



Module 1 Séquence 2

Fred de Lamotte - Montpellier
<https://orcid.org/0000-0003-4234-1172>





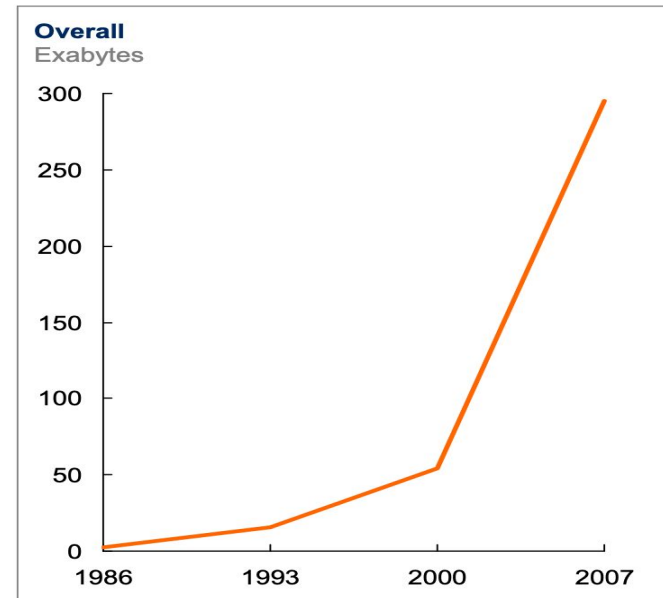
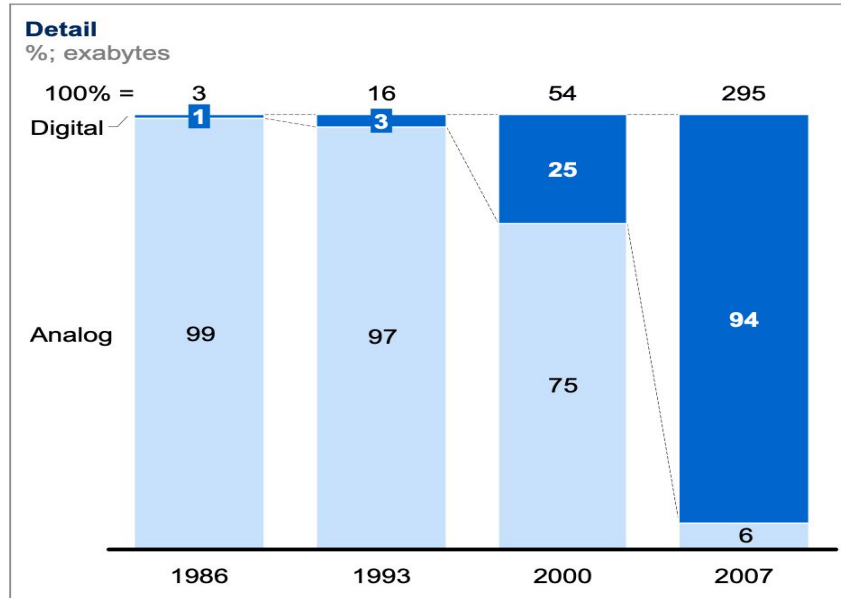
Crise de reproductibilité

30 minutes

Fred de Lamotte - Montpellier
<https://orcid.org/0000-0003-4234-1172>



La disruption numérique : une bascule brutale



Qui bouscule l'existant

- La première compagnie de taxi n'en possède aucun (Uber)
- Le premier fournisseur de logement n'en possède pas (AirBnB)
- La première compagnie de téléphonie ne possède pas de standard (Skype)
- Le premier fournisseur d'info ne crée pas de contenu (Facebook)
- Le premier diffuseur de film ne possède pas de salle de cinéma (Netflix)

Waves of Digital Disruption

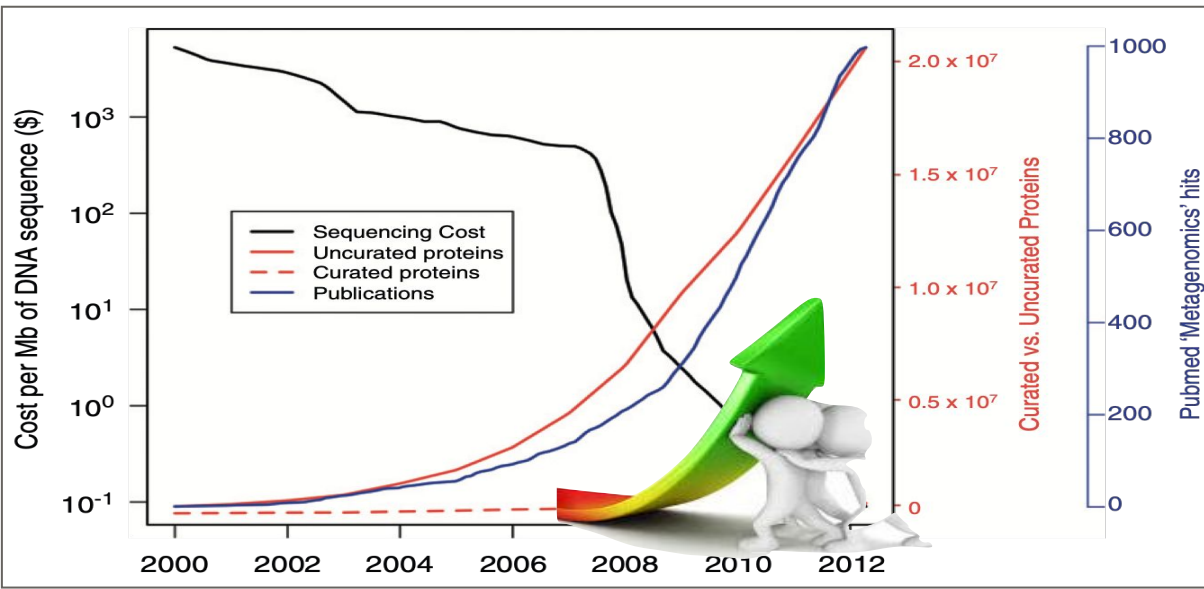


Le déluge des données en Science

Les techniques à haut débit, une révolution qui provoque un déluge de données
Génome humain :

en 1990 = **13 ans et 3 Milliards \$**

en 2015 = **quelques heures et 1000 \$**



1. La quantité de données à stocker et analyser explose
2. Le *rendement* d'analyse chute

Répondre aux questions de la Science

AVANT

- 1 Concevoir l'expérimentation
- 2 Collecter des résultats
- 3 Analyser des résultats

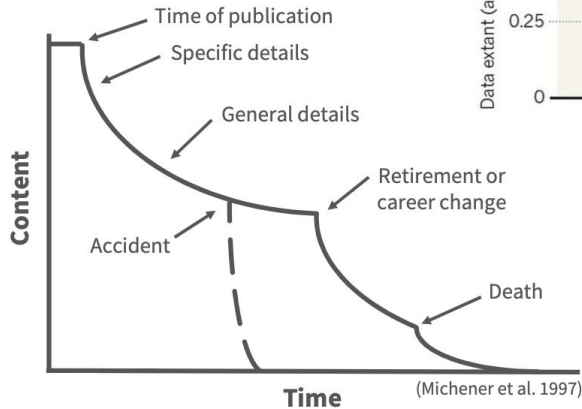
Un changement de paradigme

MAINTENANT

- 1 Générer massivement des données
- 2 Organiser (stocker, documenter, annoter)
- 3 Analyser (extraire de l'information)
- 4 Diffuser l'information

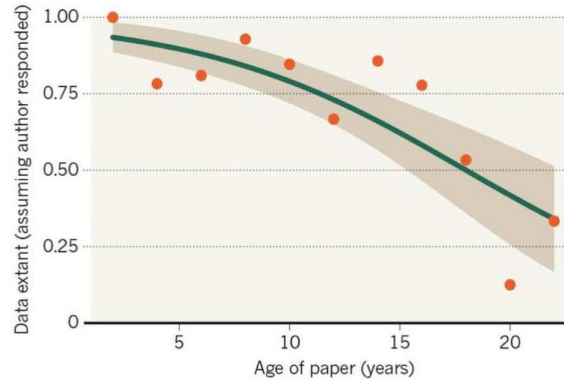
Les ravages du temps

Data Entropy



MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



Vines, T. H. et al. Curr. Biol. <http://dx.doi.org/10.1016/j.cub.2013.11.014> (2013).

Les défis de la reproductibilité

RESEARCH ARTICLE

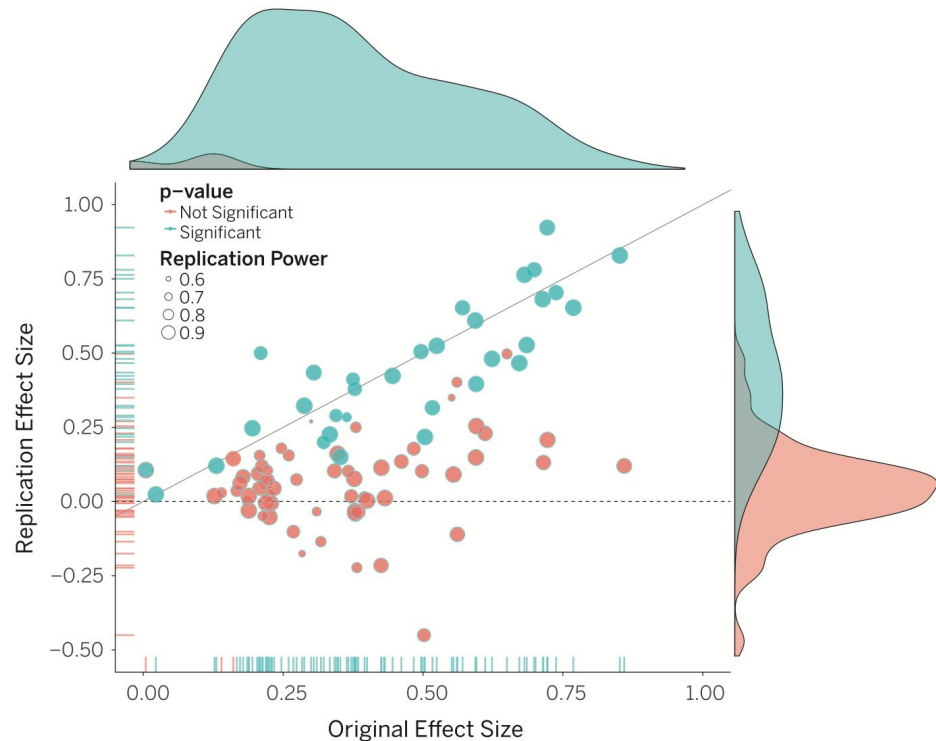
Estimating the reproducibility of psychological science

Open Science Collaboration^{*,†}
* See all authors and affiliations

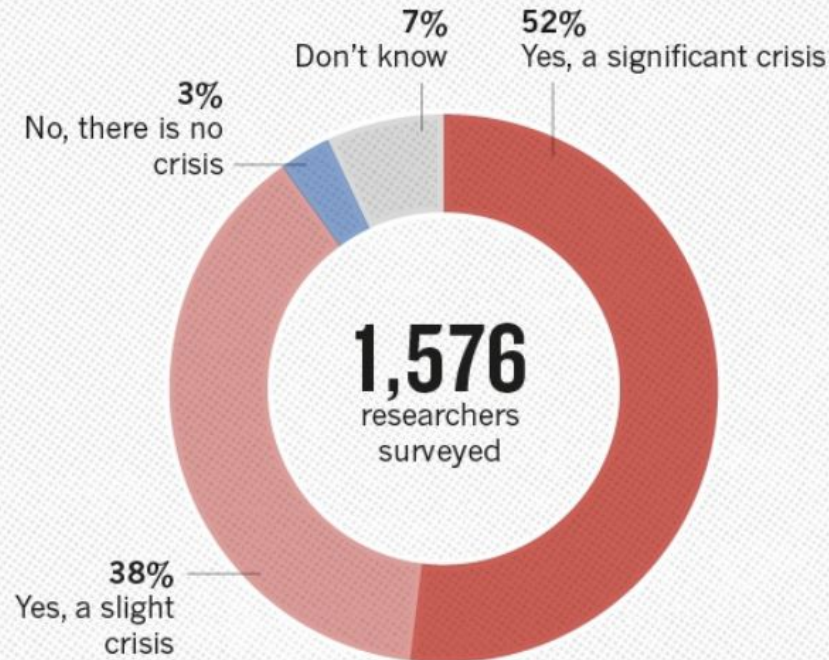
Science 28 Aug 2015;
Vol. 349, Issue 6251, aac4716
DOI: 10.1126/science.aac4716

The *Reproducibility project* set out to replicate 100 experiments published in high-impact psychology journals.

About one-half to two-thirds of the original findings could not be observed in the replication study.



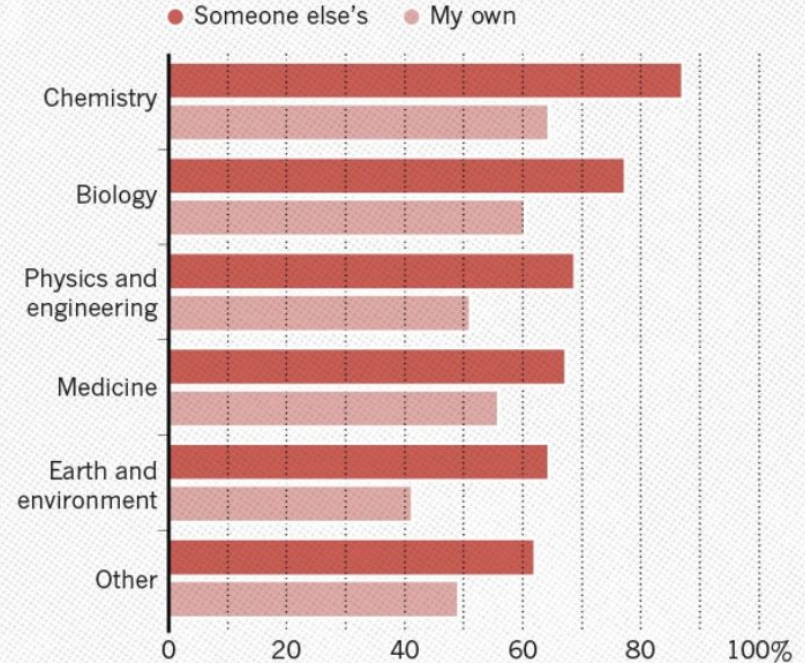
IS THERE A REPRODUCIBILITY CRISIS?



©nature

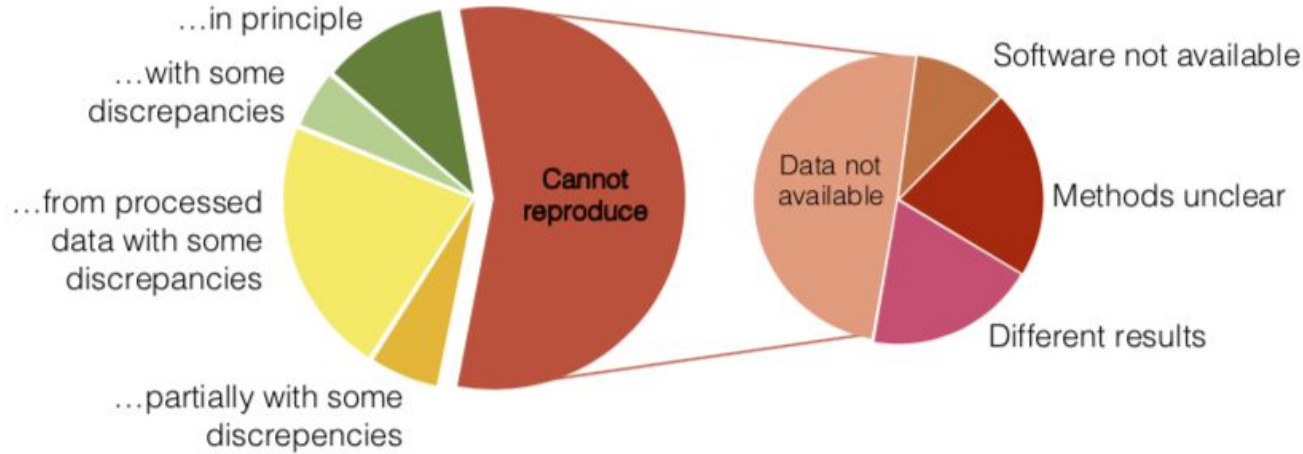
HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



Replication of data analyses in 18 articles on microarray-based gene expression profiling published in Nature Genetics in 2005–2006:

Can reproduce...



Summary of the efforts to replicate the published analyses.

Adopted from: Ioannidis et al. Repeatability of published microarray gene expression analyses
Nature Genetics **41** (2009) doi:10.1038/ng.295

Open Science

Open Data

Open Source

Open Methodology

Open Peer Review

Open Access

Open Educational Resources

Disruption + BigData + Crise

Le traitement de l'information (scientifique)
sera notre Noeud Gordien



Exercice 1.1 !

Quelle définition pour les données de la recherche ?

Pour tenter d'aborder cette question, nous allons procéder en 4 étapes :

1. Vous allez prendre un temps de réflexion individuelle de **5 minutes** pour rédiger sur le document partagé une première définition qui vous est personnelle.
2. Pendant les **5 minutes** suivantes, vous discuterez de vos définitions avec trois autres participants et proposerez une définition combinée, en gardant la trace des divergences s'il y a lieu. Donc vous finissez ces 5 minutes avec **1 définition**
3. Votre groupe prendra connaissance de l'ensemble des définitions consolidées proposées puis une discussion portera sur les divergences entre ces définitions (temps prévu **10 minutes**)
4. L'activité se terminera par la présentation des définitions les plus courantes des données de la Recherche



Définition OCDE

Les données de recherche sont les **preuves** qui sous-tendent la réponse à la question de recherche et peuvent être utilisées pour **valider** les **résultats**, quelle que soit leur forme (i.e. imprimée, numérique ou physique).

Il peut s'agir de **renseignements quantitatifs** ou d'**énoncés qualitatifs** recueillis par les chercheurs dans le cadre de leurs travaux par **expérimentation, observation, modélisation, entrevue** ou autres méthodes, ou de renseignements tirés de preuves existantes.

Les données peuvent être **brutes** ou **primaires** (par exemple, directement issues de mesures ou de collectes) ou **dérivées** de données primaires par analyse ou interprétation (e.g. nettoyées ou extraites d'un ensemble de données plus vaste), ou encore dérivées de sources existantes dont les droits peuvent être détenus par d'autres.