

Lien entre science ouverte reproductibilité et logiciels libres

Romain Caneill

JdLL, 25 mai 2025, Lyon

Qui suis-je?

- Romain Caneill
- Chercheur en océanographie physique et banquise (et oui à Grenoble)
- ENSLyon 2014-2017
- Libriste et geek
- J'aime la guitare, la musique, la montagne, la magie
- Père de 3 enfants de 1 à 4 ans

romaincaneill.fr
@rcaneill@fediscience.org



Certificat de réussite
International Institute of La RACHE



100% de réussite (ou presque) pour

Romain Caneill

Qui est capable de mener à bien n'importe quel projet, même à l'arrache.

Formé(e) à la cafèt', le 5/25/2025

Le Président
METODA Gilles

Chargé de Formation
Le stagiaire



la-rache.com

Je ne suis pas *expert en tout* donc certains détails seront sans doute simplifiés.

N'hésitez-pas à poser des questions au cours de la présentation.

Les 4 libertés du logiciel libre

(Rappel?)

- Utiliser
- Étudier
- Redistribuer
- Améliorer et publier les améliorations

Lien entre science ouverte reproductibilité et logiciels libres

Partie 1

La recherche scientifique
aujourd'hui

Partie 2

Qu'est ce que la reproductibilité d'une étude?

Partie 3

Et le rôle des logiciels libres dans tout ça?

Lien entre science ouverte **reproductibilité** et logiciels libres

Partie 1

La recherche scientifique
aujourd'hui

Partie 2

Qu'est ce que la reproductibilité
d'une étude?

Partie 3

Et le rôle des logiciels
libres dans tout ça?

Lien entre science ouverte reproductibilité et logiciels libres

Partie 1

La recherche scientifique aujourd'hui

Partie 2

Qu'est ce que la reproductibilité d'une étude?

Partie 3

Et le rôle des logiciels libres dans tout ça?

Lien entre science ouverte reproductibilité et logiciels libres

Partie 1

La recherche scientifique aujourd'hui

Partie 2

Qu'est ce que la reproductibilité d'une étude?

Partie 3

Et le rôle des logiciels libres dans tout ça?

Partie 1

La recherche scientifique
aujourd'hui

Partie 2

Qu'est ce que la reproduc-
tibilité d'une étude ?

Partie 3

Et le rôle des logiciels
libres dans tout ça ?

Comment fonctionne la recherche ?

- Recherche bibliographique : on lit l'état de l'art
- Articles dans des revues scientifiques (par exemple Nature, Sciences)
- Ces articles sont écrits par d'autres chercheurs
- Ensuite, bah on cherche...

Et comment on cherche ?

- Observations, expérience en labo, modèle numérique, etc
- Analyse des données → diagnostics
- Construction de quelques graphes, tableaux représentatifs
- On discute et échange des idées sur les phénomènes à l'origine des résultats

- Écriture d'un article pour revue scientifique
- Processus de review
- Si l'article est suffisamment bon : publication

La boucle est bouclée.

Les chercheurs sont notés sur le nombre d'articles scientifiques.

“Quand une mesure devient un objectif, elle cesse d'être une bonne mesure.”

(Loi de Goodhart)

“Crise de la reproductibilité”

- Années 2005–2010
- Surtout dans le domaine médical
- Article de John Ioannidis : “*Why most published research findings are false*”

Partie 1

La recherche scientifique
aujourd'hui

Partie 2

Qu'est ce que la reproduc-
tabilité d'une étude?

Partie 3

Et le rôle des logiciels
libres dans tout ça?

Définitions approximatives de la reproductibilité

- Pouvoir refaire une étude
- Refaire les expériences en labo ou les simulations numériques
- Relancer le code qui a produit les analyses et figures

Utile pour...

- Confiance dans les résultats
- Vérification
- Processus de revision par les pairs (review)
- Accélérer la recherche : si je peux reproduire, plus facile pour étendre

C'est donc utile pour soi (review) et pour les autres!

Quelles sont les conditions nécessaires

(De plus en plus de revues demandent à publier le code et les données avec les articles)

- Avoir accès au code
- Avoir accès aux données
- Pouvoir relancer le code

Facile!

Facile mais en fait non

Accès au code

Accès aux données

Pouvoir relancer le code

Qui a déjà essayé de compiler un gros logiciel ?

CMIP6 (modèles climat) : 40PB de donnée, 1000 millions d'heures de calcul (Adastra : 50000 CPU)

Avez-vous déjà lu le code de quelqu'un d'autre ?

Les chercheurs sont rarement formés au logiciel

Des idées ?

Facile mais en fait non

Accès au code

Accès aux données

Pouvoir relancer le code

Qui a déjà essayé de compiler un gros logiciel ?

CMIP6 (modèles climat) : 40PB de donnée, 1000 millions d'heures de calcul (Adastralab : 50000 CPU)

Avez-vous déjà lu le code de quelqu'un d'autre ?

Les chercheurs sont rarement formés au logiciel

Des idées ?

Facile mais en fait non

Accès au code

Accès aux données

Pouvoir relancer le code

Qui a déjà essayé de compiler un gros logiciel ?

CMIP6 (modèles climat) : 40PB de donnée, 1000 millions d'heures de calcul (Adastralab : 50000 CPU)

Avez-vous déjà lu le code de quelqu'un d'autre ?

Les chercheurs sont rarement formés au logiciel

Des idées ?

Facile mais en fait non

Accès au code

Accès aux données

Pouvoir relancer le code

Qui a déjà essayé de compiler un gros logiciel ?

CMIP6 (modèles climat) : 40PB de donnée, 1000 millions d'heures de calcul (Adastralab : 50000 CPU)

Avez-vous déjà lu le code de quelqu'un d'autre ?

Les chercheurs sont rarement formés au logiciel

Des idées ?

Facile mais en fait non

Accès au code

Accès aux données

Pouvoir relancer le code

Qui a déjà essayé de compiler un gros logiciel ?

CMIP6 (modèles climat) : 40PB de donnée, 1000 millions d'heures de calcul (Adastralab : 50000 CPU)

Avez-vous déjà lu le code de quelqu'un d'autre ?

Les chercheurs sont rarement formés au logiciel

Des idées ?

Que faire si le code est en Matlab, excel ?

Si les analyses sont faites avec un logiciel payant et non accessible ?

Si l'entreprise que développait le logiciel a coulé, ou bien si je n'ai pas les moyens de payer la licence ?

Partie 1

La recherche scientifique
aujourd'hui

Partie 2

Qu'est ce que la reproduc-
tibilité d'une étude ?

Partie 3

Et le rôle des logiciels
libres dans tout ça ?

Les 4 libertés du logiciel libre

- Utiliser
- Étudier
- Redistribuer
- Améliorer et publier les améliorations

Quelles sont les libertés nécessaires à la reproductibilité ?

Discussion interactive

- Utiliser
- Étudier
- Redistribuer
- Améliorer et publier les améliorations

“La reproductibilité parfaite est un horizon.”

Quelques pistes

- Ne **jamais** produire un diag en interactif
- **Toujours** utiliser des scripts
- Utiliser des logiciels libres (Python plutôt que Matlab par exemple)
- Commenter son code, fichier README pour savoir comment exécuter le bazar
- Se baser sur un outils type Make, Snakemake, etc (automatisation du workflow)
- Si tout cela est mis en place, on a le CI quasi gratuit!

Merci!