GUIDE

OPEN SCIENCE

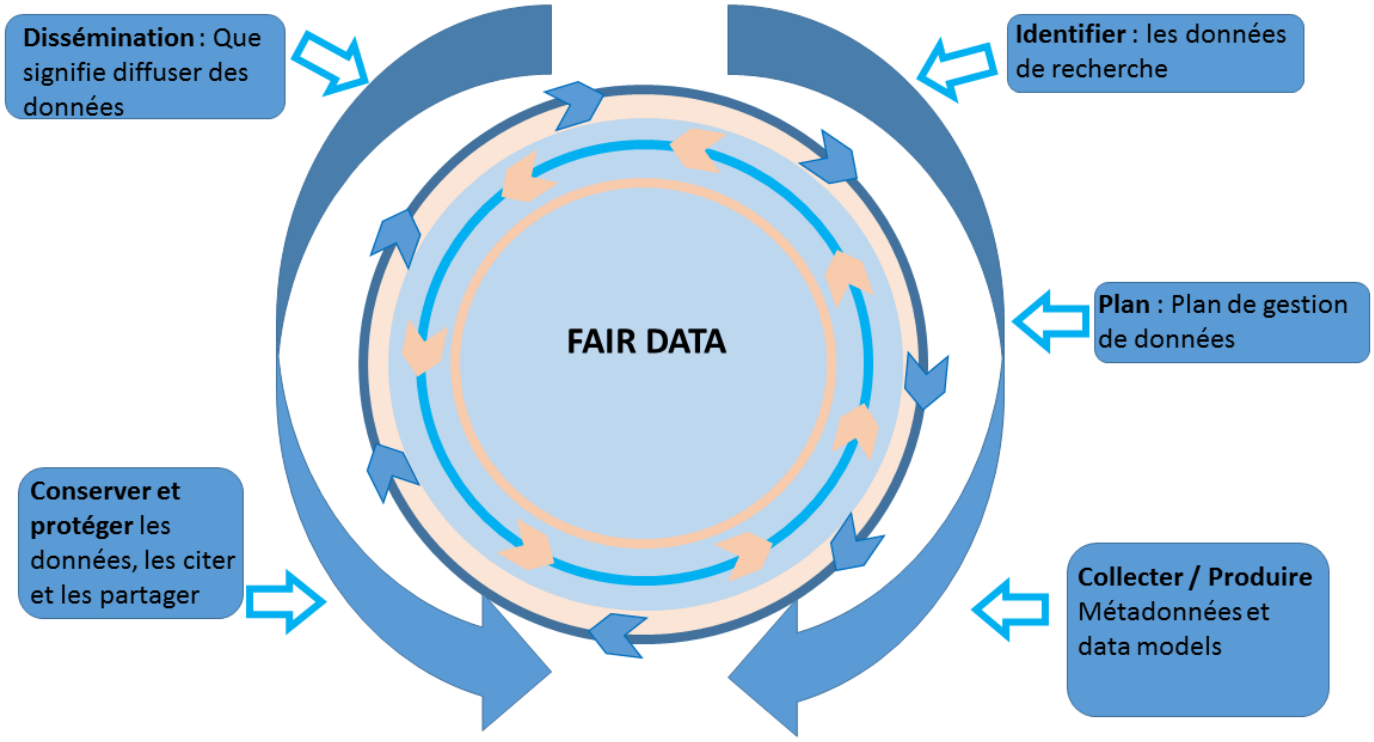
UNIVERSITÉ de Picardie Jules Verne

Table des matières

Introduction

Le développement de la science ouverte répond à des enjeux scientifiques et aussi économiques exigés par l'Union Européenne et les principaux financeurs nationaux à l'exemple de l'ANR. Cette évolution modifie « la façon » de faire de la recherche en instaurant de nouvelles méthodes de travail. Mais cette complexification de la recherche via les principes FAIR (Facile à trouver, Accessible, Interopérable et Réutilisable) n'est pas un problème en soi car mettre en place sa recherche selon ces principes est un gage de qualité et d'intégrité permettant une meilleure transparence des résultats et une recherche plus éthique.

Il est désormais établi que les principes sur l'ouverture, la communication, l'appropriation et la réutilisation des données de recherche, quelles qu'elles soient, sont applicables aux résultats de la recherche dans toutes les disciplines.



1) Identifier

Dans toutes les disciplines de recherche, les chercheurs utilisent des données, même s'ils n'en ont parfois pas toujours conscience. Bien que le terme « données » semble intuitivement plus facile à définir dans le domaine des sciences dures, les données en sciences humaines pourraient être envisagées comme des matériaux permettant de mettre en place les outils de la recherche, qu'ils soient numériques ou analogiques, comme à titre d'exemple : un livre imprimé, des notes manuscrites, des catalogues, des listes, des tableaux ou matrices, ou encore des bases de données.

Exemples : population des villes médiévales françaises, nombre des titres et des tirages des romans victoriens, liste de participants à la Society of Independent Artists exposition en 1917, PIB des pays européens avant et après le Brexit, analyses moléculaires etc… Ces données peuvent être à la fois physiques et numériques, comme des artefacts, documents numériques (y compris numérisés), images (2D ou 3D), enregistrements sonores et vidéo, des données archéologiques (relevés topographiques, photographies, dessins de fouilles), manuscrits, textes de poésie, publications sur les réseaux sociaux, peintures, scans d'architecture 3D, enregistrement d'une représentation théâtrale, etc…

Nota Bene :

- Considérez tous vos actifs de recherche comme des données de recherche qui pourraient être potentiellement réutilisées par d'autres chercheurs et leur être utiles, afin de rédiger par un exemple l'état de l'art de leur recherche.
- Utilisez des outils bien établis pour faciliter votre travail de recherche, car beaucoup d'entre eux autorisent les données partagées par ex. Bibliothèques du MIT Digital Humanities : Tools and Resource.
- Parcourez les ensembles de données de votre domaine et déterminez si vos propres actifs pourraient être publiés d'une manière similaire (par exemple Humanities Commons, UK Data Archive, ARCHE re3data.org filtré pour les sciences humaines).
- Gardez à l'esprit, lorsque vous commencez à produire des données, cette maxime de la science ouverte : les données doivent être «aussi ouvertes que possible et aussi fermées que nécessaire».

2) Plan de gestion de données

Une référence : https://opidor.fr/

Gérer les données pour leur éventuel partage ou réutilisation est un processus qui requiert attention et planification. Les chercheurs doivent planifier et allouer du temps pour la gestion des données au début de leur projet de recherche, et suivre les progrès de la recherche vers les objectifs, les livrables et les jalons du projet. Ceci implique également de comptabiliser, le plus précisément possible, en amont tous les coûts pouvant résulter de la conservation des données. La planification de la gestion des données est une démarche essentielle et de plus en plus une exigence des financeurs. Elle présente de multiples avantages pour le chercheur, car elle permet de visualiser les étapes, les périodes et les personnes clés d'un projet de recherche. Le plan de gestion de données (le PGD ou DMP) doit prévoir comment les données seront créées, collectées, gérées, documentées, décrites, partagées et préservées avant, pendant et après une activité de recherche.

Le plan de gestion de données peut être créé via le site Opidor [1]. Ce modèle a été validé par l'ANR, mais il est possible qu'un financeur autre vous impose son propre modèle.

Le PGD doit préciser également les restrictions sur l'utilisation des données, y compris les aspects de propriété intellectuelle, les données sensibles (RGPD[2]) et les questions de confidentialité. L'objectif est de s'assurer que les données sont traitées de manière appropriée tout au long de l'activité de recherche. Une bonne planification permet aux chercheurs de maximiser le potentiel d'utilisation (et de réutilisation) de leurs données de recherche actuelles et futures. Énumérer et organiser les données dès le début d'un projet engendrent des économies de temps et d'efforts. Une bonne gestion des données facilite leur réutilisation ce qui permet d'éviter la perte de données ou un travail conséquent de réorganisation.

Nota Bene :

Comment faire son plan de gestion (les questions auxquelles il faut répondre):
1. Définir les données - en termes de contenu, de format, taille, etc.
2. Méthodologie de création et / ou de collecte de données et la qualité des données collectées, veillez notamment à mettre en place des métadonnées très enrichies

[1] https://opidor.fr/ **DMP OPIDoR** est un outil d'aide à la création en ligne de plans de gestion de données (Data Management Plan ou DMP) mis à disposition de l'Enseignement Supérieur et de la Recherche. Il est hébergé et géré par l'Inist-CNRS

[2] Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données).

3. Modèles de données, comment les données seront-elles structurées /organisées et utilisées.
4. Documentation - comment les données seront-elles documentées et décrites, et quelles métadonnées allez-vous créer, et dans quelle mesure.
5. Questions juridiques et / ou éthiques.
6. Comment les données seront-elles stockées et accessibles pendant l'activité de recherche.
7. Comment les données seront-elles conservées à long terme, et toutes les politiques pertinentes régissant la conservation ou la disposition des données pour les chercheurs.
8. Comment les données seront-elles partagées - organiser les données dans des formats ouverts et normalisés, avec identifiants persistants qui facilitent leur réutilisation et lisibilité (machine) dans un environnement durable à long terme.
9. Propriété et responsabilités des données: qui est responsable de la gestion des données? Et qui possède les données ?

Une bonne gestion des données facilite la réutilisation de celles-ci, facilite la continuité entre les projets, et améliore la visibilité et l'impact des résultats de la recherche.

3) Collecter, produire, structurer et stocker

A) Types et formats

Une fois les besoins de recherche identifiés, les chercheurs doivent décider de la forme finale de leurs données, en tenant compte du type de données (c'est-à-dire quel type de données seront collectées), le format (c'est-à-dire quel format de fichier utilisé) et les typologies. En Sciences humaines par exemple, les chercheurs numérisent souvent des objets physiques, mais aussi créent des données de recherche qui ne sont pas numériques, telles que les textes annotés manuellement ou les notes de terrain sur papier. Les principes FAIR se concentrent essentiellement sur les données numériques, mais il est recommandé d'essayer de rendre accessible tous types de données de recherche.

Différentes disciplines peuvent utiliser différents types de données, mais aussi différentes typologies de données. Les types de données peuvent être distingués par le ou les médias qu'elles utilisent, à savoir le mot, l'image, le son ou combinaison de ceux-ci, par exemple, un carnet de terrain, un entretien, un enregistrement de performance, des photographies, des notes de bas de page. Les types de données peuvent également être pris en compte selon leur structure : Sont-elles structurées (base de données), semi-structurées (XML) ou non structurées (ordinaire texte). En outre, une telle structure pourrait être linéaire (par exemple table), hiérarchique (par exemple structure arborescente) ou multi-relationnelle (par exemple réseau).

Une fois le type de données sélectionné, il faut décider de son format, c'est-à-dire le type de fichier dans lequel les données seront encodées. Les choix de format de fichier sont extrêmement importants et doivent figurer dans le plan de gestion de données. Pour un même objet de recherche, il existe plusieurs formats d'encodages possibles. Par exemple, le même texte pourrait être stocké dans un TXT (texte brut), Format ODT (formaté) ou XML (structuré). Le choix du format doit refléter à la fois son type et l'utilisation souhaitée de la recherche. L'usage de logiciel libre et non propriétaire est recommandé. Le PDF pourtant format propriétaire est toléré car lisible par tous.

Les images sont mieux stockées dans un format d'image haute résolution sans perte (par exemple le TIFF).

Nota Bene

1) Pensez et recherchez les formats et les pratiques les plus communes dans votre domaine de recherche, pensez également à l'avance à toute dépendance à un logiciel créée par votre choix de format.
2) Assurez-vous que vos formats de données sont choisis et respectés par vos partenaires de projet.
3) Les mêmes informations peuvent être enregistrées sur différents types de supports et formats de données. Par exemple, une liste de notices bibliographiques peut être transcrite et enregistrée sous forme de tableau au format CSV ou XML. Avant de faire un choix, vous devez rechercher quels sont les formats utilisés par d'autres chercheurs pour des données similaires pour une meilleure diffusion et accessibilité de vos données.
4) Peu importe le choix que vous ferez, vous devez être conscients qu'un format numérique - propriétaire ou non – est susceptible de devenir obsolète ou corrompu faute de mise à jour par exemple, ce qui rendraient vos données illisibles. Par conséquent, il est recommandé d'utiliser des formats libres des formats standards adoptés depuis le début, ou si ce n'est pas le cas d'exporter les données et les résultats obtenus avec un logiciel aux formats d'enregistrement standards.
5) Les formats standards et les formats ouverts, en raison de leur large utilisation et du soutien communautaire, sont le plus souvent plus durables et accessibles à long terme.

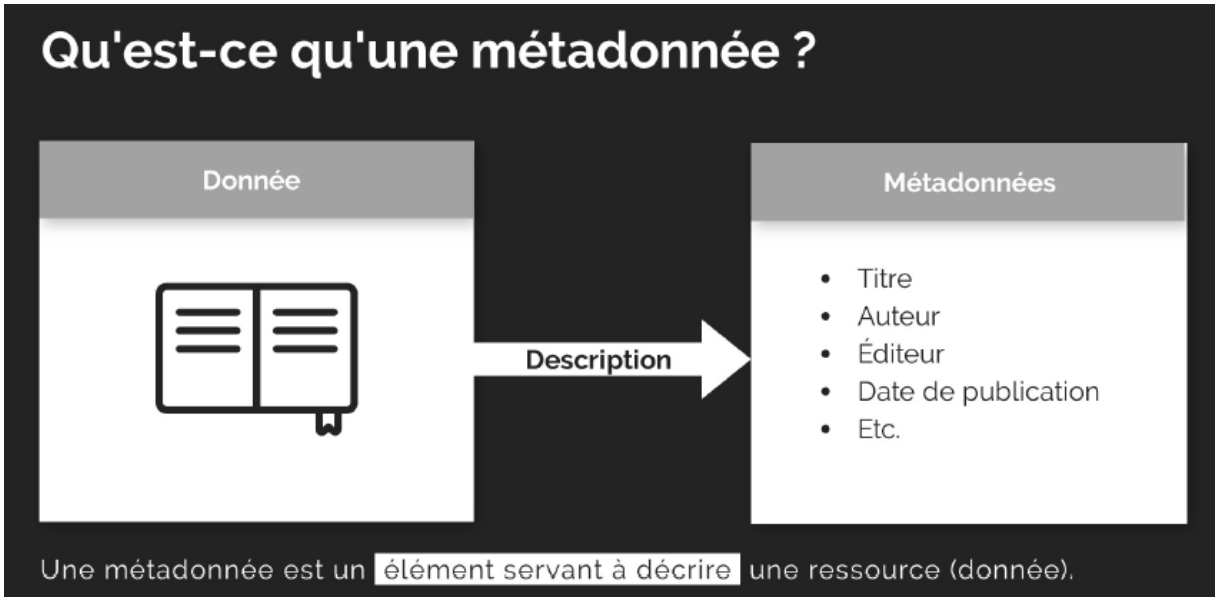
B) Les Métadonnées

Pour que les données soient réutilisables, elles doivent être accompagnées d'informations suffisantes pour communiquer le contenu de l'ensemble de données, les fins ou les circonstances dans lesquelles elles ont été créées ainsi que les modalités de leur réutilisation.

Qu'est-ce qu'une métadonnée ?

En termes simples, les métadonnées sont des « données sur les données », ou des informations utilisées pour identifier et décrire les données. Elles sont l'un des éléments clés de la pratique FAIR.

Selon les principes FAIR, l'identifiant persistant (PID) et « des métadonnées suffisamment riches » sont suffisants pour permettre à vos données d'être trouvées, utilisées et citées de manière fiable.

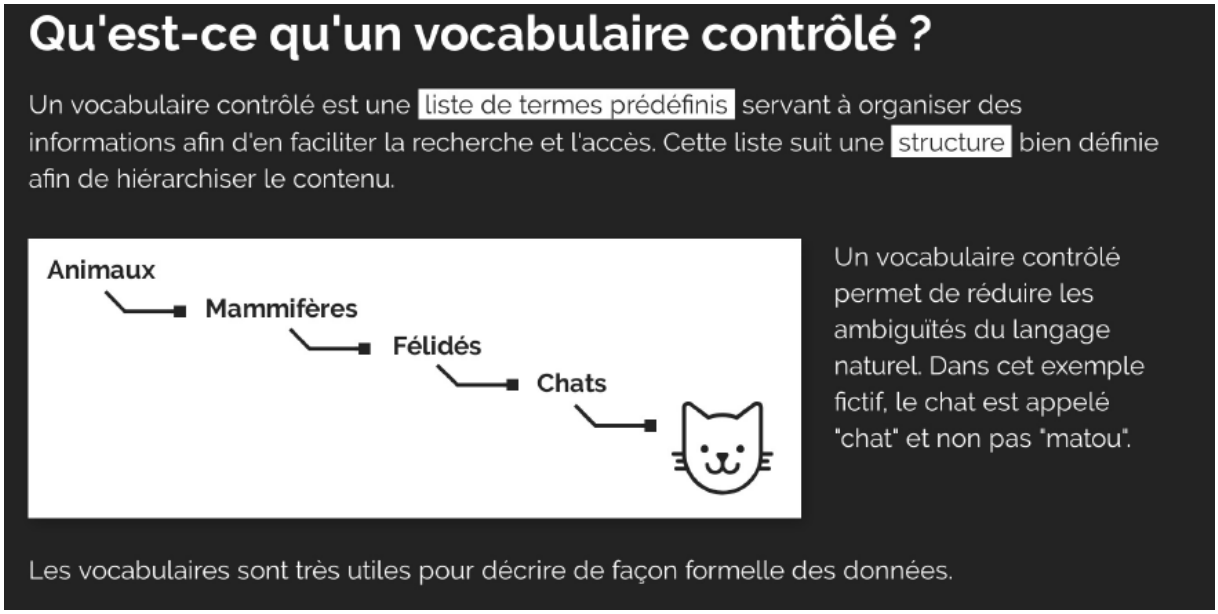


[3]

Les métadonnées sont inhérentes aux documents que vous avez créés, par exemple celles présentes dans un document texte, dans une photographie, ou des données constructeur pour une image produite par un microscope. Votre travail sera de les enrichir avec des mots clefs pour qu'elles puissent être trouvées par une simple recherche. Pour cela, vous devez créer vos métadonnées en utilisant un vocabulaire contrôlé.

[3] Cette présentation est réadaptée de ANDS | The FAIR data principles. https://www.ands.org.au/working-with-data/fairdata/training#.XNqUWpLgeTA.link

https://view.genial.ly/5d64fbbd8352350fa3d22603/interactive-content-les-principes-fair



4

Les métadonnées peuvent être une série de champs qui décrivent des données et d'autres objets de recherche de manière cohérente et standardisée, comme une notice bibliographique d'un livre de bibliothèque.

Les métadonnées enrichies sont celles que vous allez créer en les ajoutant en référence à vos documents.

Des normes de métadonnées ont été créées, souvent par différentes communautés et disciplines de recherche, pour fournir des manières optimales et adaptées de décrire les données. Au démarrage du processus de recherche, il est crucial d'identifier une norme de métadonnées appropriée pour votre discipline ou domaine, et qui est compatible avec le référentiel dans lequel vous déposez vos jeux de données.

Par exemple : Certaines des normes communément adoptées pour les données des sciences humaines sont Dublin Core, qui est flexible et largement utilisé dans les sciences humaines numériques, interMARC, un standard de catalogage de bibliothèque et EAD, qui est utilisé par les archives.

Bien que les normes varient selon la discipline, l'important est de choisir une norme existante, au lieu de créer un nouveau schéma d'organisation de vos métadonnées. De plus, les métadonnées elles-mêmes devraient être persistantes et accessibles, même si les données décrites sont restreintes ou ne sont plus disponibles. En plus des métadonnées bibliographiques

---

[4] Cette présentation est réadaptée de ANDS | The FAIR data principles. https://www.ands.org.au/working-with-data/fairdata/training#.XNqUWpLgeTA.link

https://view.genial.ly/5d64fbbd8352350fa3d22603/interactive-content-les-principes-fair

(auteurs, titre, résumé, etc.), il est important de fournir des métadonnées sur le contexte et la provenance des données. L'entrepôt choisi doit indexer les données et permettre leur recherche par les humains et les machines.

Nota Bene :

1) Un bon point de départ est de consulter le Metadata Standards Directory, un répertoire géré par la communauté hébergée par la Research Data Alliance: https://rd-alliance.github.io/metadata-directory/
2) Les métadonnées fonctionnent mieux lorsque la terminologie est cohérente, par exemple les conventions de dénomination sont respectées, l'orthographe est normalisée, etc. Selon la complexité et la taille de vos métadonnées, pensez à utiliser un outil tel qu'OpenRefine pour « nettoyer » vos métadonnées.
3) Pour des raisons de recherche, les métadonnées doivent inclure une référence claire et explicite à l'ensemble de données qu'elles décrivent, grâce à l'inclusion d'un PID dans les métadonnées.
4) Rendez vos métadonnées aussi riches que possible afin de mieux contextualiser vos données et faciliter leur réutilisation. Envisagez des descriptions plus détaillées et des informations de provenance plus complètes, ainsi que l'utilisation du spectre des champs de métadonnées disponibles. Les métadonnées doivent être lisibles par ordinateur ou système informatique.

C) Modèles de données

La modélisation des informations sur les artefacts et les concepts abstraits a toujours été une question importante dans le processus de recherche en sciences humaines. Un appareil critique d'une édition ou du catalogage des découvertes archéologiques sont des façons reconnues de représenter les connaissances dans certaines disciplines.

Le cahier de laboratoire constitue un véritable outil scientifique et ce, dès le commencement d'un projet. Les données enregistrées ont vocation à être utilisées par un tiers et ce, plusieurs années après l'enregistrement des premières données. Pour cela, il faut être le plus précis possible pour faciliter la relecture par l'auteur ou un de ses collaborateurs dans un délai inconnu (jusqu'à plusieurs années) et ainsi favoriser la réutilisation potentielle des informations. Point important, la numérisation de son contenu, peut éviter la dégradation de cet outil (encre effaçable, crayon de papier…) à condition que le format d'enregistrement soit viable et pérenne.[5]
Il existe également des cahiers de laboratoire électroniques. Les cahiers de laboratoire électroniques ne sont pas des simples équivalents numériques des cahiers papier. Ils présentent de nombreux autres avantages, pour *FAIRiser* ses données :

- En créant un lien direct vers des données déjà disponibles sous forme numérique;
- En aidant la documentation des données de recherche avec des métadonnées;
- En facilitant l'accès et la récupération des données de recherche, grâce aux fonctions de recherche et de filtrage;
- En contribuant à la reproductibilité et à la réutilisation des données de recherche.

De plus, ces cahiers sont intégrés dans un environnement de recherche numérique en réseau.

[5] Ref Guide CNRS Traçabilité des activités de recherche et gestion des connaissances page 11

Idéalement, ils doivent être en mesure de se connecter à d'autres outils de recherche : instruments de mesure, référentiels de données, etc…, ce qui facilite les flux de données et de métadonnées, et évite la perte d'informations et garantit ainsi la qualité des données.[6]
Par ailleurs, si des informations confidentielles sont inscrites dans ce cahier, elles peuvent être protégées par un embargo : la loi limite la durée des embargos à 6 mois en sciences, techniques, médecine et à 12 mois en sciences humaines et sociales après la publication ou le dépôt dans une archive ouverte. Si vos données sont particulièrement sensibles (données personnelles ou données pouvant menacer les intérêts de l'Union européenne, secrets industriels par exemple) vous n'êtes pas tenu de les publier mais vous serez obligé de vous expliquer.

NB : Les données personnelles à la fin de votre recherche doivent être anonymisées. Suite à ce processus, ces données pourront être publiées ou déposées dans une archive, ou un entrepôt généraliste ou disciplinaire.

Dans le cadre du processus de recherche, le processus de modélisation des données est donc d'une grande importance et doit être abordé de manière sérieuse, préférentiellement lors de la phase initiale de la recherche. Il est nécessaire de prévoir suffisamment de temps pour cette période, car les modifications ultérieures des structures de données ou même du modèle de données peuvent prendre beaucoup de temps.

[7]

[6] https://openscience.pasteur.fr/2019/11/14/what-role-do-electronic-lab-notebooks-play-in-the-context-of-research-data-management-and-open-science/

[7] Cette présentation est réadaptée de ANDS | The FAIR data principles. https://www.ands.org.au/working-with-data/fairdata/training#.XNqUWpLgeTA.link

https://view.genial.ly/5d64fbbd8352350fa3d22603/interactive-content-les-principes-fair

Nota Bene

1) Utiliser des normes ouvertes et, dans la mesure du possible, des technologies et procédures normalisées. Le World Wide Web Consortium W3C maintient plusieurs normes pertinentes pour les modèles de données comme XML et RDF. Au sein de XML, le TEI / MEI de Text ou Music Encoding Initiative ou leurs expressions spécifiques sont devenus des normes pour les éditions de texte ou de musique. La requête le langage SPARQL et l'outil de représentation des données liées JSON-LD sont des normes communes pour RDF.
2) Préférez les systèmes lisibles par l'homme et la machine: le codage des modèles de données et des données réelles qui sont à la fois lisibles par l'homme et la machine de manière unifiée offre une meilleure durabilité et l'accessibilité à long terme que le code lisible par machine uniquement (codes binaires), qui peut utiliser différents formats pour la description du modèle de données et les données réelles comme XML, TEI / XML, Turtle, N3, RDF / XML, tandis que les technologies basées sur SQL nécessitent des efforts plus importants.
3) Normaliser autant que possible: pour éviter les informations redondantes
4) Les modèles de données suivent le plan de gestion des données (DMP) : lors de l'établissement d'un modèle de données, les chercheurs doivent garder à l'esprit le cycle de vie complet de leurs données, comme décrit dans un DMP. Par conséquent, une documentation complète du modèle de données, de ses logiciels et outils est très pertinente et facilite le transfert de données dans un référentiel sécurisé et fiable afin de les garder accessibles.

D) Préserver et stocker

1) Aspects légaux

Le partage de données soulève inévitablement des questions de droits de propriété intellectuelle et de vie privée. Avec le numérique, les données deviennent de plus en plus accessibles, mais il est possible de les protéger et vous devez prévoir cette protection dans l'élaboration de votre plan de gestion de données en début de projet. Dans certaines disciplines, les exigences de l'obtention d'une protection par brevet peuvent restreindre la diffusion des données. La question du droit d'auteur est plus que pertinente pour les données issues des recherches en sciences humaines.

Dans la pratique, avant toutes réflexions sur la diffusion de vos données, vous devez vous poser ces questions :

• Quelle législation s'applique aux travaux de chercheurs que j'utilise dans mon projet ?

• Ai-je le droit de collecter, conserver et de transmettre les données de mon projet ?

• Vais-je collecter des informations sensibles ?

• Existe-t-il des risques d'exposer l'identité des personnes participant à mon étude ?

Nota Bene :

Clarifiez toutes les questions juridiques au début de votre projet de recherche et incluez les conclusions de ce processus dans le plan de gestion des données.

1) Utilisez des listes de contrôles adaptées à votre sujet de recherche / discipline.
2) Dans le cas de données personnelles, assurez-vous que seules les personnes concernées peuvent accéder aux données et que celles-ci sont clairement identifiées (voir RGPD).
3) Demandez le consentement des personnes impliquées dans l'étude avant de partager toute donnée même anonymisée. Mettez en place des tableaux de concordance afin d'anonymiser les données au fur et à mesure de la recherche et prévoyez ce temps dans le PGD.
4) Évitez de collecter des données personnelles (sensibles et non sensibles) lorsque cela est possible.
5) Prenez contact avec votre référent au sein de votre direction de la recherche ou votre délégué à la protection des données afin d'obtenir des conseils (RGPD, DPI, droits d'auteur, brevets, marques, etc.).
6) Si vous avez besoin de l'autorisation du détenteur des droits d'auteur pour utiliser des sources telles que des images pour votre publication, essayez d'en obtenir une qui couvre les copies imprimées et numériques.

2) Stockage

Pour que les données soient gérées et accessibles sur le long terme, il faut les déposer dans un entrepôt qui permet leur identification de manière unique et pérenne, leur accès et le téléchargement par les humains et les machines.

Les chercheurs déposant leurs données afin de faciliter leur réutilisation par d'autres chercheurs doivent s'assurer que les données sont authentiques, récupérables et annotées suffisamment pour comprendre le contexte de leur création. L'attribution des informations de licence clarifie les conditions de réutilisation.

Il existe de nombreuses façons de stocker des données pendant le processus de recherche, les chercheurs sauvegardent leurs données sur des ordinateurs personnels, disques durs externes, disques durs, clés USB, serveurs institutionnels ou le cloud via différents services de stockage. Il est tout de même préférable de stocker sur les serveurs et les clouds de votre institution que sur votre ordinateur personnel (vol, perte) ou sur un serveur privé comme Google Drive qui pose question en matière de protection des données.

Cependant, le stockage n'est pas la même chose que la conservation, parce que les données numériques sont fragiles et sujettes à la corruption et la dégradation au fil du temps. Il en va de même pour les données présentes sur des sites internet qui peuvent ne plus être accessibles en cas de suppression du site, ou en cas de lien hypertexte mort.

a) Identifiants persistants (PID)

Dans tous les cas, le principe « facile à trouver » de la démarche FAIR doit remplir l'exigence de l'attribution d'un identifiant unique à une donnée. Ces identifiants persistants permettent l'identification fiable d'un auteur ou d'une institution et favorisent et facilitent le partage, la

réutilisation des productions scientifiques et permettent leur accès sur le long terme. Ils simplifient également leur citation.[8]

Il existe 2 grands types de PID :

- les identifiants « objet » pour les productions scientifiques (publications et données) : permettent l'identification fiable des productions scientifiques et facilitent la recherche et leur accessibilité
- les identifiants « contributeur » pour les auteurs et les institutions.

De nombreux identifiants existent déjà, par exemple OrCID, DOI, IdHAL, Wosld, ArXivID, ISSN, Handle…

Nota Bene :

1) Les ensembles de données doivent se voir attribuer des identifiants persistants (PID). Pensez également à vous inscrire à ORCID[9], un service gratuit qui attribue des identifiants persistants aux individus / auteurs.
2) Pour faciliter la recherche de tous les résultats de recherche, des liens bidirectionnels doivent être créés entre les résultats liés aux publications, tels que les données (à l'aide des PID).
3) Incluez les métadonnées les plus riches possibles avec vos données déposées afin que d'autres puissent les trouver, comprendre les paramètres sous lesquels elles ont été créées et comprendre les conditions dans lesquelles elles peuvent y accéder et / ou les réutiliser.

b) Entrepôts de stockage

A la fin de votre projet, vous pouvez déposer vos données dans un entrepôt de stockage, disciplinaire ou général, notamment si les données ne sont pas rattachées à une publication particulière, ou si la taille ou le format des fichiers ne permet pas de les déposer dans une archive ouverte. Le répertoire re3data.org permet de rechercher parmi 1500 entrepôts de données, en fonction des types de fichiers à déposer, des disciplines concernées et des politiques de dépôt et de consultation. Vous pouvez également utiliser OAD, OpenDOAR, etc. Si vous cherchez un entrepôt certifié, consultez le site CoreTrustSeal.

[8] https://doranum.fr/identifiants-perennes-pid/fiche-synthetique/

[9] https://orcid.org/

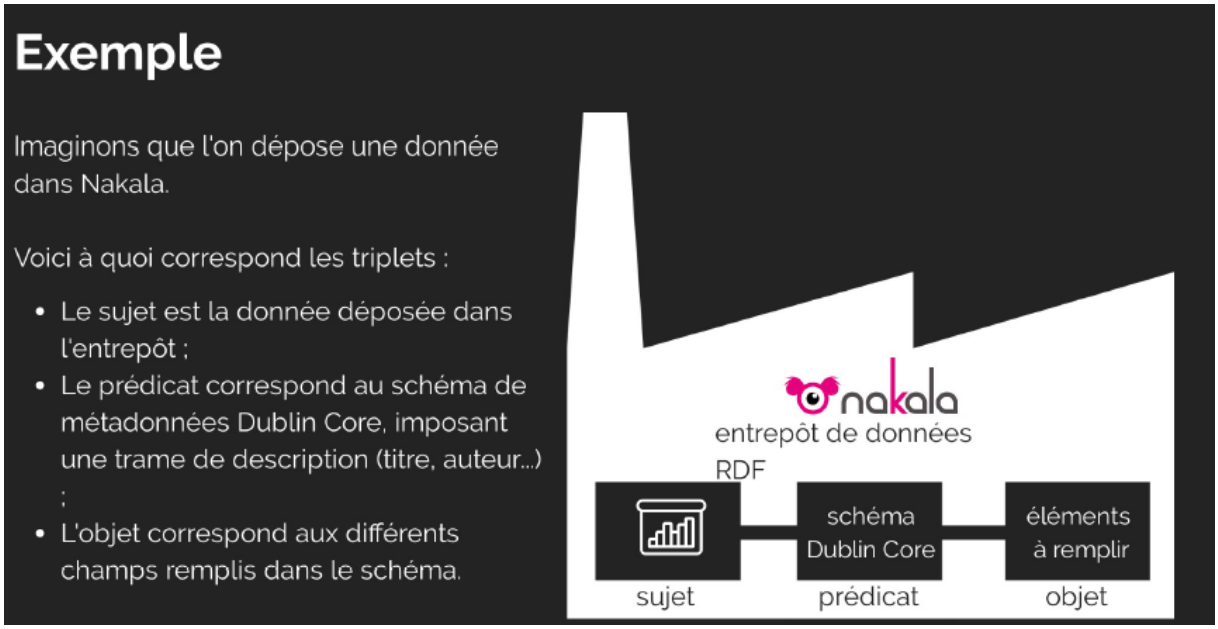
10

Par exemple, l'entrepôt NAKALA propose deux grands types de services : des services d'accès aux données elles-mêmes et des services de présentation des métadonnées. Les producteurs de données numériques ainsi soulagés de la gestion purement technique, peuvent ainsi se consacrer à la valorisation scientifique de leurs données. Les données hébergées par NAKALA peuvent être éditorialisées sur le web à l'aide du pack NAKALONA (associant Oméka et NAKALA) développé et géré par Huma-Num[11].

[10] Cette présentation est réadaptée de ANDS | The FAIR data principles. https://www.ands.org.au/working-with-data/fairdata/training#.XNqUWpLgeTA.link

https://view.genial.ly/5d64fbbd8352350fa3d22603/interactive-content-les-principes-fair

[11] https://www.huma-num.fr/services-et-outils/exposer



12

Il existe également ISODORE, moteur de recherche permettant l'accès aux données numériques des sciences humaines et sociales (SHS). ISIDORE offre également une boîte à outils (ISIDORE à la demande) qui permet, sur demande, d'enrichir, de catégoriser et d'annoter à l'aide de référentiels scientifiques des données de la recherche. ISIDORE est une réalisation de la très grande infrastructure de recherche Huma-Num (CNRS, Aix-Marseille Université, Campus Condorcet).[13].

Autre exemple, pour les sciences du vivant, l'entrepôt GnpIS est un système d'information intégratif multispécifique dédié aux ravageurs des plantes et des champignons. Il relie les données génétiques et génomiques, permettant aux chercheurs d'accéder à la fois aux informations génétiques et génomiques des plantes. GnpIS est utilisé à la fois par les grands projets internationaux et les départements de phytologie de l'Institut national français de la recherche pour l'agriculture, l'alimentation et l'environnement.[14]

---

[12] Cette présentation est réadaptée de ANDS | The FAIR data principles. https://www.ands.org.au/working-with-data/fairdata/training#.XNqUWpLgeTA.link

https://view.genial.ly/5d64fbbd8352350fa3d22603/interactive-content-les-principes-fair

[13] https://isidore.science/about https://isidore.science/

[14] https://urgi.versailles.inra.fr/gnpis

Si un entrepôt disciplinaire ne vous convient pas, vous pouvez utiliser un entrepôt de données généraliste, comme Zenodo . Zenodo vous permet de déposer n'importe quel type de fichiers, sans limitation de taille. Un identifiant DOI est attribué automatiquement aux données que vous y déposez pour faciliter leur citation. Ils permettent également de choisir les modalités de diffusion de vos différents fichiers, notamment par la mise en place d'embargo.

Nota Bene :

- Les chercheurs en sciences humaines sont encouragés à tirer parti des cadres, réseaux et ressources qui favorisent le potentiel de découverte et une réutilisation plus large de la recherche :
    - Registres de domaine, portails, moissonneuses, par ex. Re3data et FAIRsharing.org
    - Plateformes par exemple Europeana
    - Profils de chercheurs, par ex. ORCID
- Partagez en ligne vos données et tous les supports tels que les présentations, les affiches, blogs, des documents de données, etc. Tout en faisant attention à ne pas divulguer des informations valorisables
- Parlez de vos recherches en dehors du milieu universitaire, tenez compte de publics divers, tels que les journalistes, les décideurs politiques, les entreprises privées ou les citoyens scientifiques, car dans le cadre de l'Open Science, il est souhaitable qu'un public plus large puisse accéder aux résultats de voter recherche.
- Envisagez des canaux et des formats non traditionnels pour présenter vos données: infographie ou visualisations de données interactives, exposition en ligne ou visites numériques, sites Web ou applications...
- Encouragez et soutenez les approches pédagogiques qui incluent la production et la conservation de données de recherche.

3) Les publications

Si vous bénéficiez d'un financement ANR ou Européen, vos publications devront être accessibles soit en publiant directement dans une revue open access soit en déposant votre article dans une archive ouverte de type HAL.

Il existe plus de 10 000 revues scientifiques diffusées en libre accès et référencés dans plusieurs plateformes. Celles-ci proposent un moteur de recherche permettant de trouver du contenu scientifique.

- Directory of Open Access Journals (DOAJ) : Plateforme qui recense plus de 10 000 revues scientifiques disponibles en libre accès et provenant de plus de 130 pays. Cela représente plus de 2 millions d'articles.
- Persée : Portail présentant les collections rétrospectives de plus de 100 revues francophones : Annales, Bibliothèque de l'Ecole des chartes, l'Homme, Revue de l'art, Revue française de science politique, etc. Cela représente plus de 50000 articles librement accessibles.
- Revues.org : Plateforme qui propose plus de 400 revues dans les domaines de sciences humaines et sociales ; 95% des contenus sont accessibles librement, soit plus de 100000 articles. Revues.org fait partie du portail Openedition qui comprend OpendeditionBook (>2500 livres), Hypothèses (> 1300 carnets de recherche) et Calenda (agenda des événements scientifiques en SHS)
- Scientific Electronic Library Online (Scielo) : Plateforme présentant les revues scientifiques diffusées dans les pays d'Amérique du Sud. Cette plateforme permet de rechercher des revues scientifiques en Open Access

Nota Bene :

Les chercheurs cèdent souvent à leurs éditeurs un droit de diffusion exclusive sur leurs publications, ce qui les empêche de diffuser leur article dans une archive ouverte. Depuis la loi pour une République Numérique, un droit d'exploitation secondaire a été octroyé aux chercheurs dont la recherche a été financée au moins pour moitié par de l'argent public. Ce droit ne s'applique qu'aux écrits parus dans des publications périodiques éditées au moins une fois par an, c'est-à-dire, principalement aux articles publiés dans des revues scientifiques. Ce nouveau droit est valable pour les publications dont les droits de distribution ont été cédés à partir du 9 octobre 2016.

4) Les licences

Les chercheurs sont des « créateurs et des consommateurs » qui produisent et qui consomment des informations et les résultats ainsi que des recherches d'autres chercheurs. Pour cette raison, notre recommandation est d'éviter d'appliquer des restrictions légales qui ne respectent pas le principe d'ouverture. La réutilisation des données en général devrait être couverte par une licence aussi ouverte que possible.

Creative Commons (CC) propose des contrats-type ou licences pour la mise à disposition d'œuvres en ligne.

Ces licences s'adressent **aux auteurs** souhaitant :

- partager et faciliter l'utilisation de leur création par d'autres
- autoriser gratuitement la reproduction et la diffusion (sous certaines conditions)
- accorder plus de droits aux utilisateurs en complétant le droit d'auteur qui s'applique par défaut
- faire évoluer une œuvre et enrichir le patrimoine commun (les biens communs ou *Commons*)
- économiser les coûts de transaction
- légaliser le *peer to peer* de leurs œuvres.

Les licences Creative Commons sont fondées sur le droit d'auteur, alors que le régime du droit d'auteur classique vous incite à garder l'exclusivité sur la totalité de vos droits (« tous droits réservés »)

Le système CC propose quatre types de licences réglementant l'utilisation des œuvres (dérivé et aucun dérivé) en termes de copie, distribution, affichage, exécution et remixage par les titulaires de licence.

• CC BY (Attribution): L'œuvre peut être utilisée en donnant crédit aux auteurs.

• CC SA (Partage dans les mêmes conditions) : L'auteur autorise la reproduction, la diffusion et la modification de son œuvre, à condition que les utilisateurs publient toute adaptation sous les mêmes conditions que l'œuvre originale (sauf autorisation préalable).

• CC NC (Pas d'utilisation commerciale) : L'auteur autorise la reproduction, la diffusion et la modification de son œuvre, pour toute utilisation autre que commerciale, à moins que les utilisateurs obtiennent son autorisation au préalable.

• CC ND (Pas de modification) : L'auteur autorise la reproduction et la diffusion de l'œuvre originale uniquement. Si quelqu'un veut la modifier, il doit obtenir au préalable l'autorisation de l'auteur.

Vous ne pouvez attribuer une licence qu'à une œuvre dont vous êtes le titulaire des droits d'auteur. S'il y a des co-auteurs, vous devez être d'accord avec eux sur la licence. De plus, vous n'êtes pas autorisés à licencier des œuvres du domaine public. Vous devez également savoir s'il existe des exigences de licence de la part de l'organisme de financement ou du référentiel de données.

Nota Bene :

1) Déterminez le niveau nécessaire et suffisant de restrictions d'accès. Certaines données ne peuvent pas être partagées ouvertement mais peuvent toujours être partagées sous certaines restrictions tout en protégeant les données.
2) Utilisez des licences gratuites et standardisées : Afin de bénéficier de la possibilité de partager des données depuis le virage numérique et de favoriser l'Open Science, utilisez une licence la gratuite. L'Open Knowledge Foundation et l'Open Access Scholarly Publishers Association reconnaissent uniquement CC BY, CC BY-SA et CC0 comme compatibles avec Open Access. N'oubliez pas que la licence CC BY reflète un élément de bonne conduite scientifique établi de longue date : vous pouvez citer une œuvre, ou des parties de celle-ci, dans votre publication tant que vous indiquez correctement la source de votre citation, sinon c'est du plagiat.

Lexique

Articles de données : À la différence d'un article scientifique classique qui exploite, analyse et interprète les données scientifiques, un article de données décrit finement un/des jeu(x) de données de façon à en faciliter la compréhension et l'éventuelle réutilisation.
Administrateur des données (chief data officer - CDO) : Il coordonne l'action des acteurs en matière d'inventaire, de gouvernance, de production, de circulation et d'exploitation des données de la recherche. À ne pas confondre avec le Délégué à la protection des données.
Accord de consortium : contrat conclu entre les partenaires d'un projet pour préciser les modalités de leur collaboration.
Archive ouverte : plateforme où les documents déposés sont en accès ouvert. Elle peut être nationale (comme HAL2 – Hyper Article en Ligne), institutionnelle ou disciplinaire (comme arXiv3). Les chercheurs peuvent y déposer une notice décrivant leurs publications (titre, auteurs, résumé…) et éventuellement y associer un fichier selon les droits dont ils disposent.
2 https://hal.archives-ouvertes.fr/
3 https://arxiv.org/
4 https://www.ouvrirlascience.fr/presentation-du-comite/
CES : Comité d'Évaluation Scientifique. Il existe 48 CES, un par axe de recherche (35 axes de recherche disciplinaires et 13 axes transversaux). Les comités déterminent la liste des projets financés par l'ANR, en s'appuyant notamment sur les rapports remis par les experts scientifiques qui ont évalué les projets. Ces experts sont extérieurs aux CES.
Coordinateur du projet : personne responsable de la réalisation de la totalité d'un projet (à distinguer du Responsable scientifique).
CoSO : Comité pour la Science Ouverte4. Il impulse et soutient une politique nationale Science Ouverte.
CPP : Comité de protection des personnes. Il doit être sollicité pour avis en amont de toute recherche impliquant la personne humaine. Son avis est contraignant.
DataCite - Agence d'enregistrement des digital object identifier (DOI) pour les données de la recherche
Data paper : article décrivant un ou plusieurs jeux de données, notamment leur potentiel de réutilisation. Il peut être publié dans des revues spécifiques (*data journal*) ou dans des revues scientifiques traditionnelles.
DMP : *Data Management Plan*. Voir Plan de gestion des données (PGD).
DOAJ : *Directory of Open Access Journals*. Répertoire de revues en libre accès.
DOI : *Digital Object Identifier*. Identifiant unique attribué à des jeux de données ou à des publications.
Données de la recherche : ensemble des informations scientifiques produites ou collectées dans le cadre d'un projet de recherche, les données de la recherche peuvent être des photos, des mesures, des sons, etc. Elles sont nécessaires comme éléments probants afin de valider les résultats de la recherche et doivent être accompagnées d'informations qui les documentent, telles que des protocoles
DORA : San Francisco Declaration on Research Assessment5.
5 https://sfdora.org/
DPO : *Data Protection Officer*, ou DPD (Délégué à la Protection des Données).
Entrepôts : plateformes sur lesquelles sont déposés, décrits et conservés des jeux de données de la recherche. Les entrepôts sont généralistes ou disciplinaires.
EOSC – European open science cloud

FAIR : Principes d'ouverture des données, qui visent à les rendre Faciles à trouver (Findable), Accessibles (*Accessible*), Interopérables (*Interoperable*), Réutilisables (*Reusable*).
Frais de publication en libre accès : APC (Articles Processing Charges), BPC (Books Processing Charges), BCPC (Book Chapters Processing Charges).
GO FAIR – Initiative internationale visant à construire un environnement international de recherche enrichi par les données.
HAL – Archive ouverte nationale française portée par le Centre pour la communication scientifique directe (CCSD), unité mixte de service
I4OC – Initiative for Open Citations ISIDORE – Moteur de recherche sur les publications et les données des sciences humaines et sociales
Métadonnées : informations nécessaires à la description de données, en général structurées selon une norme.
OPERAS – Open access in the european research area through scholarly communication
ORCID : *Open Researcher and Contributor ID*. Identifiant unique pour les chercheurs.
OGP – Open government partnership, organisation regroupant 75 pays et des centaines d'organisations de la société civile pour la transparence de l'action publique.
***Open Access* / Accès ouvert** : mise à disposition immédiate, gratuite et permanente sur Internet des publications scientifiques issues de la recherche et de l'enseignement. On distingue plusieurs modèles ou voies de l'*Open Access* : la voie verte (*green Open Access*) et la voie dorée (*gold Open Access*).
OpenAIRE : consortium qui a pour objectif principal de soutenir le travail de recherche des scientifiques européens en créant et en exploitant une infrastructure d'accès ouvert. OpenAIRE a notamment un rôle d'agrégateur de productions de la recherche (publications, jeux de données, logiciels) et de mise en relation avec les projets financés par les financeurs européens de la recherche (H2020, ANR…).
PGD : Plan de Gestion des Données ou DMP (*Data Management Plan*). Document qui synthétise la description des données de recherche d'un projet et la manière dont elles seront gérées tout au long du projet, afin, notamment, de préparer leur partage, leur réutilisation et leur pérennisation.
Revue hybride : revue diffusée par abonnement mais dont certains articles peuvent être librement accessibles au lecteur (*Open Access*), moyennant le paiement de frais de publication en libre accès. La majorité des revues d'éditeurs comme Elsevier ou Springer sont des revues hybrides.
Revue *Open Access* : revue dont les articles sont immédiatement accessibles au public.
RSSI : Responsable de la sécurité des systèmes d'information.
ScanR – Moteur de la recherche et de l'innovation
Version acceptée pour publication (*post-print* ou *author accepted manuscript*) : version comportant les révisions issues du processus d'évaluation par un comité de lecture (*peer-reviewing)*. Fichier sans mise en page éditeur ou avec une mise en page partielle.
Version éditeur (*version of record, final version*) : article avec la mise en page finale. Version diffusée par l'éditeur.
Version soumise pour publication (*pre-print* ou *submitted manuscript*) : version envoyée par les auteurs à une revue, avant le processus de révision par les pairs.
Voie dorée (*gold Open Access*) : la voie dorée qualifie la publication d'articles dans des revues où les articles sont accessibles au lecteur en libre accès, c'est-à-dire sans barrières. Il existe plusieurs modèles économiques à l'intérieur de la « voie dorée ». **Voie verte (*green Open Access*)** : la voie verte qualifie le dépôt et la diffusion des publications dans une archive ouverte, par un auteur ou par une personne tierce.

Sources

Allea : *All European Academies Février 2020*

Guide CNRS *Traçabilité des activités de recherche et gestion des connaissances*

Esther Dzale Yeumo : *Les principes FAIR*, UAR DIST Délégation Information Scientifique et Technique, Versailles, Inra, France

Plan national pour la science ouverte, 4 juillet 2018

https://doranum.fr
https://lilliad.univ-lille.fr/chercheur/open-access
https://openaccess.couperin.org/
https://openscience.pasteur.fr/
https://www.ouvrirlascience.fr/
https://view.genial.ly/5d64fbbd8352350fa3d22603/interactive-content-les-principes-fair