

Rendre les données qualitatives réutilisables

Un petit guide pour les chercheurs et les gestionnaires de données

Travailler avec des données qualitatives

Auteurs

[Maaïke Verburg](#) (DANS)

[Ricarda Braukmann](#) (DANS)

[Widia Mahabier](#) (DANS)

Ce guide a été réalisé dans le cadre du [projet CaRe & DaRe](#), financé par le fonds NWO Open Science sous le numéro de dossier 203.001.153 et dirigé par [le prof. dr. ir. JJ Bérends](#) (VU)

Date de publication : 25-04-2023

Identifiant persistant : <https://doi.org/10.5281/zenodo.7777519>

Licence : CC-BY 4.0

Table des matières

À propos de ce guide	3
Définition des données qualitatives	3
L'importance de rendre les données qualitatives réutilisables	3
Les défis pour rendre les données qualitatives réutilisables	4
Objectifs de ce guide	5
Planifier votre projet	6
Préparer le consentement éclairé	7
Organiser et documenter les données qualitatives	8
Traitement de vos données	9
Désidentification des données personnelles	dix
Archivage et publication de données qualitatives	12
Documentation et métadonnées	12
Formats de fichiers	13
Arbre de décision pour la réutilisation des données	13
Archiver en libre accès	15
Archivage avec accès restreint	15
Environnement sécurisé	16
La solution CaRe & DaRe : Réanalyse décentralisée	16
Publier uniquement les métadonnées	17
Dernières pensées	19
Ressources additionnelles	19
Les références	20

À propos de ce guide

Ce document vise à donner un aperçu des défis associés à la réutilisation des données qualitatives ainsi qu'à fournir des conseils sur la manière dont la réutilisabilité peut être améliorée et abordée à toutes les étapes du cycle de vie des données de recherche.

Le guide comprend un [arbre de décision](#) que les chercheurs et les gestionnaires de données peuvent utiliser pour évaluer les options permettant de rendre les données qualitatives réutilisables les plus adaptées à leurs projets.

Ce guide a été réalisé par DANS ([Archivage de Données et Services en Réseau](#)), le centre national néerlandais d'expertise et de référentiel de données de recherche, dans le cadre de l' [étude de cas « Recherche d'études et réutilisation des données »](#) (CaRe & DaRe) qui vise à accroître et à mieux faciliter la réutilisation des données qualitatives en science.

Bien qu'il soit probablement applicable à un public plus large, le principal public cible de ce document est la communauté de recherche néerlandaise, en particulier dans le domaine des sciences sociales et économiques, travaillant avec des données qualitatives. Ce guide s'inspire en partie d'une petite enquête menée auprès d'un groupe de gestionnaires de données néerlandais (Verburg et al., 2023).

Définition des données qualitatives

Les données qualitatives font référence à des informations qui ne sont pas recueillies sous forme numérique, mais qui décrivent plutôt des qualités ou des caractéristiques. Les données sont exprimées en langage naturel et souvent sous forme textuelle ou visuelle. Les méthodes de collecte comprennent par exemple la photographie, les enregistrements audio, la vidéo, les entretiens, les journaux intimes ou d'autres observations non structurées (CESSDA, 2022).

Les données qualitatives sont donc assez hétérogènes et peuvent prendre diverses formes. Les types de données qualitatives les plus souvent rencontrés par les répondants à l'enquête néerlandaise sur l'administration des données étaient les entretiens, les enregistrements vidéo et audio, les transcriptions, les notes de terrain, les groupes de discussion et les observations (Verburg et al., 2023).

Les données qualitatives contiennent souvent des informations personnelles et/ou sensibles. Les informations personnelles sont définies comme toute information relative à une personne physique identifiée ou identifiable. Les informations sensibles peuvent être soit des informations personnelles sensibles (par exemple, l'orientation sexuelle, les croyances religieuses, les données de santé), mais d'autres informations telles que par exemple les données de stratégie commerciale ou les données financières peuvent également être sensibles.

L'importance de rendre les données qualitatives réutilisables

Les mouvements de données Open Science et FAIR (Findable, Accessible, Interoperable, and Realistic, Wilkinson et al., 2016) exigent que la recherche soit bien documentée et disponible pour être réutilisée par d'autres. Il est généralement proposé que les données soient « aussi ouvertes que possible, aussi fermées que nécessaire », rendant possible leur réutilisation tout en protégeant les informations personnelles et les droits de propriété intellectuelle si nécessaire (Landi et al., 2020). Conformément à ces mouvements, de nombreux bailleurs de fonds, dont le NWO néerlandais, exigent que les données soient ouvertes et demandent aux chercheurs de décrire leurs plans de stockage et d'archivage des données dans le cadre d'un plan de gestion des données (DMP, voir [la page d'informations du NWO](#)).

Rendre les données disponibles pour la réutilisation permet de poser de nouvelles questions de recherche et de réduire le coût de la collecte des données. Il a également été démontré que la mise à disposition des données améliore la visibilité des recherches existantes (par exemple Colavizza et al., 2020 ; Piwowar et al., 2007) et permet à la recherche d'être reproductible, augmentant ainsi la transparence du processus scientifique (Munafò et al., 2017).

La collecte de données qualitatives, en particulier, demande souvent beaucoup de temps et de ressources, ce qui rend sans doute encore plus important la possibilité de réutiliser les données de ce type déjà disponibles. Cela peut réduire le fardeau des participants aux études qualitatives qui pourraient être contactés par plusieurs chercheurs pour des objectifs d'étude similaires. De plus, les données qualitatives sont de nature très riche, ce qui permet de répondre à une multitude de questions de recherche potentielles lorsque ces données sont partagées pour être réutilisées.

Les défis pour rendre les données qualitatives réutilisables

Le partage de données qualitatives est souvent plus complexe, voire impossible, à faciliter par rapport aux données quantitatives. Une raison importante est que l'anonymisation des données peut souvent entraîner des pertes d'informations, rendant les données inutiles pour leur réutilisation. Le partage de données non anonymisées peut être difficile, voire impossible, en raison de contraintes juridiques (par exemple RGPD, droits d'auteur) et lorsque les informations sont particulièrement sensibles (par exemple données personnelles sensibles, informations commerciales sensibles).

Un autre aspect qui rend le partage de données qualitatives difficile est qu'il peut être difficile de capturer toutes les informations contextuelles essentielles à l'interprétation des données qualitatives dans la documentation ou les métadonnées. Il peut sembler difficile de garantir que les données sont interprétées de la bonne manière compte tenu de leur variété et de leur richesse.

Si le partage de données qualitatives est donc plus complexe, les orientations sur la gestion des données de recherche (RDM) autour des données qualitatives font souvent défaut (Verburg et al., 2023) et les données qualitatives sont facilement négligées dans le développement des données Open Science et FAIR.

Il est important de noter que la mise à disposition de données qualitatives pour réutilisation ne sera pas toujours possible en raison de ces contraintes. Pourtant, les chercheurs qualitatifs doivent évaluer soigneusement leurs projets et processus afin de maximiser la réutilisation de leurs données. Bien que certaines données qualitatives ne puissent pas du tout être mises à la disposition d'autres chercheurs, certaines peuvent être mises à disposition ouvertement après désidentification, et d'autres peuvent être mises à disposition avec des restrictions ou via des processus d'analyse alternatifs tels que ceux proposés dans le projet CaRe & DaRe.

En outre, la documentation et les informations sur l'étude peuvent probablement être mises à disposition dans tous les cas, fournissant ainsi aux chercheurs des informations sur les données existantes même s'ils ne peuvent pas accéder aux données elles-mêmes.

Les différentes possibilités et les éléments à prendre en compte lors de l'évaluation des possibilités de rendre les données qualitatives réutilisables seront décrits plus en détail dans les sections suivantes de ce guide.

Objectifs de ce guide

L'objectif de ce guide est de donner un aperçu des étapes que les chercheurs et les gestionnaires de données peuvent suivre pour améliorer la réutilisabilité des données qualitatives.

En tenant compte des défis liés à la mise à disposition de données qualitatives pour réutilisation, comme mentionné ci-dessus, nous passons en revue chaque étape du processus de recherche et décrivons comment la réutilisabilité peut être envisagée dès le départ. De cette manière, un projet peut être conçu pour maximiser le potentiel de réutilisation.

Ce guide fournit des conseils sur ce qu'il faut considérer en matière de réutilisabilité. Il répertorie également les outils ouverts existants pouvant être utilisés pour traiter des données qualitatives. Nous examinons des aspects tels que la collecte du consentement éclairé, les options de désidentification des données et fournissons un aperçu des différentes options pour archiver et publier (des parties des) données qualitatives. Enfin, nous fournissons des exemples d'ensembles de données disponibles pour la réutilisation et des pratiques de réutilisation alternatives particulièrement conçues pour les données sensibles.

Nous espérons que ce guide pourra devenir un point de départ pour de nouvelles discussions et de meilleures orientations pour rendre les données qualitatives réutilisables.

Planifier votre projet

La première étape du cycle de vie des données de recherche est la planification du projet. Au cours de cette phase, il est souvent demandé aux chercheurs de créer un plan de gestion des données (DMP) contenant toutes les informations sur l'étude qui doivent être collectées et décrites. C'est le moment idéal pour réfléchir déjà au type de données que vous allez collecter. Le [guide expert en gestion de données CESSDA](#) vous guide plus en détail à travers cette phase de recherche et [fournit également une liste de contrôle](#) avec des questions à considérer. Une checklist DMP [adaptée aux données qualitatives](#) est également [disponible](#) dans le référentiel de données qualitatives.

Les questions spécifiques que vous souhaitez aborder à ce stade et qui peuvent vous aider à évaluer si et comment vos données peuvent être rendues réutilisables sont :

Vue commune sur le partage de données entre collaborateurs :

- Quelles autres parties sont impliquées dans mon projet ? • Que pensent-ils de la possibilité de rendre les données réutilisables ? • Soutiendront-ils les efforts visant à améliorer la réutilisabilité des données ?
 - À qui appartiennent les données ? Qui devrais-je consulter avant de rendre les données disponibles pour réutilisation ? • Comment puis-je garantir un point de contact permanent pour le jeu de données, notamment si je restreint l'accès ?

Collecte de données personnelles : •

Mes données contiennent-elles des informations personnelles ?

- La collecte de ces données est-elle conforme aux exigences éthiques et juridiques (par exemple RGPD) ?
- Puis-je anonymiser les données personnelles ? Dans quelle mesure cela affecterait-il l'utilité des données ?
- La désidentification est-elle nécessaire pour mon étude ? •

Ai-je la permission/le consentement pour partager des données personnelles données sans désidentification ?

- Puis-je minimiser la quantité de données personnelles que je vais

collecter d'une manière ou d'une autre ?

- Puis-je séparer les informations personnelles des informations qui peuvent être partagées ouvertement ? • Que pensent les sujets de mes recherches de la possibilité de rendre les données réutilisables ?

• Puis-je inclure la réutilisation des données dans mes formulaires de consentement éclairé ?

Collecte de données sensibles : •

Les données que je collecte sont-elles sensibles d'une manière ou d'une autre ? • Puis-je séparer les informations sensibles des informations qui pourraient être partagées ouvertement ? • Ce niveau de sensibilité de mes données interdit-il tous les types de partage de données, ou pourrais-je envisager le partage dans des circonstances spécifiques (par exemple, accès restreint, environnement sécurisé) ?

- Quelles mesures supplémentaires devront être prises pour gérer les données conformément aux politiques institutionnelles (par exemple cryptage, anonymisation, pseudonymisation)

Utiliser des formats ouverts et des outils ouverts : •

Dans quels formats vais-je collecter mes données ? Puis-je utiliser des formats ouverts afin qu'ils puissent être lus dans différents programmes d'analyse ? • Quels outils vais-je utiliser pour traiter mes données ? Puis-je utiliser des outils ouverts pour que d'autres puissent effectuer les mêmes analyses ?

- Si vous utilisez des outils de traitement spécifiques pour des données personnelles et/ou sensibles, des accords de traitement de données doivent-ils être signés ?

Préparer le consentement éclairé

Le consentement éclairé est l'un des principes fondateurs de l'éthique de la recherche. Son objectif est que les participants puissent participer librement (volontairement) à la recherche avec des informations complètes sur ce que cela signifie pour eux de participer et qu'ils donnent leur consentement avant de participer à la recherche.

Un aspect important est également d'informer les participants sur leurs droits légaux tels que spécifiés dans le RGPD. Des informations détaillées sur les conditions du consentement éclairé peuvent être trouvées à [l'article 7](#) du RGPD.

Vous pouvez également trouver les [exigences](#) le consentement doit être respecté conformément aux règles du RGPD sur le site Web du RGPD.

Le consentement doit être obtenu avant que le participant ou le sujet ne participe à la recherche (prospectivement), et il ne doit y avoir aucune influence induite sur les participants pour qu'ils consentent. Les exigences minimales pour que le consentement soit considéré comme éclairé sont que le participant doit comprendre en quoi consiste la recherche et à quoi il consent (voir par exemple [ce guide](#)).

Le consentement doit être donné non seulement pour la participation à l'étude, mais également pour l'archivage et le partage des données. Par conséquent, il est important de réfléchir dès le départ si et comment vous souhaitez rendre les données qualitatives disponibles pour la réutilisation et de les inclure dans votre formulaire de consentement éclairé.

Il est probable que votre propre institution disposera également de conseils sur les pratiques de consentement éclairé et d'exemples de formulaires de consentement éclairé. Il est conseillé d'en discuter avec vos gestionnaires de données qui pourront également vous conseiller sur les options permettant d'inclure le partage de données dans votre formulaire de consentement éclairé.

Organiser et documenter les données qualitatives

Lorsque vous organisez et documentez vos données collectées, vous pouvez également envisager la réutilisation future de vos données.

Dans ce processus de recherche, vous souhaitez fournir suffisamment d'informations contextuelles avec vos données afin que vous et les autres puissiez toujours comprendre comment les données sont organisées et comment elles ont été collectées, traitées et présentées ultérieurement.

Le [guide expert en gestion de données CESSDA](#) donne des conseils sur les bonnes pratiques concernant les noms de fichiers et les structures de dossiers lorsque vous travaillez avec vos données, et fournit des exemples de documentation de données qualitatives telles que des fichiers de données textuelles, des données audiovisuelles et des périodiques ou magazines.

Lorsque vous nommez vos fichiers, il est généralement recommandé d'éviter d'avoir des données personnelles (par exemple le nom de la personne que vous avez interrogée) dans un nom de fichier de données ou de dossier. En général, il est sage de minimiser la quantité de données personnelles que vous utilisez et de ne conserver les informations personnelles que si elles s'ajoutent à votre question de recherche ou si elles sont nécessaires à la compréhension de la recherche. Ce dernier cas peut par exemple être le cas si vous avez interviewé des célébrités ou des artistes à propos de leur travail.

Si possible, stockez les parties de vos données qui contiennent des informations personnelles ou sensibles séparément des parties des données qui n'en contiennent pas. Il convient de noter que le cryptage peut être recommandé pour les données qualitatives sensibles ou confidentielles et que des plateformes institutionnelles spécifiques peuvent être requises. Votre gestionnaire de données local peut vous conseiller à ce sujet. Dans tous les cas, séparer les parties de vos données qui contiennent des informations personnelles ou sensibles des autres données peut vous permettre de partager plus facilement (des parties de) vos données pour une réutilisation ultérieure. Si l'on considère cette distinction dès le début du projet, l'archivage des données sera beaucoup plus facile par la suite.

Traitement de vos données

Bien que ce guide ne soit pas destiné à guider le traitement des données qualitatives, nous souhaitons encourager les chercheurs et les gestionnaires de données à réfléchir aux logiciels et aux outils utilisés pour traiter les données et à déterminer si des alternatives gratuites et ouvertes sont disponibles. L'utilisation d'outils ouverts facilite la reproduction de la recherche, car d'autres chercheurs peuvent également utiliser les outils ouverts.

Le respect des normes communautaires est également recommandé pour accroître l'interopérabilité et la réutilisabilité au sein de votre communauté. Cependant, gardez à l'esprit que les programmes utilisés dans une certaine communauté peuvent parfois ne pas être ouverts et pourraient donc constituer un obstacle pour les chercheurs extérieurs à votre communauté.

Certains outils et pointeurs connus des auteurs au moment de la publication sont abordés ci-dessous, mais nous encourageons vivement les lecteurs du guide à nous contacter si vous connaissez d'autres outils ouverts utilisés¹ pour traiter des données qualitatives que nous pourrions inclure. ici aussi.

Outils de transcription :

[CLARIAH](#), le Centre pour la technologie du langage et de la parole, notamment, travaille sur divers outils autour de la reconnaissance automatique de la parole (par exemple pour le néerlandais et l'anglais).

- Des outils de reconnaissance vocale sont disponibles en Python (par [exemple ce guide](#)).
- [oTranscrire](#) est une application web gratuite pour transcrire les enregistrements entretiens.
- Certains outils standards de la communauté qui ne sont pas ouverts, mais incluent des fonctionnalités de transcription sont Nvivo, Otter.ai, transcriptonline, amberscript et QDA Miner,

Outils d'annotation :

- [ÉLAN](#) est un outil d'annotation gratuit pour les enregistrements audio et vidéo développé par The Language Archive de l'Institut Max Planck de psycholinguistique.
- [Taguette](#) est un outil gratuit et [open source](#) pour la recherche qualitative dans lequel vous pouvez importer vos documents de recherche et mettre en évidence et étiqueter les citations.
- [Hypothes.is](#) est un outil d'annotation ouvert. [En collaboration avec le référentiel de données qualitatives, l'annotation pour une enquête transparente \(ATI\) a été développé pour accroître la transparence dans les annotations des données qualitatives.](#)

¹ Veuillez envoyer un message à info@dans.knaw.nl mentionnant « Contribution aux données qualitatives du guide »

- Certains outils standards de la communauté qui ne sont pas ouverts, mais qui incluent également des fonctionnalités d'annotation sont Atlas.ti, Nvivo et QDA Miner.

Désidentification des données personnelles

Lorsque vous travaillez avec des données personnelles, vous devez suivre certaines étapes supplémentaires lors du traitement de vos données, en particulier lorsque vous souhaitez archiver et partager les données. Il existe des considérations éthiques et des réglementations légales, comme le [Règlement Général sur la Protection des Données \(RGPD\)](#) prendre en compte. Il est généralement recommandé de consulter le responsable de la protection de la vie privée de vos institutions lorsque vous traitez des données personnelles ou sensibles.

Comme cela a été expliqué dans une section précédente, en vous assurant que votre consentement éclairé inclut de manière adéquate le stockage, l'archivage et le partage des données, vous garantirez que vous disposez d'une base juridique pour le traitement des données personnelles. Vous pouvez demander aux participants de consentir au partage de données personnelles identifiables. Ceci est recommandé lorsque vous savez à l'avance qu'il sera difficile d'anonymiser les données. Cela peut par exemple être le cas lorsque vous travaillez avec des données vidéo ou lorsque vous travaillez avec des données provenant d'organisations ou d'individus bien connus qui peuvent être facilement identifiés à partir de simples descriptions.

Pour protéger la vie privée de vos participants, la désidentification des données est un outil important. Vous pouvez anonymiser les participants à l'étude en anonymisant les données. ² Toutefois, pour travailler avec des données qualitatives, la désidentification n'est souvent ni possible ni souhaitable, car elle peut entraîner une trop grande perte d'informations. Il est important de trouver un équilibre entre confidentialité et compréhensibilité des données. Si vos données ne peuvent pas être anonymisées, vous pouvez toujours les partager dans certains cas. Vous souhaitez peut-être restreindre l'accès aux données pour protéger les participants ou vous pourrez peut-être les rendre réutilisables grâce à une réanalyse décentralisée comme dans CaRe & DaRe, comme indiqué plus loin.

Certains guides utiles sur la désidentification des données qualitatives ont été créés par l'Université [Erasmus de Rotterdam](#), [Service de données britannique](#), et le [référentiel de données qualitatives](#). L'université d'Utrecht travaille sur un [manuel sur la confidentialité des données](#) qui couvre également la désidentification des données. Tous les guides recommandent de développer une stratégie de désidentification dès le début de vos recherches.

Outils d'anonymisation des données qualitatives •

[anonymoUUS](#) est un package python créé à l'Université d'Utrecht pour pseudonymiser les dossiers et fichiers de votre documentation. • UKDS dispose d'un [outil d'aide à l'anonymisation de texte](#) qui met en évidence les termes de votre texte qui peuvent être révélateurs, afin que vous puissiez décider de les modifier en conséquence.

² Nous considérons que les données sont anonymisées si le processus de suppression des caractéristiques d'identification est irréversible, c'est-à-dire qu'il est impossible de réidentifier les participants / de revenir à un état dans lequel les participants peuvent être à nouveau identifiés. Si vous parvenez à anonymiser complètement vos données, celles-ci ne sont plus considérées comme des données personnelles.

³ Chaque fois qu'une clé est disponible pour traduire les données anonymisées en données originales, nous parlons de données pseudonymisées. Les données pseudonymisées sont toujours considérées comme des données personnelles au sens du RGPD. Il peut toujours être utile de procéder à une pseudonymisation pour protéger la vie privée des participants.

- [FAMTAFOS](#) est l'anonymisation de texte multilingue automatisée et gratuite pour la science ouverte. L'outil remplace les informations divulgatives de manière à conserver la sémantique des données. • Certains outils standards de la communauté qui ne sont pas ouverts, mais incluent également des fonctionnalités pour l'anonymisation sont Atlas.ti et Nvivo.

Archivage et publication de données qualitatives

Une fois votre projet de recherche terminé, il est important d'archiver vos données pour garantir leur conservation à long terme. Si possible, vous pouvez également publier vos données pour les rendre disponibles pour réutilisation. Compte tenu de la valeur des données qualitatives et des défis liés à leur partage et à leur préservation appropriée, il est important que les données soient archivées dans un référentiel fiable spécifique à un domaine comme la [Station de données DANS Sciences sociales et humaines. \(SSH\)](#), [qui conservera et conservera les données à long terme](#) (Mannheimer et al., 2018). Alternativement, votre propre institution peut disposer d'un référentiel de données recommandé, par exemple [DataverseNL](#). Votre gestionnaire de données local sera [en mesure de](#) vous fournir plus d'informations.

Documentation et métadonnées

Pour les données qualitatives, la documentation et les métadonnées détaillant le contexte de la recherche sont essentielles pour favoriser la compréhension et la réutilisation potentielle. Comme mentionné précédemment, il peut être assez difficile de capturer toutes les informations nécessaires.

De nombreux référentiels fournissent des indications sur la documentation requise et recommandée pour des disciplines scientifiques ou des types de recherche particuliers :

- DANS a

développé un modèle pour les métadonnées qui doivent accompagner une transcription des données d'entretien (audiovisuel). Le modèle est répertorié dans le guide DANS sur [la préparation des données pour le dépôt qui fournit également des](#) informations sur les métadonnées requises et la documentation en général.

- Le Guide Expert en Gestion de Données CESSDA comporte une section dédiée à [la Documentation et aux Métadonnées](#), [qui contient des conseils](#) sur la documentation au niveau du projet et au niveau des données pour les données qualitatives.
- Bien que les archives finlandaises de données sur les sciences sociales n'archivent généralement pas toutes sortes de données qualitatives, elles fournissent [des conseils](#) sur l'organisation et la documentation des informations de base avec des exemples utiles.

Lors du dépôt de données dans un référentiel, il est essentiel de fournir autant de métadonnées que possible pour garantir la possibilité de trouver les données. Cela permettra d'informer les gens sur le contexte de la recherche.

L'utilisation d'un standard de métadonnées permet de documenter les données de manière structurée et interopérable. Pour les sciences sociales, le manuel de codes de la Data Documentation Initiative, [DDI 2.5](#), est le plus souvent utilisé. Cette norme, bien que créée pour les données quantitatives, est également applicable à la recherche qualitative et aux méthodes mixtes (Mannheimer et al., 2018).

Les référentiels peuvent utiliser plusieurs normes de métadonnées adaptées à des communautés et disciplines scientifiques spécifiques. Les référentiels, comme DANS Data Station SSH, qui font partie du Consortium of European Social Science Data Archives (CESSDA) suivent le [modèle de métadonnées CESSDA \(CMM\)](#), [pour décrire](#) leurs banques de données. Quelques autres exemples :

- Le [service de données britannique QualiBank](#) utilise DDI 2.5, le schéma Qualitative Data Exchange (QuDEX) et le schéma Text Encoding Initiative (TEI) pour leurs données

⁴ Nous considérons qu'un référentiel est digne de confiance s'il est certifié par une norme internationalement acceptée qui évalue les processus de ce référentiel. Une norme bien établie est la certification [CoreTrustSeal](#).

- Le [Référentiel de Données Qualitatives](#) utilise DDI 2.5 et le noyau de métadonnées DataCite 3.1

Vous n'avez pas besoin de connaître les tenants et les aboutissants de ces schémas de métadonnées. Tant que vous déposez dans un référentiel fiable et spécialisé dans votre domaine, les conservateurs vous aideront à fournir les bonnes métadonnées pour votre ensemble de données.

Formats de fichiers

Pour faciliter la réutilisation, les données doivent être archivées dans des formats de fichiers avec lesquels d'autres personnes peuvent facilement interagir. Cela revient à dire qu'il y a toujours une préférence pour les formats de fichiers ouverts, car ils sont durables à long terme et ne nécessitent pas que quelqu'un installe ou achète un logiciel spécifique pour accéder, modifier ou traiter les données. .

Parfois, il existe des normes communautaires pour les formats de fichiers qui ne sont pas des formats ouverts (par exemple, parce qu'un certain logiciel est très couramment utilisé dans cette communauté, ou parce qu'il n'existe aucune alternative ouverte disponible), auquel cas vous pouvez déposer dans ce format à la place. Gardez toutefois à l'esprit que les réutilisateurs potentiels de vos données peuvent provenir de communautés très différentes et qu'ils se heurteront à des obstacles pour accéder à vos données. Si vous utilisez un logiciel spécial, il sera utile pour les futurs utilisateurs de vos données que le logiciel soit décrit, y compris sa version.

Certains référentiels fournissent à leurs utilisateurs des listes de formats de fichiers préférés, qui prennent en compte les normes communautaires et les formats ouverts. Le dépôt dans ces formats garantira la conservation de vos données à long terme, car le référentiel se charge de convertir les fichiers à mesure que la technologie progresse.

- DANS organise une [liste de formats de fichiers préférés](#) pour les [Data Stations](#) avec des explications sur le raisonnement derrière ces choix. Cela inclut les formats de textes, d'images, d'audio, de vidéo, d'analyse de données qualitatives (CAQDAS) et bien d'autres encore. • DANS et d'autres référentiels peuvent transférer des données pour vous dans des formats préférés afin que l'accessibilité à long terme puisse être assurée.

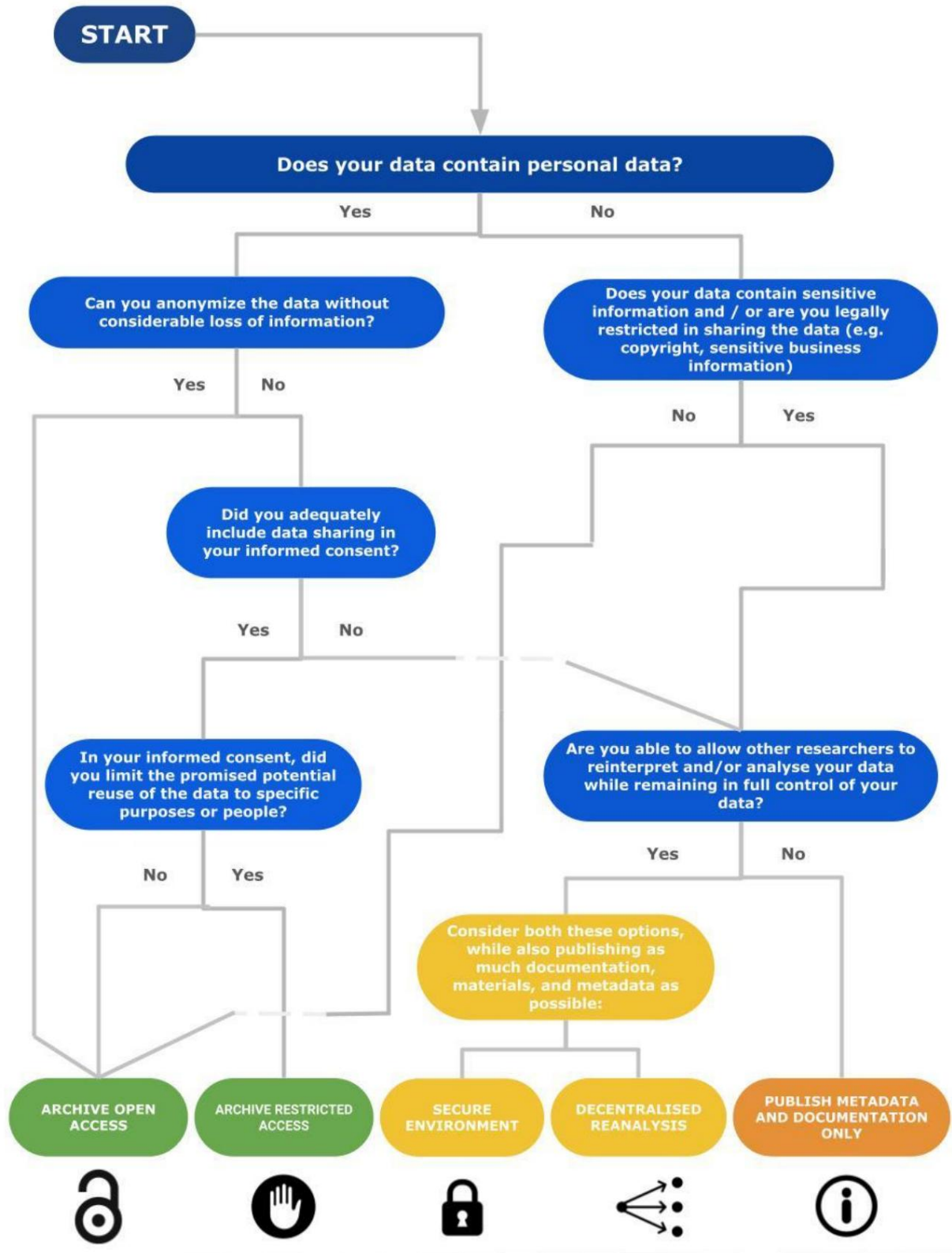
Arbre de décision pour la réutilisation des données

Savoir que la réutilisation des données qualitatives est précieuse s'accompagne également de certains défis et choix. Il peut être difficile de comprendre quelles options s'offrent à vous et de décider laquelle convient le mieux à votre situation. Pour faciliter le processus de prise de décision, nous avons créé un arbre de décision présentant toutes les différentes manières par lesquelles les données qualitatives peuvent être mises à disposition pour être réutilisées dans le paysage archivistique actuel. L'arbre de décision est également [disponible séparément sur Zenodo](#).

Ci-dessous, nous vous fournissons un contexte supplémentaire pour les différentes options de réutilisation, classées de la plus ouverte à la moins ouverte.

Qualitative data reuse decision tree

as open as possible, as closed as necessary



Afin de permettre aux autres de prendre connaissance de vos recherches, nous vous recommandons dans tous les cas de publier votre documentation et autant de métadonnées que possible sur vos recherches même lorsque vous n'êtes pas en mesure de partager les données elles-mêmes.

Certains référentiels acceptent uniquement les enregistrements contenant des métadonnées (voir ci-dessous), mais nous vous conseillons de publier tous les fichiers et documentations pertinents, ainsi que toutes les données (anonymisées) qui peuvent être partagées avec les métadonnées en tant qu'ensemble de données dans un référentiel digne de confiance, comme la DANS Data Station SSH.

Archiver en libre accès Si vous

n'avez aucune donnée personnelle ou sensible dans votre ensemble de données, vous pouvez envisager d'archiver les données en libre accès. Il est recommandé d'utiliser une [licence CC0](#) ou une autre licence Creative Commons et choisissez la licence la moins restrictive possible. Si vous parvenez à anonymiser complètement votre ensemble de données, celui-ci n'est plus considéré comme une donnée personnelle au sens du RGPD. Cela signifie que vous pouvez théoriquement partager vos données ouvertement, mais tenir compte de vos participants reste important. Lisez la section sur la désidentification dans ce guide pour plus d'informations.

Si vous envisagez de publier vos données en libre accès, nous vous recommandons d'archiver vos données dans un référentiel spécialisé dans votre domaine et certifié en fiabilité, par exemple le [DANS Data Station SSH](#). Alternativement, votre propre institution peut disposer d'un référentiel de données recommandé, par exemple [DataverseNL](#).

Archivage avec accès restreint Lorsque vous

déposez des données dans un référentiel, vos données ne deviennent pas automatiquement accessibles au public. Dans de nombreux référentiels, vous avez le choix du niveau d'accès que vous souhaitez attribuer à votre ensemble de données. Dans de nombreux référentiels fiables spécialisés dans les sciences sociales, un accès restreint aux données est une possibilité. Cela peut vous permettre de partager des données sous certaines conditions, si vous les avez spécifiées à vos participants dans le consentement éclairé, ou si ces conditions ont été stipulées par un bailleur de fonds, un éditeur ou une autre partie prenante. Cela pourrait signifier que seuls certains groupes de personnes peuvent accéder à vos données, ou que la réutilisation ne peut avoir lieu qu'à des fins spécifiques.

Dans la plupart des cas, un accès restreint signifie que les personnes intéressées à accéder à vos données devront faire une demande que vous examinerez personnellement. Il est important que vous disposiez d'un protocole d'accès aux données clair pour accepter ou refuser les demandes d'accès aux données sur la base d'un raisonnement bien fondé, afin que vous puissiez examiner objectivement les demandes à long terme.

Un autre niveau d'accès souvent regroupé sous accès restreint est l'application d'un embargo, ce qui signifie que les données ne seront accessibles au public qu'après une période de temps spécifiée. Un embargo peut par exemple être utilisé pour que les données soient disponibles en même temps que la publication scientifique.

Si vous envisagez de publier vos données en accès restreint, nous vous recommandons d'archiver vos données dans un référentiel spécialisé dans votre domaine et certifié en fiabilité,

⁵ Si vous souhaitez plus d'informations sur les licences et laquelle choisir, vous pouvez trouver des informations dans les conseils de DANS

par exemple la [DANS Data Station SSH](#). Alternativement, votre propre institution peut disposer d'un référentiel de données recommandé, par exemple [DataverseNL](#).

Environnement sécurisé

Dans certains cas (par exemple, lorsque vous devez faire face à des contraintes juridiques spécifiques), vous ne pouvez pas archiver vos données dans un référentiel mais vous pouvez permettre à d'autres chercheurs d'analyser vos données dans un environnement sécurisé où vous gardez le contrôle de toutes les données qui entrent et les résultats qui en ressortent. Dans certains cas, vous pouvez même exiger une analyse « aveugle », dans laquelle le chercheur ne voit pas les données originales mais ses algorithmes peuvent effectuer des analyses.

L'utilisation d'environnements sécurisés est une pratique courante pour les chercheurs travaillant avec des microdonnées de Statistics Nederland (CBS), qui sont très sensibles. Le projet [Secure ANalysis Environment \(SANE\)](#) travaille sur une infrastructure d'environnement sécurisé qui s'appuie sur cette expertise mais qui peut être utilisée par divers chercheurs et fournisseurs de données pour analyser des données sensibles.

Considérez que ces environnements nécessitent des efforts financiers et personnels et une formation pour les chercheurs qui souhaitent les utiliser ainsi que pour les propriétaires de données qui doivent préparer les données et potentiellement évaluer les résultats que les chercheurs souhaitent exporter.

La solution CaRe & DaRe : Réanalyse décentralisée

Cette option innovante de partage de données a été testée dans le projet CaRe & DaRe et constitue une solution dans les situations où d'autres formes de partage de données ne peuvent pas s'appliquer. Cette approche de réutilisation des données laisse les données entre les mains des créateurs d'origine. Les réutilisateurs potentiels peuvent approcher les chercheurs sur la base des informations (métadonnées et pièces justificatives) qu'ils partagent sur leurs données et leur demander de réanalyser leurs données sur la base d'un protocole d'analyse qu'ils ont créé. Les chercheurs originaux ne partageront alors que les résultats de cette analyse, pas les données elles-mêmes. En échange, les chercheurs participants pourraient devenir co-auteurs ou obtenir une autre forme de reconnaissance pour leur contribution. Cette approche du partage de données peut faciliter les études de cas croisées et autres réutilisations de données, tout en minimisant le risque d'interprétation erronée et de décontextualisation car c'est le chercheur d'origine qui effectue l'analyse.

Gardez à l'esprit que cette forme de réutilisation des données impose un fardeau de travail futur aux chercheurs d'origine, qu'ils doivent évaluer par rapport aux avantages qu'ils en retirent, et qu'il est toujours nécessaire de disposer d'une documentation et d'informations adéquates sur les données pour garantir que les réutilisateurs peuvent interpréter correctement leurs propres données (étant donné que vous pourriez recevoir des demandes plusieurs années dans le futur).

Le projet CaRe & DaRe a présenté l'approche pilote de réanalyse décentralisée.

Résultats initiaux - qui ont été [présentés lors d'une réunion du groupe d'intérêt Data Stewards](#) - faire preuve d'un enthousiasme général pour l'approche auprès des chercheurs et des data stewards de multiples disciplines qui autrement ne seraient pas en mesure de faciliter la réutilisation de leurs données.

Il reste encore beaucoup à développer en termes de règles d'engagement, de protocoles et d'explications plus approfondies du processus, mais les premiers résultats sont prometteurs. Le projet [OPEN-QUAL](#), Des méthodes innovantes pour la science ouverte dans la recherche en gestion qualitative ,

se concentrer sur l'élaboration plus approfondie du processus et de la théorie ainsi que sur la création d'une plate-forme où l'échange d'informations et de protocoles d'analyse peut avoir lieu. Si vous souhaitez en savoir plus sur cette approche de partage de données qualitatives, vous pouvez contacter le porteur du projet : prof. dr. ir. JJ Bérends⁶.

Publier uniquement les métadonnées

Lorsqu'aucune des options permettant de rendre vos données réutilisables n'est applicable à votre situation, vous pouvez toujours envisager de publier de la documentation, du matériel et autant de métadonnées que possible sur votre recherche afin de la rendre découvrable par d'autres. De cette façon, d'autres personnes sauront au moins que la recherche a été effectuée, ce qui pourrait être utile dans des revues ou des études documentaires. De plus, votre matériel et votre documentation pourraient être utiles pour être réutilisés dans d'autres études, quelles que soient les données.

Certains référentiels acceptent uniquement les enregistrements contenant des métadonnées, mais nous vous conseillons de publier tous les fichiers et documentations pertinents ainsi que toutes les données (anonymisées) qui peuvent être partagées avec les métadonnées en tant qu'ensemble de données dans un référentiel fiable tel que DANS Data Station. SSH.

Exemples d'ensembles de données

Pour vous inspirer sur la façon de mettre en pratique les conseils de ce livret, jetez un œil à certains de ces ensembles de données qualitatives déposés dans la station de données DANS, qui sont disponibles avec des catégories d'accès ouvertes ou restreintes. Réfléchissez à la manière dont ces chercheurs ont documenté et partagé leurs travaux et à la manière dont ceux-ci peuvent être réutilisés efficacement.

Exemples de jeux de données qualitatives disponibles au DANS avec la catégorie d'accès Libre accès :

- [Hanzon, MA CA \(JSO\) \(2019\) : Leren van Herhaald beroep dans le JeugdzorgPlus](#)
- [Evers, Drs. JC \(Université Erasmus de Rotterdam/Evers Recherche & formation\) \(2014\) : Collection thématique : Analyse qualitative : kunst én kunde](#)
- [Hoogsteder, dr MHH \(Amsterdam UMC, localisation VUmc\) \(2020\) : Samen Gezond Groot - optimisation du conseil en formation dans le JGZ 0-4](#)
- [Damen, Dr. AM \(Universiteit voor Humanistiek\) \(2022\) : Palliatief Landelijk Onderzoek Eerstelijns Geestelijke verzorging \(PLOEG\) projet de projet 3 : « Integratie GV eerste lijn vanuit 3 multidisciplinaire praktijken »](#)
- [Ducimetière, BA MD \(Université d'Amsterdam\) ; Soulioti, BA AS \(Université d' Amsterdam\) \(2020\) : Interview Eva Gonggrijp sur Grenzeloze Inhoud](#)

Exemples d'ensembles de données qualitatives disponibles au DANS avec catégorie d'accès Accès restreint où les utilisateurs doivent demander l'autorisation du propriétaire des données pour utiliser les données :

⁶ Pour contacter l'approche de réanalyse décentralisée, contactez jjberends@vu.nl.

- [Vries, Dr. DH de \(Université d'Amsterdam\) \(2021\) : L'impact social de l'impact social sur les populations vulnérables pendant le COVID-19 \(2020\) : interviews de haut niveau avec des personnes âgées et des personnes âgées](#)
- [Hofhuizen, Drs. CJM \(Niveau\) \(2022\) : Zorgen en zwaaien : étude longitudinale sur les mesures restrictives de lutte contre le COVID-19 pour les hommes avec un aperçu complet et un soutien à l'assistance](#)
- [Hulst, Dr. AAL van der \(2015\) : Wegbereiders, Roma et Sinti aux Pays-Bas et en Tsjechië sur le profit d'onderwijs, 1950-2020 Interview 05](#)
- [Ducimetière, BA MD \(Université d'Amsterdam\) ; Vooqd, BA MV \(Université d'Amsterdam\) ; Stigter, Dr. S. \(Université d'Amsterdam\) \(2020\) : Interview Monique Laros sur Roel QoQo et Grenzeloze Inhoud](#)
- [Mol, Drs. T. \(Rijksmuseum\) ; Stigter, Dr. S. \(Université d'Amsterdam\) \(2017\) : Interview Alessandro Mendini](#)

Dernières pensées

Ceci conclut le guide « Rendre les données qualitatives réutilisables ». Nous espérons que ce guide aidera les chercheurs et les gestionnaires de données travaillant avec des données qualitatives à envisager les opportunités de partage et de réutilisation de leurs données au cours des différentes phases d'un projet. Même si les données qualitatives semblent souvent mal adaptées à la compréhension et aux politiques actuelles entourant la gestion et le partage des données, des options de partage de données sont disponibles même pour les ensembles de données les plus complexes. Étant donné que les données qualitatives sont si particulières dans la manière dont elles sont collectées et dans la valeur qu'elles peuvent avoir pour un large éventail de disciplines scientifiques, elles peuvent faire une grande différence si elles deviennent plus facilement disponibles pour être réutilisées. La portée et les implications des données pourraient même s'étendre bien au-delà des communautés envisagées par les chercheurs initiaux eux-mêmes. L'expression « aussi ouverte que possible, aussi fermée que nécessaire » s'applique sûrement également aux données qualitatives.

Ce guide est une première approche d'un aperçu des ressources et des informations permettant de rendre les données qualitatives réutilisables. Si vous souhaitez partager plus d'informations, des outils ouverts ou d'autres ressources pertinentes à ce sujet, nous serions très reconnaissants si vous nous contactez avec ces informations.

7.

Ressources additionnelles

Outre les guides, outils et conseils rassemblés dans ce guide, il existe également des ressources plus longues à considérer si vous souhaitez mieux vous familiariser avec la gestion des données de recherche qualitatives.

Ressources pour les données qualitatives : •

- [UK Data Service](#) contient des enregistrements de webinaires et du matériel de présentation répertoriés sur le thème des données qualitatives
- [Le Référentiel de Données Qualitatives](#) a une liste de conférences, d'ateliers et de formations liés aux données qualitatives sur leur site Web.
- Le cours en ligne [Gestion des données qualitatives en sciences sociales](#) a été créé par le Social Science Research Council et le Qualitative Data Repository. • Le manuel de cours sur les données de recherche circulaire sur [les données qualitatives FAIR](#) contient le [matériel de cours d'un atelier](#) créé à l'Université de Maastricht.

Ressources pour les données personnelles ou sensibles : •

- [Research Data Pays-Bas](#) a créé un cours sur le RGPD pour les partisans des données. Vous pouvez vous inscrire pour suivre le cours de manière formelle ou parcourir librement les documents en ligne (actuellement disponibles uniquement en néerlandais) • [Le manuel sur la confidentialité des données](#) a été créé par l'Université d'Utrecht et contient connaissances, outils et cas d'utilisation pour travailler avec des données personnelles.
- [EOSC-Futur](#) a organisé une session « demandez-moi n'importe quoi » sur les données sensibles, discutant des défis et des questions fréquemment posées sur le sujet. Vous pouvez revoir l'enregistrement de ce webinaire.

⁷ Veuillez envoyer un message à info@dans.knaw.nl mentionnant « Contribution aux données qualitatives du guide »

Les références

Verburg, M., Braukmann, R., Mahabier, W. et Khadjavi Pour, K. (2023). Réutilisation de données de recherche qualitatives - développements et besoins de la communauté (version 1). Zénodo. <https://doi.org/10.5281/zenodo.7704480>

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., McGillivray, B. (2020) Le avantage de citation de relier les publications aux données de recherche. PLoS ONE 15(4) : e0230416. <https://doi.org/10.1371/journal.pone.0230416>

Piwowar, HA, Day, RS, Fridsma, DB (2007) Le partage de données de recherche détaillées est Associé Augmentation Citation Taux. PLoS UN 2(3) : <https://doi.org/10.1371/journal.pone.0000308>

Mannheimer, S., Pienta, A., Kirilova, D., Elman, C. et Wutich, A. (2018). Données qualitatives Partage : les référentiels de données et les bibliothèques universitaires, partenaires clés dans la résolution des problèmes Défis. Américain Scientifique comportemental, 63(5). <https://doi.org/10.1177/0002764218784991>

Munafò, M., Nosek, B., Bishop, D. et al. Un manifeste pour une science reproductible. Nat Hum Comportement 1, 0021 (2017). <https://doi.org/10.1038/s41562-016-0021>

Wilkinson, M. et coll. (2016). Les principes directeurs FAIR pour la gestion des données scientifiques et l'intendance. Données Sci 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Landi, A. et coll. (2020) ; Le « A » de FAIR – Aussi ouvert que possible, aussi fermé que nécessaire. Intelligence des données 2 (1-2) : 47-55. est ce que je: https://doi.org/10.1162/dint_a_00027