

# DATA PAPER UN NOUVEAU FORMAT ÉDITORIAL DE COMMUNICATION SCIENTIFIQUE : OPPORTUNITÉS POUR LES SHS ?

Journée d'étude – « Un data journal interdisciplinaire pour les sciences humaines et sociales. Enjeux scientifiques et mise en œuvre pratique »

Nancy, MSH Lorraine, 10 mars 2023

# BONJOUR



## Florence Thiault

[Florence.thiault@univ-rennes2.fr](mailto:Florence.thiault@univ-rennes2.fr)

Maîtresse de conférences en sciences de l'information et de la communication  
Laboratoire PREFICS  
Université Rennes 2



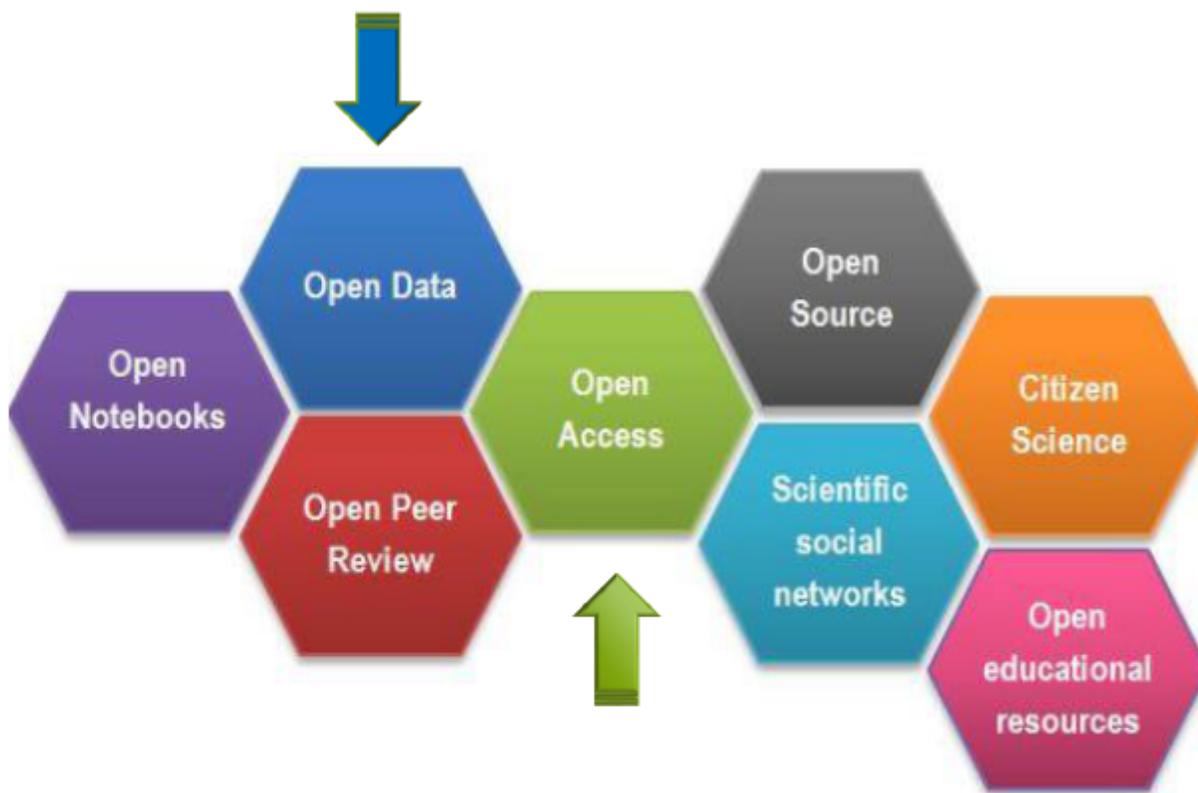
**PREFics**  
EA 4246

Pôle de recherche  
Francophonies, Interculturel  
Communication, Sociolinguistique

# DATA PAPER ET SCIENCE OUVERTE

Définitions

# Open Science

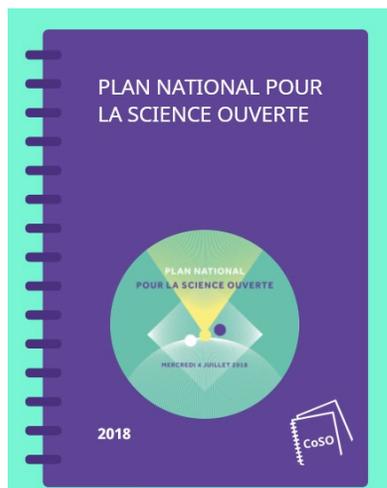


« La science ouverte : une nouvelle façon de faire de la science, en ouvrant les processus, les codes, les méthodes ou encore les protocoles. L'accès ouvert rend possible le mouvement de diffusion de la production scientifique sur internet. Il lève les barrières de l'accès en maintenant toutes les protections du droit d'auteur. »

<https://www.science-ouverte.cnrs.fr/le-mouvement-pour-la-science-ouverte/>

« AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY »

# DATA PAPER ET SCIENCE OUVERTE



## AXE 2 : STRUCTURER ET OUVRIR LES DONNÉES DE LA RECHERCHE

Dans le cadre du soutien public aux revues, la France recommandera l'adoption d'une politique de données ouvertes associées aux articles et le développement des **data papers**.  
**(PNSO 1, 2018)**

## AXE 1 : GENERALISER L'ACCES OUVERT AUX PUBLICATIONS

Construire la bibliodiversité

→ Développer et soutenir des innovations éditoriales : prépublications, évaluation ouverte par les pairs, articles exécutables (Jupyter notebooks), articles de données (**data papers**), overlay journals, etc.  
**(PNSO 2, 2021)**



# DÉFINITION D'UN DATA PAPER



« Un *data paper* est un article dans une revue à comité de lecture, décrivant les données d'un projet de recherche. Ce type d'article décrit des données liées à une publication (*underlying research data*) ou indépendantes d'une publication. »

Reymonet, 2017



Data in Brief  
Volume 31, August 2020, 105880



Data Article

## Impact of the Covid-19 pandemic on perceptions and behaviors of university students in Vietnam

Duy Van Nguyen <sup>a</sup>, Giang Hoang Pham <sup>a</sup> , Dat Ngoc Nguyen <sup>b</sup>

Show more 

+ Add to Mendeley  Share  Cite

<https://doi.org/10.1016/j.dib.2020.105880>

[Get rights and content](#)

Under a Creative Commons license

[open access](#)

### Abstract

This article presents a novel data set on perceptions and behaviors of university students collected after the beginning of the Covid-19 outbreak in Vietnam. Our questionnaire design is based on employing both qualitative interview with students and survey of SARS literature, probing into the sensitivity of students toward the crisis in making crucial decisions of daily routines, as well as future travel plans in presence of a grave health concern. The data set consists of 440 valid responses from Vietnamese university students through Internet platforms (Facebook, Google Form). Besides descriptive statistics, this article also includes the results of explanatory factor analysis, which may serve as a good reference for future studies.

# INTERÊT D'UN DATA PAPER

**EXPOSER les données dans un data paper permet :**

- de répondre aux exigences des financeurs de la recherche en termes de visibilité et d'accessibilité des résultats scientifiques ;
- de répondre aux recommandations de la communauté scientifique internationale concernant les données ;
- de présenter un accès aux données avec un lien vers l'entrepôt choisi, et les rendre intelligibles en les décrivant ;
- de fournir une référence citable, car il est publié dans une revue scientifique (peer-reviewed journal) ;
- la génération de citations par des services tels que le Data Citation Index de Thomson Reuters ;
- la reconnaissance du travail réalisé par l'équipe de recherche qui a produit les données décrites en matérialisant ces données par un article publié.

(Reymonet, 2017)

**Une nouvelle forme de publication scientifique au service de la reproductibilité.**

**Un data paper en SHS : pourquoi, pour qui, comment ?**

---

# COMMENT PUBLIER UN DATA PAPER ?

## **Choisir une revue : data journal ou revue classique**

- Vérifier que la revue effectue le peer review du data paper

## **Choisir un entrepôt et déposer les données, accompagnée d'un DOI**

- L'éditeur peut en recommander un

## **Rédiger le datapaper et la cover letter**

- Suivre les instructions aux auteurs, modèles mis à dispositions ou en s'appuyant sur des articles existants

## **Lier l'article à l'entrepôt**

- Insérer le DOI du jeu de données

## **Soumettre le datapaper**

- sera évalué selon différents critères

## **Attribuer un DOI au datapaper**

- devient citable et lien réciproque avec le jeu de données



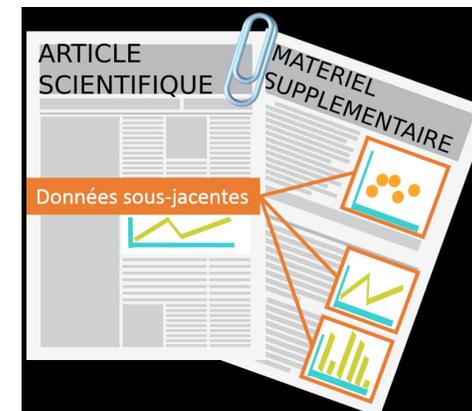
# STRUCTURE ET CONTENU D'UN DATA PAPER

- **Structure très variable** selon les revues : structure assez simple et d'autres une structure beaucoup plus détaillée
- Trois **composantes communes** à tous les data papers :
  - ✓ des **éléments communs** aux autres types d'article : titre, résumé, mots clefs...
  - ✓ un **accès aux données** (dans l'article, matériel supplémentaire, déposées dans un entrepôt)
  - ✓ une **description** des données.

Le *Data paper* est généralement structuré en 3 points :

- description du jeu de données et de son contexte,
- description des méthodes d'obtention,
- explication du potentiel de réutilisation des données.

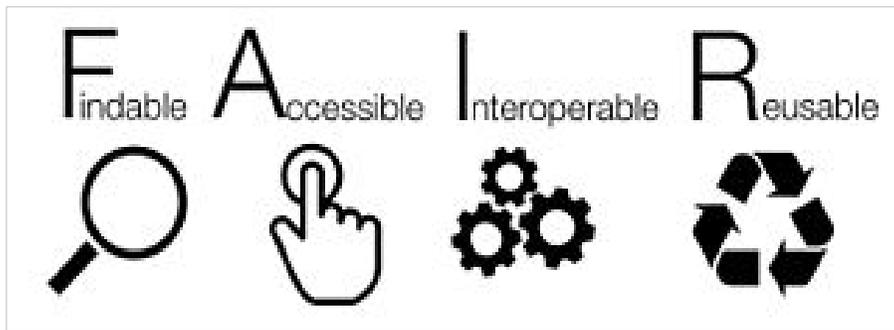
Les instructions aux auteurs de la revue proposent souvent un modèle (*template*) de *Data paper*.



# PROXIMITÉ AVEC LA RÉDACTION DE PLAN DE GESTION DE DONNÉES

## Données partagées et publiées

Données publiées = accessibles de façon pérenne (entrepôt), citables, documentées avec des métadonnées, évaluées.



// avec le Data Management Plan (DMP)

Agence Nationale de la Recherche (ANR):  
ANR - modèle de DMP (français)

### informations générales

Renseignements administratifs  
Acronyme  
Code décision  
Titre  
Nom du coordinateur  
Prénom du coordinateur  
Affiliation  
Contact concernant le PGD  
Version du PGD  
Date

1. Description des données et collecte ou réutilisation de données existantes
2. Documentation et qualité des données
3. Stockage et sauvegarde pendant le processus de recherche
4. Exigences légales et éthiques, codes de conduite
5. Partage des données et conservation a long terme
6. Responsabilités et ressources en matière de gestion des données

# EXPÉRIENCE DE PUBLICATION ET D'ÉVALUATION D'UN DATA PAPER

Revue RFSIC

24 | 2022

## Data Paper : émergence d'une nouvelle donne scientifique

<https://doi.org/10.4000/rfsic.12219>



Sous la direction de **Gérald Kembellec et Olivier Le Deuff**



Informations sur cette image

La production scientifique fait bien entendu référence aux publications, telles que les articles de revues, les ouvrages ou encore les actes de colloque. Mais depuis quelques années, les données de la recherche sont de plus en plus considérées, à raison, comme faisant partie intégrante de la production scientifique. A ce titre, tout comme les publications, les données doivent, quand cela est possible et pertinent, être préservées, signalées, partagées ; afin d'encourager la transparence, l'éthique et la reproductibilité de la recherche mais aussi de valoriser des jeux de données de qualité. [En savoir +](#)

## Dossier

**Gérald Kembellec et Olivier Le Deuff**

[Poétique et ingénierie des data papers](#) [Texte intégral]

**Laetitia Bracco**

[Mesurer l'ouverture de la science : le cas de l'Université de Lorraine](#) [Texte intégral]

**Florence Thiault et Marie-Laure Malingre**

[Corpus d'enquêtes sur les pratiques d'information scientifique des chercheurs. Constitution et exploitation des données](#) [Texte intégral]

**Alaric Tabariès**

[Vers une métrique pour évaluer les métadonnées de documents scientifiques](#) [Texte intégral]

Étude de cas de l'Université de Toulon et l'archive ouverte HAL

**Jean-Louis Pernin**

[Intention d'abandonner ses études chez les étudiants et Covid-19 : une application de la théorie du comportement planifié](#) [Texte intégral]

**Audrey Knauf**

[Une mise en perspective bibliométrique de l'intelligence économique reposant sur sa revue de référence : la R2IE](#) [Texte intégral]

**David Reymond, Kouamvi Couao-Zotti, Alaric Tabariès, Amandine Lebourgeois et Lauren Campos**

[Analyse lexicale d'une page web](#) [Texte intégral]

Extractions du message hypertextuel pour comparaisons

*Lexical analysis of a web page : verification of hypertextual communicative coherence*

<https://journals.openedition.org/rfsic/12219>

---

# CRITÈRES D'ÉVALUATION D'UN DATAPAPER

Les data papers sont évalués par le comité de lecture de la revue choisie. Les critères d'évaluation dépendent de la revue, on peut dégager des éléments communs :

Certaines revues affichent leurs critères d'évaluation de data papers. Exemples : [Journal of Environmental Quality](#)

[Geoscience Data Journal](#)

Les critères d'évaluation des data papers,  
Doranum

- **Importance et originalité** des données ;
  - **Potentiel et valeur de réutilisation** des données ;
  - **Qualité et fiabilité** des données : structure de la base de données, organisation logique des données, intégrité des données (vérification des erreurs potentielles) ;
  - **Accès** aux données : point important. L'auteur doit s'assurer que ses données sont toujours accessibles dans l'entrepôt. Si les données sont retirées de l'entrepôt, la rétractation du data paper pourra être décidée par l'éditeur ;
  - Qualité et rigueur de la **méthode de collecte** des données : méthode appropriée, actuelle, suffisamment claire pour permettre la reproductibilité ;
  - Choix des **métadonnées descriptives** et **formats** : présentation, complétude, degré de précision, etc. ;
  - **Autres critères « classiques »** : qualité générale du manuscrit, citations appropriées, respect des instructions, etc.
-

# SPÉCIFICITÉS DE LA RÉDACTION D'UN DATA PAPER EN SHS

- Prise en compte de l'épistémologie des sciences de l'information et de la communication (citation auteurs)
- Chercheurs plutôt en sciences de l'information, collaboration avec professionnels IST, IE...
- Usages de méthodes numériques dans le processus de recherche, traitement de l'information (BDD, visualisation, ...)
- consensus sur accès aux données (intégrité scientifique, reproductibilité)

## Épistémologie des SHS

- spécificités des données de la recherche en SHS / Science ouverte,
- discipline qui travaillent sur corpus (proches Humanités Numériques) / approche multidisciplinaire
- habitude d'écriture plus longue en SHS
- question de la rigueur scientifique dans la démarche de recherche (approche qualitative différente des STM)

# FORMATION SUR LES DATA PAPERS

Stages et ateliers

# OFFRE DE FORMATION

Public mixte (EC et doctorants) :

questions des apprenants :

- où publier ?
- dans quels entrepôts ?
- Quel reviewing ?
- Publication reconnu dans ma discipline ?

## Trois types de formation :

- Stages URFIST (en distanciel) : publics EC, doctorants, personnel d'appui à la recherche, professionnels de l'IST – 3H (4 sessions depuis 2021, 46 stagiaires)
- Atelier Ecole doctorale en SHS : présentiel 3H (juin 2022, 20 doctorants)

*Ex. de datapaper à découvrir (différents champs scientifiques et divers datajournal) : repérage du plan et du lieu de stockage des données*

- Ateliers du portail SOcle (Clés pour la science ouverte) Rennes 2 : « Publier un article sur ses données : rédiger un data paper » (1H, pause méridienne)

*« j'ai bien compris les motivations et l'intérêt de faire un data paper. J'ai aussi une idée plus précise de comment je pourrais l'écrire. L'idée de travailler sur des exemples de data paper à la fin était très bonne, on a bien vu qu'il est important de bien mettre en avant l'accès aux données dans le data paper » (questionnaire évaluation)*

# PERSPECTIVES

Pistes et points d'attention pour la création d'un data journal en SHS

Quel format pour un data paper en SHS ?

- longueur et structuration (template spécifique)
- différence de format et proximité avec STM (cadre de rédaction)
- data journal sans APC (différent Elsevier) / entrepôt de données de la revue ?

Opportunités et limites ?

- lien avec DMP des projets ANR, européen
- intérêt de la lecture de ce type de publication scientifique ?