

Guideline rédaction Data paper – revue DEMC *version juillet 2024*

Le guideline suivant a pour but d'apporter quelques éclairages à celles et ceux qui souhaiteraient s'engager dans la rédaction d' "articles de données", c'est-à-dire présentant une base de données originale ou un matériel empirique original. L'originalité est définie de deux manières : création d'informations nouvelles ou rassemblement original d'informations préexistantes mais disparates ; présentation d'une méthode de production / d'extraction de données. Cette rubrique correspond au format émergent du "data paper". Vous êtes encouragé.e à traiter les points suivants :

Objectifs de la recherche sous-tendant la création de la base de données

- contours généraux de la recherche sur laquelle repose la constitution d'une base de données, d'un corpus de documents ou d'archives, etc. (origine, cadre, financement, durée, lieux, etc.)
- éléments sur les controverses méthodologiques ou théoriques ayant guidé la création des données
- explicitation de la question de recherche et de la manière dont les données y répondent
- éléments renseignant sur la manière dont les producteurs des données se situent dans ce courant de recherche ; ces éléments peuvent, *si cela est pertinent*, relever d'un registre plus personnel, afin de donner à comprendre comment les opérations de constitution des données ont pu être menées (éventuels contacts que les auteurs de la recherche auraient préalablement noués avec les groupes enquêtés, éléments biographiques faisant écho au thème de l'enquête, etc.)

Méthode de constitution et/ou sources

- les grandes lignes de la méthode d'enquête suivie peuvent être exposées ; des informations pourront être apportées sur le contexte de création des données
- des précisions pourront être apportées quant au choix de la "population" / des "cas"
- des précisions pourront être apportées sur les procédures de constitution d'un échantillon, que l'échantillon porte sur un grand N avec un objectif de représentativité statistique, ou qu'il porte sur un petit N avec une démarche d'échantillonnage théorique visant la diversification des "cas"
- des informations générales pourront être apportées sur les modes de contact et de "recrutement" des individus, la fréquence de collecte, les méthodes de pondération, etc.
- des précisions pourront être apportées sur certaines spécificités du questionnaire ou du guide d'entretien ou d'observation
- si cela est pertinent, des informations pourront être apportées pour décrire le ou les guide(s), grille(s) ou protocoles d'analyse ayant permis d'exploiter et d'interpréter les données

Présentation du contenu et de l'organisation du jeu de données

- informations synthétiques sur les types de données (base de donnée créée à partir de sources préexistantes, mise en oeuvre de questionnaires, collection de documents ou d'archives, collecte d'entretiens ou d'observations, etc.), leur volumétrie, ainsi que les formats de conservation
- organisation de la base de données ; s'il existe une arborescence de fichiers, décrire son principe,

Guideline rédaction Data paper – revue DEMC

version juillet 2024

les liens entre les éléments ; à titre d'illustrations il est possible de fournir des captures d'images

- indiquer les traits significatifs des éléments du corpus, par exemple : combien d'entretiens ont été réalisés avec quelle sous-population ? quelle est la durée moyenne des entretiens ? les transcriptions sont-elles disponibles en format texte ? les enregistrements audio sont-ils également disponibles ? si cela est pertinent, préciser les langues utilisées dans les documents composant le corpus (français, anglais, autres langues)
- donner des précisions sur les opérations réalisées sur les données, comme par exemple les recodages, les "nettoyages", l'utilisation de scripts, etc.

Modalités d'accès aux données

- indiquer un lien pérenne de type DOI, et préciser l'entrepôt hébergeant les données
- indiquer les modalités d'accès à la base de données ; si certains éléments du corpus sont en open access et d'autres protégés, voire en embargo, préciser les raisons justifiant le degré d'ouverture des données ; indiquer également si d'autres éléments peuvent être consultés dans un service d'archives
- dans le cas d'enquêtes comparatives où seule la partie française des données a été gérée par les auteurs du Data paper, fournir des informations minimales sur l'état des bases de données produites et éventuellement archivées et/ou mises à disposition dans d'autres pays
- s'il s'agit de séries, préciser quels éléments sont accessibles au sein de la série
- si la base de données est amenée à être incrémentée, actualisée ou modifiée, l'information pourra également être précisée

L'exploitation des données : analyses et interprétations

- donner des indications sur les démarches d'analyse mobilisées dans la recherche initiale
- si le corpus mis à disposition comprend des documents constituant des traces des opérations d'analyse, indiquer où ils se situent dans le corpus et comment ils doivent être manipulés
- si cela est pertinent, mentionner les logiciels utilisés et les modalités de leur utilisation
- proposer un résumé des principaux résultats produits dans le cadre de l'exploitation initiale des données – y compris, et notamment, lorsque ces résultats ont été exposés en ordre dispersé dans plusieurs publications. Ces informations doivent rester très synthétiques et servir avant tout à un cadre de compréhension global de la recherche initiale

Perspectives de réutilisation

- les données ont-elles déjà servi ? si oui indiquer des projets de réutilisation plus ou moins aboutis (y compris ayant donné lieu à des publications, qu'il convient de citer)
- donner des indications sur le potentiel de réutilisation des données à des fins de recherche ou d'enseignement – ou pour d'autres publics si les données sont en open access
- dans le cas de bases de données assises sur un échantillon réduit mais localisé, il est possible de

Guideline rédaction Data paper – revue DEMC
version juillet 2024

souligner leur potentiel de réutilisation pour des recherches à plus large échelle (nationale, internationale), pour des besoins portant sur des populations spécifiques ou difficiles à atteindre

- évoquer les protocoles d’anonymisation des données personnelles / sensibles qui ont pu avoir lieu en vue de la mise à disposition du jeu de données ; si cela est pertinent, indiquer si et comment les consentements des enquêtés ont pu être obtenus

Les auteurs et autrices sont libres dans le choix de la problématique et de l’organisation des contenus. Il est recommandé d’optimiser le degré de précision des informations qui seront fournies sur la base de données ou le matériel empirique. Si certaines informations (techniques, résultats de la recherche, etc.) sont déjà développées dans les publications existantes (articles de recherche, articles méthodologiques, littérature grise, etc.) ou la documentation (métadonnées, documentation ad hoc) accompagnant le jeu de données, il est souhaitable de se contenter de rappels très synthétiques et de privilégier les renvois vers ces sources.