

Data papers : pourquoi? Comment ? l'expérience de *Annals of Forest Science*



Marianne PEIFFER

Data paper : c'est quoi?

- Le ***data paper*** est une publication qui décrit un jeu de données scientifiques (*data, dataset*), notamment à l'aide d'informations précises, appelées métadonnées (*metadata*). Les données décrites doivent être accessibles par un lien pérenne (DOI) vers « l'entrepôt de données » en ligne (*data repository*) où elles sont déposées et correctement formatées.
- Les **métadonnées** détaillent pourquoi, par qui et comment ces données ont été collectées, qui en est propriétaire, sous quel format elles sont stockées, etc.
- Le ***data paper*** est publié sous la forme d'un article examiné par les pairs dans une revue scientifique classique publiant différentes formes d'articles dont des *data papers* ou dans un *data journal*, c'est-à-dire une revue contenant exclusivement des *data papers*.
- Source : basée sur : <http://coop-ist.cirad.fr/aide-a-la-publication/rediger/data-paper/1-qu-est-ce-qu-un-data-paper>



Un nouveau besoin pour les chercheurs

- OPEN SCIENCE

-> Augmentation croissante du nombre de *data papers* Candela et al. 2014

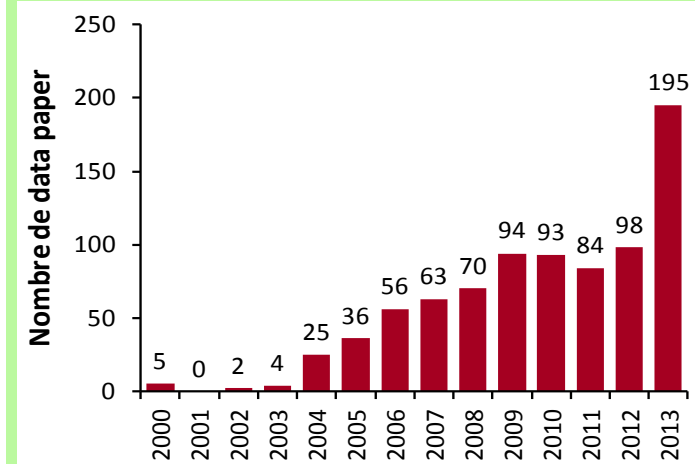
- REPONSE A UN BESOIN DES CHERCHEURS

Annals of Forest Science
DOI 10.1007/s13595-014-0375-3

EDITORIAL

The open data debate: a need for accessible and shared data in forest science

Bruno Fady • Alain Benard • Christian Pichot •
Marianne Peiffer • Jean Michel Leban • Erwin Dreyer



Nombre de *data papers* publiés entre 2000 et 2013, d'après Candela et al. 2014

- BESOIN DE CERTIFICATION DES PUBLICATIONS/QUALITE

-> AFS lance un nouveau type d'article: les *data papers*

-> qualité garantie par le workflow et le processus de *peer reviewing*



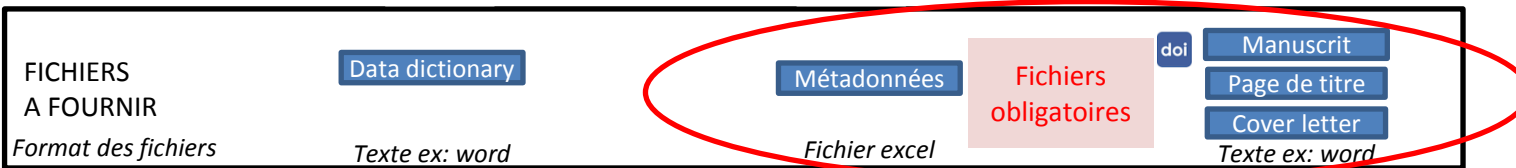
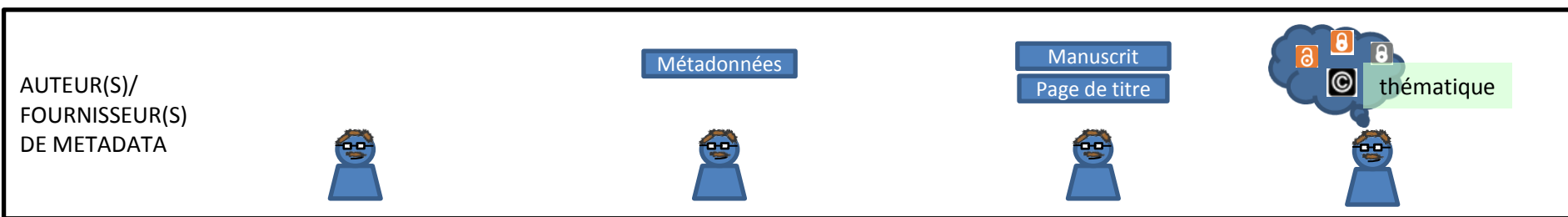
Organisation du *workflow* pour publier des *data papers*

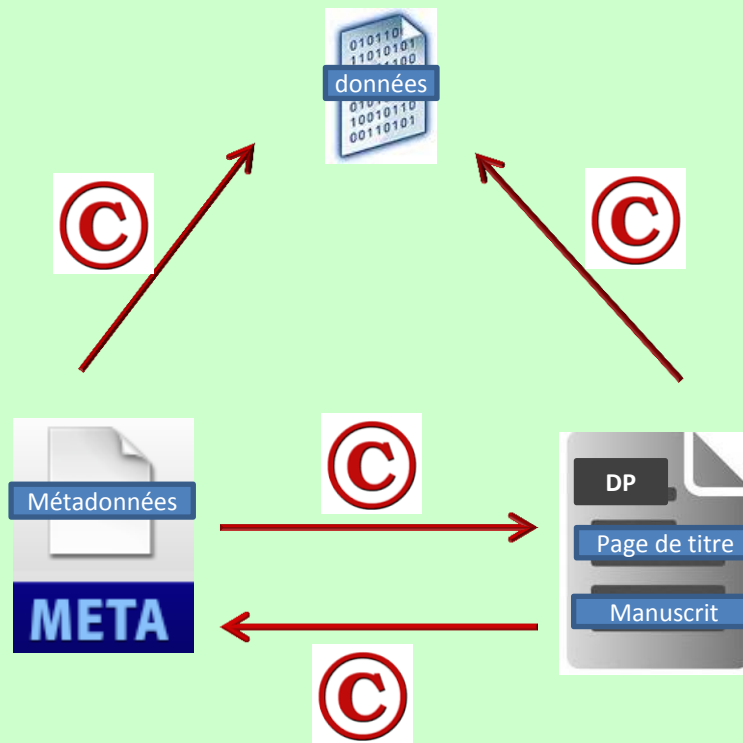


Obtenir Modèles

Obtenir DOI doi

Déposer données





le processus de peer reviewing

Reviewers will carefully consider: (a) the quality of the manuscript, (b) the quality of the data, (c) the data reusability and (d) the relevance of the data set and its potential contribution to the progress of science. They will not review the whole data set, whose quality is under the responsibility of the authors.

- **Quality of the manuscript:** Usual criteria for assessing manuscript quality including style, consistency, clarity. The review will address following questions:

- Is a DOI provided for the database?
- Is the metadata template filled out as recommended?
- Do title and key message accurately reflect the content of the data paper?
- Are relevant non-textual media (e.g. tables, figures) appropriate?
- Have abbreviations and symbols been properly defined?
- Is the context of prior research properly described, citing relevant articles and datasets?
- Are conflicts of interest, relevant permissions and other ethical issues addressed in an appropriate manner?

SOURCE :

<http://ist.blogs.inra.fr/afs/2015/11/23/guidelines-for-reviewing-a-data-paper/>

- **Quality of the dataset :** Although publication of a data paper does not guarantee the overall quality of the dataset, there is a need to check whether suitable and reproducible methods have been used to obtain the data, and whether the data are displayed in a sensible way. The review will address following questions:

- Is anything missing in the manuscript or the data resource itself that would prevent replication of the measurements, or reproduction of the figures or other representations of the data?
- Are the methods used to generate the data (including calibration, code and suitable controls) described in sufficient detail?
- Have possible sources of error been appropriately addressed in the protocols and/ or the paper?
- Are the data consistent internally and described using applicable standards (e.g. in terms of file formats, file names, file size, units and metadata)?
- Does the data resource cover a scientifically important region(s), time period(s) and/or group(s) of taxa to be worthy of a separate publication?
- Are the methods used to process and analyse the raw data, thereby creating processed data or analytical results, sufficiently well documented that they could be repeated by third parties? Accepted formats are: 1) datasets, deposited additionally to scientific datasets in the repository; 2) links to online published papers; 3) a section in the body of the manuscript dedicated to material and method.
- Are the data plausible, given the protocols? Authors are encouraged to report any test undertaken to address this point.
- Are the data completely and consistently recorded within the dataset(s)?

- **Reusability of the data set**

- Does the manuscript provide an accurate description of the data?

- **Utility and contribution of data**

- Is the dataset sufficiently original to merit publication as a data paper?
- Are the use cases described in the data paper consistent with the data presented? Would other possible use cases merit comments in the paper?
- Are all claims made in the data paper substantiated by the underlying data?

Conclusions

- De « Peer-reviewed » vers « Peer-reproductibility » :
Reproductibilité par les pairs?

González-Beltrán, A. et al. From Peer-Reviewed to Peer-Reproduced in Scholarly Publishing: The Complementary Roles of Data Models and Workflows in Bioinformatics. PLoS ONE 10, e0127612 (2015)

- L'introduction dans le workflow de Springer n'a pas été difficile jusque ici (On va bientôt accepter le 1^{er} data paper)
- Obstacle: la phase de dépôt des données sur un entrepôt
→ nouvelles compétences à avoir pour accompagner le dépôt

Perspectives/Réflexions

- Le data paper pourrait-il être utilisé pour publier des résultats négatifs? → pas besoin de créer un nouveau type de journal?



Merci pour votre attention

Marianne PEIFFER, PhD

INRA Nancy

Email ann.for.sci@nancy.inra.fr

<http://www.springer.com/13595>

Researcher ID

ORCID: [0000-0001-5326-3590](https://orcid.org/0000-0001-5326-3590)