



# 4<sup>ème</sup> Rencontres de l'édition en SHS

## 15-16 novembre 2022

### Atelier sur les politiques de données des revues

Romain Féret

# Proposition de programme

- Brève présentation sur les données de recherche
- Enjeux autour des données de recherche pour les revues en SHS
- Présentation des *Recommandations aux revues souhaitant définir une « politique de données » liées aux publications*
- Retour d'expérience : la revue *Population*
- De la politique à sa mise en œuvre
- Quiz : quel data policy officer êtes-vous ?

# Données sources, données produites

## Données sources

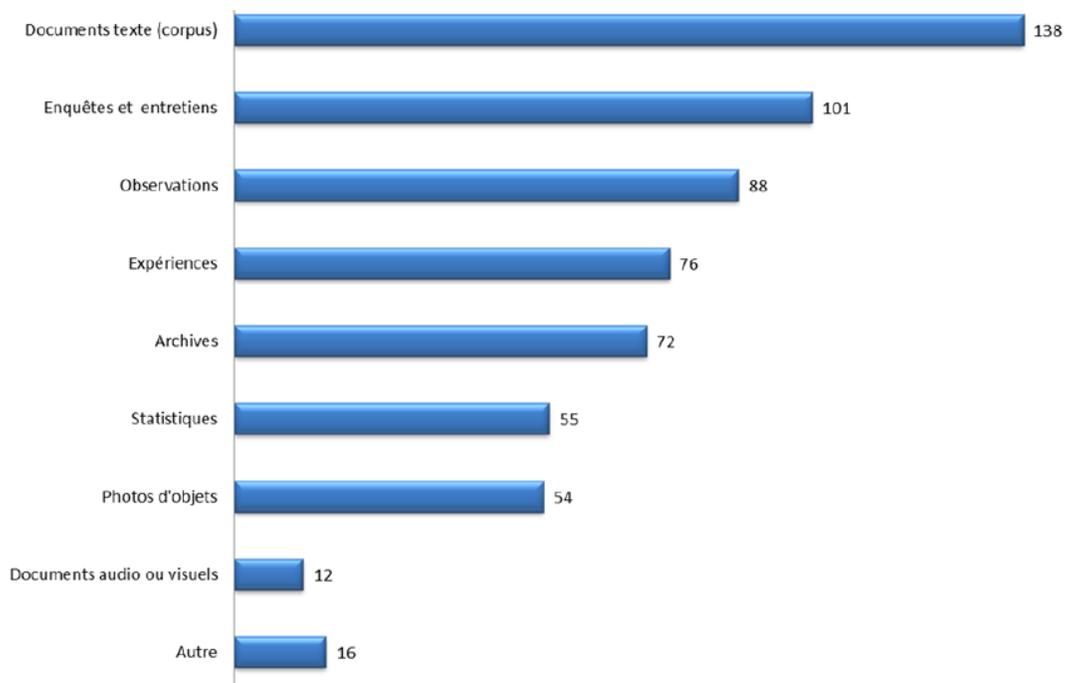


Figure 4 : Données sources (N=214)

## Données produites

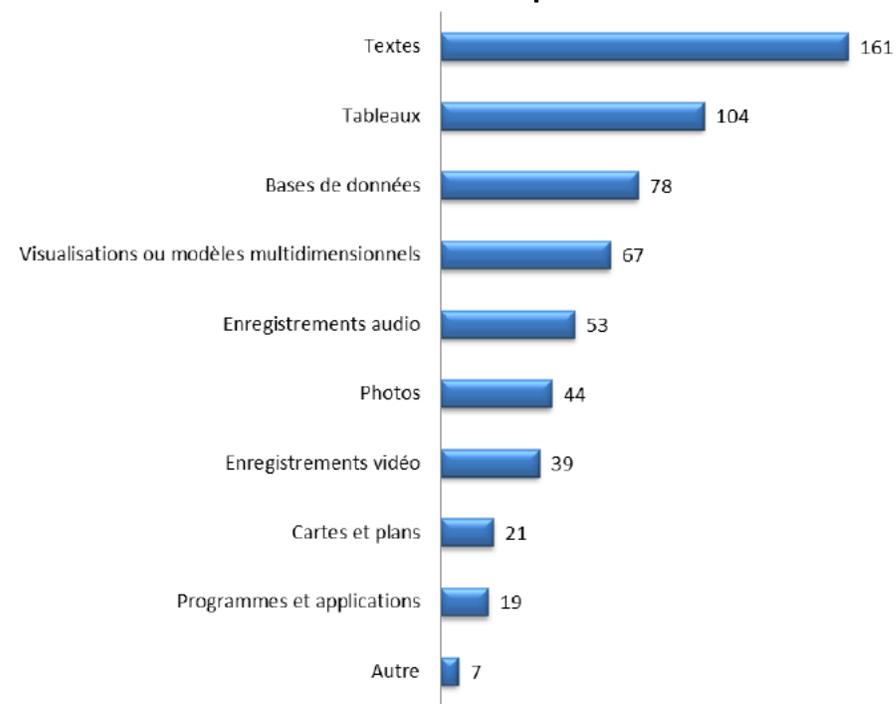


Figure 5 : Données résultats (N=211)

Les données de la recherche en SHS. Une enquête à l'Université de Lille 3. : Rapport final, Hélène Prost et Joachim Schöpfel, 2015 (p. 15-16). Disponible sur HAL : <https://hal.archives-ouvertes.fr/hal-01198379>

# ■ Données brutes, données traitées

## Témoignage :

*« C'est difficile de savoir où commence le traitement. On a le fichier brut, on a le fichier traité, et après en sortie, on a des positions au cours du temps, des gros tableaux, c'est comme ça que ça se lit, et ensuite on fait nos calculs avec MATLAB ».*

Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2. Rapport de recherche. Alexandre Serres, Marie-Laure Malingre, Morgane Mignon, Cécile Pierre, Didier Collet. Université Rennes 2. 2017. Disponible sur HAL : <https://hal.archives-ouvertes.fr/hal-01635186v2> (p.36)

# ■ Exemples de données et d'espaces de conservation/diffusion /accès

Corpus de textes, entretiens. Ex : [beQuali](#)

Données statistiques. Ex. [Portail Quetelet-PROGEDO-Diffusion](#)

Enregistrements audios ou vidéos. Ex. [ORTOLANG](#)

Images : photos, cartes. Ex. [Services IGN](#)

Code source, logiciels. Ex. [Software Heritage](#)

Variété des espaces de conservation/diffusion de données

Certains entrepôts accueillent aussi des données de différentes natures et éventuellement de différentes disciplines : [Recherche Data Gouv](#), [Nakala](#)

# Quels enjeux autour des données de recherche pour les revues en SHS ?

Quels enjeux pour quels acteurs ?

**Revue :**

**Auteurs :**

**Lecteurs :**

# ■ Présentation des *Recommandations...*

## **Constats de départ:**

- Les revues en STM sont de plus en plus nombreuses à adopter des politiques de données
- Peu de revues en SHS ont adopté ce type de politique : structuration différente de l'écosystème de l'édition, besoin de sensibilisation des comités éditoriaux, sujet encore émergent dans certaines disciplines

## **Objectifs :**

- Proposer un document d'entrée dans le sujet qui en couvre les enjeux principaux et permette de faire émerger le sujet auprès des revues intéressées
- Montrer l'étendue des positionnements possibles pour les revues et la possibilité d'adopter une approche incrémentale

## **Forme :**

- Segmentation des enjeux d'une politique en 7 thématiques
  - Pour chaque thématique : identification des enjeux et questions à se poser ; proposition de formulations pour rendre les choses plus concrètes
- Pas de modèle de politique ni de recommandation sur les choix à faire

# ■ Sections des *Recommandations...*

## 7 sections

1. Définition des données couvertes par la politique
2. Précisions sur les formats des données et standards de métadonnées
3. **Modalités d'hébergement** : Dans quel entrepôt ?
4. **Modalités de mise à disposition** : degré d'ouverture ? Moment de la mise à disposition, licence...
5. Accompagnement des auteurs et relecteurs
6. **Liens prévus entre les publications et les données** : systématisation des *data availability statements*, identifiants pérennes, recensement des jeux de données liés à la revue
7. Gestion des cas ne respectant pas la politique

## Autres points d'attention :

- Anticiper l'évaluation de la politique
- Réfléchir à un phasage de la mise en œuvre de la politique

# De la politique à sa mise en œuvre



Normandie Université

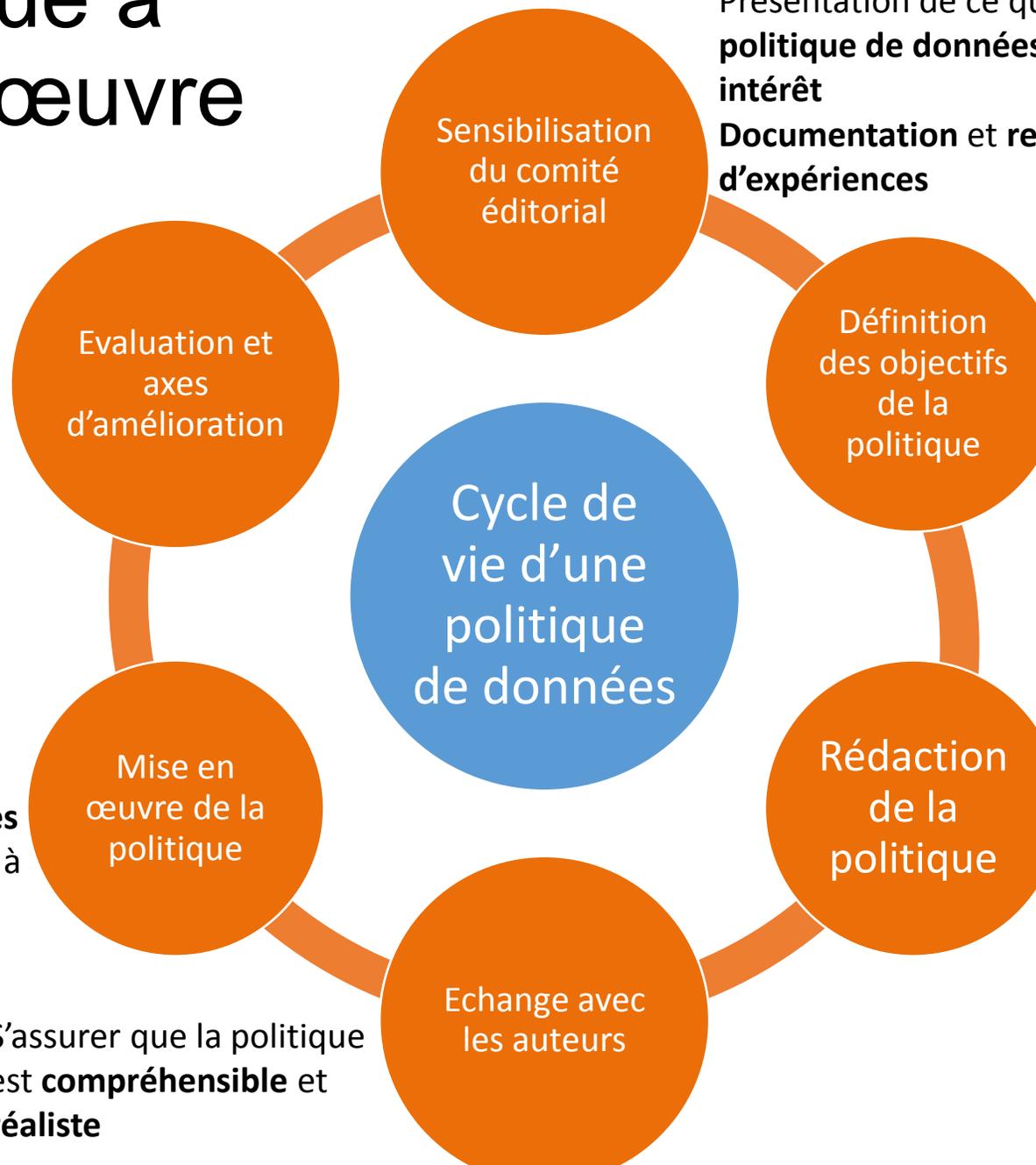
Présentation de ce qu'est une **politique de données** et de son **intérêt**

**Documentation et retour d'expériences**

Vérifier la **mise en œuvre** de la politique (critères d'évaluation)  
**Echanger** avec d'autres revues engagées dans la démarche ou des **spécialistes** des données

Déterminer les **modalités d'accompagnement**  
Proposer des **modèles de citation de jeux de données**  
**Valoriser** les données liées à la revue

S'assurer que la politique est **compréhensible** et **réaliste**



Sensibilisation du comité éditorial

Définition des objectifs de la politique

Rédaction de la politique

Cycle de vie d'une politique de données

Evaluation et axes d'amélioration

Mise en œuvre de la politique

Echange avec les auteurs

Niveau de **prescription**  
Niveau de **précision**  
Niveau **d'éditorialisation** des données  
**Chronologie** de mise à disposition des données  
Intégration au **processus d'évaluation** par les pairs

Structurer et **formaliser** une politique  
Prévoir les **modalités d'évaluation**  
**Phaser** la mise en œuvre

# Focus : enjeux de chronologie

## **Enjeux lors de la définition de la politique**

- Être très clair sur le moment où les données devront être mises à disposition de la revue : dès la soumission-évaluation, lors de l'acceptation de l'article, à sa publication...
- Envisager un phasage dans la montée en exigence de la politique

## **Enjeux de mise en œuvre de la politique**

- Anticiper et annoncer (aux auteurs et relecteurs) la mise en œuvre de la politique avant sa prise d'effet
- Préciser si la date d'effet s'applique aux articles nouvellement soumis ou publiés

# ■ Exemple de politique de données - *Population*

Politique disponible sur : [http://www.revue-population.fr/pour-les\\_auteur%C2%B7es/politique-de-donnees/](http://www.revue-population.fr/pour-les_auteur%C2%B7es/politique-de-donnees/)

Conformément à la politique de l'Ined en faveur d'un partage, aussi ouvert que possible, des données de la recherche, et dans un souci de reproductibilité et de transparence des analyses, la revue encourage les auteur·es à donner accès aux sources, données et codes sur lesquels s'appuie leur contribution.

Les données de la recherche qui seraient nécessaires pour vérifier les résultats présentés dans l'article doivent être déposés, chaque fois que possible, dans un entrepôt de données public répondant aux standards de la science ouverte (FAIR).

Toutes les informations nécessaires sur les différents entrepôts disciplinaires et généralistes sont présentées dans les répertoires suivants :

- [Cat OpiDoR](#) ou [re3data](#)
- ou [Repository Finder](#)

Les citations des données dans l'article doivent inclure les informations minimales recommandées par [DataCite](#) : Auteur, année, titre, éditeur (nom de l'entrepôt ou de l'archive), identifiant (DOI).

# Autre exemple de politique de données - *Bulletin de la Société préhistorique française*

Extrait de la politique disponible sur : [https://www.prehistoire.org/515\\_p\\_57105/donnees-supplementaires.html](https://www.prehistoire.org/515_p_57105/donnees-supplementaires.html)

« Le Bulletin de la Société préhistorique française permet désormais d'intégrer des données supplémentaires en ligne à vos articles publiés. Ils seront hébergés sur la plateforme institutionnelle NAKALA. C'est un service mis en place et maintenu par la TGIR Huma-Num. Il a pour vocation la publication de données issues de projets de recherche en SHS en s'efforçant de mettre en œuvre les principes FAIR et les valeurs de la "science ouverte".

Chaque dépôt de données supplémentaires se verra attribuer un identifiant numérique pérenne (DOI / ex : ID : 10.34847/nkl.c0e5rm2w) qui garantit la durabilité des données publiées. L'URL d'accès ainsi que le DOI seront publiés dans les versions papiers et numérique du Bulletin (ex : <https://doi.org/10.34847/nkl.c0e5rm2w>)

Les données supplémentaires peuvent prendre la forme de bases de données, d'images, de modèles 3D ou de tout type de fichier qui permet d'illustrer et de soutenir l'argumentation développée dans l'article. [...] »

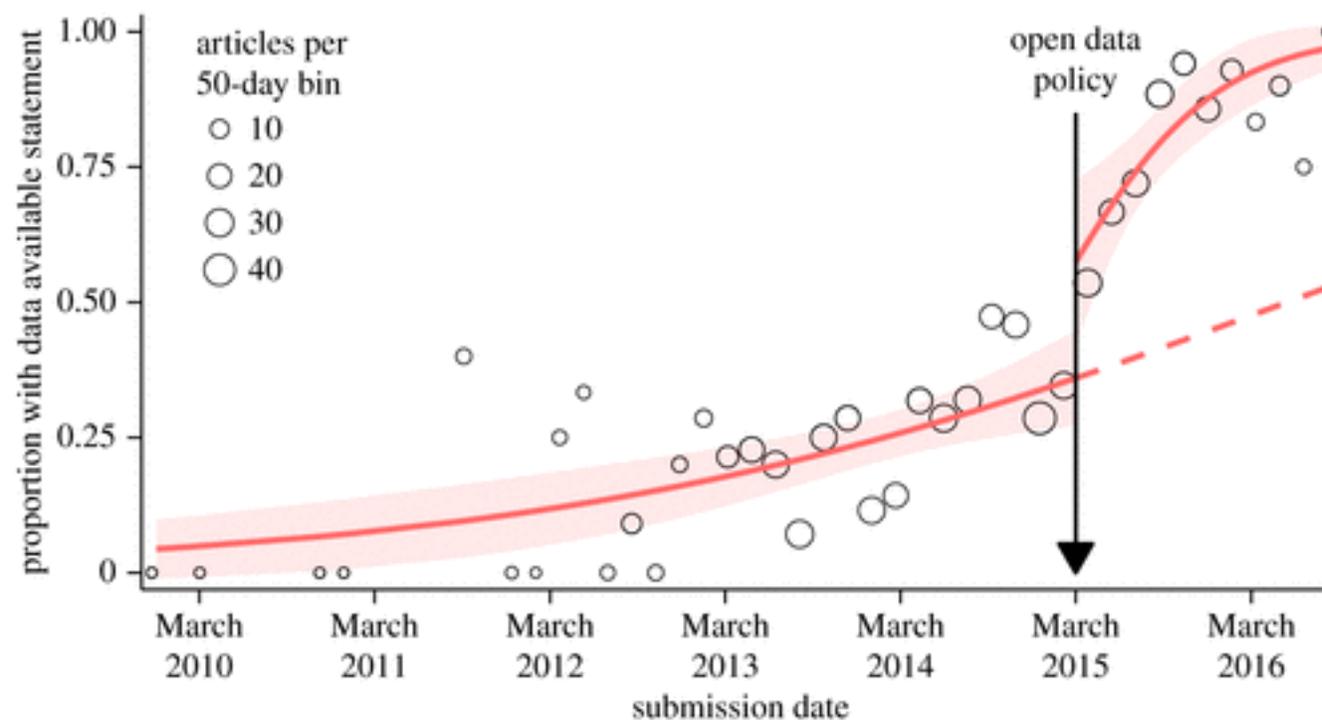
# ■ Pour aller plus loin...

L'article [Developing a Research Data...](#) ([lien](#)) propose 6 modèles de politique de données (voir l'annexe [Policy feature standard texts for policy template construction](#)), qui vont de la simple incitation (modèle 1) à une obligation de mise à disposition des données (modèle 6).

- **Modèle 1** : encouragement à la mise à disposition des données sans obligation ; processus de publication non affecté par la non mise à disposition des données
- **Modèle 2** : encouragement à la mise à disposition des données, avec ***data availability statement (DAS)***
- **Modèle 3** : **obligation** de proposer un *DAS* ; pas de contrainte de mise à disposition des données
- **Modèle 4** : obligation de proposer un *DAS* et de **partager les données en cas de demande** (via un entrepôt), a minima pour les éditeurs et relecteurs ; dans certains cas, **l'acceptation de l'article est conditionné au partage des données** Les relecteurs s'assurent du respect de la politique
- **Modèle 5** : **toutes les données doivent être partagées** dans un entrepôt et signalées via un *DAS* ; l'acceptation de l'ensemble des articles est conditionné au partage des données
- **Modèle 6** : les données doivent être partagées et **font l'objet d'une évaluation**

# Exemple d'évaluation d'une politique

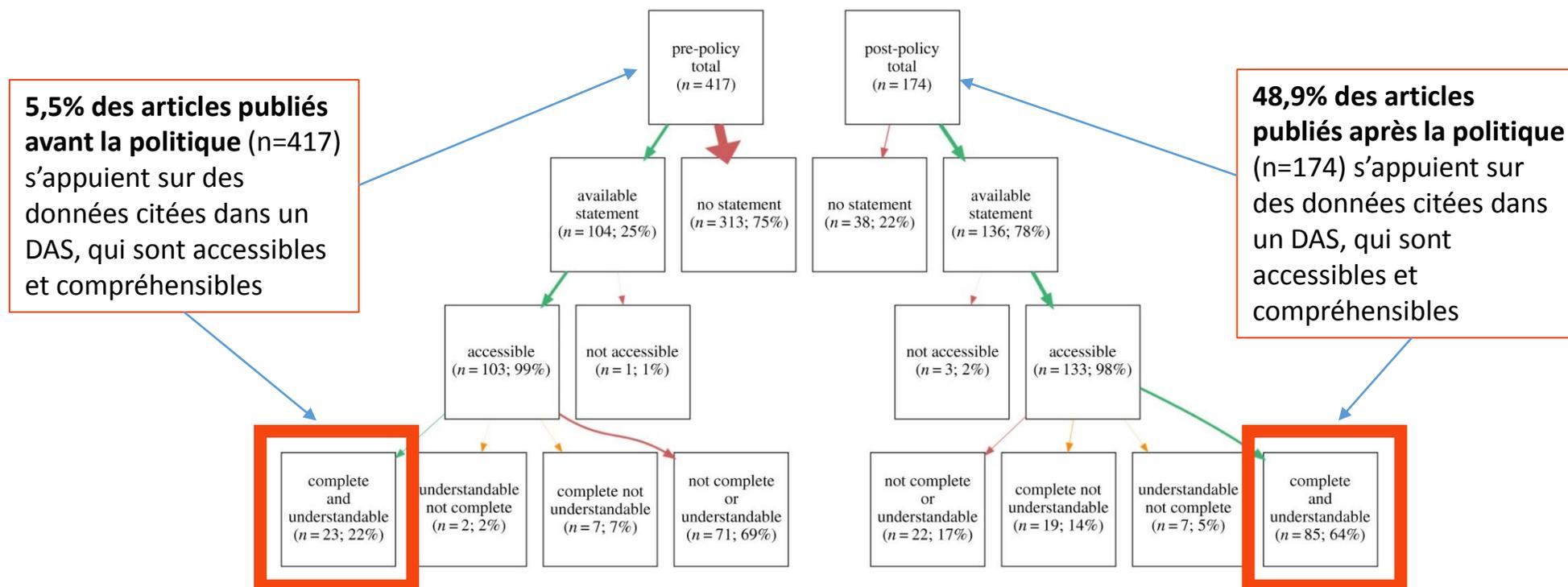
Evolution de l'utilisation de *data availability statements* après la mise en place d'une politique de données par la revue *Cognition*



Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal *Cognition*, Volume: 5, Issue: 8, DOI: (10.1098/rsos.180448)

# Exemple d'évaluation d'une politique

Evolution de la disponibilité et de l'accessibilité des données liées aux articles de la revue *Cognition* après la mise en place d'une politique de données



Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal *Cognition*, Volume: 5, Issue: 8, DOI: (10.1098/rsos.180448)

# ■ Quizz : quel data policy officer êtes-vous ?

1. La diffusion des données par vos auteurs c'est :
  - a) Tranquille, on va leur dire que c'est bien de le faire et puis on verra après
  - b) La moindre des choses
  
2. L'utilisation des *data availability statements* par les auteurs :
  - a) Si ça leur chante, pourquoi pas ?
  - b) S'ils souhaitent lier des données à leur article, autant qu'ils en utilisent un
  - c) Cela doit être systématique, même si c'est pour dire qu'il n'y a pas de données ou qu'elles ne sont pas accessibles
  
3. Accepter un article sans que les relecteurs aient eu accès aux données :
  - a) Après tout, on a toujours fait comme ça
  - b) Ce serait mieux d'avoir les données, surtout si certains résultats sont étonnants
  - c) Hors de question de publier un article sans que les relecteurs aient eu accès aux données
  - d) Hors de question de publier un article sans que les relecteurs aient eu accès aux données, et quitte à exploiter les relecteurs, autant leur demander d'évaluer les données aussi

# Résultats

## Réponses :

- 1a ; 2a ; 3a : modèle 1
- 1a ; 2b ; 3a : modèle 2
- 1a ; 2c ; 3a : modèle 3
- 1b ; 2c ; 3b : modèle 4
- 1b ; 2c ; 3c : modèle 5
- 1b ; 2c ; 3d : modèle 6

# L'essentiel à retenir

- Démarrer progressivement car le sujet est complexe et peut générer des crispations
- Les modèles de politique de donnée sont des guides et ne sont pas à réutiliser tel quel : l'essentiel est que la politique soit bien comprise par le comité éditorial, assumée, et qu'elle soit fidèle aux objectifs et souhaits de la revue
- Tester la politique auprès de publiants ou d'autres chercheurs pour s'assurer qu'elle est rédigée de manière compréhensible
- Trouver des interlocuteurs sur ce sujet : autres revues, plateformes de diffusion, services d'appui à la publication dans les établissements...

# Bibliographie (1/2)

## Données de recherche en shs

- Les données de la recherche en SHS. Une enquête à l'Université de Lille 3. Hélène Prost et Joachim Schöpfel. 2015. Disponible sur [HAL](#)
- Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2. Rapport. Alexandre Serres, Marie-Laure Malingre, Morgane Mignon, Cécile Pierre, Didier Collet. 2017. Disponible sur [HAL](#)

## Politique de données de revues

- Developing a Research Data Policy Framework for All Journals and Publishers. *Data Science Journal*, Iain Hrynaszkiewicz, Natasha Simons et *al.* 2020. Disponible [en ligne](#)
- Proposition de modèle de politique pour les revues et éditeurs quant aux données de la recherche. Chloée Fabre, Françoise Gouzi. 2020. Disponible sur [HAL](#)
- Recommandations aux revues souhaitant définir une « politique de données » liées aux publications. Collège Données de recherche du Comité pour la science ouverte. 2021. Disponible sur [HAL](#)

# Bibliographie (2/2)

## En pratique

- Exemple d'évaluation de l'effet de la politique de données d'une revue : Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal *Cognition*. *The Royal Society Publishing*. Tom E. Hardwicke, Maya B. Mathur et *al.*, 2018. [Lien](#)
- Analyse de la compréhension des politiques de données par les publiants et des différences de perception avec les éditeurs, avec des recommandations pour limiter les incompréhensions : Exploring how the understanding of editors and authors corresponds to the policies themselves. *PLOS*. Thu-Mai Christian, Amanda Gooch et *al.* 2020. Disponible [en ligne](#)

## Accompagner la gestion de données

- Partager les données liées aux publications. Collège Données de recherche du Comité pour la science ouverte. 2022. Disponible [en lien](#)
- Documentation disponible sur [Recherche Data Gouv](#)

# Glossaire

**Data availability statement** : déclaration accompagnant un article précisant les modalités d'accès aux données liées à un article scientifique

**Data paper** : article de recherche qui consiste en la présentation d'un jeu de données

**Données FAIR** : données qui répondent aux exigences suivantes **Findable** (trouvables), **Accessible** (accessibles), **Interoperable** (interopérables), **Reusable** (réutilisables) [Plus d'informations](#).

**Identifiant pérenne** : identifiant alphanumérique permettant d'accéder à une ressource même si son URL change (ex DOI, handle)

**Journal data policy** (ou politique de données d'une revue) : politique d'une revue définissant les attendus vis-à-vis des auteurs en ce qui concerne la gestion et la mise à disposition des données de recherche liées aux articles publiés dans la revue

# Merci de votre attention

[wanda.romanowski@ined.fr](mailto:wanda.romanowski@ined.fr)

[romain.feret@normandie-univ.fr](mailto:romain.feret@normandie-univ.fr)