traçabilité complète permet de *refaire* à l'identique

Mêmes données

Redo – Réexecute

But : capter le maximum d'*informations*

permettant d'expliquer un résultat



Drummond C Replicability is not Reproducibility: Nor is it Good Science, online
Peng RD, Reproducible Research in Computational Science *Science* 2 Dec 2011: 1226-1227.

Niveau 2

Replicate

On s'autorise quelques variations

Mêmes résultats – données similaires

But : on teste les limites d'une approche

Niveaux de reproductibilité : cadre (proposition)

Niveau 1

Repeat

La traçabilité complète permet de *refaire* à l'identique

Mêmes données

Redo – Réexecute

But : capter le maximum d'*informations*

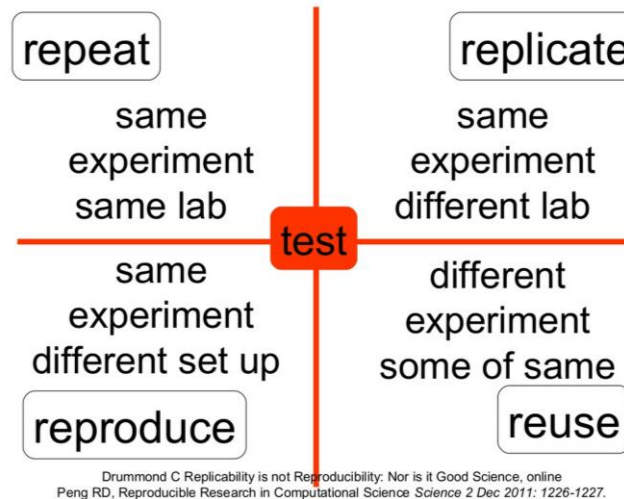
permettant d'expliquer un résultat

Niveau 3

Reproduce

Même résultat – même *inférence*

Mais les moyens/procédures/méthodes/données peuvent avoir changé



Niveau 2

Replicate

On s'autorise quelques variations

Mêmes résultats – données similaires

But : on teste les limites d'une approche

Reuse

On s'adapte à de nouveaux besoins

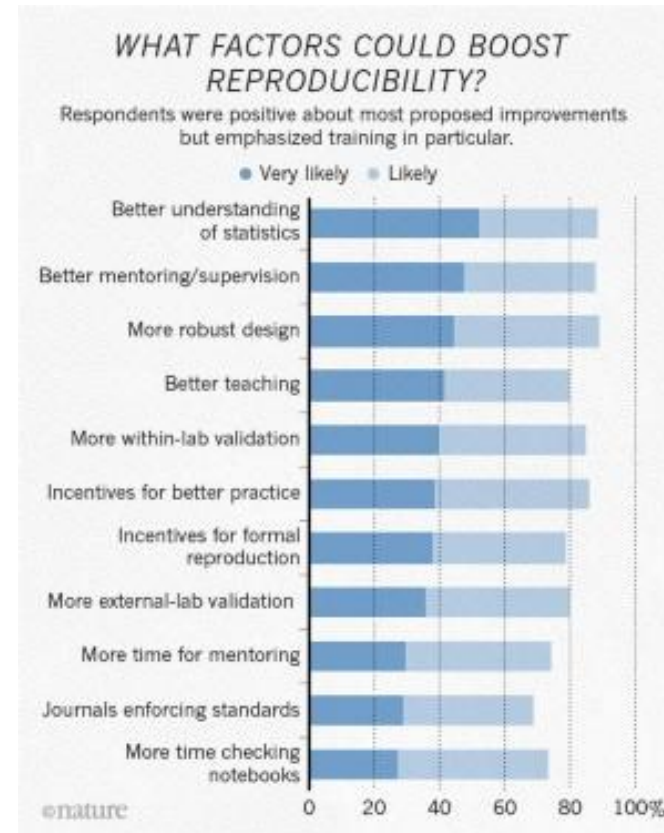
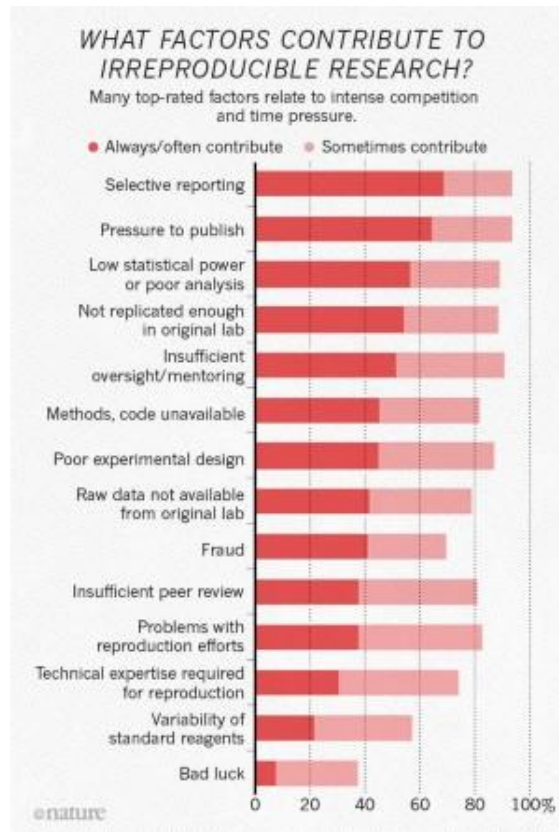
On réutilise en partie – dans un autre contexte

On peut obtenir un résultat différent

→ **Science Cumulative**

Survey Nature – Les raisons et les pistes de solutions

Besoins



Ralentir

Moins de pression à publier
Plus de temps pour des résultats plus robustes

Être formé

Compétences à acquérir
Bonnes pratiques
Echanges

Valoriser (*incentives*)

Journaux
Institutions

Les réseaux - *Reproducible Networks*

Global Networks

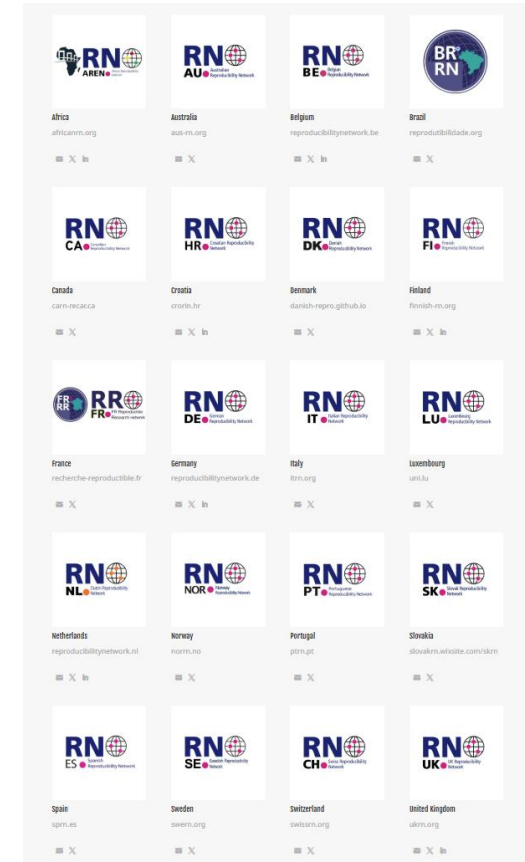
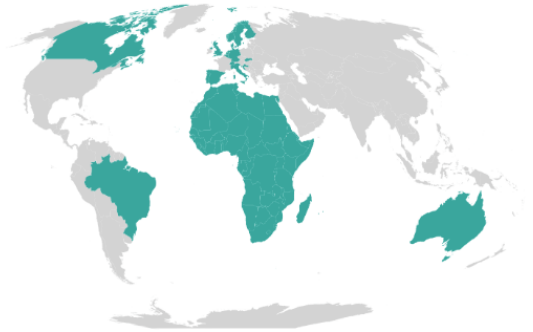
Outside the UK? Find a Reproducibility Network in your area

[See full Global Networks Statement](#)

Global Reproducibility Networks

A Reproducibility Network (RN) is a national, peer-led consortium of researchers that aims to promote and ensure rigorous research practices by establishing appropriate training activities, designing and evaluating research improvement efforts, disseminating best practice and working with stakeholders to coordinate efforts across the sector. RNs aim for broad disciplinary representation and an intensive interdisciplinary dialogue (e.g., with funding agencies, publishers, learned societies and other sectoral organisations, as well as researchers from all disciplines and across all career stages).

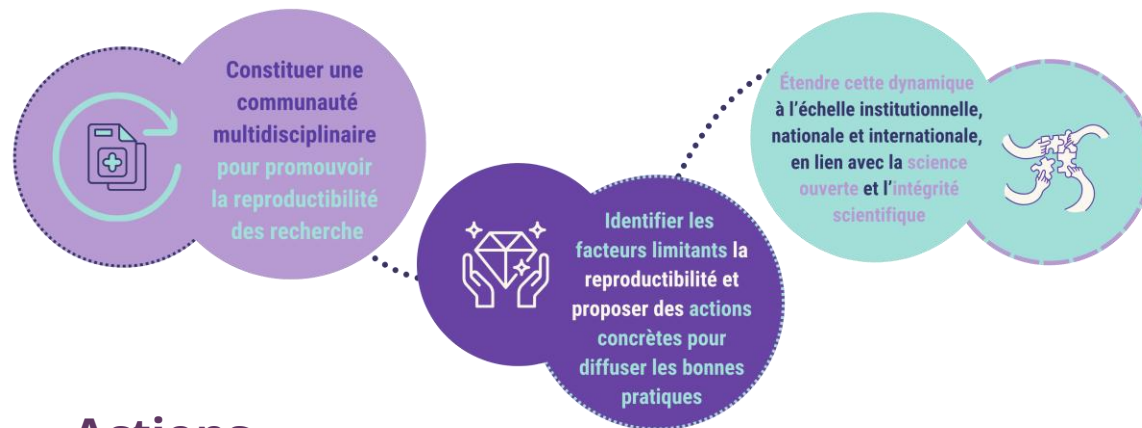
To reach as many researchers as possible, and to operate as efficiently as possible, we are keen to support other countries interested in creating similar networks. If you are interested in setting up a national RN, or finding out who in your country is working towards this, please email: contact@ukrn.org.



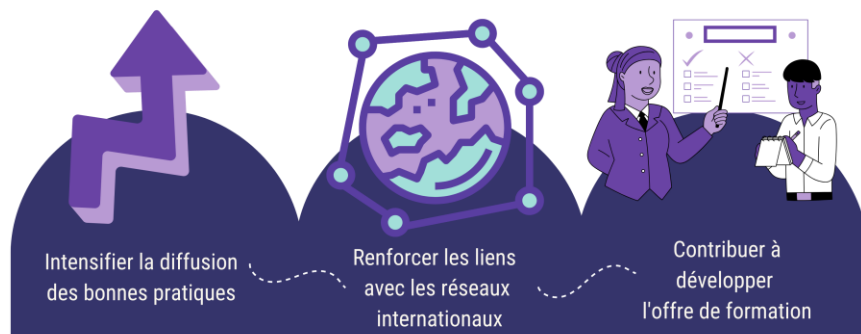
Le réseau Recherche Reproductible



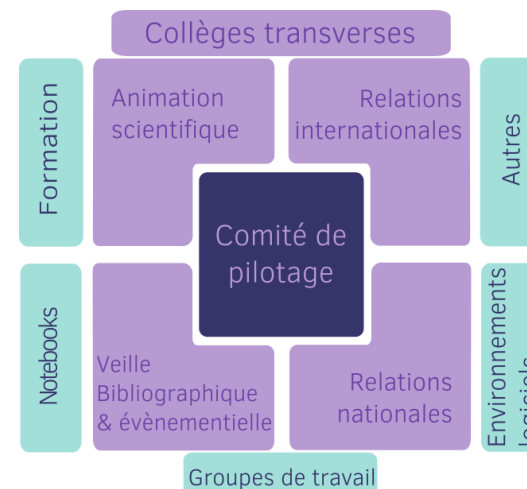
Objectifs



Actions



Structuration



Comité de pilotage



Céline Acary-Robert — Laboratoire Jean Kuntzmann / GRICAD



Sarah Cohen-Boulakia — Laboratoire Interdisciplinaire des Sciences du Numérique



Raphaëlle Krummeich — Laboratoire IDEES UMR6266



Arnaud Legrand — LIG



Frédéric Lemolne — Institut Pasteur



Dominique Muller — Laboratoire Interuniversitaire de Psychologie



Sébastien Rey-Coyrehourcq — Laboratoire IDEES UMR6266



François Ric — Laboratoire de Psychologie UR 4139



Nicolas P. Rougier — Institute of Neurodegenerative Diseases

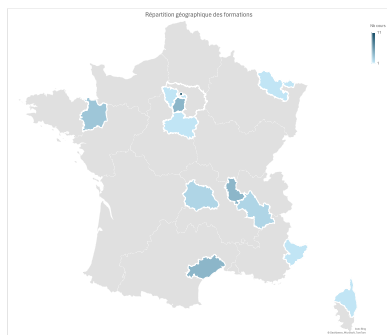
Notre réseau c'est **300+ membres**
30 disciplines : Informatique 20% - Bioinfo 10% - Physique 8% - Neurosciences 8%....

Exemple Action – Cartographie formations

Questionnaire envoyé aux membres du réseau

16 modules + 2 MOOC de reproductibilité computationnelle

Enseignés en pluridisciplinarité (pas informatique pure)



Objectifs

Cartographier les formations existantes

Partage de contenus

Créer un réseau de formation des formateurs

Site web dédié avec ressources disponibles



La vie du réseau – Actions en cours et à venir

Nombreux événements sur le site web

Webinaires

Dec 12, 2025 The African Reproducibility Network: Building Grassroots Capacity in Africa

Oct 10, 2025 The Brazilian Reproducibility Network: past, present, and future.

Conférences

Oct 3, 2025 Replication Games - Paris

SCIENCES • RECHERCHE SCIENTIFIQUE

Les « jeux de la réplication », ou quand la science s’amuse à se reproduire

Une vingtaine de chercheurs ont traqué, le 3 octobre, les failles de plusieurs articles de sciences sociales parus dans des revues de renom.



Le Monde

<https://www.recherche-reproductible.fr>

ANF “Workflows et reproductibilité en bioinformatique”, Paris, 25 au 27 novembre 2025.

Le réseau MétiER en bioinformatique **MERIT** et l’Institut Français de Bioinformatique (**IFB**) organisent une Action Nationale de Formation (ANF) sur les principes FAIR appliqués aux workflows bioinformatiques. Cette formation, portée par le CNRS (INS2I), vise à former des bioinformaticiens sur les bonnes pratiques de reproductibilité. Les participants acquièrent des compétences avancées en gestion de workflows (Nextflow et Snakemake), et maîtrisent les outils de dépôt (SWH, HAL) et d’intégration continue pour garantir la qualité de leur code. Le réseau français de la recherche reproductible est ainsi impliqué, renforçant l’adoption de méthodes rigoureuses en bioinformatique.



Conclusion – Science ouverte et reproductibilité

- La **science ouverte** pose un cadre
 - Rendre les résultats, les méthodes, les données et les outils de la recherche **aussi accessibles que possible - FAIR**
 - Sans science ouverte, la reproductibilité reste **théorique**
- **Reproductibilité** : un critère de qualité dans la science ouverte ?
- La recherche reproductible donne **de la valeur** à l'ouverture
 - Ce qui est ouvert **doit** être exploitable
 - Elle pousse à standardiser données, formats, métadonnées
 - Elle met en évidence les zones d'ombre (scripts manquants, doc insuffisante, dépendances oubliées...)
 - Les exigences de reproductibilité **structurent** la science ouverte