

# Enjeux et défis de la recherche reproductible

Konrad HINSEN

Centre de Biophysique Moléculaire, Orléans, France

13 juin 2017

# La science doit être vérifiable

## La nature humaine

- L'erreur est humaine.
- Les biais personnels sont en conflit avec la recherche de la vérité.

## La méthode scientifique

La science est un processus social pour éliminer les erreurs et les biais grâce à une vérification continue par les pairs.

## Une règle fondamentale

**Les affirmations non vérifiables sont irrecevables.**

- Décrire les expériences en détail pour qu'on puisse les refaire.
- Expliquer les raisonnements assez bien pour qu'un lecteur puisse les suivre pas par pas.

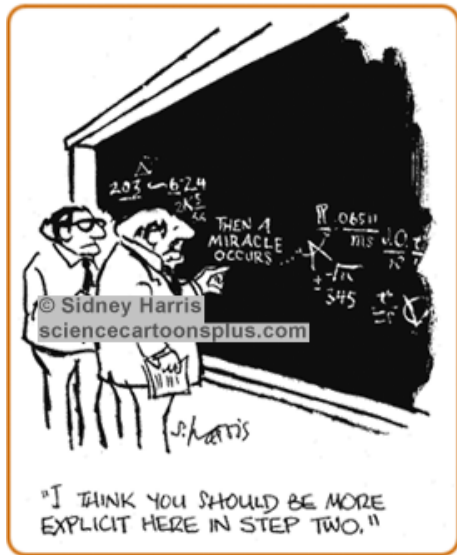
# Jamais croire sur parole

## Royal Society, 17ème siècle



*"It is an expression of the determination of Fellows to withstand the domination of authority and to verify all statements by an appeal to facts determined by experiment."* ([source](#))

# Se méfier des miracles



(S. Harris)

"Then a miracle occurs..."

*“Pour analyser nos observations, nous avons appliqué des techniques statistiques sophistiquées implémentées en langage R.”*

*“Nous supposons qu'il s'agit d'artefacts introduits par les algorithmes de pré-traitement intégrés dans les capteurs, qui ne sont pas documentés.”*

Peu de recherches se font sans calcul aujourd'hui, mais le calcul est très difficile à vérifier.

- Votre smartphone vous espionne? Comment savoir?
- Votre voiture triche lors des contrôles d'émissions?
- Que font Google/Apple/Facebook/Amazon/... avec vos données?

# Publier un calcul de façon vérifiable

## Le code source des logiciels *y compris les workflows*, ...

- Vérification par relecture - pénible mais au moins possible.
- Permet de trouver des erreurs et des différences par rapport aux modèles décrits dans l'article.

## Les données d'entrée *du paramètre au Big Data*

- Vérification par inspection ou analyse par ordinateur.
- Permet d'identifier des incohérences, traces de manipulation, etc.

## Des instructions pour refaire les calculs

- Permet de vérifier que le code source et les données fournis produisent réellement les résultats publiés.

# La “recherche reproductible”

Etiquette provocatrice proposée par [Jon Claerbout](#) dans les années 1990.

*“An article about a computational result is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result.”*

25 ans plus tard, la recherche reproductible reste l'exception!

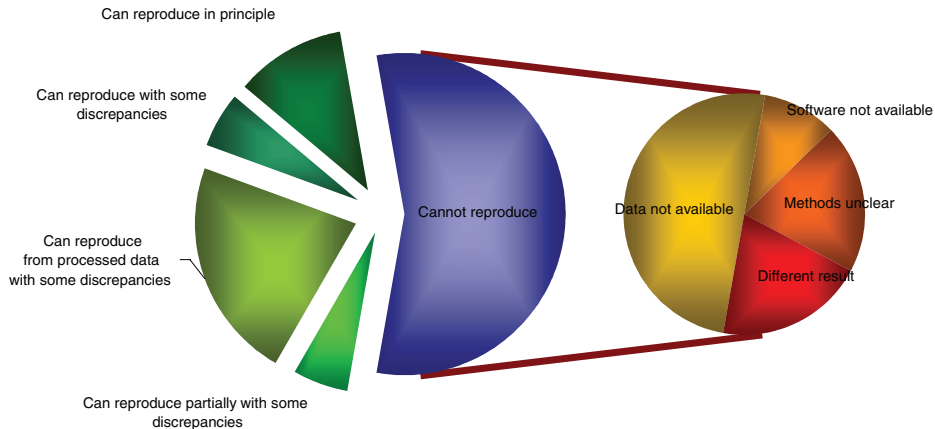


# Analyse de données en génétique (2009)

Ioannidis *et al.*

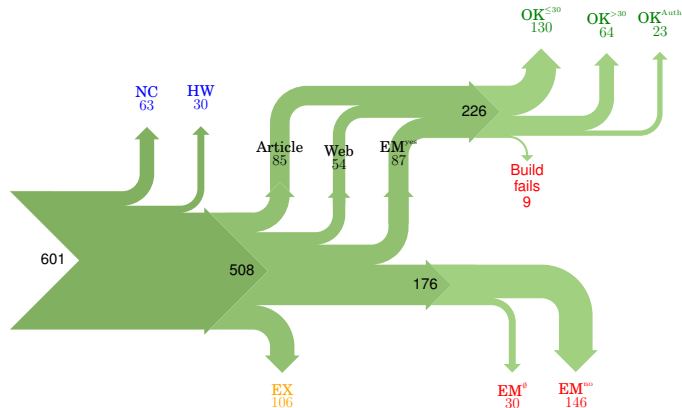
Repeatability of published microarray gene expression analyses

[Nature Genetics 41, 149-155 \(2009\)](#)



# Publications en informatique (2015)

Peut-on obtenir et faire fonctionner le code qui va avec une publication en informatique ?



Collberg, Proebsting & Warren, University of Arizona TR 14-04, Feb 2015  
<http://reproducibility.cs.arizona.edu/v2/RepeatabilityTR.pdf>

- Beaucoup d'outils, beaucoup de bons conseils...
- ... tous jeunes, immatures, en pleine évolution.
- Rien d'universellement applicable.
- Tout dépend de la nature des données, des algorithmes, de la nature de la recherche, etc.

## Un bon point de départ:

*The Practice of Reproducible Research*

*Case Studies and Lessons from the Data-Intensive Sciences*

Justin Kitzes, Daniel Turek, Fatma Deniz (Eds.)

<https://www.practicereproduciblesearch.org/>

## SCIENTIFIC PUBLISHING

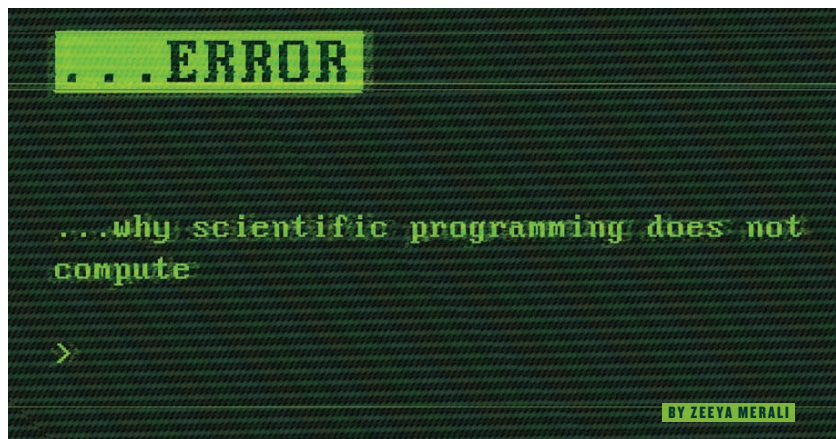
# A Scientist's Nightmare: Software Problem Leads to Five Retractions

Until recently, Geoffrey Chang's career was on a trajectory most young scientists only dream about. In 1999, at the age of 28, the protein crystallographer landed a faculty position at the prestigious Scripps Research Institute in San Diego, California. The next year, in a ceremony at the White House, Chang received a

2001 *Science* paper, which described the structure of a protein called MsbA, isolated from the bacterium *Escherichia coli*. MsbA belongs to a huge and ancient family of molecules that use energy from adenosine triphosphate to transport molecules across cell membranes. These so-called ABC transporters perform many

[G. Miller, Science 314 1856 \(2007\)](#)

... et omniprésente



Z. Merali, Nature **467**, 775 (2010)

- Les auteurs ont-ils fourni une description précise et complète?  
→ reproductibilité
- Y a-t-il des erreurs dans le code?  
→ vérification du code
- Les logiciels ont-ils été utilisés correctement?  
→ vérification du protocole de calcul

La reproductibilité n'est que le premier pas, et elle reste difficile.

La seule technique que nous avons aujourd'hui pour vérifier le code et le protocole de calcul est la réimplémentation par une équipe indépendante.

Un [journal en ligne](#) entièrement dédié à la publication de réplifications d'études computationnelles.

The Re**Science** Journal

[ABOUT](#) [READ](#) [WRITE](#) [EDIT](#) [BOARD](#) [FAQ](#)

**Reproducible Science is good. Replicated Science is better.**

Re**Science** is a peer-reviewed journal that targets computational research and encourages the explicit [replication](#) of already published research, promoting new and open-source implementations in order to ensure that the original research is [reproducible](#).