

Data sharing : trouver, réutiliser et citer des données de recherche

- TP et corrigé des quizz

Pour toutes les questions des quizz, la source est l'enquête menée pour l'édition 2019 du rapport *State of Open Data*. Cette enquête a reçu 9500 réponses de chercheurs du monde entier.

Digital Science, Fane, B., Ayris, P., Hahnel, M., Hrynaszkiewicz, I., Baynes, G., & Farrell, E. (2019). *The State of Open Data Report 2019*. Digital Science. <https://doi.org/10.6084/m9.figshare.9980783.v2>

Nature Research, Penny, D., Fane, B., Goodey, G., & Baynes, G. (2019). State of Open Data 2019 [Data set]. In *Figshare*. <https://doi.org/10.6084/m9.figshare.10011788.v2>

Sommaire

Les liens ne sont pas cliquables dans le fichier PDF.

- [Sommaire](#)
- [Quizz 1, enjeux du partage des données de recherche](#)
 - [A votre avis, à quelle fréquence les chercheurs mettent-ils à disposition des autres leurs données de façon ouverte?](#)
 - [Quels sont ou quels seraient vos problèmes et préoccupations concernant le partage de vos données ?](#)
 - [Quelles circonstances vous inciteraient-elles à partager vos données ?](#)
- [Quizz 2, réutiliser des données](#)
 - [A votre avis, quelle part des chercheurs ont-ils déjà réutilisé des données ouvertes?](#)
 - [De quelle manière pensez-vous que les données partagées par d'autres vous ont été ou pourraient vous être utiles ?](#)
- [TP1, trouver des entrepôts de données](#)
- [TP2, trouver et évaluer un jeu de données](#)
 - [Recherche de données](#)
 - [Evaluation des données trouvées](#)

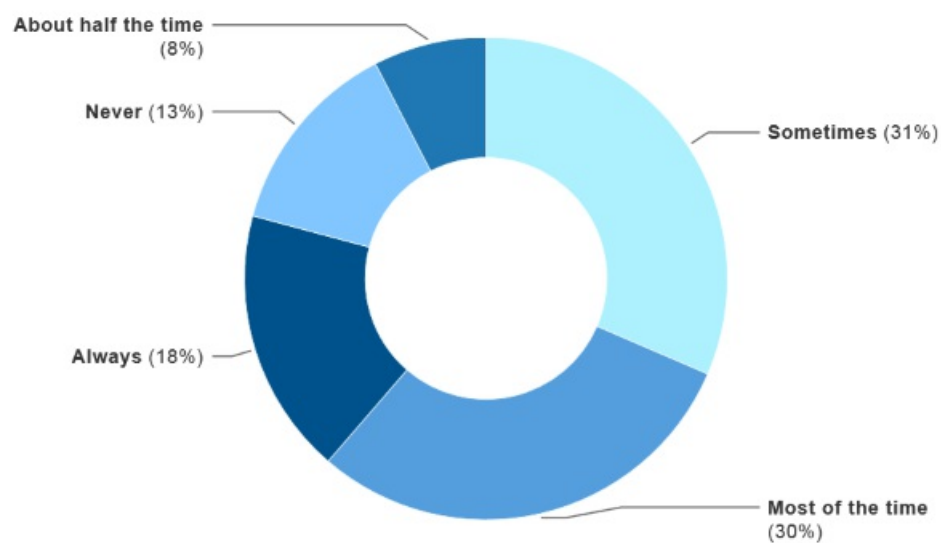
Quizz 1, enjeux du partage des données de recherche

A votre avis, à quelle fréquence les chercheurs mettent-ils à disposition des autres leurs données de façon ouverte?

Une seule réponse possible

- Toujours
- La plupart du temps
- Environ la moitié du temps
- Parfois
- Jamais

How often do you make data openly available to others?



n= 6820

[Get the data](#) • Created with Datawrapper

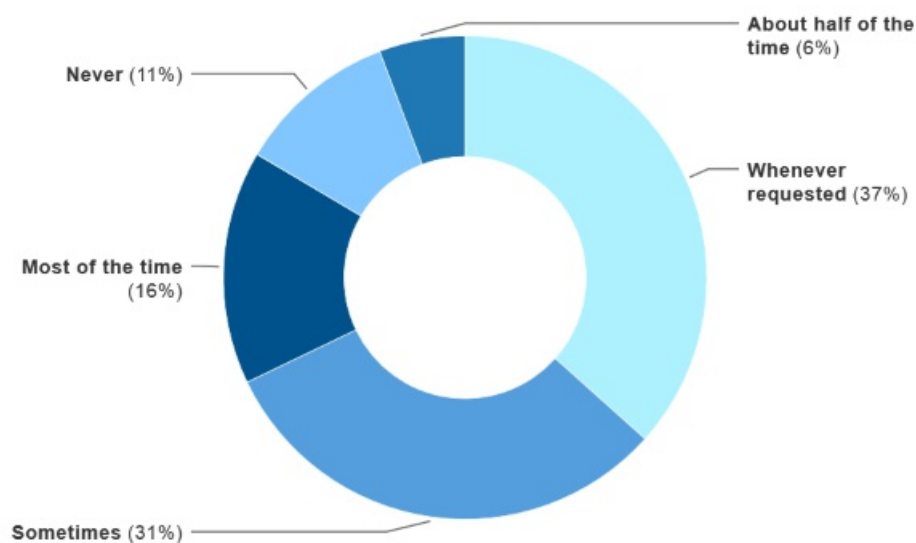
How often do you make data openly available to others?

n = 6820 - réponses triées par fréquence

Fréquence	Nombre de réponses	% des réponses
Parfois - <i>Sometimes</i>	2147	31
La plupart du temps - <i>Most of the time</i>	2036	30
Toujours - <i>Always</i>	1217	18
Jamais - <i>Never</i>	904	13
Environ la moitié du temps - <i>About half the time</i>	516	8

L'enquête comportait également une question concernant la mise à disposition **dans un cadre privé**.

How often do you make your research data privately available to others?



n = 6805

[Get the data](#) • Created with Datawrapper

How often do you make your research data privately available to others?

n = 6805

réponses triées par fréquence

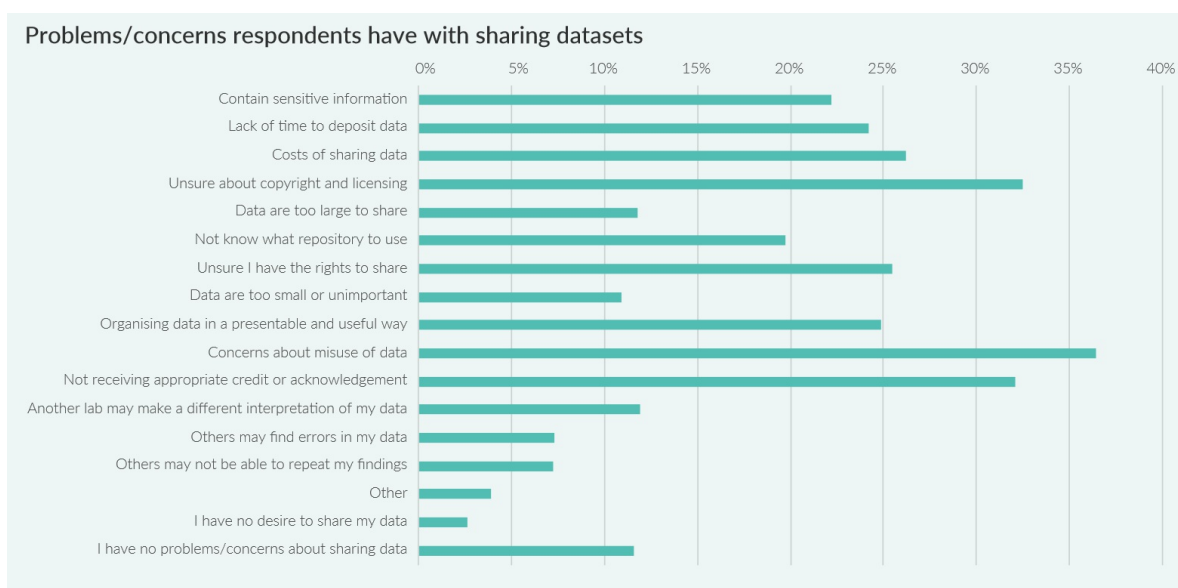
Fréquence	Nombre de réponses	% des réponses
Whenever requested	2493	37
Sometimes	2126	31
Most of the time	1064	16
Never	731	11
About half of the time	391	6

Quels sont ou quels seraient vos problèmes et préoccupations concernant le partage de vos données ?

Plusieurs réponses possibles

- D'autres pourraient trouver des erreurs dans mes données
- Préoccupations concernant un usage impropre des données
- Manque de temps pour déposer les données
- Ne pas recevoir le crédit ou la reconnaissance appropriés
- Je n'ai pas envie de partager mes données
- Autre (veuillez préciser)

What problems/concerns do you have with sharing datasets?



Le tableau ci-dessous est déduit de la lecture de la figure ; les pourcentages sont approximatifs

n= 8423

Modalité de réponse	% des réponses
Préoccupations concernant un usage impropre des données - <i>Concerns about misuse of data</i>	37
Ne pas recevoir le crédit ou la reconnaissance appropriés - <i>Not receiving appropriate credit or acknowledgement</i>	33
Incertitude sur les droits d'auteur et la licence de publication à utiliser - <i>Unsure about copyright and licensing</i>	33
Je ne suis pas sûr d'avoir le droit de partager - <i>Unsure I have the rights to share</i>	26
Coûts associés au partage des données - <i>Costs of sharing data</i>	26
Organiser les données de manière présentable et utile - <i>Organising data in a presentable and useful way</i>	25
Manque de temps pour déposer les données - <i>Lack of time to deposit data</i>	24
Les données contiennent des informations sensibles - <i>Contain sensitive information</i>	22
Je ne sais pas quel entrepôt de données utiliser - <i>Not know what repository to use</i>	19
Je n'ai pas de problèmes/préoccupations concernant le partage des données / <i>I have no problems/concerns about sharing data</i>	12
Les données sont trop volumineuses pour être partagées - <i>Data are too large to share</i>	12
Les données sont trop peu nombreuses ou sans importance - <i>Data are too small or unimportant</i>	12
Un autre laboratoire pourrait faire une interprétation différente de mes données - <i>Another lab may make a different interpretation of my data</i>	12
D'autres ne seraient peut-être pas en mesure de répéter mes résultats - <i>Others may not be able to repeat my findings</i>	7

D'autres pourraient trouver des erreurs dans mes données - Others may find errors in my data	7
Autre (veuillez préciser) / Other (please specify)	4
Je n'ai pas envie de partager mes données - I have no desire to share my data	3

Quelles circonstances vous inciteraient-elles à partager vos données ?

Plusieurs réponses possibles

- Exigence d'un financeur
- Bénéfice public
- Il s'agissait d'une attente de mon domaine de recherche/d'une industrie
- Augmentation de l'impact et de la visibilité de mes recherches
- Autre (veuillez préciser)
- Je ne partagerais jamais mes données

What circumstances would motivate you to share your data?

Which circumstances would motivate you to share your data?	2019 RANK	2019	Count	2018 RANK	2018	Count
Increased impact and visibility of my research	1	62%	3,659	1	62%	841
Public benefit	2	60%	3,522	2	59%	802
Getting proper credit for sharing data	3	54%	3,172	4	46%	621
Journal/publisher requirement	4	51%	3,009	5	44%	599
Transparency and re-use	5	48%	2,817	3	48%	652
Funder requirement	6	47%	2,767	9	33%	453
Institution/organization requirement	7	44%	2,592	7	38%	522
Trust the person requesting my data	8	43%	2,510	6	41%	561
It was made easy and simple to do so	9	36%	2,102	8	36%	485
Freedom of information request	10	30%	1,789	10	26%	352
It was a field/industry expectation	11	22%	1,272	n/a	n/a	n/a
Other (please specify)	12	3%	191	11	5%	63
I would never share my data	13	2%	94	12	1%	17
TOTAL		100%	5,886		100%	1,359

Quizz 2, réutiliser des données

A votre avis, quelle part des chercheurs ont-ils déjà réutilisé des données ouvertes?

C'est-à-dire ont répondu "Oui" à la question suivante: *Have you ever reused open data?*

n = 6764

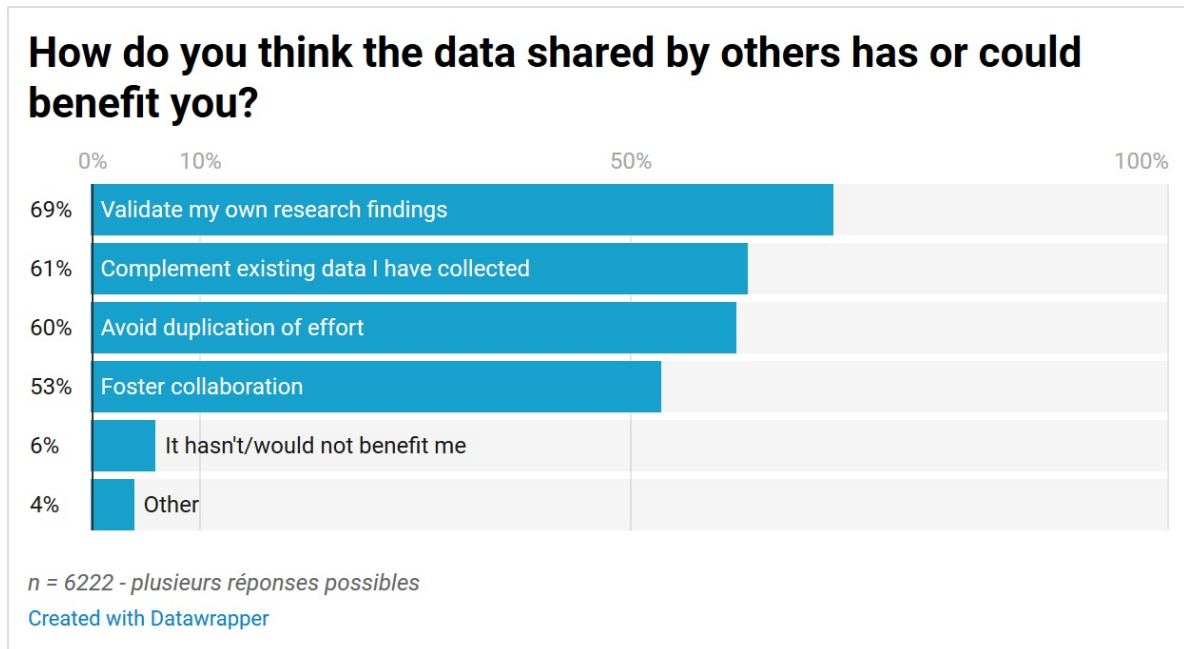
Modalité de réponse	Nombre de réponses	% des réponses
Yes	3117	46
I don't know	432	6,5
No	3215	47,5

De quelle manière pensez-vous que les données partagées par d'autres vous ont été ou pourraient vous être utiles ?

Plusieurs réponses possibles

- Favoriser des collaborations
- Valider mes propres résultats de recherche
- Compléter des données existantes que j'ai collectées
- Éviter de dupliquer des efforts
- Autre (veuillez préciser)
- Cela ne m'a pas été ou ne me serait pas utile

How do you think the data shared by others has or could benefit you?



TP1, trouver des entrepôts de données

A partir des services et outils mentionnés précédemment, il s'agit d'identifier un ou plusieurs entrepôts de données pertinents pour votre domaine de recherche.

Si vous ne trouvez pas d'entrepôt spécifique pour votre domaine de recherche, consultez les fiches re3data et interrogez sommairement les 3 entrepôts de données généralistes suivants.

URL re3data : <https://www.re3data.org/>

- **Zenodo** : <https://zenodo.org/>
- **Dryad** : <http://www.datadryad.org/>
- **B2SHARE** : <https://b2share.eudat.eu/>

Les utiliseriez-vous? Pourquoi?

Quelles différences repérez-vous entre ces 3 entrepôts?

TP2, trouver et évaluer un jeu de données

Recherche de données

A partir des services et outils mentionnés précédemment, il s'agit d'identifier des données ou des jeux de données pertinents pour votre projet de recherche.

Comment procédez-vous?

Parmi tous ces services et outils, lesquels utilisez-vous?

/!\ Vous pouvez retenir des données non librement accessibles, pour lesquelles seule la description est librement accessible.

Evaluation des données trouvées

Une fois identifié un jeu de données qui vous semble pertinent, considérez les points suivants.

- Pensez-vous que ce jeu de données puisse vous être utile? Pourquoi?
- Les **métadonnées** sont-elles suffisantes? La **documentation** (fichier README, etc.) vous paraît-elle suffisante?
- Qu'en est-il de la licence de publication? Les conditions de **réutilisation** sont-elles claires?
- Sauriez-vous comment **citer** ce jeu de données?



Allez plus loin : calculez le score FAIR (Findable, Accessible, Interoperable, Reusable) de ce jeu de données

avec le service en ligne suivant.

Australian Research Data Commons. (s. d.). *FAIR self-assessment tool*. Australian Research Data Commons.

<https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/>

/!\ Concentrez-vous sur les lettres F, A et R