



**HAL**  
open science

## Executable papers : vers une science plus ouverte, transparente et réutilisable

Michael Nauge

► **To cite this version:**

Michael Nauge. Executable papers : vers une science plus ouverte, transparente et réutilisable.  
Progedo : Semaine DATA SHS, Dec 2020, Poitiers, France. hal-03128918

**HAL Id: hal-03128918**

**<https://hal.science/hal-03128918>**

Submitted on 2 Feb 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Executable papers : vers une science plus ouverte, transparente et réutilisable

Michael Nauge, Laboratoire FoReLLIS/MIMMOC, Université de Poitiers (2020)



## C'est quoi en quelques mots ?

C'est un article scientifique qui permet de combiner habilement : - textes - données primaires et enrichies - codes - visualisations dynamiques

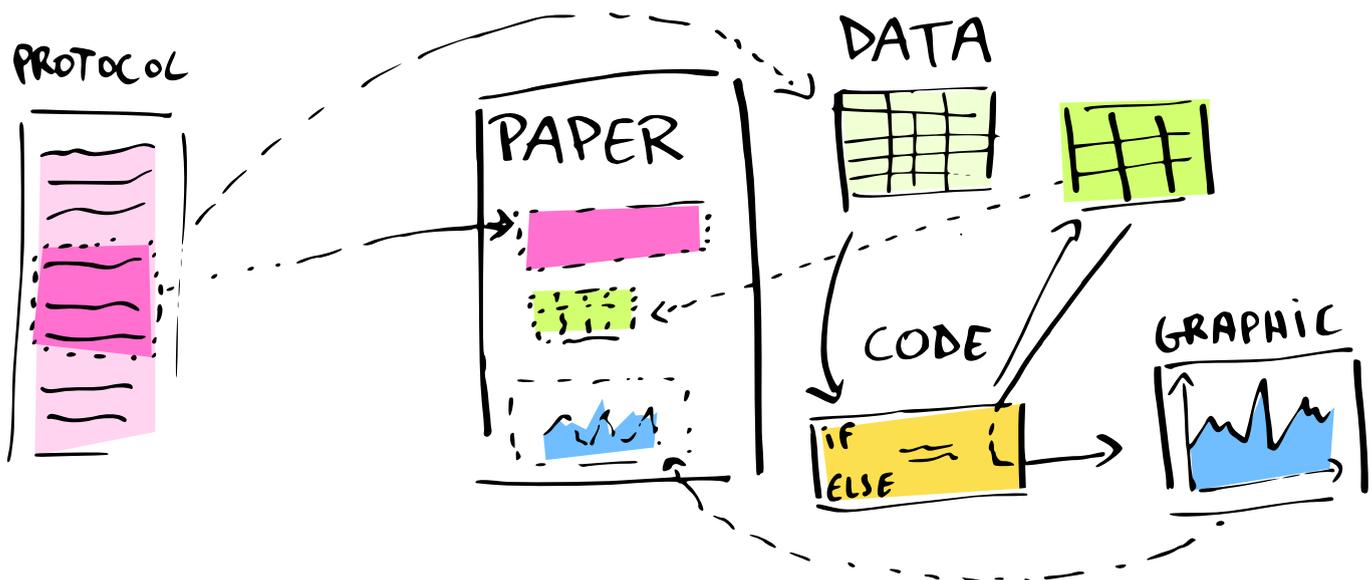


Figure 1 : schéma des éléments utilisés pour la construction d'un article scientifique traditionnel. Illustration CC-BY-4.0 par [R.Nuage](<https://rnuage.com/>)

L'idée n'est plus de partager uniquement une "image" statique et partielle d'un résultat de recherche mais bien de partager un environnement dynamique et complet où il est possible d'interagir directement sur les multiples données et leurs visualisations pour faciliter l'analyse critique et constructive.

## Pourquoi en avons nous besoin ?

Ci-dessous une compilation humoristique d'échanges inspirés de faits réels :

**Cher Professeur X., j'ai adoré votre dernier article ! Pouvez-vous me partager les données et les codes que vous avez utilisés pour cette publication pour que je m'en inspire pour mes prochains travaux ?**

"Merci de l'intérêt que vous portez à mes travaux. Malheureusement c'est mon doctorant qui a réalisé le travail technique, juste avant de soutenir sa thèse. Je n'ai plus de nouvelle depuis qu'il est chercheur contractuel à temps partiel dans une start-up." Professeur X.

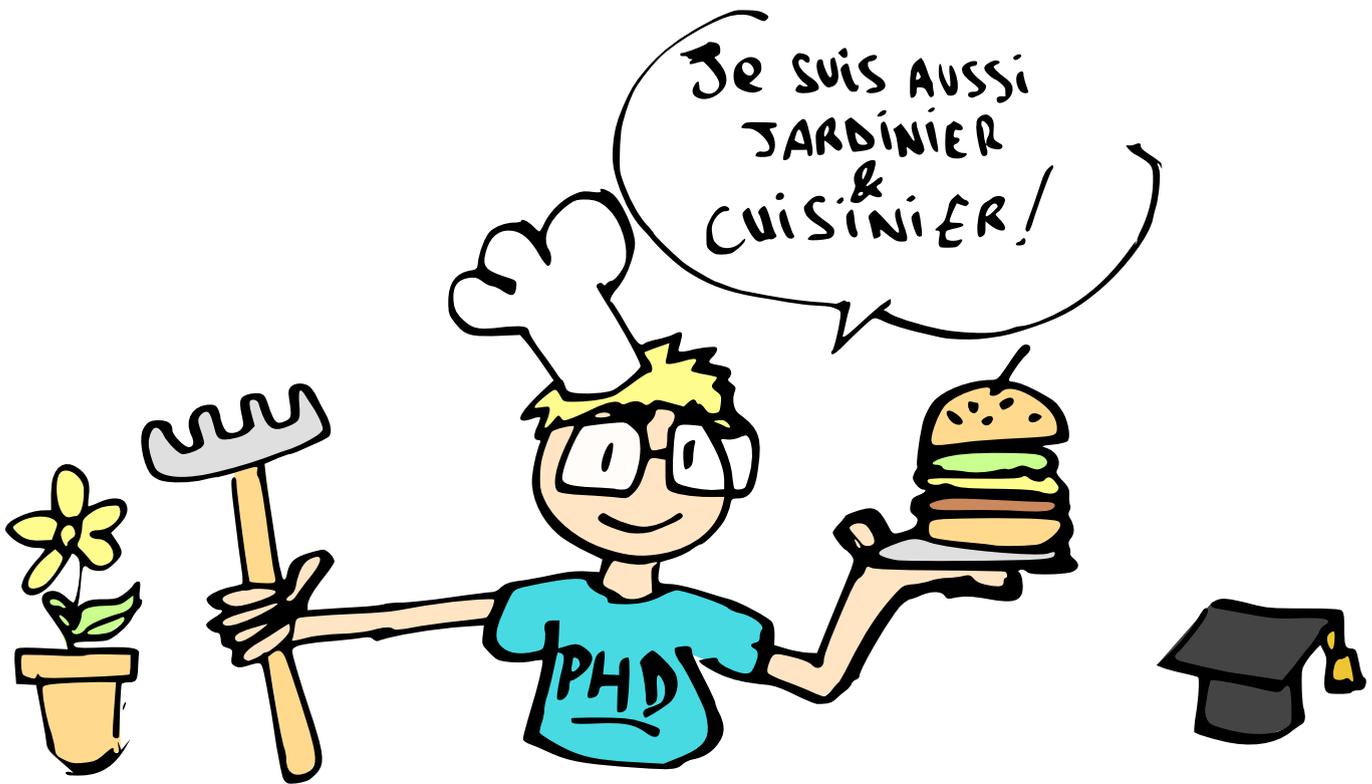


Figure 2 : Illustration d'un chercheur contractuel avec plusieurs activités professionnelles. Illustration CC-BY-4.0 par [R.Nuage](<https://rnuage.com/>)

**Chère Professeure A., j'ai adoré votre dernier article ! Pouvez-vous me donner plus de détails sur la ou les valeurs que vous avez utilisées pour le paramètre  $\alpha$  dans votre formule expliquant votre mesure novatrice de *distance inter-stroke-logistique-multivariée-bayésienne* ?**

“Merci de l'intérêt que vous portez à mes travaux. Malheureusement les valeurs utilisées sont pour le moment empiriques et confidentielles en respect des accords avec notre partenaire industriel international”. *Professeure A.*

**Cher Professeur M., j'ai adoré votre dernier article ! Pouvez-vous partager l'implémentation informatique dans le domaine discret de votre métrique novatrice dont vous évoquez uniquement l'équation dans le domaine continu ?**

“Merci de l'intérêt que vous portez à mes travaux. Malheureusement la dernière opération de mise à jour du système d'exploitation de notre ordinateur de calcul a écrasé le dossier contenant le code matlab de notre implémentation.”  
*Professeure M.*

**Chère Professeure N., j'ai adoré votre dernier article ! Pouvez-vous partager le tableau de données ayant permis de produire la représentation graphique sous forme de moyenne en trois sous-groupes d'âges ?**

“Merci de l'intérêt que vous portez à mes travaux. Malheureusement le disque dur externe contenant les données sous cette forme a été volé par un individu masqué la semaine dernière.” *Professeure M.*



Figure 3 : Illustration d'un disque dur perdu. Illustration CC-BY-4.0 par [@R.Nuage](https://rnuage.com/)

## Autres exemples

De nombreux autres exemples sont à lire ici :

L. Desquilbet, S. Granger, B. Hejblum, A. Legrand, P. Pernot, et al.. (2019) *Vers une recherche reproductible : Faire évoluer ses pratiques.* ([en ligne](#))

Voici un extrait : > “J’ai récemment participé à la rédaction d’un article scientifique et nous venons de recevoir les commentaires des relecteurs : je dois modifier les couleurs d’une figure afin que celle-ci soit lisible en noir et blanc. Comme je n’arrivais pas à remettre la main sur mon script R ayant généré la figure en question, j’ai ré-écrit le programme correspondant. Le seul problème, c’est que cette nouvelle figure est un peu différente de la précédente et remet en cause les conclusions de l’article. Je ne comprends pas ce qui a pu se passer.” *Jeanne A., jeune méthodologiste en biostatistique*

## A quoi ça ressemble un article exécutable ?

*Démonstration en direct*

Ci-dessous un bouton pour lancer et interagir de chez vous avec votre navigateur internet sur une instance de l’article interactif de démonstration :



Pendant que ça charge, vous pouvez toujours consulter la version statique de l’[article de démonstration](#)

L’article interactif présenté manipule des données structurées sous formes de tableurs. Mais en science humaine il est courant d’initier ses travaux avec des données textuelles non structurées. Pour illustrer ce cas voici le lien vers un autre article interactif :

NAUGE M. (2019), Marmython : Les bases d’une cuisine de corpus électoral. *Atelier annuel du consortium CAHIER* ([en ligne](#))

## Comment les fabriquer et les partager ?

- [JupyterLab](#) + [GitLab Huma-Num](#) + [MyBinder](#)
- Collab ([Google](#))
- R + [R Shiny](#)
- [org-mode](#)

- [Iodide](#)

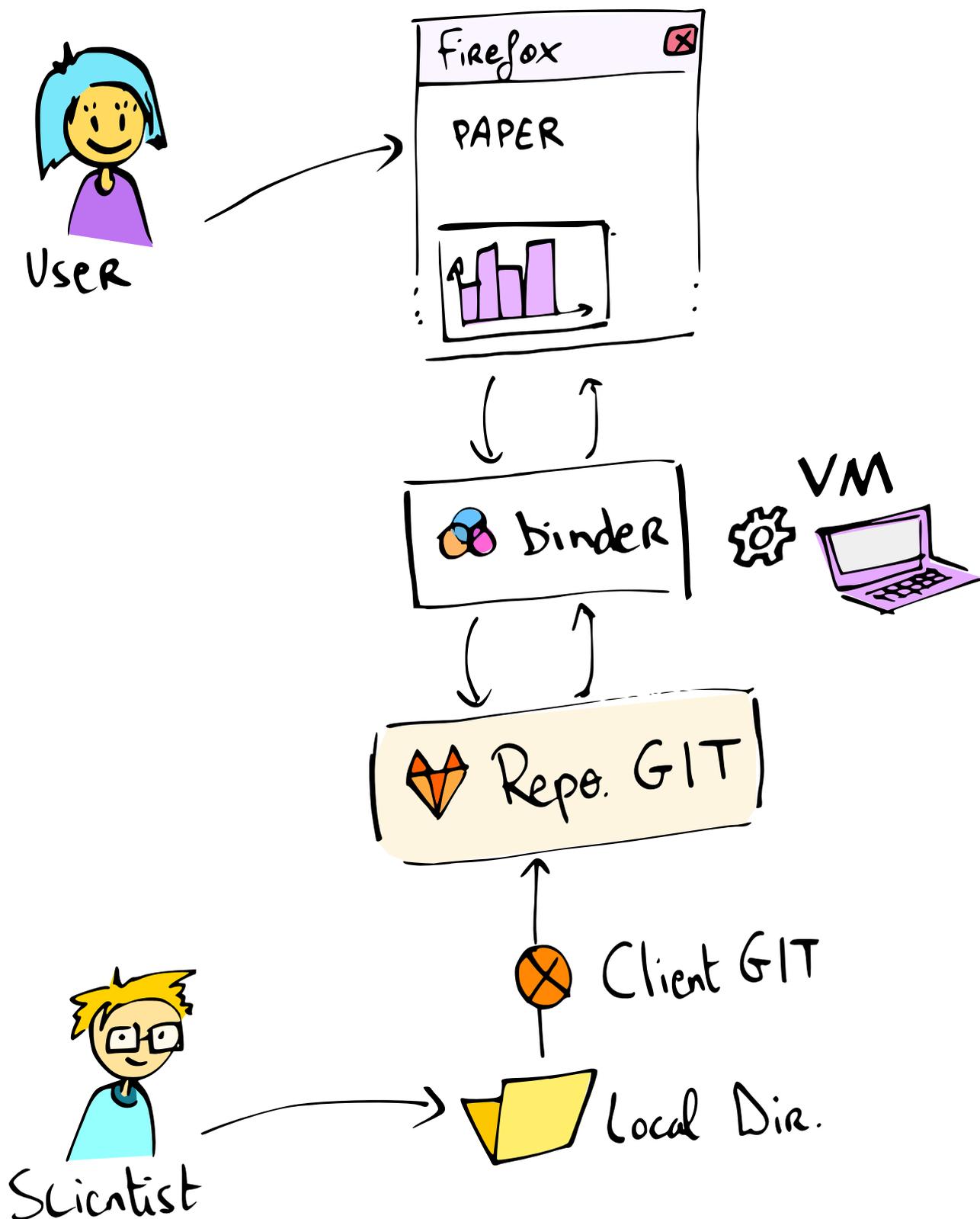


Figure 4 : schéma des composants logiciels utilisés pour la gestion des versions de ses fichiers de recherche et la diffusion d'un article scientifique interactif. Illustration CC-BY-4.0 par [[@R.Nuage](https://rnuage.com/)](<https://rnuage.com/>)

## Installer jupyter lab sur sa machine

Le plus simple est d'installer l'environnement complet pour "data scientist" [Anaconda](#).

Cela permet entre autre de disposer en plus de Jupyter d'un ensemble de bibliothèques de code python très souvent utiles (sans avoir besoin de les ajouter manuellement un à un).

## Utiliser un gestionnaire de version

La recherche transparente et reproductible passe par l'utilisation d'un gestionnaire de version local et un dépôt distant.

### Pourquoi ?

- Besoin de garder les traces des modifications des données et des codes
- Besoin d'ubiquité des données
- Besoin de pouvoir rejouer les opérations de manipulation des données

### Dépôt distant pour les SHS

[GitLab Huma-Num](#)

### Gestionnaire de version local (avec interface graphique)

[GitHub Desktop](#)

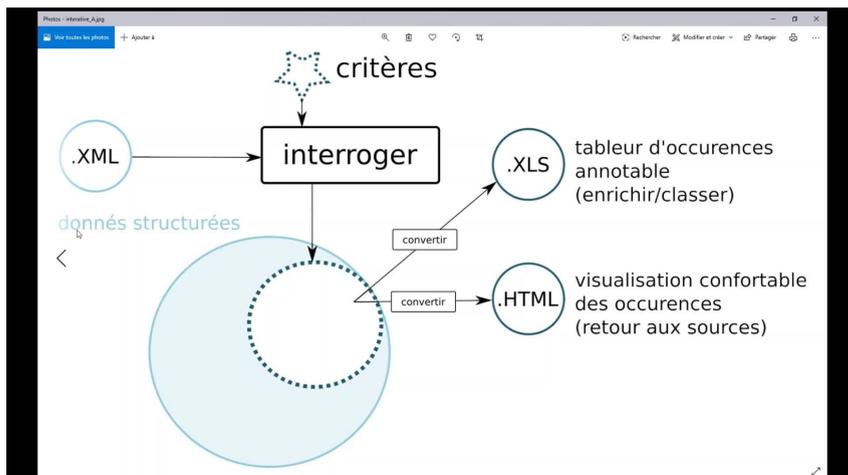
### LE MOOC à suivre

[Recherche reproductible : principes méthodologiques pour une science transparente](#)

## Produits dérivés

Cette technologie permet également de créer à coût réduit des appareillages de recherche spécifiques pour faciliter l'exploration et l'interrogation de corpus.

### Exemple pour la recherche linguistique diachronique sur dictionnaires



<https://videotheque.univ-poitiers.fr/video.php?id=ar1hj59pjzh8vw5zjtc&link=5swajkajh4klb3a7sw4cl2nnszbvzbz>