

Opening Research Data: What Does It Mean for Social Sciences?

(Ouvrir les données de la recherche : Quelles implications pour les sciences sociales?)

Héloïse Berkowitz, CNRS, LEST, Aix Marseille Université, heloise.berkowitz@univ-amu.fr

Hélène Delacour, Université de Lorraine, helene.delacour@univ-lorraine.fr

Résumé

Les tendances internationales œuvrent pour l'ouverture de la science sur l'ensemble de son cycle de vie, c'est-à-dire depuis les données de la recherche jusqu'à la publication des résultats. En sciences sociales, si le processus de publication est déjà largement ouvert ou en voie de l'être, la situation est différente concernant l'ouverture des données de la recherche qui soulève de nombreuses questions. Afin de participer à ce débat, nous cherchons, dans cet article, à expliquer ce que signifie l'ouverture des données de la recherche en sciences sociales, à comprendre ses motivations ainsi que les conditions de son opérationnalisation. En effet, alors que l'écosystème d'acteurs, d'infrastructures, de standards et de principes directeurs (FAIR et CARE) se structurent en France et à l'étranger, plusieurs éléments entravent actuellement le processus d'ouverture des données en sciences sociales : 1) une perception erronée des motivations pour l'ouverture des données, que sont la patrimonialisation, la mise en commun des données et les synergies potentielles, l'instauration d'un climat de confiance, un engagement plus large, plutôt que le simple exercice d'un contrôle sur la liberté académique et les équipes de recherche, 2) un système de concurrence et de starification de la recherche, 3) un manque de ressources qui pourrait davantage aggraver les inégalités entre les sexes, les communautés, les institutions et les pays, et 4) des risques potentiels liés à l'ouverture des données et les contraintes spécifiques des sciences sociales. Dans ce contexte, nous proposons des solutions pour rendre opérationnels l'ouverture des données en sciences sociales, au-delà du simple respect des principes FAIR et CARE, avant de présenter la nouvelle politique d'ouverture des données de la recherche de *M@n@gement*.

Mots clefs

Données ouvertes ; Open science; Principes FAIR ; Principes CARE ; Sciences sociales

Citation : Berkowitz H. & Delacour H. (2022) Opening research data: What does it mean for social sciences? *M@n@gement*, 25(4): 1–15
- <http://dx.doi.org/10.37725/mgmt.v25.9123>.

1. Introduction

Lors de la dernière décennie, le paysage institutionnel et réglementaire scientifique a connu de nombreuses transformations. Le Règlement Général européen sur la Protection des Données (RGPD) adopté en 2016 et mis en œuvre en 2018, a profondément modifié la collecte et la gestion des données de la recherche. Parallèlement, la multiplication des plans, des engagements et des fonds en faveur de la science ouverte, notamment le fonds national français pour la science ouverte, la Déclaration de San Francisco sur l'évaluation de la recherche (DORA) publiée en 2013, les principes H2020 FAIR data, la Coalition S, le projet de recommandations 2021 de l'UNESCO sur la science ouverte, et plusieurs expérimentations récentes de modèles alternatifs de publication viennent bousculer les acteurs, les normes et les pratiques. Ces tendances mondiales ont commencé à aborder différentes questions, comme la nécessité de faciliter l'accès à la science en tant que bien public mondial en réponse aux critiques croissantes de la marchandisation de la science et du modèle à but lucratif de la publication scientifique (Berkowitz & Delacour, 2020), ou encore la nécessité de faire face au nombre toujours plus important de scandales éthiques ou de fraudes scientifiques ainsi qu'un besoin accru d'évaluer la qualité des données (Peng et al., 2021).

Dans ce contexte général, la science ouverte ou la recherche ouverte décrit la communication et le partage transparents et ouverts de la recherche sur l'ensemble du cycle de vie et de l'écosystème de la recherche, des parties prenantes et des données de la recherche aux méthodes, logiciels, outils et publications utilisés (Willinsky, 2005). Le degré d'adoption de la science dite ouverte varie selon les domaines et les disciplines ainsi que selon le cycle de vie des données de la recherche et l'écosystème disponible des données ouvertes. En effet, dans de nombreuses disciplines, la publication ouverte devient de plus en plus fréquente et s'institutionnalise progressivement. Cela se traduit, par exemple, par des infrastructures et des acteurs clairement identifiés et légitimes comme OpenEdition ou OJS, des normes et des labels reconnus comme DOAJ, ou encore une variété de modèles et d'initiatives qui favorisent la biodiversité. La biodiversité peut être définie comme la diversité des supports d'écriture, de révision, de publication et des pratiques socio-matérielles (par exemple, les revues en accès libre, l'évaluation ouverte par les pairs, *Peer Community In*, les monographies ouvertes). Dans le cas très spécifique des données ouvertes de la recherche, la situation reste cependant différente.

Dans le cadre des données ouvertes, les principes FAIR (Findable, Accessible, Interoperable, Re-usable) sont les principaux qui encadrent et guident cette ouverture des données (The FAIR data principles - FORCE11, 2014 ; Wilkinson et al., 2016). Publiée quelques années plus tard, en 2019, la Global Indigenous Data Alliance (GIDA) a ensuite proposé un cadre complémentaire, les principes CARE (Collective benefit, Authority to control, Responsibility, Ethics) pour la gouvernance des données liées aux communautés autochtones. Si les principes FAIR sont généralement connus, appliqués voire même exigés par certains organismes de financement (Roche et al., 2015 ; Wilkinson et al., 2016), les principes CARE restent, eux, plus confidentiels et moins mis en œuvre, malgré leur importance. Par ailleurs, dans le cadre spécifique des sciences sociales, il n'est pas toujours évident de savoir ce que signifie réellement les données ouvertes de la recherche,

ni les principes FAIR et CARE. Ces questions sont d'autant plus importantes que les incitations à l'ouverture des données de la recherche ont suscité des inquiétudes, des réticences, voire une résistance active en sciences sociales (Bonneville et al., 2021 ; Longley Arthur & Hearn, 2021).

Nous considérons que les principaux arguments en faveur de l'ouverture des données devraient être de renforcer la science en tant que bien public mondial grâce à la patrimonialisation, à la mise en commun des données et aux synergies, à la vérification et à l'instauration de la confiance, à un engagement plus large, à la valorisation de la production de données, ainsi qu'à la facilitation et à l'accélération de la recherche orientée sur le changement climatique et les 'grands challenges'. Nous soutenons également que si l'écosystème d'acteurs, d'infrastructures et de normes et standards se structure progressivement en France et à l'étranger, plusieurs facteurs entravent actuellement le processus d'ouverture des données en sciences sociales et devraient être questionnées collectivement : 1) une perception erronée des motivations à l'ouverture des données, c'est-à-dire la construction d'un bien public mondial plutôt que le simple exercice d'un contrôle sur la liberté académique et les équipes de recherche, 2) un système de concurrence caractérisé par la domination des classements et une starification de la recherche, 3) un manque ou un accès inégal aux ressources qui pourrait aggraver les inégalités entre les sexes, les communautés, les institutions et les pays, et 4) des risques potentiels liés à l'ouverture des données, notamment en raison des spécificités des sciences sociales qui s'appuient sur une variété des données (entretiens, observations sur le terrain, photos, ensemble de données économiques et autres bases de données quantitatives ou archives, par exemple), avec différents degrés de sensibilité et de possibilité de normalisation.

Dans cet article, nous cherchons à contribuer à une meilleure compréhension de ce que signifient l'ouverture des données de la recherche en sciences sociales. A cette fin, nous commençons par proposer des définitions afin de clarifier certains termes ainsi qu'un aperçu de l'écosystème des acteurs, des infrastructures et des principes soutenant l'ouverture des données de la recherche. Nous présentons ensuite brièvement les principales motivations à l'ouverture des données, en particulier en sciences sociales. Puis, en nous référant à la littérature émergente sur la science ouverte et en particulier les données ouvertes de la recherche, nous mettons en évidence plusieurs éléments qui pourraient entraver le processus d'ouverture des données en sciences sociales. Enfin, nous nous intéressons à la mise en œuvre opérationnelle et aux questions soulevées par l'ouverture des données de la recherche, avant de présenter la nouvelle politique de *M@n@gement* en matière de données ouvertes et de conclure.

2. Présentation de l'écosystème des données ouvertes et du cycle de vie des données de la recherche

2.1. Définitions et présentation de l'écosystème

Avant de décrire l'écosystème des données ouvertes de la recherche, nous souhaitons définir des termes spécifiques liés aux données ouvertes, tels que les différents types d'accès aux données, les métadonnées, le référentiel de données et le plan de gestion des données (PGD).

Premièrement, il existe différents types d'accès aux données, qui varient en fonction de leur degré d'ouverture : 1) accès totalement ouvert, sans aucune barrière d'accès, 2) accès sous embargo, ce qui se traduit par un accès aux données seulement à la fin de l'embargo, 3) accès restreint, avec certaines barrières qui permettent de limiter l'accès sous condition aux utilisateurs externes, 4) accès fermé, avec une non-accessibilité aux données. Il existe également différentes catégories de données, ce qui aura un impact sur leur accessibilité : données tièdes, chaudes, froides, en fonction de la fréquence de leur accessibilité. En parallèle, la sensibilité des données désigne les informations qui doivent être protégées et/ou restreintes en raison de leur nature sensible (santé, sexe, croyance, orientation sexuelle, etc.).

Deuxièmement, les métadonnées décrivent les informations clés relatives à ces données qui facilitent leur découverte et ainsi, conditionnent leur accessibilité. La structure et le contenu des métadonnées sont généralement guidés par une norme de métadonnées et varient souvent d'une discipline ou d'un dépôt à l'autre. Il existe diverses normes comme DDI, COAR, DCMI. L'ensemble des normes est disponible dans le référentiel *Fair Sharing* (<https://fairsharing.org/>).

Troisièmement, un autre terme couramment utilisé est celui de dépôt de données qui correspond à l'endroit où les données sont déposées. Les données publiées ont besoin d'un identifiant permanent comme un DOI (*Digital Object Identifier*), souvent fourni par le dépôt. Il existe plusieurs dépôts, comme Zenodo et NAKALA pour les données générales, ELIXIR pour les sciences de la vie ou CESSDA ou DARIAH pour les sciences sociales et humaines, respectivement. Des sites tels que Re3data (<https://www.re3data.org/>) offrent un aperçu des dépôts existants. D'autres types de méta-plateformes, comme ISIDORE, récoltent les données produites en sciences humaines et sociales, les agrègent et les traitent, afin de fournir un accès unifié et enrichi. De même, OpenAIRE est une méta-organisation européenne d'accès ouvert qui agrège des dépôts, des archives et des revues en libre accès. OpenAIRE fournit également des services de données ouvertes, comme Amnesia, un outil d'anonymisation des données (<https://www.openaire.eu/amnesia-guide>). En France, Huma-Num est une infrastructure de recherche dédiée aux sciences humaines et sociales, qui est impliquée, par exemple, dans DARIAH et à l'origine de la création des dépôts NAKALA ou de la plateforme Isidore, entre autres services.

Quatrièmement, un plan de gestion des données (PGD ou *Data management plan*), exigé par la plupart des organismes de financement, décrit où, comment et quand les données

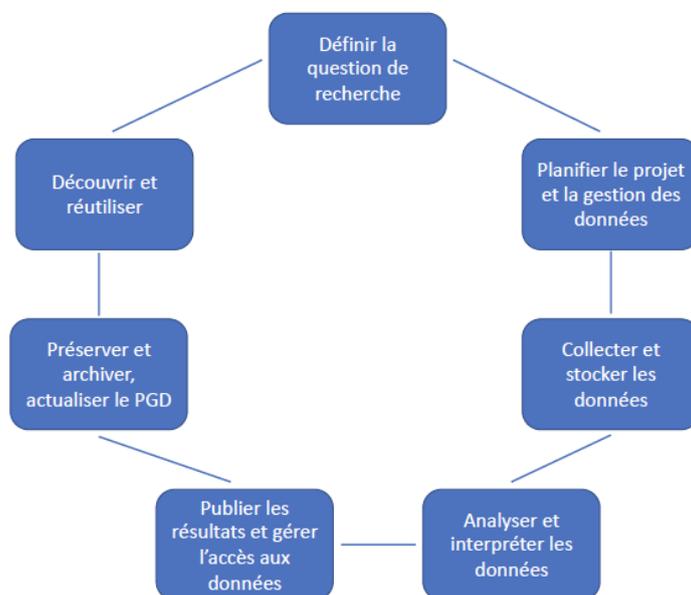
du projet sont collectées, stockées et accessibles. Le PGD Opidor, par exemple, a été développé pour aider à la définition de plans de gestion des données. Nous aborderons ce qu'est un PGD de manière plus détaillée dans la section 4.

En complément de ces définitions, l'écosystème des données ouvertes peut être décrit à partir de ses parties prenantes : depuis la production de données (par les équipes de recherche, par exemple, les ingénieur·es, le personnel technique, les chercheur·es), jusqu'à l'édition et la conservation de données (par exemple, les organismes de normalisation, les dépôts et autres plateformes), les organismes de financement qui imposent généralement des exigences en termes d'accessibilité, et les utilisateurs de données (par exemple, les universitaires, les décideurs, les citoyen·nes) qui utilisent ces données accessibles.

2.2. Cycle de vie des données et flux de travail

Les objectifs de réutilisation et de patrimonialisation des données impliquent que l'ensemble des données de la recherche a une durée de vie plus longue que le projet de recherche lui-même. Le cycle de vie des données de la recherche peut être illustré comme suit (Figure 1).

Figure 1. Cycle de vie des données (adapté de Princeton Research Lifecycle Guide)



A partir d'une question de recherche posée, un plan de gestion des données (PGD) consiste à identifier les types de données nécessaires, qui seront ensuite collectées puis gérées. À ce stade, une première version du PGD est généralement créée, dans le cadre d'une demande de financement par exemple. Elle comprend certaines questions telles que l'endroit où les données seront stockées, si elles seront accessibles et de quelle manière.

Ensuite, pendant la phase empirique, les chercheur.es collectent des données, mènent des enquêtes ou des entretiens, consultent des archives et autres formats de données. Ils stockent et décrivent les méthodes de collecte des données.

Une fois les données collectées, elles doivent être traitées, nettoyées, combinées, sélectionnées ou supprimées. Le traitement des données doit également être documenté afin de garantir le contrôle de la qualité et sa réplique. Ce n'est qu'ensuite que les données peuvent être analysées et interprétées.

À des fins de publication, il peut être nécessaire de donner accès aux données brutes et traitées, à leur visualisation, à leur documentation, via, par exemple, des plateformes de données (avec des droits d'accès fermés ou ouverts). À ce stade, la validation des données peut avoir lieu, soit par le biais d'un processus normal d'évaluation, soit par une validation spécifique des données.

Vers la fin du projet, le partage des données peut nécessiter des procédures de préservation et d'archivage appropriées qui garantiront leur accessibilité et leur réutilisation à long terme. La préparation de la préservation et de l'archivage des données peut être différente de la collecte ou du traitement. Cette étape implique, entre autres, la description des métadonnées, l'attribution d'un identifiant permanent qui permet la citation, comme un numéro DOI, l'octroi de licences pour la réutilisation des données, le stockage dans un dépôt de données et la gestion de l'accès aux données, en fonction de leur sensibilité.

Il peut exister des variations tout au long de ce cycle de vie schématique, car la sensibilité et l'accessibilité des données peuvent avoir des effets sur leur ouverture ou leur diffusion. Pour illustrer davantage cette question, nous nous référons à la description détaillée des flux de publication de données tels que décrits par Austin et al. (2015) (figure 2). Leurs schémas sont particulièrement illustratifs pour souligner la complexité croissante des flux de travail ou *workflow*, notamment dans le cas des articles sur les données (*data research paper*), c'est-à-dire les articles qui décrivent et accompagnent les jeux de données.

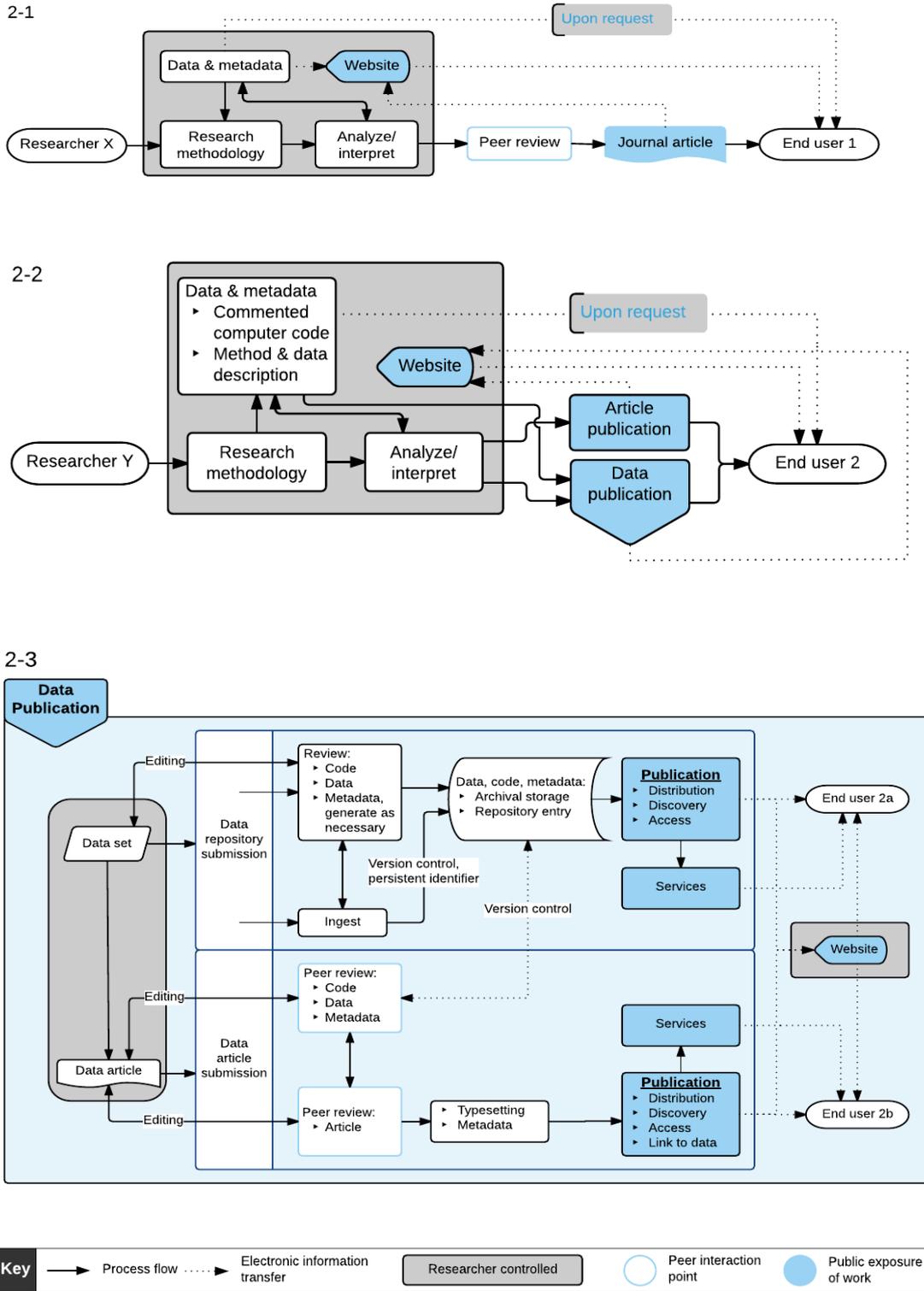


Figure 2. Research data publication workflows reproduit d’après Austin et al. (2015, p. 16). Flux de travail pour les articles de recherche traditionnels (Fig. 2-1), Flux de travail pour les articles de recherche reproductibles (Fig. 2-2), et Flux de travail pour les publications avec données ouvertes (Fig. 2-3).

Après cette présentation du cycle de vie des données, nous présentons les principaux guides qui structurent les pratiques de données ouvertes.

2.3. Cadre directeur et principes guidant l'ouverture des données de la recherche

En 2014, une déclaration commune de citation des données (JDDC) a été publiée par FORCE11. FORCE11 est une communauté composée de bibliothécaires, de chercheurs, d'éditeurs ainsi que de diverses institutions qui s'attachent à l'évolution des pratiques vers une création et un partage de la recherche plus ouverts, transparents et collaboratifs. Cette déclaration est l'une des premières à considérer les données comme un résultat de recherche qui a de la valeur en soi. Cette déclaration visait également à encourager les meilleures pratiques en matière de citation et de réutilisation des données, en définissant un ensemble de principes directeurs pour la citation des données (Data Citation Synthesis Group : Déclaration conjointe des principes de citation des données, 2014) (tableau 1).

Tableau 1. La déclaration commune des principes de citation des données

1. Importance

Les données doivent être considérées comme des produits légitimes et citables de la recherche. Les citations de données doivent se voir accorder la même importance dans un dossier scientifique que les citations d'autres objets de recherche, tels que les publications.

2. Crédit et attribution

Les citations de données doivent faciliter l'attribution du crédit scientifique, d'un point de vue légal, à tous ceux qui ont contribué aux données, tout en reconnaissant qu'un mécanisme d'attribution unique peut ne pas être applicable à toutes les données.

3. Preuves

Dans la littérature scientifique, chaque fois qu'une affirmation repose sur des données, les données correspondantes doivent être citées.

4. Identification unique

Une citation de données doit inclure une méthode d'identification stable qui soit exploitable par une machine, unique au niveau mondial et largement utilisée par une communauté.

5. Accès

Les citations de données doivent faciliter l'accès aux données elles-mêmes ainsi qu'aux métadonnées, à la documentation, au code et aux autres éléments associés, qui sont nécessaires aussi bien aux humains qu'aux machines pour utiliser les données référencées en connaissance de cause.

6. Persistance

Les identifiants uniques et les métadonnées décrivant les données et leur disposition doivent durer dans le temps, même au-delà de la durée de vie des données qu'ils décrivent.

7. Spécificité et vérifiabilité

Les citations de données doivent faciliter l'identification, l'accès et la vérification des données spécifiques qui soutiennent une affirmation. Les citations ou les métadonnées de citation doivent inclure des informations sur la provenance et la stabilité pour faciliter la

vérification que la période considérée, la version, et/ou la partie granulaire des données récupérées ultérieurement sont les mêmes que celles citées initialement.

8. Interopérabilité et flexibilité

Les méthodes de citation des données doivent être suffisamment souples pour s'adapter aux différentes pratiques des communautés, mais ne doivent pas différer au point de compromettre l'interopérabilité des pratiques de citation des données entre les communautés.

Source : <https://force11.org/info/joint-declaration-of-data-citation-principles-final/>

Sur cette base, une série de principes a émergé pour encadrer et guider le fonctionnement des données ouvertes de la recherche. Plus connus sous le nom des principes FAIR (Findable, Accessible, Interoperable, Reusable), ils ont également été développés par FORCE11, en 2016. Selon les principes FAIR, les données ouvertes de la recherche doivent respecter les exigences suivantes (tableau 2).

Tableau 2. Les principes FAIR

Etre trouvable (Findable)

- F1. Les (méta)données se voient attribuer un identifiant unique au monde et éternellement persistant.
- F2. Les données sont décrites par des métadonnées riches.
- F3. Les (méta)données sont enregistrées ou indexées dans une ressource consultable.
- F4. Les métadonnées spécifient l'identifiant des données.

Etre accessible (Accessible)

- A1. Les (méta)données sont récupérables par leur identifiant en utilisant un protocole de communication standardisé.
 - A1.1. Le protocole est ouvert, gratuit et universellement applicable.
 - A1.2. Le protocole permet une procédure d'authentification et d'autorisation, si nécessaire.
- A2. Les métadonnées sont accessibles, même lorsque les données ne sont plus disponibles.

Être interopérable (Interoperable)

- I1. Les (méta)données utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances.
- I2. Les (méta)données utilisent des vocabulaires qui suivent les principes FAIR.
- I3. Les (méta)données incluent des références qualifiées à d'autres (méta)données.

Etre réutilisable (Reusable)

- R1. Les (méta)données ont une pluralité d'attributs précis et pertinents.
 - R1.1. Les (méta)données sont publiées avec une licence claire et accessible d'utilisation des données.
 - R1.2. Les (méta)données sont associées à leur provenance.
 - R1.3. Les (méta)données répondent à des normes communautaires pertinentes pour le domaine.

Source : <https://www.go-fair.org/fair-principles/>

Quelques années plus tard, en 2019, la Global Indigenous Data Alliance (GIDA) a exposé un cadre complémentaire aux principes FAIR, appelé les principes CARE (Collective benefit, Authority to control, Responsibility, Ethics) pour la gouvernance des données autochtones. Ces principes ont été établis pour faire progresser les droits en matière de données dans le contexte de la Déclaration des Nations Unies sur les droits des peuples autochtones (UNDRIP) et pour garantir que la production, la collecte, le stockage, la diffusion et la citation des données profitent aux peuples autochtones (Carroll et al., 2021). Ces principes sont repris dans leur entièreté dans le tableau 3.

Tableau 3. Principes CARE

Bénéfice collectif (Collective benefit)

Les écosystèmes de données sont conçus et fonctionnent de manière à permettre aux peuples autochtones de tirer profit des données.

C1. Pour un développement et une innovation inclusifs

Les gouvernements et les institutions doivent soutenir activement l'utilisation et la réutilisation des données par les nations et les communautés autochtones en facilitant l'établissement des bases de l'innovation indigène, de la création de valeur et de la promotion des processus de développement local auto-déterminé.

C2. Pour une meilleure gouvernance et un meilleur engagement des citoyen·nes

Les données enrichissent les processus de planification, de mise en œuvre et d'évaluation qui répondent aux besoins des communautés autochtones en matière de services et de politiques. Les données permettent également un meilleur engagement entre les citoyen·nes, les institutions et les gouvernements afin d'améliorer la prise de décision. L'utilisation éthique des données ouvertes a la capacité d'améliorer la transparence et la prise de décision en fournissant aux nations et aux communautés autochtones une meilleure compréhension de leurs peuples, de leurs territoires et de leurs ressources. De même, elle peut permettre de mieux comprendre les politiques et les programmes de tiers qui affectent les peuples autochtones.

C3. Pour des résultats équitables

Les données autochtones sont fondées sur les valeurs communautaires, qui s'étendent à la société dans son ensemble. Toute valeur créée à partir de données autochtones doit bénéficier aux communautés autochtones de manière équitable et contribuer aux aspirations autochtones au bien-être.

Autorité de contrôle (Authority to control)

Les droits et intérêts des peuples autochtones en matière de données autochtones doivent être reconnus et leur autorité pour contrôler ces données doit être renforcée. La gouvernance des données autochtones permet aux peuples autochtones et aux instances dirigeantes de déterminer comment les peuples autochtones, ainsi que les terres, territoires, ressources, connaissances et indicateurs géographiques autochtones, sont représentés et identifiés dans les données.

A1. Reconnaître les droits et les intérêts

Les peuples autochtones ont des droits et des intérêts à la fois sur les connaissances autochtones et sur les données autochtones. Les peuples autochtones ont des droits collectifs et individuels au consentement libre, préalable et éclairé pour la collecte et l'utilisation de ces données, y compris l'élaboration de politiques et de protocoles de collecte de données.

A2. Données pour la gouvernance

Les peuples autochtones ont le droit de disposer de données qui correspondent à leur vision du monde et leur permettent de s'autodéterminer et de s'autogouverner efficacement. Les données autochtones doivent être rendues disponibles et accessibles aux nations et communautés autochtones afin de soutenir la gouvernance autochtone.

A3. Gouvernance des données

Les peuples autochtones ont le droit de développer des protocoles de gouvernance culturelle pour les données autochtones et d'être des leaders actifs dans la gestion et l'accès aux données autochtones, en particulier dans le contexte du savoir autochtone.

Responsabilité (Responsibility)

Ceux qui travaillent avec des données autochtones ont la responsabilité de partager la manière dont ces données sont utilisées pour soutenir l'autodétermination et le bénéfice collectif des peuples autochtones. La responsabilité exige des preuves significatives et ouvertement disponibles de ces efforts et des avantages qui en découlent pour les peuples autochtones.

R1. Pour des relations positives

L'utilisation des données autochtones n'est viable que si elle est liée à des relations basées sur le respect, la réciprocité, la confiance et la compréhension mutuelle, telles que définies par les peuples autochtones auxquels ces données se rapportent. Les personnes travaillant avec des données autochtones doivent s'assurer que la création, l'interprétation et l'utilisation de ces données respectent la dignité des nations et des communautés autochtones.

R2. Pour le développement des capacités et des compétences

L'utilisation de données autochtones implique la responsabilité réciproque d'améliorer la connaissance des données au sein des communautés autochtones et de soutenir le développement d'une main-d'œuvre autochtone spécialisée dans les données et d'une infrastructure numérique permettant la création, la collecte, la gestion, la sécurité, la gouvernance et l'application des données.

R3. Pour les langues et les visions du monde autochtones

Des ressources doivent être fournies pour générer des données ancrées dans les langues, les visions du monde et les expériences vécues (y compris les valeurs et les principes) des peuples autochtones.

Éthique (Ethics)

Les droits et le bien-être des peuples autochtones doivent être la préoccupation première à toutes les étapes du cycle de vie des données et dans tout l'écosystème des données.

E1. Pour minimiser les dommages et maximiser les avantages

Les données éthiques sont des données qui ne stigmatisent pas les peuples autochtones, leurs cultures ou leurs connaissances et ne les dépeignent pas négativement. Les données éthiques sont collectées et utilisées de manière à s'aligner sur les cadres éthiques autochtones et sur les droits affirmés dans l'UNDRIP. L'évaluation des avantages et des inconvénients éthiques doit se faire du point de vue des peuples, nations ou communautés autochtones auxquels les données se rapportent.

E2. Pour la justice

Les processus éthiques traitent des déséquilibres en matière de pouvoir et de ressources, et de la manière dont ils affectent l'expression des droits autochtones et des droits de l'homme. Les processus éthiques doivent inclure une représentation des communautés autochtones concernées.

E3. Pour une utilisation future

La gouvernance des données doit tenir compte de l'utilisation future potentielle et des dommages futurs, sur la base de cadres éthiques fondés sur les valeurs et les principes de la communauté autochtone concernée. Les métadonnées doivent mentionner la provenance et la finalité des données, ainsi que les limitations ou obligations liées à leur utilisation secondaire, y compris les questions de consentement.

Source : <https://www.gida-global.org/>

L'opérationnalisation de ces principes est cruciale. L'Integrated Data Infrastructure (IDI) de Nouvelle-Zélande a réfléchi à cette question et a élaboré un protocole garantissant que les données sont conformes aux principes FAIR et CARE (Carroll et al., 2021). Ce protocole peut être une source d'inspiration. Nous le reproduisons dans la figure 3.



Figure 3. Ngā Tikanga Paihere guidelines (Stats NZ, 2020, p. 10)

Bien que ces principes soient connus et acceptés, ils ne sont pas toujours directement appliqués. Les exigences relatives à l'ouverture des données de la recherche varient généralement d'un financeur à l'autre, sans qu'il y ait d'orientation claire à ce sujet. Par exemple, l'Agence Nationale de la Recherche française (ANR) exige qu'un plan de gestion des données (PGD) soit fourni six mois après le début du projet. Dans certains cas, l'utilisation des principes FAIR est obligatoire, par exemple pour les financements européens.

En ce qui concerne les revues en sciences sociales, peu d'entre elles ont une politique d'ouverture des données. Cependant, nous pouvons remarquer que ces revues exigent de plus en plus de fournir des ensembles de données pour les processus d'évaluation et permettent aux auteur·rices de connecter des ensembles de données en ligne pour les articles acceptés. Nous pouvons supposer que cette tendance émergente observée va progressivement se généraliser à la plupart des revues et, dans certains cas, devenir même une obligation. À cette fin, quelles sont les motivations et quels sont les défis réels posés par l'ouverture des données en sciences sociales ?

3. Raisons et motivations soutenant l'ouverture des données de la recherche

La justification de l'ouverture des données a été détaillée et défendue avant tout par des acteurs publics, nationaux, transnationaux ou internationaux, gouvernementaux ou non gouvernementaux, comme la Commission Européenne, le Centre National de la Recherche Scientifique français (CNRS), FORCE11 ou l'OCDE, afin de promouvoir une plus grande transparence, accessibilité et responsabilité.

Les principales raisons avancées à l'ouverture des données de la recherche financée par des fonds publics sont les suivantes : permettre une vérification plus précise des données, réduire les risques et les coûts de duplication dans la collecte et la production des données, faciliter la mise en commun des données et donc leur réutilisation, favoriser un engagement plus large non seulement des autres chercheur·es mais aussi des citoyen·nes, des gouvernements et des organisations, créer la confiance dans la science (OCDE 2016), faciliter la recherche multidisciplinaire pour répondre aux défis sociétaux complexes comme ceux liés au climat (Gauquelin et al., 2017), valoriser ce qui est parfois considéré comme un travail invisible (c'est-à-dire la collecte et la production de données) et, finalement, assurer la patrimonialisation, c'est-à-dire le stockage et la préservation en toute sécurité. La patrimonialisation est peut-être l'une des raisons les moins connues et pourtant l'une des plus importantes pour promouvoir le développement des données ouvertes de la recherche. En effet, considérer les jeux de données actuels comme de futures archives disponibles représente une réelle opportunité pour la recherche. Les données collectées et produites aujourd'hui détiennent une valeur patrimoniale, surtout lorsqu'elles sont financées par des fonds publics et en tant que telles, elles doivent être préservées et mises à disposition sous certaines conditions (sensibilité, temps, etc.) afin de pouvoir être ré-utilisées par d'autres chercheur·es.

En complément de cette liste de premières raisons évoquées, d'autres arguments peuvent également être avancés. OpenAIRE, l'infrastructure d'accès ouvert pour la recherche en Europe, s'attache à spécifier les avantages des données ouvertes, en fonction des parties prenantes concernées. Pour les chercheur·es, OpenAIRE met en avant une plus grande découvrabilité, une efficacité accrue, de nouvelles collaborations et davantage de financement. Pour les bailleurs de fonds, il s'agit d'une visibilité et d'une réutilisation plus importantes de la recherche financée, d'un impact plus fort ainsi que d'un meilleur retour sur investissement. Pour le grand public, l'ouverture des données de la recherche garantit l'autonomisation, une plus grande transparence et un engagement plus important dans la science et la recherche. Pour les organisations et les ONG, les avantages consistent en un meilleur accès à la recherche, un partage d'informations et un lobbying plus efficace. Enfin, pour les gouvernements, l'ouverture des données de la recherche est la promesse d'un processus décisionnel basé sur les données, d'une réduction de coûts et de services publics plus efficaces et efficients.

Toutefois, ces arguments reposent sur deux hypothèses fortes. Tout d'abord, la science et les données de la recherche devraient contribuer à la performance et à l'efficacité, plutôt qu'à assurer la justice sociale et écologique ou le bien commun. Cela soulève plus généralement la question de l'utilisation commerciale des données ouvertes de la

recherche, sur laquelle nous reviendrons plus tard. Ensuite, les avantages mis en avant par OpenAir impliquent que toutes les parties prenantes citées sont en capacité d'utiliser et de réutiliser les données ouvertes, en particulier les gouvernements et les organisations. Cependant, l'utilisation des données et la prise de décision basée sur lesdites données nécessitent des capacités et des infrastructures spécifiques qui ne sont pas encore toutes disponibles. La simple ouverture des données ne garantit pas en soi que les services gouvernementaux, par exemple, seront plus efficaces.

De notre point de vue, en tant que chercheuses et co-rédactrices en chef de *M@n@gement*, la patrimonialisation, la mise en commun des données et les synergies, la vérification et l'instauration de la confiance, un engagement plus large, la valorisation de la production de données, ainsi que la facilitation et l'accélération de la recherche orientée sur le changement climatique et les problèmes sociétaux complexes constituent les principaux arguments en faveur de l'ouverture des données en sciences sociales, car ils contribuent à renforcer la science en tant que bien public global. Partant de ce constat, nous allons maintenant analyser les défis et obstacles auxquels fait face l'ouverture des données en sciences sociales.

4. Défis et obstacles à l'ouverture des données de la recherche en sciences sociales

Cette section n'a pour objectif d'être le 'Bingo de l'excuse des données' ou les 'Préoccupations concernant l'ouverture des données, et réponses qui se sont avérées efficaces', ou 'Comment se faire des amis et les amener à vous donner leurs données', mis en évidence par Gutteridge et Dutton. Cette liste, plutôt exhaustive, offre un aperçu intéressant des préoccupations courantes concernant l'ouverture des données et de leurs solutions, notamment concernant les risques d'utilisation à des fins terroristes, la possible mauvaise interprétation des données, les lois sur la protection des données, l'autorisation et la propriété, ou encore la (mauvaise) qualité des données collectées. Dans cette section, notre objectif est plutôt de souligner les défis et les obstacles qui nous semblent les plus importants à l'heure d'ouvrir les données en sciences sociales.

4.1. Perception erronée des motivations

L'un des principaux obstacles à l'ouverture des données de la recherche dans le domaine scientifique en général et dans celui des sciences sociales en particulier, repose sur les raisons et justifications de l'ouverture des données. Les chercheur-es en sciences sociales ignorent souvent ou perçoivent mal les motivations de l'ouverture des données. Nous (et en tant qu'auteurs de cet article, nous nous incluons dans ce 'nous') considérons souvent l'ouverture des données comme un moyen pour les autorités d'exercer un contrôle sur la liberté académique et sur les chercheur-es, ou simplement comme une charge supplémentaire de travail. Cependant, les autres raisons mentionnées dans la section précédente, comme la patrimonialisation, la mise en commun des données et les synergies, la vérification et l'instauration de la confiance, la facilitation d'un engagement plus large ou le soutien à une science orientée vers une mission comme le changement climatique, nous apparaissent comme des justifications bien plus importantes et méritent notre considération. De plus, nous partageons la philosophie qui consiste à considérer les données 'aussi ouvertes que possible et aussi fermées que nécessaire'. Cela se traduit par

beaucoup de liberté et de variations et permet de ne plus considérer cette ouverture comme une éventuelle menace.

4.2. Domination de la concurrence et de la starification

La science en général, et les sciences sociales et les sciences de gestion en particulier, oscillent entre un processus de starification de la recherche et un appel à la collaboration, i.e. aux relations dites horizontales, ces dernières étant reconnues comme plus innovantes et ayant le plus d'impact à long terme par rapport aux équipes hiérarchiques dans la recherche (Xu et al., 2022). Toutefois, dans certains domaines en particulier, les classements des écoles, des revues, etc. ont beaucoup d'influence (Osterloh & Frey, 2015). Les sciences de gestion ont même la spécificité d'un double système composé de recherche publique dans les institutions publiques et de recherche privée dans les écoles de commerce. Dans ces écoles, les chercheur·es bénéficient de bonus financiers pour les publications, ce qui introduit de fortes inégalités entre ces deux systèmes. Les données ouvertes de la recherche, en revanche, sont collaboratives, souvent co-construites et facilitent les réutilisations, ce qui pourrait être contradictoire avec un système basé sur la compétition et la starification. Ce paradoxe est lié à l'évaluation académique actuelle qui s'appuie encore massivement sur les publications individuelles dans des revues classées ou à fort facteur d'impact, malgré les changements en cours encouragés par le CNRS ou l'Université d'Utrecht qui soutient les valeurs de la science ouverte (voir Woolston, 2021).

Cela renvoie aussi plus généralement à une question d'attribution de la propriété intellectuelle ou de reconnaissance dans la production de données (Gauquelin et al., 2017). Certains chercheur·es pourraient tout simplement ne pas être disposés à partager leurs données, qu'ils considèrent comme leur propriété. Pour certains, les données ne peuvent pas être partagées ouvertement et mises gratuitement à disposition de tiers, qu'ils soient considérés comme des concurrents ou non. Dans d'autres cas, ceux qui ont réellement produit les données, comme les ingénieurs ou les assistants de recherche, ne sont souvent pas reconnus, ni considérés comme des co-auteur·rices, ce qui soulève des questions déontologiques. L'ouverture des données pourrait encore modifier les relations de pouvoir, car les chercheur·es pourraient être reconnus pour la production de données, au détriment des ingénieur·es et autres collègues du support technique qui pourraient devenir (encore plus) invisibles si leur participation n'est pas reconnue (Gauquelin et al., 2017 ; Millerand & Bowker, 2008).

4.3. Coûts de mise en oeuvre et manque de ressources

Un autre défi face à l'ouverture des données de recherche concerne leurs coûts de mise en œuvre ainsi que le manque de ressources pour répondre aux nouvelles exigences. En effet, les jeux de données sont souvent produits en sciences sociales par un seul chercheur·e ou une équipe de taille réduite (Burgelman et al., 2019). Alors que la collecte de données est une activité très chronophage, la gestion des données ouvertes requiert du temps et des efforts supplémentaires. Un découplage entre les exigences demandées et les perceptions du travail par les chercheur·es pourrait apparaître (Longley Arthur & Hearn, 2021). Par ailleurs, si certains organismes de financement permettent de couvrir une partie des coûts de préparation des données, ce soutien n'est pas répandu, ni généralisé, malgré

des exigences croissantes en matière de données ouvertes. Cette situation peut entraîner, voire aggraver les inégalités entre les chercheurs et les institutions, entre ceux qui disposent des ressources ou du soutien institutionnel et ceux qui ne disposent pas ces capacités.

Par ailleurs, les équipes éditoriales demandent de plus en plus régulièrement des déclarations de disponibilité des données (*data availability statements*) pour indiquer où et comment les jeux de données sont accessibles. Or le soutien de la plupart des institutions reste, à cet égard, encore limité (Longley Arthur & Hearn, 2021). D'ailleurs, Austin et al. (2015) regrettent que l'on demande aux équipes de s'adapter aux infrastructures émergentes, au lieu de construire des plateformes autour des équipes de recherche et de la manière dont elles produisent des données.

4.4. Contraintes et risques spécifiques aux sciences sociales

Parmi les divers défis auxquels sont confrontées les sciences sociales, la pratique de non-partage des pré-publications en ligne rend plus difficile le changement des pratiques (Longley Arthur & Hearn, 2021), notamment concernant le partage des données.

Plus important encore, l'une des spécificités des sciences sociales repose sur le large éventail de types de données collectées et de leurs formats. Cela peut varier d'un jeu de données quantitatives à un jeu de données qualitatives, étayées par des entretiens, des journaux de bord, des notes personnelles, des dossiers confidentiels, ainsi que des archives des organisations et institutions observées, pour n'en citer que quelques-uns. Les méthodes mixtes sont également de plus en plus utilisées. Cela soulève de nombreuses questions concernant les pratiques de normalisation des données, les coûts liés au travail de formatage et ainsi que ceux liés à la question de la propriété. Ce dernier point suppose, par exemple, que les institutions publiant elles-mêmes des archives, adoptent plus largement une politique de données ouvertes et partagent leurs données selon les principes FAIR et CARE. L'utilisation des licences *Creative Commons Attribution* pourrait représenter un moyen de soutenir cet accès plus large aux données ouvertes et faciliter ainsi le travail en sciences sociales.

Une autre particularité des sciences sociales concerne l'anonymat et la désidentification des données, termes sont souvent confondus avec celui de confidentialité (Borgerud & Borglund, 2020). Si le remplacement des marqueurs d'identité par un identifiant qui n'a de sens pour personne est une pratique traditionnelle dans la recherche, à l'ère du *big data*, cela ne peut garantir la protection de l'identité de la personne concernée et donc son anonymat. En parallèle, lorsque des données personnelles sensibles telles que des notes personnelles ou des carnets de bord ne peuvent pas être garanties par l'anonymat, il apparaît nécessaire de restreindre l'accès aux données pour garantir leur intégrité. Grâce à une protection par mot de passe et la signature d'accords éthiques, seules les personnes autorisées peuvent avoir accès et utiliser ou réutiliser lesdites données.

En ce qui concerne la confidentialité, de nombreuses règles et recommandations ont déjà été élaborées par l'OCDE (2016) par exemple. En effet, la question de la confidentialité et de la protection de la vie privée est complexe, avec des aspects à la fois juridiques et

éthiques qui doivent être abordés dans le PGD comme nous le détaillerons ci-après. Clarifier la manière dont la protection de la vie privée sera assurée lors de l'utilisation de données revêt une importance capitale. La notion de consentement est également au cœur de cette approche éthique. En ce qui concerne les données ouvertes, cela concerne non seulement la manière dont les individus sont informés de la façon dont les données seront collectées et consentent à leur utilisation, mais aussi à leur réutilisation potentielle par d'autres usagers. De plus, en sciences sociales, les données mais aussi les personnes interrogées doivent être protégées (Bonneville et al., 2021). Un risque non négligeable pour la recherche, dans ce cas, est de perdre l'accès au terrain, soit parce que les personnes interrogées sont moins disposées à participer à des projets dont les données de la recherche seront ouvertes, soit en raison des coûts de réalisation de ces projets.

5. Conditions opérationnelles pour l'ouverture des données de la recherche

Nous allons à présent explorer les conditions opérationnelles et les questions concrètes à se poser pour ouvrir les données de la recherche. Nous adopterons tout d'abord le point de vue des chercheuses et chercheurs puis celui des revues indépendantes.

5.1. Pour les équipes de recherche

Au-delà des questions liées au cycle de vie des données et aux flux de travail mentionnés dans la section 1, au moins cinq questions clés sont à se poser lors de l'ouverture des données : 1) Les données peuvent-elles être partagées et dans quelles conditions (éthiques) ? 2) Comment préparer les données pour le partage ? 3) Comment trouver et choisir un dépôt de données ? 4) Qu'est-ce qu'une déclaration d'accessibilité des données et qu'est-ce qui doit être détaillé dans un article ? 5) Comment lier les jeux de données à un article ? Ces questions sont généralement abordées dans les plans de gestion des données (PGD). D'une manière générale, ces questions doivent être abordées collectivement avec toutes celles et ceux qui ont été impliqués dans la collecte, la production et le traitement des données, c'est-à-dire l'intendance des données. Cela signifie que la question de la co-paternité des données doit également être abordée.

Premièrement, il faut garder à l'esprit que des données ouvertes de la recherche signifient 'aussi ouvertes que possible, et aussi restreintes que nécessaire'. Ainsi, la question des conditions du partage des données amène à réfléchir à la manière dont les données sont ou seront collectées, aux diverses limites et barrières à la mise à disposition des données (par exemple, si elles sont sensibles ou non). L'intégration des principes FAIR et CARE nécessite une réflexion supplémentaire concernant les relations avec la communauté, comme le synthétise la figure 4 (Carroll et al., 2021).

Utiliser les principes 'CARE' dans la collecte de données	Engager les principes 'CARE' dans la gestion des données	Mettre en oeuvre les principes 'CARE' dans une communauté de données	Utiliser les principes 'FAIR' et 'CARE' dans les répercussions sur les données
- Définir les métadonnées culturelles - Préciser la provenance dans les métadonnées	- Utiliser les modèles de gouvernance appropriés - Rendre les données 'FAIR'	- Respecter les éthiques autochtones pour l'accès aux données - Utiliser les outils pour garantir la transparence, l'intégrité et la provenance des données	- Équité - Responsabilité - Transparence

Figure 4. Mise en oeuvre des principes CARE (Carroll et al., 2021, p. 4)

Deuxièmement, la préparation des données en vue de leur partage va varier en fonction de leur type et des règles et attentes spécifiques à chaque dépôt de données et discipline. Le COAR (Controlled Vocabularies for Repositories) propose une description normalisée des types de données. Comme déjà souligné, les sciences sociales utilisent une grande variété de données, des images aux corpus de texte, en passant par les données des médias sociaux, les enquêtes et autres bases de données quantitatives. Les données qualitatives qui peuvent être ouvertement mises à disposition sur des dépôts peuvent inclure des entretiens structurés, semi-structurés ou non structurés, des groupes de discussion, des histoires orales sous forme d'enregistrements audio ou vidéo, des transcriptions, des résumés, des notes de terrain (y compris de l'observation participante ou de l'ethnographie), des cartes, images satellites ou données géographiques, des documents officiels, des dossiers internes, des rapports, des compte-rendus de réunions, des transcriptions de discours publics, des chroniques, des autobiographies, des mémoires, des carnets de voyage, des journaux intimes, des brochures, des affiches, des prospectus, des documents personnels tels que des lettres, des émissions de radio sous forme de fichiers audio, des programmes télévisés, etc. En général, cinq champs de métadonnées inspirés de Dublin-Core, qui propose des normes spécifiques, sont obligatoires : titre, auteur·rice, date, type et licence. D'autres champs Dublin-Core peuvent être ajoutés en fonction des types de données : contributeur, couverture, créateur, description, langue, éditeur, relation, droits, source, sujet, etc.

Troisièmement, le choix d'un dépôt de données dépend de la discipline et du type de données, mais surtout de plusieurs critères qui garantissent à minima la fiabilité et les principes FAIR (Science Europe, 2021). Les dépôts dignes de confiance doivent fournir 1) des identifiants uniques et permanents afin de permettre la recherche, l'identification, la citation et la récupération des données ; 2) des méta-données organisées, faciles à trouver, normalisées et récupérables ; 3) des licences d'accès et d'utilisation des données avec des conditions d'accès bien spécifiées, garantissant l'authenticité et l'intégrité des

données, dans le respect de la confidentialité et des droits des personnes concernées et des créateurs de données ; 4) des mécanismes de préservation qui garantissent la stabilité, tout en étant transparents sur leur mission, leur modèle et leur durabilité, le tout en conformité avec les principes FAIR de partage et d'accès aux données. Certains référentiels, comme NAKALA ou le Qualitative Data Repository (QDR), permettent de stocker la plupart, voire tous les formats de données qualitatives.

Quatrièmement, l'inclusion d'une déclaration d'accessibilité des données est de plus en plus exigée par les revues, même lorsqu'aucune donnée n'a été utilisée. Cette déclaration doit être ajoutée aux manuscrits avant leur soumission. Pour les articles où toutes les données associées sont décrites et présentées dans le manuscrit, la déclaration peut indiquer que 'toutes les données sous-jacentes aux résultats sont disponibles dans le cadre de l'article et qu'aucune source supplémentaire n'est requise.' Lorsque les données sont stockées dans un dépôt, les déclarations doivent généralement mentionner le titre et le DOI (pour les dépôts comme Zenodo ou QDR) ou le titre, le DOI et le code pour l'outil de réanalyse interactif dans le cas de dépôts de données et de code mixtes (Code Ocean), par exemple. Si les données ne peuvent être partagées, les métadonnées peuvent toujours être déposées et mentionnées dans la déclaration d'accessibilité des données. Les exceptions incluent la confidentialité, les secrets commerciaux, les règles de sécurité, les intérêts concurrentiels de l'Union Européenne les droits de propriété intellectuelle, l'impossibilité de 'désidentifier' les données relatives aux personnes, auquel cas la déclaration d'accessibilité des données précise cette raison.

Cinquièmement, une fois qu'un manuscrit est accepté et publié, il est recommandé de mettre à jour les informations du PGD et du dépôt d'archives avec un lien vers l'article final, et inversement. Cette connexion réciproque est importante pour assurer la visibilité et l'accessibilité des données.

Tous ces éléments sont abordés dans le plan de gestion des données (PGD). De nombreuses ressources disponibles en ligne fournissent des conseils sur la conception des PGD. Dans le cadre de cet article, nous présentons brièvement les sections clés d'un PGD en relation avec les questions précédentes et le guide étendu développé par Science Europe (2021) (tableau 4).

Un PGD débute généralement par des informations administratives générales telles que le nom du demandeur, le numéro du projet, le programme de financement et la version du PGD.

La section suivante nécessite une attention particulière car elle concerne la description, la collecte et la réutilisation des données existantes. Cette section traite de la manière dont les données seront collectées ou produites, du type de données, des formats et des volumes, ainsi que des mesures de contrôle de la qualité des données qui seront utilisées.

Ensuite, le PGD détaille la documentation et la qualité des données, c'est-à-dire les métadonnées fournies, les normes utilisées, la méthodologie décrivant la collecte des données et les mesures visant à assurer le contrôle de la qualité.

Les questions relatives au stockage et les solutions de sauvegarde pendant le processus de recherche lui-même, y compris la sécurité des données et la protection des données sensibles, sont abordées.

Le PGD précise ensuite les exigences légales et éthiques et les codes de conduite, indiquant comment les données personnelles sont traitées, en conformité avec le GDPR et d'autres lois de protection. Les questions relatives aux droits de propriété intellectuelle et de la propriété sont évoqués ainsi que les questions éthiques et les codes de conduite à adopter.

Ensuite, le PGD détaille les mécanismes de partage des données et de conservation à long terme, les éventuelles restrictions ou raisons d'un embargo, les outils d'utilisation et de réutilisation, ainsi que l'attribution des identifiants uniques et permanents (DOI).

La dernière section traite généralement des responsabilités et des ressources en matière de gestion des données, c'est-à-dire qui est responsable de la gestion des données et quelles ressources y seront consacrées.

Tableau 4. Plan de Gestion des Données (PGD) et questions soulevées pour les équipes de recherche (Science Europe, 2021)

Section	Questions soulevées
<i>Information générale</i>	Nom du demandeur, numéro du projet, programme et numéro de financement, version du DMP
<i>Description des données et collection et réutilisation des données existantes</i>	Comment les données seront-elles collectées ou produites ? Quel type de données ? Quel format ? Quel volume ? Quelles mesures de contrôle de la qualité des données utilisées ?
<i>Documentation et qualité des données</i>	Métadonnées, normes, méthodologie décrivant la collecte des données, mesures de contrôle de la qualité
<i>Stockage et solutions de sauvegarde (pendant la recherche)</i>	Stockage, sauvegarde, sécurité des données et protection des données sensibles
<i>Exigences légales et éthiques et codes de conduite</i>	Comment les données personnelles sont traitées, en conformité avec le GDPR et d'autres lois de protection ? Comment les droits de propriété intellectuelle et la propriété sont gérés ? Lesquels ? Quelles sont les questions éthiques et les codes de conduite à se poser ? Comment les gérer ?

<i>Règles de partage des données et de préservation à long terme</i>	Restrictions possibles ou raisons de l’embargo, utilisation et réutilisation, attribution d’un identifiant unique et permanent
<i>Responsabilités et ressources en matière de gestion des données</i>	Qui est responsable de la gestion des données ? Quelles ressources y seront consacrées ?

Ces différentes sections et les questions afférentes des PGD peuvent sembler *a priori* laborieuses. Cependant, l’anticipation, la réflexivité, le temps, la formation et le soutien des institutions contribuent à faire du partage ouvert des données une tâche beaucoup plus réalisable et gratifiante.

5.2. Pour les revues

Nous allons maintenant aborder les conditions et limites à l’ouverture des données du point de vue des revues en sciences sociales. Une première question se pose. Pourquoi se concentrer sur les revues, et non, par exemple, sur les sociétés académiques, les laboratoires de recherche ou encore les universités qui ont un rôle important dans ce processus ? Nous considérons que les revues sont à l’avant-garde de ces questions. En effet, à l’heure d’élaborer la politique d’ouverture des données pour la revue *M@n@gement*, nous nous sommes posées nous-mêmes plusieurs questions clés : Quel type de directives une revue peut-elle élaborer ? Les jeux de données ouvertes doivent-ils être examinés ? Dans l’affirmative, par qui ? Qui devrait supporter le coût de l’examen des données ? Comment traiter les articles sur les données (*data research paper*) ? À ce jour, certaines de ces questions restent encore ouvertes.

Notre questionnement est motivé par une problématique majeure : quels sont les rôles et quelles sont les responsabilités d’une revue comme la nôtre dans le processus d’ouverture des données de la recherche ? A *M@n@gement*, nous reconnaissons l’importance d’explorer les moyens d’améliorer la science ouverte, et cela inclut notamment de faciliter l’ouverture des données. Cependant, en tant que journal en libre accès, indépendant et interne, l’une de nos premières préoccupations reste le coût humain et financier de la mise en œuvre et du portage d’une politique d’ouverture des données. Si nous pouvons définir des lignes directrices et rendre incitative l’ouverture des données, nous ne pouvons toutefois pas développer notre propre infrastructure par exemple. Ce rôle se situe en dehors de notre champ de responsabilité.

En observant les politiques déjà mises en place par d’autres revues, nous constatons que peu d’entre elles ont déjà adopté une politique d’ouverture des données. Certaines ont des politiques de données plus ou moins obligatoires. En économie par exemple, l’*American Economic Review* fait figure de référence et sa politique en matière de données a été adoptée par d’autres revues, comme *Journal of Political Economy*. Ces revues économiques rendent le partage des données obligatoire lors de la soumission et utilisent *Harvard Dataverse*, un dépôt de données de la recherche où les revues peuvent créer leur

propre espace. Dans le cas de *M@n@gement*, nous pensons que l'utilisation de dépôts publics comme NAKALA permettrait une plus grande biodiversité dans les sources de données et éviterait que les données soient stockées aux États-Unis (ce qui pose évidemment des problèmes spécifiques de sécurité et de propriété). À moyen et long terme, la durabilité des dépôts et la concurrence entre les dépôts pourraient devenir un problème. Le choix d'un dépôt renommé et institutionnalisé n'en est que plus important.

Le tableau 5 synthétise les problèmes et les questions à aborder lors de la définition d'une politique de données ouvertes pour les revues (Feret et al., 2021 ; Hrynaszkiewicz et al., 2020). L'objectif d'une telle politique est d'identifier quelles sont les données concernées, à quelle étape du processus, ou encore quels protocoles de normes et de méta-données doivent être utilisés.

Tableau 5. Feuille de route pour définir la politique de données d'une revue

Thème général	Questions soulevées
<i>Définition des données de recherche et des exceptions</i>	Quelles sont les conditions d'accès et d'embargo ? Quelles sont les données concernées par la politique ? À quel stade du processus ? Soumission ? Acceptation ?
<i>Normes et formats des données et des métadonnées</i>	Quelles sont les normes à respecter ? Quels sont les protocoles à suivre pour les métadonnées ?
<i>Accès aux données, hébergement et publication</i>	Quels sont les protocoles et les directives pour le dépôt des données et le choix d'un dépôt ?
<i>Procédures quant à la disponibilité des données</i>	Quelles sont les procédures de mise à disposition des données (délai, étapes, etc.) ? Par qui sont examinées les données ? des pairs ? D'autres personnes ? Existe-t-il une déclaration de disponibilité des données ?
<i>Accessibilité des données</i>	Comment les données seront-elles reliées à la publication ? Des identifiants permanents sont-ils utilisés ?
<i>Non-conformité</i>	Que se passe-t-il lorsque les auteur·rices ne respectent pas la politique de la revue en matière de données ?
<i>Soutien aux auteur·rices, lecteur·rices et éditeur·rices</i>	Existe-t-il un soutien de la part de la revue ? D'une association académique ?

En ce qui concerne la disponibilité des plateformes et des dépôts, il existe, au niveau des revues, certaines tensions entre les revues en libre accès et les systèmes de données ouvertes de la recherche. Les revues disposant de ressources importantes de la part de

maisons d'éditions commerciales pourraient avoir plus de facilité à s'adapter à ces changements, ce processus étant déjà en cours dans certaines revues. En revanche, les revues en libre accès de type 'diamant', soutenues par des organisations indépendantes ou à but non lucratif, pourraient avoir plus de difficultés, s'il est attendu qu'elles traitent les données ouvertes de la recherche par elles-mêmes. Pour les revues, en particulier les revues indépendantes, il est difficilement concevable de développer l'infrastructure nécessaire au traitement ou même à l'examen des données ouvertes. L'intendance des données est une profession spécifique, différente de l'édition scientifique. Cependant, par rapport à ces éléments, il est à noter que l'un des arguments en faveur de la science ouverte est précisément de s'écarter du modèle de commodification, de marchandisation et de profitabilité de la publication. Cela soulève alors plusieurs questions : Est-il légitime de permettre qu'une entreprise fasse des profits à partir de données ouvertes produites par la communauté de recherche ? Si oui, dans quelles conditions ? Devrions-nous envisager que les utilisations commerciales profitent avant tout aux organisations publiques, aux universités et aux citoyen·nes ?

6. Politique en termes d'ouverture de données de la recherche de M@n@gement

M@n@gement a décidé d'encourager les auteur·rices à rendre toutes les données associées à leur recherche en libre accès, dans la mesure du possible et avec toutes les conditions de restriction nécessaires, conformément aux principes FAIR et CARE. Notre politique n'est qu'incitative, mais nous espérons sensibiliser à ces questions et en particulier, à la nécessité d'intégrer les principes FAIR et CARE.

Pour développer la politique d'ouverture des données de M@n@gement, nous nous sommes appuyés sur le guide de référence développé par Science Europe (2021). En parallèle, nous nous sommes également inspirés des politiques existantes des revues (par exemple, *Designs for learnings*) et des lignes directrices développées par le groupe d'intérêt sur la standardisation et la mise en œuvre des politiques de données de la Research Data Alliance (RDA) (Fabre & Gouzi, 2020 ; Hrynaszkiewicz et al., 2020) et le Collège des données de la recherche du comité français pour la science ouverte (Ferret et al., 2021).

Nous considérons cette politique de données comme une première étape qui sera sujette à évolutions, à la fois en réponse aux évolutions institutionnelles, de l'écosystème de la science ouverte, et de notre propre domaine de recherche. Notre objectif est avant tout d'ouvrir le débat avec notre communauté sur les besoins, les défis et les perspectives des données ouvertes de la recherche. Comme déjà souligné, notre philosophie en matière d'ouverture de données rejoint celle du comité français pour la science ouverte : 'aussi ouvert que possible, aussi restreint que nécessaire'.

6.1. Définition des données ouvertes de la recherche

Cette politique s'applique aux données de la recherche nécessaires pour vérifier les résultats présentés dans les publications de la revue. Par données de la recherche, nous entendons les données produites par les auteur·rices ainsi que les données provenant d'autres sources qui sont analysées par les auteur·rices dans leur étude. Ces données

peuvent être présentées sous différentes formes : images, vidéos, textes, codes, tableaux statistiques, etc. Ces données doivent être produites et partagées dans le respect des principes FAIR et CARE (<https://www.rd-alliance.org/implementing-care-principles-care-full-process>). Les données de la recherche qui ne sont pas nécessaires aux résultats démontrés dans la publication ne sont pas couvertes par cette politique.

Cette politique est limitée par les exceptions légitimes réglementées par la loi, comme tout ce qui a trait au secret professionnel, industriel et commercial, aux données personnelles ou encore aux contenus protégés par les droits d'auteur·rice. Par ailleurs, *M@n@gement* n'examine pas et ne publie pas les articles de données (*data research paper*). Cependant, nous encourageons les auteur·rices à ajouter tous les documents liés à leurs jeux de données, comme ces articles de données, afin de clarifier les procédures, les conditions et les limites à la réutilisation des données.

6.2. Normes et formats des données et métadonnées

M@n@gement encourage les auteur·rices à utiliser des formats ouverts et standards. Les métadonnées descriptives doivent être structurées à l'aide de normes reconnues, à minima celles du Dublin Core. Les normes utilisées par les équipes de recherche peuvent être soit disciplinaires, soit plus génériques (https://en.wikipedia.org/wiki/Dublin_Core).

L'utilisation de vocabulaires 'contrôlés' (ou de référence), disciplinaires ou plus génériques, pour détailler ces métadonnées, est recommandée (par exemple, pour référencer un auteur·rice, voir : <https://orcid.org> ; pour référencer un lieu, voir : <https://www.geonames.org>, ou pour les concepts de données, voir : Controlled Vocabularies for Repositories). La conformité des formats de fichiers de données aux recommandations du CINES pour la conservation à long terme est également recommandée (<https://facile.cines.fr>).

6.3. Accès et hébergement des données

Les données ayant contribué à la rédaction de la publication doivent être déposées dans un dépôt de données qui garantira un stockage et un accès sécurisé aux données, notamment par l'attribution d'un identifiant permanent, tel que le DOI. Nous conseillons aux auteur·rices d'éviter l'utilisation de dépôts privés dont la feuille de route n'est pas transparente en termes de modèle économique, de gouvernance ou de pérennité (voir par exemple, Figshare).

M@n@gement recommande l'utilisation de dépôts dignes de confiance et conformes aux principes FAIR, qu'ils soient généralistes (cf. Zenodo), institutionnels (cf. Data INRAE ou plateformes universitaires) ou disciplinaires (cf. QDR ou NAKALA pour les sciences sociales). Dans tous les cas, les auteur·rices doivent vérifier que le dépôt choisi répond à des critères de qualité minimum (voir <https://zenodo.org/record/4915862>).

Dans cette perspective, nous encourageons les auteur·rices à contacter les services d'assistance de leur institution concernant les bonnes pratiques de gestion et de partage des données, ainsi que la conception et le développement de plans de gestion des données.

6.4. Procédure de mise à disposition des données

La procédure de mise à disposition des données se déroule en trois étapes.

Phase de soumission. Les auteur·rices sont encouragés à transmettre les données lorsqu'ils soumettent leurs contributions, dans la mesure du possible. Cela peut se faire, soit au sein de l'article, soit en annexe (option *dataset* sur notre plateforme) mais de préférence avec un accès restreint ou contrôlé via un dépôt. Il reste nécessaire que les données soient anonymisées et suivent les recommandations générales en la matière. Cela inclut l'anonymisation du projet de recherche lui-même, de l'équipe de recherche et des participant·es à la recherche.

Phase de relecture par les pairs. Si les rédacteur·rices et les relecteur·rices le jugent nécessaire, les auteur·rices doivent mettre à la disposition des relecteurs les données qui étayent les résultats présentés dans leur contribution. Le refus de fournir les données demandées entraînera le rejet de l'article.

Phase d'acceptation. Lorsque cela est possible et dans les limites nécessaires, les données doivent être mises à disposition sans embargo, ou avec la période d'embargo la plus courte possible, au moment de l'acceptation de l'article. Les conditions de partage doivent permettre la réutilisation, avec un lien explicite entre les données et la publication qu'elles soutiennent, et ce, dans le respect des directives sur les données personnelles et la protection des personnes interrogées.

Si *M@n@gement* encourage les auteur·rices à partager leurs données sous des licences ouvertes permettant leur libre réutilisation, les auteur·rices doivent veiller à utiliser les licences recommandées par le dépôt où les jeux de données ont été déposés.

En publiant dans *M@n@gement*, et dans la mesure du possible, les auteur·rices s'engagent à rendre les données et/ou métadonnées accessibles au public pendant au moins 5 ans après la publication de leur contribution, soit par le biais d'une plateforme, soit par une mise à disposition individuelle si les données ne peuvent pas être partagées librement.

Les alternatives au partage en libre accès de données personnelles ou sensibles sont :

- L'anonymisation ou la pseudonymisation des données avant leur publication en libre accès ;
- Les données disponibles sur demande à des fins de recherche uniquement ;
- La disponibilité des métadonnées des données uniquement, ce qui est l'objectif minimal pour toutes les publications de *M@n@gement*.

5.5 Déclaration d'accessibilité aux données

Les auteur·rices sont tenus de citer les jeux de données sous-jacents à leur publication dans une déclaration spécifique d'accessibilité aux données. Cette section doit décrire les données disponibles, comment y accéder et fournir un lien permanent vers les données.

La déclaration peut inclure l'une, ou une combinaison, des options suivantes :

- Les ensembles de données générés et/ou analysés au cours de l'étude actuelle sont disponibles dans le référentiel [NOM] ; [DOI].
- Les ensembles de données générés pendant et/ou analysés pendant l'étude actuelle ne sont pas disponibles en accès libre et les raisons doivent en être précisées, mais restent disponibles auprès de l'auteur·rice sur demande raisonnable.
- Le partage des données ne s'applique pas à cet article car aucun jeu de données n'a été généré ou analysé au cours de l'étude actuelle.
- Les jeux de données sur lesquels repose l'étude actuelle n'ont pas été générés par les auteur·rices. Ils sont disponibles en ligne : Créateur (année de publication). Titre. Version. [Nom du dépôt]. [DOI].

7. Conclusion

Les tendances internationales récentes cherchent à 'ouvrir' la science sur l'ensemble de son cycle de vie et de son écosystème, c'est-à-dire des données de la recherche à la publication des résultats. Cependant, l'ouverture des données de la recherche soulève des questions et des préoccupations spécifiques en sciences sociales (Bonneville et al., 2021 ; Gauquelin et al., 2017). Dans cet article, après avoir présenté l'écosystème existant d'acteurs, d'infrastructures, de standards et de principes directeurs (FAIR et CARE) ainsi que les raisons préconisant l'ouverture des données, nous avons souligné que l'ouverture des données de la recherche se heurte à plusieurs obstacles en sciences sociales : 1) une mauvaise perception des motivations (l'accent est généralement sur l'exercice d'un contrôle sur la liberté académique et les équipes de recherche plutôt que sur la patrimonialisation, la mise en commun des données et des synergies potentielles, l'instauration de la confiance, et un engagement plus large), 2) un système de concurrence et un processus dominant de starification de la recherche qui encourage les individus à se concentrer sur les publications et à conserver la propriété des données, 3) un manque de ressources qui pourrait aggraver davantage les inégalités entre les sexes, les communautés, les revues, les institutions et les pays, et 4) des risques potentiels de l'ouverture des données et les contraintes spécifiques des sciences sociales (coûts d'utilisation d'une variété de formats de données ou perte d'accès aux terrains, entre autres). Face à ce constat, nous avons évoqué les conditions opérationnelles et les questions à se poser lors de l'ouverture des données de la recherche, du point de vue du chercheur ou de la chercheuse ainsi que du point de vue de la revue. Enfin, nous avons présenté la nouvelle politique d'ouverture des données de *M@n@gement*.

Avec cet article, nous souhaitons ouvrir le débat au sein de notre communauté sur les besoins et les défis liés à l'ouverture des données de la recherche en sciences sociales en général, et en sciences de gestion et théories des organisations en particulier.

Nous pensons qu'une tension subsiste entre la science ouverte et ses utilisations commerciales. Même si nous considérons la science comme un bien public mondial (Berkowitz & Delacour, 2020), la science ouverte permet la réutilisation des données par

les organisations, en particulier à des fins de performance financière, ce qui soulève de nombreuses questions. Pour répondre à ce paradoxe, *M@n@gement* utilise une licence *Creative Commons CC BY-NC 4.0*. Cela signifie la liberté de partager et d'adapter sous certaines conditions (par exemple en donnant le crédit approprié) mais aussi à des fins non commerciales. En cela, nous nous distinguons des recommandations générales de la science ouverte qui préconisent l'absence de toute restriction, y compris à des fins commerciales.

Ensuite, nous pensons que dans un contexte d'urgence socio-écologique et d'inégalités croissantes, l'ouverture des données ('aussi ouvertes que possible, et aussi restreintes que nécessaire') revêt un enjeu important. Cependant, nous sommes conscientes que développer des projets de recherche nécessite des temps longs et de nombreux efforts, auxquels viennent s'ajouter la conception et la mise en œuvre des plans de gestion des données.

À cet égard, nous soutenons d'autant plus la nécessité d'une *sustainable academia* et d'une 'science lente', comme nous l'avons préconisé il y a quelques années (Berkowitz & Delacour, 2020), mais aussi la nécessité de réfléchir aux risques de 'subordination de la recherche' (Vijay, 2021a, p. 56). Cela signifie qu'il faut explorer et offrir la possibilité de penser d'une manière différente de l'ordre ou des ordres dominants et ce, d'où qu'ils viennent (Vijay, 2021b). Réfléchir aux principes CARE pour la gouvernance des données autochtones nous guide dans cette réflexion et nous aide à reconnaître que nous produisons et reproduisons des relations de pouvoir et des hiérarchies de connaissances dans la production, l'utilisation et la gestion des données de la recherche. En ce sens, nous croyons que l'ouverture n'est pas simplement une question de transparence, de reproductibilité ou de patrimonialisation, mais qu'il s'agit avant tout d'assurer un accès inclusif et équitable (dans le sens de juste) à la connaissance pour toutes et tous.

Remerciements

Nous remercions Anne Bonneville pour son aide précieuse sur notre politique de données, les participant-es du colloque 'L'action publique des données et les données de l'action publique' 2022 à Strasbourg pour leurs commentaires, ainsi que Lionel Maurel et Astrid Aschehoug pour leur soutien et leurs conseils.

Références

- Austin, C. C., Bloom, T., Dallmeier-Tiessen, S., Khodiyar, V. et al. (2015). Key components of data publishing: Using current best practices to develop a reference model for data publishing. *International Journal on Digital Libraries*, 18, 77–92. doi: [10.5281/zenodo.34542](https://doi.org/10.5281/zenodo.34542)
- Berkowitz, H. & Delacour, H. (2020). Sustainable academia: Open, engaged, and slow science. *M@n@gement*, 23(1), 1–3. doi: [10.37725/mgmt.v23.4474](https://doi.org/10.37725/mgmt.v23.4474)
- Bonneville, A., Tucci, I., Vion, A. & Giglio, L. (2021). *Données de la recherche: Pratiques et besoins dans un laboratoire pluridisciplinaire SHS* [PhD thesis]. Laboratoire d'Economie et Sociologie du Travail (LEST).
- Borgerud, C., & Borglund, E. (2020). Open research data, an archival challenge? *Archival Science*, 20(3), 279–302. doi: [10.1007/s10502-020-09330-3](https://doi.org/10.1007/s10502-020-09330-3)
- Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R. et al. (2019). Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. *Frontiers in Big Data*, 2, 43.
- Carroll, S. R., Herczog, E., Hudson, M., Russell, K. et al. (2021). Operationalizing the CARE and FAIR principles for indigenous data futures. *Scientific Data*, 8(1), 108. doi: [10.1038/s41597-021-00892-0](https://doi.org/10.1038/s41597-021-00892-0)
- Fabre, C. & Gouzi, F. C. (2020). *Proposition de modèle de politique pour les revues et éditeurs quant aux données de la recherche*. Retrieved from <https://hal-univ-tlse2.archives-ouvertes.fr/hal-03026731>
- Feret, R., Gouzi, F. C., Guignonis, S., Jouguet, H. et al. (2021). *Guidelines for journals that wish to establish a 'data policy' related to their publications* (10 p.) [Report, Comité pour la science ouverte]. doi: [10.52949/19](https://doi.org/10.52949/19)
- Gauquelin, C., Lutoff, C. & La Branche, S. (2017). Quelle place pour les données issues des sciences humaines et sociales dans le développement de services climatiques régionaux? *Vertigo*, 17(3). doi: [10.4000/vertigo.18891](https://doi.org/10.4000/vertigo.18891)
- Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R. et al. (2020). Developing a research data policy framework for all journals and publishers. *Data Science Journal*, 19(1), 5. doi: [10.5334/dsj-2020-005](https://doi.org/10.5334/dsj-2020-005)

Longley Arthur, P. & Hearn, L. (2021). Toward open research: A narrative review of the challenges and opportunities for open humanities. *Journal of Communication*, 71(5), 827–853. doi: [10.1093/joc/jqab028](https://doi.org/10.1093/joc/jqab028)

Marlet, O., Baudoin, B., Bernard, L., Bézard, L. et al. (2022). *Le livre blanc du consortium 'Mémoires des archéologues et des sites archéologiques': Guide des bonnes pratiques numériques en archéologie*. Consortium MASA <halshs-03561376v2>, 39.

Martone M. (Ed.). (2014). *Data Citation Synthesis Group: Joint Declaration of Data Citation Principles*. FORCE11. doi: [10.25490/a97f-egyk](https://doi.org/10.25490/a97f-egyk)

Millerand, F. & Bowker, G. C. (2008). Metadata, trajectoires et 'énaction'. In C. Rosental (Ed.), *La cognition au prisme des sciences sociales* (pp. 277–303). Editions des Archives Contemporaines.

OECD (2016). Research ethics and new forms of data for social and economic research. OECD Science, Technology and Industry Policy Papers n° 34. doi: [10.1787/5jln7vnpxs32-en](https://doi.org/10.1787/5jln7vnpxs32-en)

Osterloh, M. & Frey, B. S. (2015). Ranking games. *Evaluation Review*, 39(1), 102–129. doi: [10.1177/0193841X14524957](https://doi.org/10.1177/0193841X14524957)

Peng, G., Lacagnina, C., Ivánová, I., Downs, R. R. et al. (2021). *International community guidelines for sharing and reusing quality information of individual earth science datasets*. Open Science Framework Preprints. doi: [10.31219/osf.io/xsu4p](https://doi.org/10.31219/osf.io/xsu4p)

Roche, D. G., Kruuk, L. E. B., Lanfear, R. & Binning, S. A. (2015). Public data archiving in ecology and evolution: How well are we doing? *PLoS Biology*, 13(11), e1002295. doi: [10.1371/journal.pbio.1002295](https://doi.org/10.1371/journal.pbio.1002295)

Science Europe. (2021). *Practical guide to the international alignment of research data Management*. Retrieved from https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf

Stats NZ. (2020). *Ngā Tikanga Paihere: A framework guiding ethical and culturally appropriate data use*. Retrieved from [www.data.govt.nz; https://data.govt.nz/assets/data-ethics/Nga-Tikanga/Nga-Tikanga-Paihere-Guidelines-December-2020.pdf](https://data.govt.nz/assets/data-ethics/Nga-Tikanga/Nga-Tikanga-Paihere-Guidelines-December-2020.pdf)

The FAIR Data Principles – FORCE11. (2014, September 3). Retrieved from <https://force11.org/info/the-fair-data-principles/>

Vijay, D. (2021a). Vulnerability as praxis in studying social suffering. In E. Bell & S. Sengupta (Eds.), *Empowering methodologies in organisational and social research* (pp. 55–74.). Routledge. doi: [10.4324/9780429352492-4](https://doi.org/10.4324/9780429352492-4)

Vijay, D. (2021b). Settled knowledge practices, truncated imaginations. *Organization*, 0(0), onlinefirst. doi: [10.1177/13505084211057261](https://doi.org/10.1177/13505084211057261)

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G. et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

Willinsky, J. (2005). *The access principle: The case for open access to research and scholarship*. MIT Press.

Woolston, C. (2021). Impact factor abandoned by Dutch university in hiring and promotion decisions. *Nature*, 595(7867), 462–462. doi: [10.1038/d41586-021-01759-5](https://doi.org/10.1038/d41586-021-01759-5)

Xu, F., Wu, L. & Evans, J. (2022). Flat teams drive scientific innovation. *PNAS*, 119(23), e2200927119. doi: [10.1073/pnas.2200927119](https://doi.org/10.1073/pnas.2200927119)