

Nouvelles formes d'articles : data paper, notebook, et plus si affinités

Konrad Hinsen

Centre de Biophysique Moléculaire, Orléans, France
Synchrotron SOLEIL, Saint Aubin, France

15 novembre 2023

Pourquoi publier ?

Pour que d'autres puissent :

1. Comprendre

- les questions posées
- les hypothèses
- les méthodes
- les outils
- les données
- les conclusions

Pourquoi publier ?

Pour que d'autres puissent :

1. Comprendre

2. Juger

- le bien-fondé des hypothèses
- la pertinence des méthodes
- l'adéquation des outils
- la qualité de l'exécution
- la validité des conclusions

Pourquoi publier ?

Pour que d'autres puissent :

1. Comprendre

2. Juger

3. S'approprier

- les idées
- les méthodes
- les outils
- les données

- Techniques : supports, diffusion, archivage

Les contraintes

- Techniques : supports, diffusion, archivage
- Économiques : coût par rapport aux bénéfices

- Techniques : supports, diffusion, archivage
- Économiques : coût par rapport aux bénéfices
- Sociales : audience, évaluation

La publication au 21ème siècle

- ~~Contraintes techniques~~
- ~~Contraintes économiques~~
- Contraintes sociales
 - L'audience inclut des non-scientifiques, non-experts.
 - La bibliométrie nous empêche d'abandonner les revues traditionnelles.

La publication au 21ème siècle

- ~~Contraintes techniques~~
- ~~Contraintes économiques~~
- Contraintes sociales
 - L'audience inclut des non-scientifiques, non-experts.
 - La bibliométrie nous empêche d'abandonner les revues traditionnelles.
- Le **volume des données** et la **complexité des méthodes automatisées** (calcul, IA) rendent le récit traditionnel **insuffisant**.

La publication au 21ème siècle

En plus du récit traditionnel, au minimum :

- Les données d'observation brutes
- Les méthodes scientifiques automatisées
- Les étapes de traitement, les résultats intermédiaires et finaux

La publication au 21ème siècle

En plus du récit traditionnel, au minimum :

- Les données d'observation brutes
- Les méthodes scientifiques automatisées
- Les étapes de traitement, les résultats intermédiaires et finaux

Idéalement aussi :

- Les outils informatiques, pour réutilisation
- Des logiciels d'exploration et de visualisation, dans un but pédagogique

Une première réussite : le “data paper”

Data paper = jeu de données numérique + récit explicatif

Une première réussite : le “data paper”

Data paper = jeu de données numérique + récit explicatif

- Rendu possible par les entrepôts de données
- Le récit explicatif est un article traditionnel

Une première réussite : le “data paper”

Data paper = jeu de données numérique + récit explicatif

- Rendu possible par les entrepôts de données
- Le récit explicatif est un article traditionnel
- Documente un jeu de données numérique publié
 - Motivation pour la collecte
 - Méthodes appliquées lors de la collecte
 - Paramètres techniques (format, ...)
 - Conseils pour la réutilisation

Une première réussite : le “data paper”

Data paper = jeu de données numérique + récit explicatif

- Rendu possible par les entrepôts de données
- Le récit explicatif est un article traditionnel
- Documente un jeu de données numérique publié
 - Motivation pour la collecte
 - Méthodes appliquées lors de la collecte
 - Paramètres techniques (format, ...)
 - Conseils pour la réutilisation

Pour en savoir plus :

Dedieu, L. 2022. Publier un Data paper, en 5 points.

Montpellier (FRA) : CIRAD, 5 p.

<https://doi.org/10.18167/coopist/0057>

En phase expérimentale : les notebooks

Notebook : récit qui intègre du code et des résultats de calcul

En phase expérimentale : les notebooks

Notebook : récit qui intègre du code et des résultats de calcul

Rôles multiples

- Outil interactif pour le calcul exploratoire
- Support pour des cours en science computationnelle
- Documentation d'une étude scientifique intégrant des calculs
- Support pour permettre l'expérimentation

En phase expérimentale : les notebooks

Notebook : récit qui intègre du code et des résultats de calcul

Rôles multiples

- Outil interactif pour le calcul exploratoire
- Support pour des cours en science computationnelle
- Documentation d'une étude scientifique intégrant des calculs
- Support pour permettre l'expérimentation

Foisonnement technologique

- Logiciels : Jupyter, RStudio, Emacs, Quarto, ...
- Formats : Jupyter, RMarkdown, Org-Mode, ...
- Langages de programmation : Python, R, ...
- Couplage étroit entre formats et outils

- The Journal of Digital History
- Computo

Il n'y a pas que les notebooks !

- Objectif : partager du code bien documenté

Il n'y a pas que les notebooks !

- Objectif : partager du code bien documenté
- Il y a beaucoup de façons d'y arriver
- **Le bon choix de la technologie dépend du domaine !**

Il n'y a pas que les notebooks !

- Objectif : partager du code bien documenté
- Il y a beaucoup de façons d'y arriver
- **Le bon choix de la technologie dépend du domaine !**
- Bon exemple : [IPOL](#)
- Approche généraliste (ReScience C) : article qui renvoie vers le code hébergé ailleurs

L'état de l'art des “papiers exécutables”

Notebooks

- Bien adapté si l'exposé a naturellement une structure linéaire
- Publication en mode “statique” (code visible mais non-exécutable)
- Rendre le code exécutable reste un défi :
 - Binder : simple mais souvent lent, environnement rigide
 - Toute autre technologie : lourd à gérer pour les lecteurs

La **rédaction** d'un notebook publiable et l'**exécution** de notebooks publiés restent réservées aux **pionniers**.

D'autres approches

- Beaucoup d'expérimentation, peu de retours
- Souvent réservées aux **pionniers**

L'état de l'art des “papiers exécutables”

1. Comprendre

- Mieux vaut un notebook qu'un article sans code
- ... mais parfois, article + script valent autant
- ... et peuvent faciliter la prise en main pour l'exploration

L'état de l'art des “papiers exécutables”

1. Comprendre

2. Juger

- Avoir le code est un progrès important
- Le jugement reste souvent difficile :
 - Piles logicielles complexes
 - Environnements computationnels mal gérés
 - Exécution trop pénible

L'état de l'art des “papiers exécutables”

1. Comprendre

2. Juger

3. S'approprier

- Exécuter le code dans un environnement personnel est difficile

Tâche principale :

- développer
- maintenir
- faire évoluer

une **infrastructure** de publication pour l'ère du numérique

Tâche principale :

- développer
- maintenir
- faire évoluer

une **infrastructure** de publication pour l'ère du numérique

Ceci nécessite des **institutions** dont c'est la mission et qui ont des **moyens** adéquats.

Tâche principale :

- développer
- maintenir
- faire évoluer

une **infrastructure** de publication pour l'ère du numérique

Ceci nécessite des **institutions** dont c'est la mission et qui ont des **moyens** adéquats.

Il faut impliquer

- des documentalistes
- des spécialistes en interaction homme-machine
- des chercheurs dans les domaines visés
- des informaticiens (mais pas trop)