

La physique médicale et l'approche "science reproductible"

Ludovic Ferrer

Département de physique médicale
ICO René Gauducheau, Nantes

15 juin 2018



Outline

- 1 Introduction
- 2 Science reproductible
- 3 Conclusion

Topic

1 Introduction

2 Science reproductible

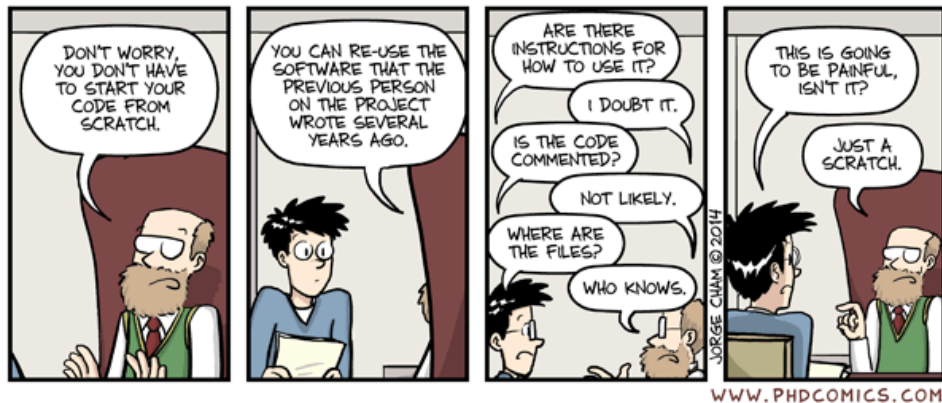
3 Conclusion

Conflits d'intérêt

- Je déclare n'avoir aucun conflit d'intérêt ces cinq dernières années.

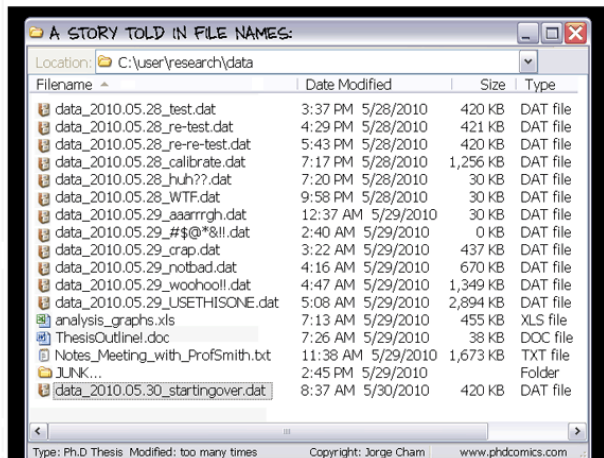
Des situations familières ?

Un nouveau stagiaire doit continuer un travail antérieur réalisé par un autre stagiaire...



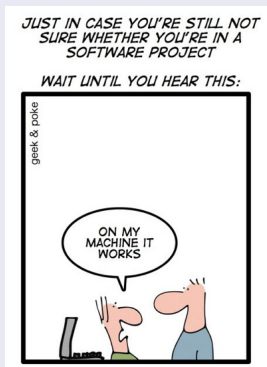
Des situations familières ?

Une gestion des versions de fichiers peu efficace ...



Des situations familières ?

Des développements informatiques facétieux



Des conséquences fâcheuses ...

Menant à de nombreuses rétractations

*However, while most scientists are careful to validate their laboratory and field equipment, most **do not know how reliable their software is** [4,5]. This can lead to serious errors impacting the central conclusions of published research [6]: **recent high-profile retractions, technical comments, and corrections** because of **errors in computational methods** include papers in Science [7,8], PNAS [9], the Journal of Molecular Biology [10], Ecology Letters [11,12], the Journal of Mammalogy [13], Journal of the American College of Cardiology [14], Hypertension [15], and The American Economic Review [16].*

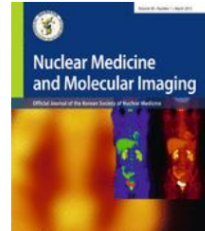
— Wilson et al. - 2014 - Best Practices for Scientific Computing. Plos Biol. 12:1

Des conséquences fâcheuses ...

Menant à de nombreuses rétractations

Major methods error prompts retraction of lung radiation paper

One of the authors of a 2014 case series on lung disease following radiation in *Nuclear Medicine and Molecular Imaging* is retracting the paper for what the the journal is calling “honest error.” That may be true, but it’s a big error — so big, it’s amazing no one detected it sooner.



Topic

- 1 Introduction
- 2 Science reproducible
- 3 Conclusion

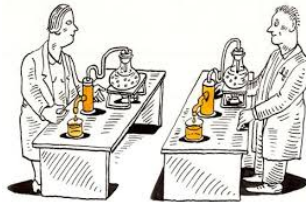
Replication vs Reproductibilité

Replication, the practice of independently implementing scientific experiments to validate specific findings, is the cornerstone of discovering scientific truth.

*Related to replication is **reproducibility**, which is the calculation of quantitative scientific results by **independent scientists using the original datasets and methods**.*

*Reproducibility can be thought of as a different standard of validity because it forgoes independent data collection and **uses the methods and data collected by the original investigator**.*

— Victoria Stodden, Friedrich Leisch, Roger D. Peng **Implementing Reproducible science**, CRC Press, 14 avr. 2014 - 448 pages



Les bonnes pratiques



La littérature

- The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences. Kitzes et al, Oakland, CA: University of California Press, 2018
- Best Practices for Scientific Computing, Wilson et al, Plos Biology, 2014
- Ten Simple Rules for Reproducible Computational Research, Sandve et al, PLoS Computational Biology, 2013

Traitements informatiques fréquents, choix des outils, se former



- Traitements informatiques prédominant
- Introduire **culture du monde du dev. info.** en **science**
 - Choix des outils, des méthodes, ...

Les bonnes pratiques



La littérature

- The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences. Kitces et al, Oakland, CA: University of California Press, 2018
- Best Practices for Scientific Computing, Wilson et al, Plos Biology, 2014
- Ten Simple Rules for Reproducible Computational Research, Sandve et al, PLoS Computational Biology, 2013

Quelques règles récurrentes ...

- Utiliser un **cahier de notes**
- **Eviter** les manipulations manuelles des données
 - ou bien, garder une trace dans le cahier de notes
- **Eviter les tableurs**
- Automatiser avec des **script** : Python, R, ...
- Utiliser un **système de contrôle de version**
- **Tester** vos codes et vos données
- **Partager** vos données et codes

Éviter les tableurs

La littérature

- Spreadsheets in the cloud- not ready yet, McCullough et al, Journal of Statistical Software, 2013
- On the numerical accuracy of spreadsheets, Almiron et al, Journal of Statistical Software, 2010
- A critical review of the literature on spreadsheet errors, Powell et al, Decision Support Systems, 2008

Met en garde d'une utilisation des tableurs dans les domaines scientifiques.

Éviter les tableurs

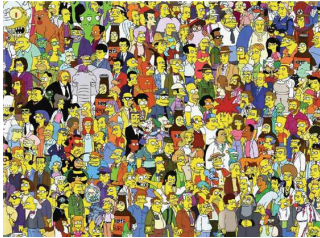
Le travail de groupe réduit les erreurs Kruck et al doi:10.1016/j.infsoc.2005.04.005

| | | | |
|-------------------------|------|--|----|
| Panko and Halverson Jr. | 1994 | Business students | |
| | | Working alone | 81 |
| | | Working in groups of 2 | 71 |
| Panko and Halverson Jr. | 1997 | Business students | |
| | | Working alone | 79 |
| | | Working in groups of 2 | 78 |
| Panko and Halverson Jr. | 1995 | Accounting students | 68 |
| | | General business students working alone | 82 |
| | | General business students working in groups of 3 | 27 |

Un groupe structuré est encore mieux Kruck et al doi:10.1016/j.infsoc.2005.04.005

| | | | |
|----------------------|------|--|-----------------|
| Janvrin and Morrison | 1996 | Upper- and masters-level accounting and business administration students | |
| | | Ad hoc development group | 14 ^a |
| Janvrin and Morrison | 1996 | Structured systems group | 7 ^a |
| | | Upper- and masters-level accounting and business administration students | |
| | | Ad hoc development group | 18 ^b |
| | | Structured systems group | 9 ^b |

Éviter les modifications manuelles



Le diable est dans les détails

- Tables, figures (arrondi, format)
- Récupérer des données d'internet (clic sur les liens url)
- Déplacer les données, couper ou reformater les fichiers de données
- "Aller, juste cette fois . . ."
- **Précisément** noter **chaque** modification dans le cahier de notes

Développer vos talents : scripts

Caractéristiques d'un bon langage de script

- Connaissance ou expérience en programmation minimum
- Facile à apprendre
- Cycle de développement court
- Des tâches complexes rendues faciles
 - Bibliothèques, ...

Anonymisation fichier dicom en 6 lignes de codes

```
import pydicom as dcm
dcm_filename = './a_dicom_file.dcm'
new_dcm = './my_new_file.dcm'
dataset = dcm.read_file(dcm_filename)
dataset.PatientName = 'a_new_name'
dcm.write_file(new_dcm, dataset)
```

Utiliser un cahier de notes

Pour noter

- Toutes modifications manuelles
- **Graines aléatoires** si analyse avec traitements stochastiques
- Le type et les versions de votre environnement informatique
- Vos réflexions sous-tendues par vos résultats

Le cahier de notes électronique: notebook Jupyter www.jupyter.org

- **A web application**: a browser-based tool for interactive authoring of documents which combine explanatory text, mathematics, computations and their rich media output.
- **Notebook documents**: a representation of all content visible in the web application, including inputs and outputs of the computations, explanatory text, mathematics, images, and rich media representations of objects.

Utiliser un cahier de notes

Test of a jupyter notebook

```
In [1]: 1 %matplotlib inline
        2 import matplotlib.pyplot as plt
        3 import pandas as pd
```

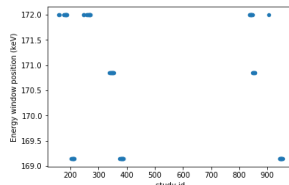
```
In [2]: 1 df = pd.read_csv('./t_energy_window.csv')
        2 df.head(2)
```

```
Out[2]:
```

| | min | max | etude_id | patient_id | type_window | peak |
|---|---------|---------|----------|------------|-------------|----------|
| 0 | 225.015 | 261.504 | 204 | 432 | EM | 243.2595 |
| 1 | 181.826 | 195.357 | 204 | 432 | SC | 188.5915 |

We can create some textual content around the analysis

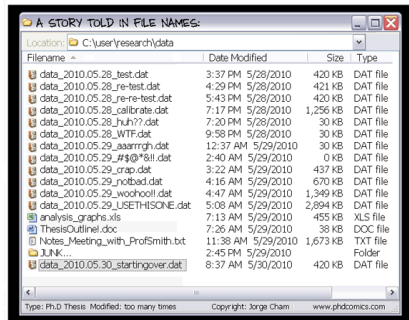
```
In [3]: 1 * plot = df[["etude_id", "peak"]][df['type_window'].str.contains('EM')] &\
        2         (df['peak'] < 200)].plot.scatter(x='etude_id', y='peak', marker='o')
        3 plot.set_xlabel('study id')
        4 plot.set_ylabel('Energy window position (keV)')
        5 plt.savefig('172_kev_peak.png')
```



Développer vos talents : utiliser un VCS

Version control system (VCS)

Renommer les fichiers ... un cauchemar



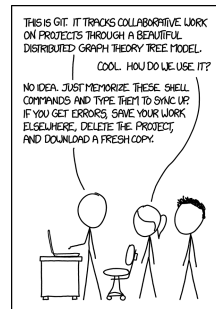
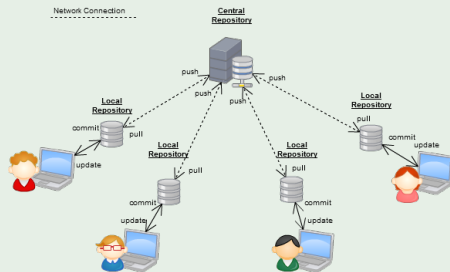
Multiples révisions ... un cauchemar



Développer vos talents : utiliser un VCS

Version control system (VCS): Git, Mercurial, ...

- **Collaboration** avec des collègues
- Trace l'**historique des fichiers** associés à des commentaires
- Synchronisation du travail sur **plusieurs ordinateurs**



Partager vos codes et vos données

Codes et documentations

- Meilleures pratiques de codages
- Documentation plus pertinente
- GitHub, GitLab, Bitbucket, ...

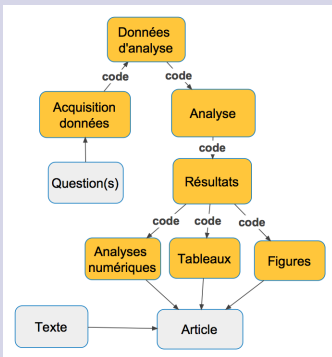
Données

- Anonymisation
- Dépôt public: www.alward.org

Pour résumer

Un flux de travail scientifique typique

La production automatisée des résultats



Le partage du travail



Topic

- 1 Introduction
- 2 Science reproductible
- 3 Conclusion

Conclusion

La démarche "science reproductible"

- Démarche **qualité** dans le traitement informatique des données
- Amélioration du savoir commun
 - Partage des codes et des données

DQPRM (nov. 2017, mai 2018)

- Formation optionnelle à Toulouse: MB, LF, DS
 - Scripts python, GIT
- Intérêt pour un EPU ?

EPU traitement d'images: présent futur

- Date: 3-5 Octobre 2018
- Lieu: Port-Bourgenay
- Mots clés: Traitement d'images, Machine learning, Reproducible science