La mise en œuvre de mécanismes de citation de données

Edward J. Gray, Nicolas Larrousse

▶ **To cite this version:**

Edward J. Gray, Nicolas Larrousse. La mise en œuvre de mécanismes de citation de données. #dhnord2021 : Publier, partager, réutiliser les données de la recherche : les data papers et leurs enjeux, Nov 2021, Lille, France. . hal-03435034

HAL Id: hal-03435034
https://hal.science/hal-03435034

Submitted on 18 Nov 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.

Distributed under a Creative Commons Attribution 4.0 International License

#dhnord2021

"La mise en œuvre de mécanismes de citation de données : Un socle solide pour bâtir des *data papers*"

Edward J. Gray et Nicolas Larrousse
TGIR Huma-Num CNRS

Méthodologie : De la citation de données aux entrepôts

Quelles recommandations pour la création d'une "bonne" citation ?

1) Evaluation des pratiques[1] de citation en SHS
2) Développement de Recommandations[2] adaptées aux SHS à partir des principes de Force11
3) Analyse des entrepôts de données en SHS à partir des Recommandations[2]

Quels critères pour évaluer un entrepôt du point de vue de la citation ?

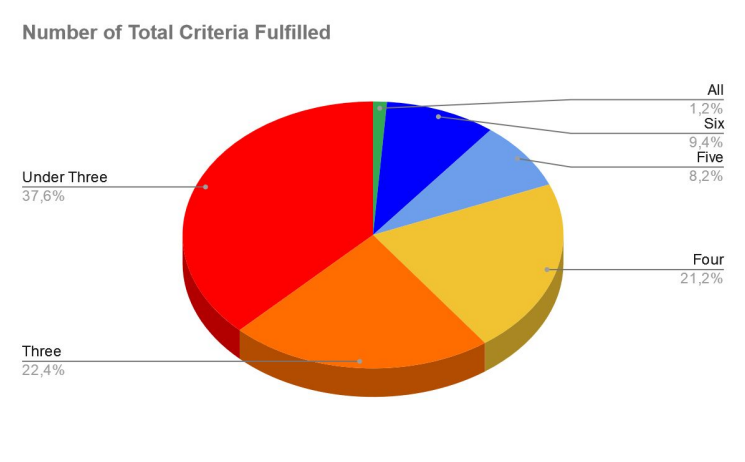
- **Présence d'un Identifiant Pérenne (PID *Persistant IDentifier*)**
- **Présence d'une page d'accueil**
- **Présence des métadonnées structurés actionnables dans le page d'accueil**
- **Présence d'une fonctionnalité "Cite As"**
- Utilisation de vocabulaires standardisés pour la description (e.g. ORCID)
- Présence d'une fonctionnalité d'accès aux différentes versions (versionning)
- Présence de liens vers les publications liées à ces données

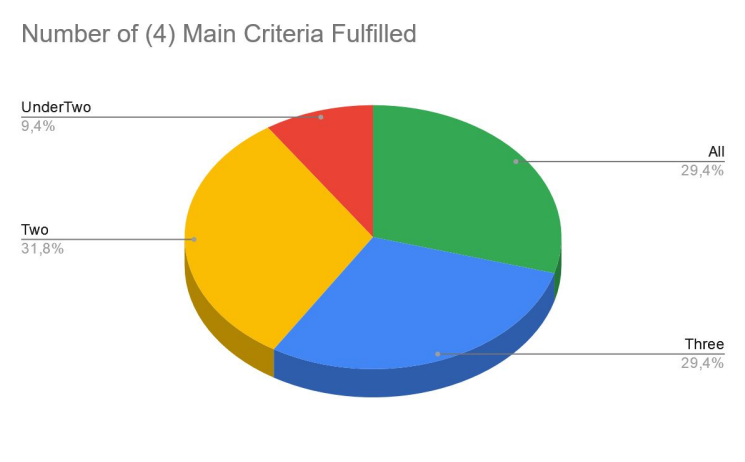
Ces 4 critères représentent les critères principaux

1. Nicolas Larrousse, Daan Broeder, Jan Brase, Cesare Concordia, and Vasso Kalaitzi. "SSHOC D3.2 Inventory of SSH Citation Practices, and Choice for SSHOC Citation Formats and Implementation Planning". Zenodo, December 31, 2019. https://doi.org/10.5281/zenodo.4436736
2. Nicolas Larrousse, and Edward J. Gray. "Recommendations for FAIR Data Citation in the Social Sciences and Humanities". Zenodo, September 1, 2021. https://doi.org/10.5281/zenodo.5361718

Entrepôts : Résultats de l'évaluation

Analyse de 85 Entrepôts de données en SHS sur **7 critères** dont 4 principaux[3]





Nombreuses pistes d'amélioration de la qualité des entrepôts pour mettre en oeuvre la citation de données, en particulier :

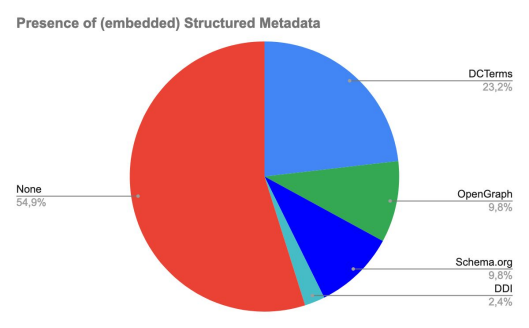
- La fonctionnalité "CiteAs" n'est présente que dans 50% des entrepôts et les chaînes générées ne sont en général pas exploitables par des machines
- Une meilleure structuration des métadonnées en particulier dans les pages d'accueils
- L'utilisation de vocabulaires standardisés

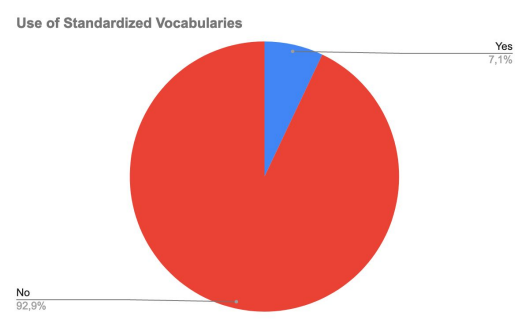
3. Nicolas Larrousse, Edward Gray, Daan Broeder, Cesare Concordia, Jan Brase, and Athina Papadopoulou. "D3.5 Report on Citation Enabled SSH Catalogues and SSH Citation Exploitation". Zenodo, September 16, 2021. https://doi.org/10.5281/zenodo.5603306

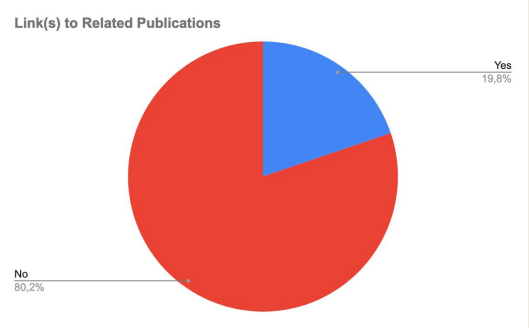
De la citation aux *data papers*

Les **critères** nécessaires à la construction d'une bonne citation vont permettre d'alimenter les informations requises dans un *data paper* (e.g. Titre, Auteurs etc.)

Le fait de rendre les citations "**actionnables**", en particulier par la normalisation des contenus (e.g. structuration de l'information dans un format standard et utilisation de vocabulaires comme ORCID), permet d'envisager **d'automatiser** le transfert de données descriptives pour un *data paper* à partir d'un entrepôt







Les données extraites de l'évaluation des entrepôts montrent que ces aspects sont encore à développer