

Réaliser un plan de gestion de données “ FAIR ” : modèle

Nathalie Reymonet, Magalie Moysan, Aurore Cartier, Renaud Délémontez

► To cite this version:

Nathalie Reymonet, Magalie Moysan, Aurore Cartier, Renaud Délémontez. Réaliser un plan de gestion de données “ FAIR ” : modèle . 2018. sic_01690547v2

HAL Id: sic_01690547

https://archivesic.ccsd.cnrs.fr/sic_01690547v2

Submitted on 15 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial| 4.0 International License

Réaliser un plan de gestion de données « FAIR » : guide de rédaction

Ce document a pour vocation d'accompagner les chercheurs et chargés de projets dans la rédaction de plans de gestion de données de la recherche (*Data Management Plans, DMP*). La structure proposée tient compte du cycle de vie des données, de leur création dans le cadre d'un projet de recherche jusqu'à leur archivage. Elle s'appuie également sur divers modèles de plans de gestion de données tels que celui de la *National Science Foundation (NSF)* ou de l'*Interuniversity Consortium for Political and Social Research (ICPSR)*.

Ce guide intègre les attendus de la Commission européenne dans le cadre d'Horizon 2020 et du *FAIR data management* : les champs requis par la Commission européenne sont signalés par un drapeau dans le texte.

Ce document est composé d'une partie introductive aux notions de données de la recherche et d'accès ouvert (*open access, open research data*). Une deuxième partie constitue le guide proprement dit, composé d'un tableur qui permet de réaliser la saisie d'un plan de gestion de façon guidée. Il détaille champ par champ les attendus, outils et contacts pour la description et la gestion des données. Les exemples mentionnés dans ces documents sont issus de guides existants.

Vous pouvez également réaliser un plan de gestion de données selon ce modèle sur l'outil OPIDOR : <https://dmp.opidor.fr/>



V2 (2018)

TABLE DES MATIERES

LES DONNEES DE LA RECHERCHE : DEFINITIONS	3
INTRODUCTION AU DATA MANAGEMENT PLAN (DMP)	4
<i>Open Access et Data Management Plan dans les appels À projets Horizon 2020</i>	<i>5</i>
<i>Open access</i>	<i>5</i>
<i>Open access des publications</i>	<i>5</i>
<i>Open access des données de la recherche</i>	<i>5</i>
○ Le périmètre thématique de l'Open research data pilot	5
○ Les conditions de l'Open research data pilot	6
○ Les attendus de la description FAIR des données	7
Les étapes de rédaction et de validation du DMP dans le cadre d'Horizon 2020	8
DMP : acteurs et services ressources	9
Références bibliographiques	10
Contacts Sorbonne Paris Cité	11
STRUCTURE DU DMP	12
Versioning et mises à jour	12
GUIDE DE SAISIE	16
PHASE 1 – Contexte, acteurs et ressources	17
Section 1 : Description du projet	17
Section 2 : Acteurs du projet	18
Section 3 : Ressources nécessaires	18
PHASE 2 – Stockage, partage, protection et diffusion au cours du projet	20
Section 4 : Informations générales sur les données	20
Section 5 : Stockage et partage des données au cours du projet	23
Section 6 : Risques, sécurité et éthique des données	24
Section 7 : Dissémination et archivage des données	26
PHASE 3 – Dissémination et archivage à l'issue du projet	27
Section 8 : Identification des jeux de données	27
Section 9 : Protection – exception de diffusion	27
Sections 10 et 11 : Description et diffusion des jeux de données	28
Section 12 : Tri et archivage des données	32

Conception et réalisation : Aurore CARTIER (Service Commun de la Documentation, Université Paris Descartes), Renaud DELEMONTEZ (stage ENSSIB), Magalie MOYSAN (Bureau des archives, Université Paris Diderot) et Nathalie REYMONET (Direction de la Recherche (DRIVE), Université Paris Diderot)

LES DONNEES DE LA RECHERCHE : DEFINITIONS

Les données de la recherche recouvrent un périmètre large, dont la définition varie selon les approches ou les contextes, donnant lieu à une pluralité de définitions dont voici quelques exemples.

- **Définition de l'OCDE**

Les données de la recherche sont des enregistrements factuels (chiffres, textes, images et sons) utilisés comme sources principales pour la recherche scientifique et généralement reconnus par la communauté scientifique comme nécessaires à la validation des résultats de recherche¹.

- **Définition de la section « Aurore » de l'Association des archivistes français**

Les données de la recherche sont des informations, spécimens et matériaux produits, recueillis et documentés. Elles sont collectées ou exploitées à des fins de recherche et de preuve par les chercheurs et leurs équipes. À ce titre, elles constituent une partie des archives de la recherche².

- **Introduction des données de la recherche dans le Code de la recherche**

La loi sur l'enseignement supérieur et la recherche de juillet 2013 modifie l'article L 112-1 du Code de la recherche en introduisant la notion d'« organisation de l'accès libre aux données scientifiques³ ».

- **Loi pour une République Numérique**

L'article 38 de la loi mentionne : « ... fichiers produits au terme des activités de recherche pour lesquelles elles ont été produites ; ces fichiers constituent des données de la recherche⁴ ».

- **Définition de la Commission européenne pour Horizon 2020⁵**

1. the 'underlying data' (the data needed to validate the results presented in scientific publications), including the associated metadata (i.e. metadata describing the research data deposited), as soon as possible

2. any other data (for instance curated data not directly attributable to a publication, or raw data), including the associated metadata, as specified and within the deadlines laid down in the DMP – that is, according to the individual judgement by each project/grantee.

La Commission européenne précise⁶ "In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form."

En conséquence, le présent document considère les données de la recherche comme un ensemble d'informations factuelles enregistrées sur des supports, produites ou collectées, selon divers procédés au cours d'un processus de recherche.

¹ OCDE. *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics* (2007)

² « Archivistes ! n°120 », disponible sur www.archivistes.org, 30 janvier 2017 (consulté le 19 mars 2017)

³ Loi n° 2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche, NOR: ESRJ1304228L

⁴ Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique, NOR: ECFI1524250L

⁵ European Commission. *H2020 Program Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020* (Version 3.1, 25 August 2016), p.9

⁶ http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

INTRODUCTION AU DATA MANAGEMENT PLAN (DMP)

Un *Data Management Plan* (DMP) ou plan de gestion de données est un document formalisé explicitant la manière dont seront obtenues, documentées, analysées, disséminées et archivées les données produites au cours et à l'issue d'un processus ou d'un projet de recherche. Ce guide est conçu pour répondre aux exigences des financeurs, mais il est aussi un outil pour gérer les données tout au long du projet en intégrant la notion de cycle de vie. Le DMP s'étend donc de la production (ou la collecte) des données à leur diffusion et/ou leur archivage, en passant par leur stockage, leur traitement/curation, leur analyse et leur description.

Le DMP s'appuie sur le cycle de vie des données/documents et décrit les choix réalisés en termes de métadonnées⁷, formats des bases de données, méthodes et sécurité d'accès, durées d'archivage, ainsi que les coûts associés à la gestion des données. Une mention particulière doit être apportée aux données venant en appui des publications et qui doivent à ce titre rester disponibles et accessibles au plus grand nombre⁸.

L'établissement d'un *data management plan* est de plus en plus demandé dans les appels à projets financés sur fonds publics, notamment européens. Dans son guide portant sur le *FAIR data management*, la Commission européenne précise : "*Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/ or generated by a Horizon 2020 project*"⁹.

L'objectif est ainsi de documenter la manière dont les données seront produites ou collectées au cours et à l'issue d'un processus de recherche, en s'attachant notamment à définir comment elles seront décrites, partagées, protégées puis conservées. La gestion des données n'est pas une fin en soi, mais le moyen de conduire à la découverte de connaissances et d'innovations par l'intégration et la réutilisation des connaissances produites.

La vocation d'un DMP n'est pas de contraindre la pratique par l'exigence de nouveaux éléments mais plutôt d'inciter à la **réflexion** ainsi qu'à l'**anticipation** et de **formaliser** au sein d'un **document unique, prospectif, descriptif et évolutif**, un ensemble d'éléments et d'informations, auparavant dispersés entre divers acteurs et divers documents, éléments utiles au suivi du projet et à la bonne gestion des résultats obtenus.

⁷ Métadonnées : les métadonnées sont un ensemble de données structurées décrivant des ressources physiques ou numériques. Elles sont classiquement divisées en métadonnées descriptives, administratives ou de structure (<http://openaccess.inist.fr>).

⁸ CNRS, DIST. *Schéma d'orientation stratégique de l'information scientifique et technique* (Nov. 2013)

⁹ European Commission. *H2020 Program Guidelines on FAIR Data Management in Horizon 2020* (Version 3.0, 26 July 2016), p.4

OPEN ACCESS ET DATA MANAGEMENT PLAN DANS LES APPELS À PROJETS HORIZON 2020

Open access

L'*open access* désigne un accès en ligne à l'information scientifique (publications et données), qui soit gratuit. Cet accès ouvert doit permettre la réutilisation des informations¹⁰.

Open access des publications

L'*open access* des publications désigne un accès gratuit, immédiat, permanent, en ligne, aux travaux publiés dans les éditions à comité de lecture¹¹.

Avec son programme cadre « Horizon 2020 », la Commission européenne fait de l'*open access* un principe général pour les publications scientifiques : elle rend obligatoire la publication en *open access* « en respect des intérêts légitimes des participants ». Il s'agit donc de rendre les publications scientifiques accessibles sans barrière financière, juridique ou technique. Les bénéficiaires s'engagent ainsi à respecter l'article « 29.2-Open access to scientific publications » de la convention de subvention¹². Ils disposent toujours de la possibilité de publier leurs résultats de recherche auprès de revues à comité de lecture, mais le cas échéant, doivent le faire en *open access* selon différentes modalités : soit directement dans des revues en *open access* (*gold open access*) soit en déposant, parallèlement ou sous réserve d'embargo, leurs publications dans un serveur d'archives ouvertes (*green open access*)¹³.

Open access des données de la recherche

La libre diffusion participe d'un cercle vertueux visant à améliorer la qualité des données, réduire la duplication des efforts de recherche, accélérer l'innovation et contribuer à la lutte contre la fraude scientifique¹⁴.

○ Le périmètre thématique de l'*Open research data pilot*

La Commission européenne a mis en place en 2014 un dispositif test intitulé "Open research data pilot" pour encourager et optimiser la gestion et la réutilisation des données de recherche générées par les projets qu'elle finance. Depuis le programme de travail 2017 de Horizon 2020, l'*Open research data pilot* est étendu à tous les projets financés et **l'accès libre devient le paramètre par défaut pour les données de recherche générées**¹⁵. Toutefois, le porteur peut se retirer du pilote, à trois moments-clés du projet :

- pendant la phase de candidature,
- pendant la préparation de la convention de financement (*grant agreement*),
- après la signature de la convention de financement.

¹⁰ European Commission. *H2020 Program Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020* (Version 3.1, 25 August 2016), p.3

¹¹ *Open access* : Budapest Open Access Initiative (2001), <http://www.opensocietyfoundations.org/openaccess>

¹² "Each beneficiary must ensure open access (free of charge, online access for any user) to all peer-reviewed scientific publications relating to its results". Commission européenne. *H2020 Programme AGA – Annotated Model Grant Agreement* (Version 2.2, 25 November 2016)

¹³ *Guidelines on Open Access* (2016), p.6

¹⁴ Commission européenne. *Recommandation de la commission du 17.7.2012 relative à l'accès aux informations scientifiques et à leur conservation* (17.7.2012) C(2012) 4890 final

¹⁵ *Guidelines on Open Access* (2016), p.8

Les projets faisant le choix de sortir du pilote sont toutefois encouragés à produire un plan de gestion de données sur une base volontaire : produire un DMP n'entraîne pas automatiquement l'ouverture des données, selon le principe "*as open as possible, as closed as necessary*".

Il faut noter que le choix de sortir du pilote n'a pas d'implication sur l'évaluation du projet¹⁶.

Dans le cas particulier des ERC, deux situations¹⁷ :

- ERC Work Programme 2016 : le plan de gestion des données n'est pas un livrable. Les bénéficiaires peuvent intégrer le pilote sur la base du volontariat, en se soumettant de ce fait à l'article 29.3 de la convention de financement¹⁸ (cf. ci-dessous).
- ERC Work Programme 2017 : par défaut, l'article 29.3 s'applique et le bénéficiaire doit produire un DMP. Il peut toutefois sortir du pilote *Open research data*.

Il existe un *template* pour le DMP des ERC, basé sur les *guidelines* de la Commission européenne¹⁷.

○ Les conditions de l'*Open research data pilot*

Les bénéficiaires de financements européens s'engagent à :

- ✓ **accepter l'article « 29.3-Open access to research data » de la convention de subvention** les engageant à déposer les données (et métadonnées correspondantes) produites au cours du projet dans un réservoir ouvert de données (*research data repositories*)¹⁹. Toutefois, en cas de sortie du dispositif, l'article 29.3 peut être retiré de la convention²⁰.
- ✓ **produire dans les 6 premiers mois du projet de recherche un plan de gestion** qui est appelé à être mis à jour tout au long du projet : des versions plus élaborées du DMP pourront être livrées au cours du projet, en cohérence avec les évaluations du projet (*periodic review*) ainsi que lors du **rapport final**²¹.

Des exceptions à l'article 29.3 sont prévues, notamment si les données produites sont des données sensibles (données personnelles, données devant faire l'objet d'une exploitation industrielle, etc.). Ces exceptions sont négociées avec la Commission européenne au moment de la rédaction de la convention de subvention (*grant agreement*).

Attention : les projets qui ne donnent pas lieu à diffusion des données doivent tout de même comporter un plan de gestion. Celui-ci concerne la gestion de l'ensemble des données, y compris celles qui n'ont pas vocation à être diffusées.

Pour atteindre son objectif d'ouverture des données, la Commission européenne met des **moyens à la disposition des équipes de recherche** :

- éligibilité au remboursement des frais relatifs à l'ouverture des données pendant la durée du projet²² ;
- mise à disposition d'un réservoir ouvert de données [Zenodo](https://zenodo.org/) ;

¹⁶ European Commission. *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020* (Version 3.0, 26 July 2016), p.4

¹⁷ European Research Council. *ERC data management plan, template ; ERC Guidelines on the implementation of open access to scientific publications and research data in projects supported by the European Research Council under Horizon 2020*. Version 1.1 (7 avril 2017)

¹⁸ A l'exception des *ERC proof of concept*, qui sont exclus du dispositif *Open research data pilot*

¹⁹ *Annotated Model Grant Agreement* (2016), p.222

²⁰ *Guidelines on Open Access* (2016), p.8

²¹ *Guidelines on FAIR Data Management* (2016), p.5

²² "*costs related to open access to research data and related costs, such as data maintenance or storage.*" *Annotated Model Grant Agreement* (2016), p.88

- promotion des licences libres permettant l'exploitation des données et leur réutilisation. Pour faciliter le choix de la licence, des outils comme le [License selector](#) sont disponibles ;
- promotion de l'outil numérique d'aide à la rédaction des plans de gestion [DMP Online](#) ;
- outil en ligne d'aide à la rédaction de plan de gestion [EUDAT B2share](#) (en cours).
- projet d'[Open Science Cloud](#) européen garantissant un environnement sécurisé.

Par ailleurs, l'Inist met à disposition l'outil [DMP Opidor](#) qui repose sur le DMP Roadmap, développé en collaboration entre le *Digital Curation Center* et l'Université de Californie à partir de leurs outils respectifs : [DMP Tool](#) et [DMP Online](#).

La trame proposée dans ce document est également intégrée dans [DMP Opidor](#).

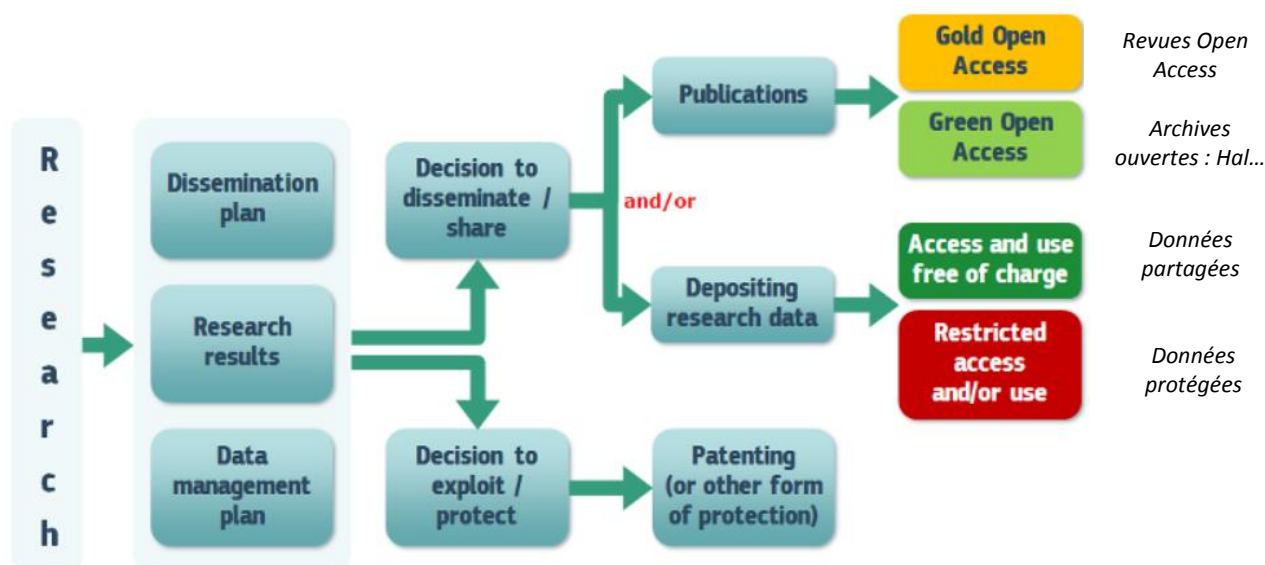
○ Les attendus de la description FAIR des données

La Commission européenne, se reposant sur des travaux du collectif FORCE11²³, propose de décrire les données de façon qu'elles soient « FAIR » : *Findable, Accessible, Interoperable and Re-usable*. Le plan de gestion doit indiquer en particulier :

- le traitement des données de recherche pendant et après la fin du projet
- les données qui seront collectées, traitées et/ou générées
- la méthodologie et les normes appliquées
- si les données seront partagées / rendues accessibles
- comment les données seront organisées et conservées (y compris après la fin du projet).

Open access in context: dissemination & exploitation of research results

Le schéma ci-dessous présente l'*open access* des publications et des données de la recherche dans le contexte de la dissémination et de l'exploitation des résultats²⁴.



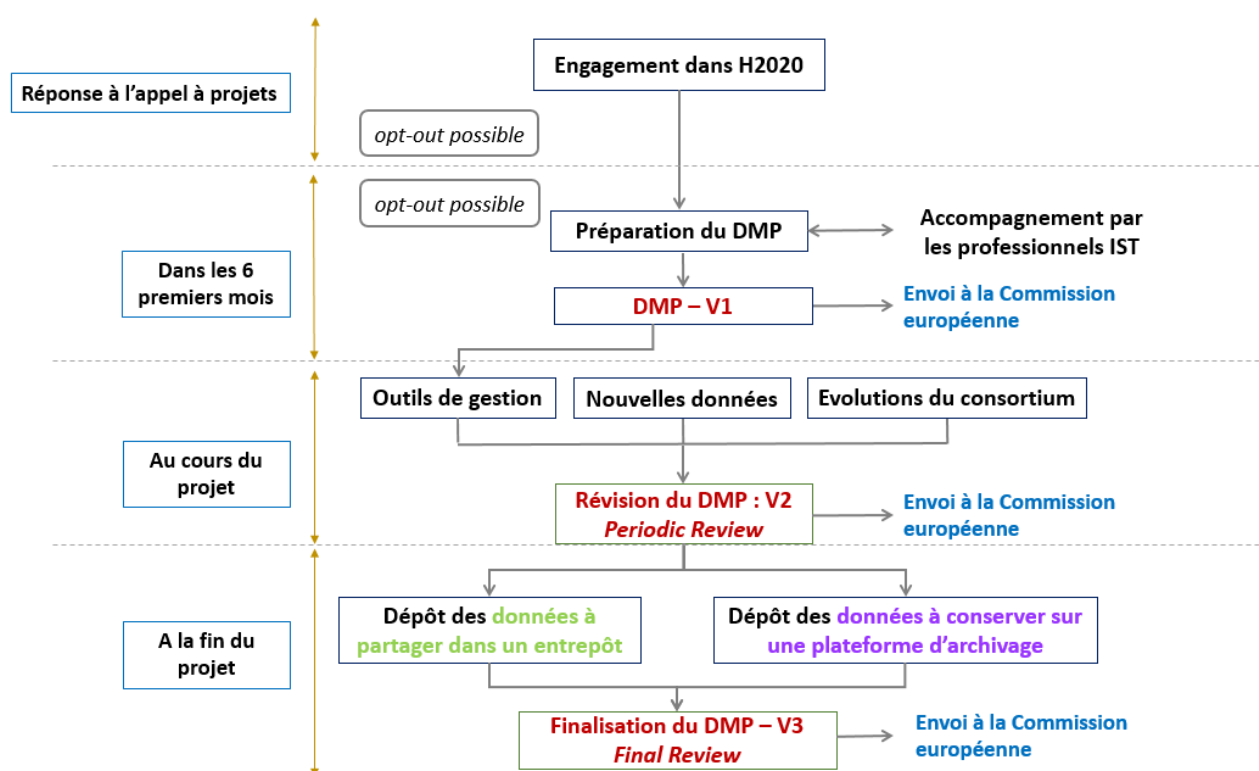
²³ FORCE11 est une communauté d'académiques, de bibliothécaires, d'archivistes, d'éditeurs et de bailleurs de fonds de la recherche organisés afin de faciliter une meilleure création et partage des connaissances <https://www.force11.org/group/fairgroup/fairprinciples>.

²⁴ http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

LES ETAPES DE REDACTION ET DE VALIDATION DU DMP DANS LE CADRE D'HORIZON 2020

Le DMP est un livrable requis par la Commission européenne, s'inscrivant dans un calendrier précis. Sa rédaction et sa mise en œuvre reviennent à l'équipe de recherche lauréate, aidée des ingénieurs-projets européens et des professionnels de l'information scientifique et technique (IST). La mise en place d'un plan de gestion de données est obligatoire depuis le *work program* 2017, en dehors d'exceptions justifiées. Une version initiale du plan de gestion de données doit être remise dans les six premiers mois suivant l'acceptation du projet. Le DMP peut alors être un document succinct d'une à deux pages mais devra être mis à jour au cours, puis à l'issue du projet.

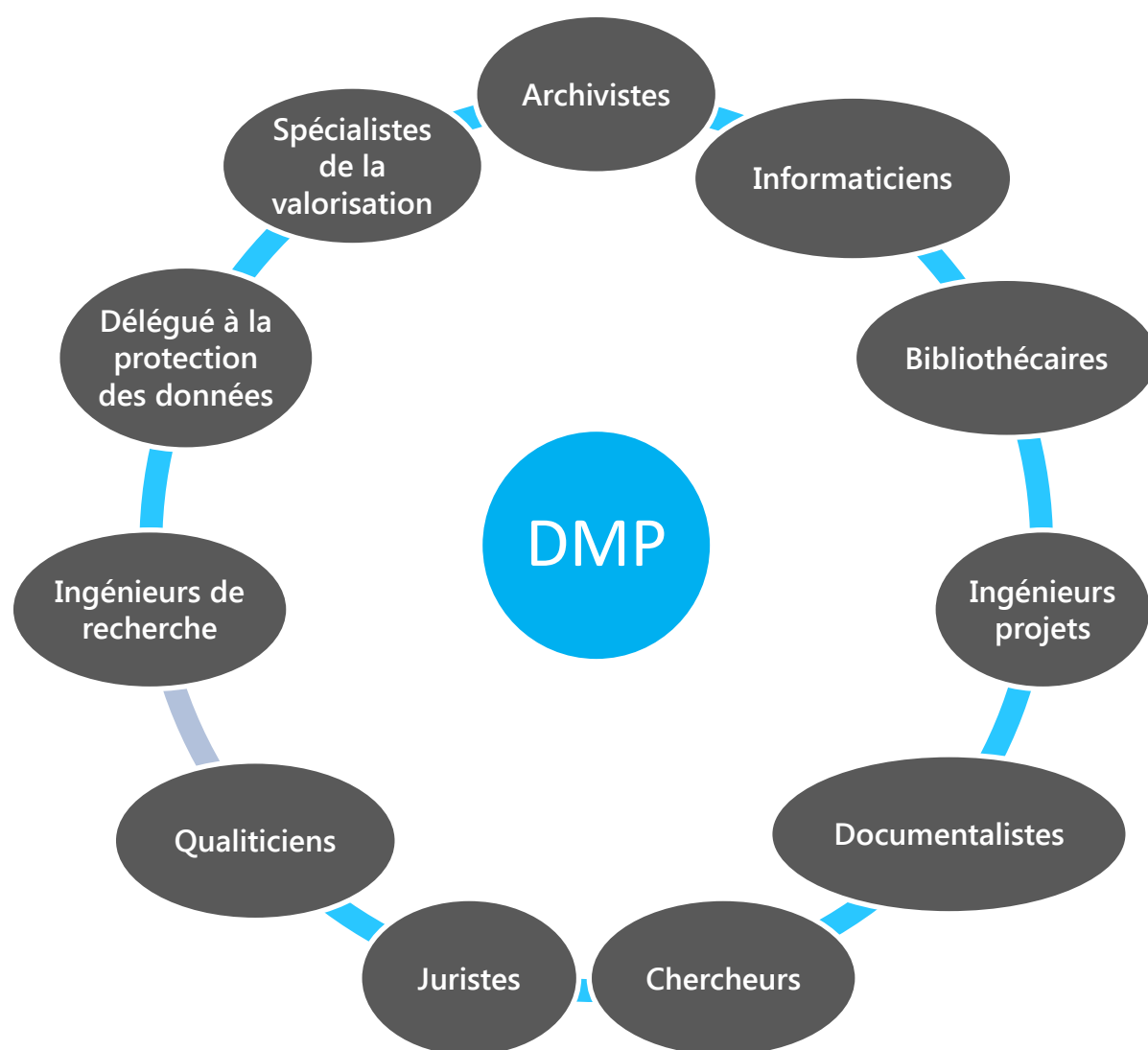
Éléments de calendrier pour la réalisation d'un *Data Management Plan* dans H2020



A. Cartier, M. Moysan, N. Reymonet, d'après les *Guidelines on FAIR data management*

DMP : ACTEURS ET SERVICES RESSOURCES

La rédaction de *data management plans* est une pratique née dans les pays anglo-saxons qu'il nous semble important d'adapter aux structures et acteurs français pour la rendre applicable. Si la rédaction d'un DMP relève davantage de la compilation et de la formalisation d'informations préexistantes que d'une démarche totalement inédite, force est de reconnaître qu'elle reste relativement coûteuse en temps pour les équipes de recherche. C'est pourquoi, à travers ce guide, notre démarche est non seulement de proposer une liste de champs applicables mais également d'identifier au sein de nos établissements d'enseignement supérieur et de recherche les différents **acteurs susceptibles d'accompagner les chercheurs** dans la rédaction de leur DMP, ainsi que les **outils disponibles** dans la communauté académique. La rédaction d'un DMP fait en effet appel à diverses compétences bien souvent présentes au sein de nos universités mais généralement disséminées entre plusieurs acteurs, voire plusieurs services qui n'étaient pas, jusqu'ici, nécessairement habitués à travailler ensemble.



Références bibliographiques

Agence des droits fondamentaux de l'Union européenne. *Manuel de droit européen en matière de protection des données*, 2014.

<http://fra.europa.eu/fr/publication/2014/manuel-de-droit-europeen-en-matiere-de-protection-des-donnees>

Agence nationale de la sécurité des systèmes d'information. *Expression des besoins et identification des objectifs de sécurité (EBIOS®)*, http://www.securite-informatique.gouv.fr/gp_article82.html

Budapest Open Access Initiative (2001) : <http://www.opensocietyfoundations.org/openaccess>

CNRS, DIST. *Schéma d'orientation stratégique de l'information scientifique et technique* (nov. 2013)

<http://www.cnrs.fr/dist/docs/131119-orientation.pdf>

Commission européenne. *Recommandation de la commission du 17.7.2012 relative à l'accès aux informations scientifiques et à leur conservation* (17.7.2012) C(2012) 4890 final

http://medoanet.sciencesconf.org/conference/medoanet/pages/recommendation_access_and_preservation_scientific_information_fr_copie.pdf

European Commission. *H2020 Programme Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020* (Version 3.1, 25 August 2016)

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

European Commission. *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020* (Version 3.0, 26 July 2016) http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Commission. *H2020 Programme AGA – Annotated Model Grant Agreement* (Version 2.2, 25 November 2016) http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

European Research Council. *ERC Guidelines on the implementation of open access to scientific publications and research data in projects supported by the European Research Council under Horizon 2020*. Version 1.1 (7 avril 2017)

European Research Council. *ERC data management plan, template*

EDUCAUSE ACTI Data Management Working Group. *Developing an institutional research data management plan service* (jan. 2013) <https://net.educause.edu/ir/library/pdf/ACTI1301.pdf>

EUDAT. *Data access and management in the EUDAT collaborative data infrastructure* (2014) <http://www.eudat.eu/system/files/Open-Access-and-Data.pdf>

ICPSR. *Guidelines for Effective Data Management Plans* <http://www.icpsr.umich.edu/files/datamanagement/DataManagementPlans-All.pdf>

Institut national de recherche agronomique. *Rapport du groupe de travail sur la gestion et le partage des données* (juin 2012) http://www.pfl-cepia.inra.fr/uploads/gdp_docs/Rapport-GestionDonnees-web.pdf

NSF. *Data Management Plan Requirements*. <http://www.nsf.gov/eng/general/dmp.jsp>

OCDE. *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics* (2007) <http://www.oecd.org/fr/science/sci-tech/38500823.pdf>

République française. *LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique*, NOR: ECFI1524250L <https://www.legifrance.gouv.fr/eli/loi/2016/10/7/ECFI1524250L/jo/texte>

République française. *LOI n° 2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche*, NOR: ESRJ1304228L <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000027735009>

UC San Diego. *Data management plan* de Laurie Smith <http://rci.ucsd.edu/files/DMP%20Example%20Laurie%20Smith.pdf>

UC San Diego. *Data management plan* d'Elsa Cleland <http://idi.ucsd.edu/files/DMP%20Example%20Cleland>

Contacts Sorbonne Paris Cité

- Aurore CARTIER aurore.cartier@parisdescartes.fr
- Magalie MOYSAN magalie.moysan@univ-paris-diderot.fr
- Nathalie REYMONET nathalie.reymonet@univ-paris-diderot.fr

STRUCTURE DU DMP

Cette trame a pour objectif de recenser toutes les informations nécessaires à la gestion des données, de leur production à leur archivage. Elle est également disponible dans une version numérique sur Opidor accompagnée des éléments d'accompagnement indiqués plus loin dans ce document.

Pour accéder à ce modèle de DMP sur [Opidor](#) vous devez :

1. Ouvrir un compte sur la plateforme (gratuit)
2. Vous affilier à une des organisations suivantes : *Université de Paris 5 – Paris Descartes* ou *Université de Paris 7 – Paris Diderot*
3. Vous pourrez ensuite créer votre plan. Pour sélectionner le présent modèle, vous devez cliquer, à la question « If applying for funding, select your research funder », sur la mention “Not applicable/not listed” présente sous le menu déroulant, *y compris s’il s’agit d’un DMP intervenant dans le cadre d’un projet H2020.*
4. A la question “To see institutional questions and/or guidance, select your organisation”, cliquez ensuite sur *Université de Paris 5 – Paris Descartes* ou *Université de Paris 7 – Paris Diderot*
5. Cliquez ensuite sur « Create plan » et confirmez dans pop-up qui s’affiche.

Versioning et mises à jour

La trame se compose de trois versions, à renseigner en fonction de l’avancement du projet : certaines rubriques peuvent ne pas être complétées dès le début du projet et sont donc facultatives dans les deux premières versions (ou phases).

DMP – Phase 1 : Contexte, acteurs et ressources

Nom de la section	Délais de Màj			
	Dans les 6 premiers mois	Régulièrement tout au long du projet	Pour l'évaluation finale	
1/ Description du projet - 8 questions - Indiquez l'identifiant interne du projet (facultatif) ? - Quel est l'organisme financeur du projet ? - Dans quel programme de recherche s'inscrit le projet ? - Indiquez la référence de la convention de financement. - Indiquez l'acronyme ou le titre du projet. - Rappelez les objectifs et les attendus du projet.	<u>nécessaire</u>	<u>nécessaire</u>	<u>nécessaire</u>	

- Indiquez les thèmes ou mots-clés décrivant le projet. - Indiquez les références et la localisation de la documentation associée.			
2/ Acteurs du projet - 3 questions - Qui est le coordinateur scientifique du projet ? - Qui sont les partenaires du projet ? - Qui sont le/les responsable(s) de la gestion des données et de la mise à jour du DMP au cours du projet ?	<u>nécessaire</u>	<u>nécessaire</u>	<u>Nécessaire</u>
3/ Ressources nécessaires – 4 questions - La gestion des données au cours du projet (de la production à la diffusion) nécessite-t-elle l'acquisition ou l'installation de matériel spécifique (serveurs, machines virtuelles, infrastructures, etc.) ? Si oui, précisez lesquels ? - La gestion des données au cours du projet (de la production à la diffusion) nécessite-t-elle des recrutements ou des actions de formations ciblées ? Si oui, précisez lesquels ? - Sur la base des besoins identifiés ci-dessus, à combien estimez-vous le surcoût attaché à la gestion des données ? à leur diffusion ? à leur archivage ? - Comment seront couverts les besoins financiers ?	recommandé	optionnel	<u>nécessaire</u>

DMP – Phase 2 : Stockage, partage, protection et diffusion au cours du projet

Nom de la section	Délais de Màj	Dans les 6 premiers mois	Régulièrement tout au long du projet	Pour l'évaluation finale
4/ Informations générales sur les données - 6 questions - Quels sont les méthodes et processus de production des données utilisés au cours du projet ? - Le projet s'appuie-t-il sur la réutilisation de données existantes ? Si oui, lesquelles et de quelle origine ? - Comment se définit la propriété des données produites et collectées au cours du projet entre les différents partenaires ? - Précisez le type et la nature des données qui seront produites au cours du projet. - Quels seront les principaux formats de production, de traitement ou de conversion des données utilisés au cours du projet ainsi que les outils et logiciels de lecture associés ?		<u>nécessaire</u>	recommandé	<u>nécessaire</u>

- Avez-vous identifié des standards et/ou des formats de métadonnées permettant de décrire vos données ?			
5/ Stockage et partage des données - 5 questions - Sur quels supports les données sont-elles produites ou échangées au cours du projet ? - Quels seront les lieux et les modalités de stockage et d'hébergement des données du projet ? - De quelle manière seront gérés l'accès et le partage des données entre les partenaires au cours du projet ? - Indiquez le volume prévisionnel de données produites au cours du projet ? - Indiquez les règles de nommage et de classement des informations et données produites sur les espaces de partage du projet.	recommandé	recommandé	<u>nécessaire</u>
6/ Risques, sécurité et éthique des données - 3 questions - Quels sont les risques et menaces qui pèsent sur les données ? - Quelles seront les mesures et les solutions mises en œuvre pour parer les risques identifiés ? Quels sont les précautions et les moyens prévus pour garantir la protection des données, notamment en matière de confidentialité ? - Le projet pose-t-il des questions éthiques ? Quelles sont les dispositions prises pour y répondre ?	recommandé	recommandé	<u>nécessaire</u>
7/ Dissémination et archivage - 4 questions - Quel potentiel de réutilisation verriez-vous pour ces données en dehors de votre projet ? - Les données produites au cours du projet ont-elles vocation à être publiées ? Sont-elles soumises à un principe général de diffusion édicté par le financeur, un programme, un engagement etc. ? - Quelles sont les modalités de diffusion envisagées pour la dissémination des données et jeux de données ? - Quelles sont les modalités et garanties d'archivage envisagées pour vos données à l'issue du projet (plateformes, partenaires, durée de conservation etc.)	optionnel	recommandé	<u>nécessaire</u>

DMP – Phase 3 : Dissémination et archivage à l'issue du projet

Nom de la section	Délais de Màj	Dans les 6 premiers mois	Régulièrement tout au long du projet	Pour l'évaluation finale
8/ Identification des jeux de données - 2 questions - Combien de jeux de données destinés à être archivés et/ou diffusés comporte votre projet ? - Existe-t-il des liens ou des relations spécifiques entre certains jeux de données ? Si oui, précisez lesquels (parenté, complémentarité, etc.) ?		optionnel	recommandé	<u>nécessaire</u>
9 / Protection – Exception diffusion – 1 question - Indiquez les raisons pour lesquelles certains jeux de données (préciser lesquels) pourraient ne pas être diffusés ?		optionnel	recommandé	<u>nécessaire</u>
10/ Choix pour la description et la diffusion des jeux de données techniquement homogènes – section répliquable, 11 questions - Intitulé du jeu de données - Identifiant du jeu de données - Auteurs et contributeurs - Editeur (institution) - Mots-clés, vocabulaires et ontologies (description des contenus) - Nature ou type de données - Format et outils de lecture associés - Standards et formats de métadonnées - Modalités de diffusion et localisation du jeu de données - Durée de l'embargo (si nécessaire) - Type de licence		optionnel	optionnel	<u>nécessaire</u>
11/ Choix pour la description et la diffusion des jeux de données techniquement hétérogènes (cohérence intellectuelle) – section répliquable, 11 questions - Intitulé du jeu de données - Identifiant du jeu de données - Auteurs et contributeurs - Editeur (institution) - Mots-clés, vocabulaires et ontologies (description des contenus) -----		optionnel	optionnel	<u>nécessaire</u>

- Nature ou type de données - Format et outils de lecture associés - Standards et formats de métadonnées - Modalités de diffusion et localisation du jeu de données - Durée de l'embargo (si nécessaire) - Type de licence			
12/ Tri et archivage des données- 4 questions - Quelles données ont vocation à être conservées sur le long terme (plus de 10 ans) ? La sélection peut être effectuée à partir de plusieurs critères : potentiel de réutilisation scientifique, valeur de preuve, valeur historique, etc. - Combien de temps chaque jeu de données doit-il être conservé ? - Quel est le volume des données à archiver ? - Quelle est la plateforme d'archivage choisie ?	optionnel	recommandé	<u>nécessaire</u>

GUIDE DE SAISIE

Dans le cadre d'Horizon 2020, il est spécifié que le livrable consiste en **un unique plan de gestion** pour l'ensemble d'un projet. Ce guide a pour vocation de vous accompagner dans la rédaction de ce livrable et de constituer un « pense-bête » sur tous les aspects de la gestion de vos données, de leur production à leur archivage.

- Les zones « **Soyez FAIR !** » donnent des indications sur les attendus de la Commission européenne pour faciliter la diffusion et la réutilisation des données offrant cette possibilité.
-  Les zones marquées d'un drapeau européen doivent être renseignées dans le cadre des projets H2020.

Nom du champ	Précisions et recommandations
Version du DMP, date	Plusieurs personnes pouvant être amenées à intervenir sur le DMP, cette information sert à identifier la version utilisée.
Historique des mises à jour	La tenue d'un historique des mises à jour permet de connaître le dernier rédacteur, la portée et la date des mises à jour du document. C'est une garantie d'intégrité du document et de son contenu.

PHASE 1 – CONTEXTE, ACTEURS ET RESSOURCES

Section 1 : Description du projet

Nom du champ	Précisions et recommandations
Identifiant interne de l'appel à projet	Avant l'attribution d'un financement, le projet peut recevoir en interne de l'institution qui le porte un identifiant permettant son suivi. Il peut être attribué par les porteurs du projet, par le service en charge de la recherche ou être généré automatiquement par un logiciel.
Organisme financeur du projet	Le financeur est l'organisme qui porte les programmes de recherche. Il peut s'agir d'une institution internationale, nationale ou locale (Union européenne, ANR, région, etc.), publique ou privée.
Programme de financement	Le programme de recherche est le nom du dispositif dont sont issus les financements (Horizon 2020, etc.).
Référence de la convention de financement	Après acceptation du projet par un financeur, la convention signée entre toutes les parties comporte une référence qui est aussi un identifiant utilisé dans des bases publiques (CORDIS, le portail de la Commission européenne consacré aux résultats des projets de recherche financés par l'UE ; Hal, le portail des publications françaises en open access, ...).
Acronyme ou titre du projet	Le titre ou l'acronyme du projet mentionné ici doit être un identifiant reconnu par tous les acteurs et demeurer pérenne.
Objectifs et attendus du projet	Ce champ comporte la description de la nature du projet, ses objectifs et son déroulement. Il permet de comprendre le contexte et les types de données qui seront produites ou collectées au cours du projet. L'objectif du projet est décrit dans le dossier de candidature à l'appel à projets.
Thématiques et mots-clefs du projet	Liens et outils : Domaines ERC http://www.must-cofund.ch/pdf/ERC_keywords.pdf
Documentation 	Ce champ vise à recueillir toute information nécessaire à l'intelligibilité à long terme des données : codes, abréviations, version des logiciels de lecture, etc. La documentation associée permet également d'identifier les sources d'informations qui décrivent le projet. Cette documentation pourra être diffusée en libre accès en complément des jeux de données ou être archivée avec les données. Le dépôt de documentation complémentaire relatif à la production des données (code source logiciel, algorithme de traitements et d'analyse etc.), peut se révéler utile à la compréhension, la réutilisation et la conservation des données.

Section 2 : Acteurs du projet

Nom du champ	Précisions et recommandations
Nom, affiliation et coordonnées du coordinateur scientifique du projet	Le coordinateur scientifique du projet est chargé de coordonner les différentes équipes et leurs actions pour atteindre les objectifs du projet.
Partenaires du projet	Ce champ comporte la liste des établissements et partenaires du projet, notamment ceux impliqués dans la production et la gestion de données. Dans le cas d'un projet associant des partenaires sur d'autres sites ou dans d'autres institutions, les responsabilités entre les partenaires doivent être précisées.
Responsabilités sur les données et mise à jour du DMP  FAIR Data	Pour garantir une gestion et un partage optimum des données et veiller à la mise à jour du DMP par l'ensemble des acteurs, il est important que les responsables des données soient clairement désignés et identifiés par l'ensemble des partenaires. Le ou les rédacteurs de DMP doivent rester impliqués dans le pilotage du projet de sa création à son achèvement, coordonner les actions nécessaires à la mise en œuvre du DMP, effectuer les modifications nécessaires à sa mise à jour et assurer sa transmission au financeur.

Section 3 : Ressources nécessaires

Le DMP ne figure pas dans la candidature : il ne sert donc pas à budgéter le projet (cette démarche doit intervenir en amont) mais à identifier ce qui est éligible au remboursement

Nom du champ	Précisions et recommandation
Moyens matériels mis en œuvre pour la gestion des données	Il s'agit de mettre en évidence des garanties de réussite du projet dans le domaine de la gestion des données en insistant sur les moyens mis en œuvre : moyens à disposition, support et accompagnement institutionnel, outils, etc. Si des faiblesses persistent dans le dispositif, il est également intéressant de les indiquer et de prévoir (champ suivant) les mesures nécessaires à leur consolidation.
Estimation des besoins humains et de formation	Le fait de disposer de personnels formés ou à défaut sensibilisés à la gestion des données, à la rédaction ou aux enjeux d'un plan de gestion de données peut faciliter la collaboration des partenaires autour de ces sujets. Selon la nature du projet, des formations plus spécialisées à la conception de bases de données ou à la manipulation de données disciplinaires (<i>big data</i> , algorithmes de traitement, méthodes d'enrichissement, ontologies, etc.) peuvent également se révéler pertinentes à défaut de compétences identifiées parmi les partenaires ou de services supports dédiés au sein du projet.

	Indiquez ici le nombre de personnes (en ETP) et type de compétences dédiées (archivistiques, informatiques, documentaires, curation de données etc.) et/ou les types de formations souhaitées. ⇒ Personnes ressources : services informatiques ; bibliothécaires ; archivistes ; documentalistes ; les URFIST ⇒ Liens et outils : ○ Plateformes d'apprentissage numérique sur les données de recherche : http://doranum.fr/ ○ Les URFIST (Unité Régionale de Formation à l'Information Scientifique et Technique) : https://urfistinfo.hypotheses.org/ ○ Webinar (english) sur le Data management plan : https://www.youtube.com/watch?v=qj3TkO4AjC4&feature=youtu.be
Estimation financière des besoins de gestion matériels et humains	Les coûts de stockage et de traitement des données sont également à prendre en compte dans cette estimation. Si les prix baissent, les volumétries ont tendance à exploser. Il ne faut pas penser uniquement en termes de stockage mais aussi en termes de services associés (entrées et sorties de données, maintenance, fréquence d'accès, sécurité, etc.), ces derniers pouvant représenter des coûts masqués importants. Concernant les solutions de stockage externe, il n'existe pas de grilles tarifaires fixes. Elles varient selon le type de données, de services et de stockage (cloud, SATA, SAS etc.). Des solutions institutionnelles internes existent parfois, renseignez-vous auprès de vos établissements. ⇒ Personnes ressources : services informatiques ; archivistes ; bibliothécaires
Estimation financière des opérations de dissémination	La publication et la dissémination des données peuvent avoir un coût : maintenance d'une base de données, création d'un site web, frais de publication (ou <i>articles processing charges</i> - APC) pour les <i>data papers</i>, frais de <i>supplementary data</i> d'un article etc. Les coûts peuvent fortement varier d'une discipline à l'autre (en général ils sont moins élevés en sciences humaines et sociales qu'en sciences du vivant) et se situent généralement dans une fourchette de 100 à 2000 €.
Estimation financière des opérations d'archivage 	Le coût de l'archivage peut être évalué sur la base d'une volumétrie prévisionnelle en demandant des devis auprès des plateformes d'archivage agréées. ⇒ Personnes ressources : bibliothécaires ; archivistes ; documentalistes ⇒ Liens et outils : Répertoire des plateformes d'archivage agréées : https://francearchives.fr/fr/article/26287437
Couverture des besoins financiers 	Selon le financeur ou le programme de financement, les coûts liés à l'ouverture des données peuvent-il être éligibles, dans des conditions précisées par la convention de financement ?

PHASE 2 – STOCKAGE, PARTAGE, PROTECTION ET DIFFUSION AU COURS DU PROJET

Section 4 : Informations générales sur les données

Nom du champ	Précisions et recommandations
Méthodes et processus de production de données  FAIR Data	Pour permettre la réutilisation des données et garantir leur intégrité, il est nécessaire de comprendre la méthode et les protocoles selon lesquels elles ont été produites et collectées : appareils et instruments de mesure, logiciels, enquête, analyse, observation, compilation, simulation, etc. Indiquez les normes qualité utilisées le cas échéant. La méthode de production est essentielle pour les données de simulation pour lesquelles la modélisation a plus de valeur que la donnée elle-même ou pour les données d'observations uniques par leur contexte. Ce champ fournit également des informations influant sur la compréhension, l'intégrité, la valeur et la lisibilité des données sur le long terme (enjeu d'archivage et de réutilisation).
Réutilisation de données existantes 	Ce champ précise si le projet s'appuie sur la réutilisation de données déjà existantes et éventuellement produites par d'autres équipes de recherche. L'objectif est de : 1/garantir l'intégrité des données produites en établissant une filiation entre les jeux de données : la science étant un processus cumulatif, cela permet d'éviter les plagats en citant correctement ses sources ; 2/ permettre l'établissement de liens entre jeux de données si le signalement et la mise en libre accès des données sont prévus, notamment si des possibilités de <i>linked data</i> existent ; 3/ repérer les jeux de données de référence et s'assurer de leur archivage au cours du processus de sélection. ⇒ Personnes ressources : documentalistes ; bibliothécaires ; archivistes ⇒ Liens et outils : Datacite : http://search.datacite.org ; Isidore (pour les SHS) : http://www.rechercheisidore.fr/ ; Zenodo : https://zenodo.org/ ; pour les données de santé https://www.cnil.fr/fr/snds-systeme-national-des-donnees-de-sante
Propriété des données FAIR Data	Ce champ précise à qui appartiennent les données produites et/ou réutilisées, ainsi que les droits qui régissent les bases de données et leurs contenus. La propriété des données est généralement fixée par l'accord de consortium définissant le rôle des partenaires du projet. La négociation des termes de l'accord en amont du projet est essentielle car elle conditionne en partie les modalités de partage et de dissémination ultérieures des données. Ce champ doit reprendre les éléments définis dans l'accord de consortium et être conforme au document établi entre les partenaires lors de la réponse à l'appel à projets. L'établissement d'un accord de consortium est obligatoire dans le cadre des projets collaboratifs européens. Prenez contact avec les ingénieurs-projets Europe pour la rédaction de cet accord. ⇒ Personnes ressources : l'ingénieur ou le chargé de projet ; les services juridiques, qui sont chargés de la rédaction des conventions et accords ; les directions de recherche ; les SATT (Sociétés d'Accélération du Transfert de Technologies).

	⇒ Liens et outils : <i>European IPR Helpdesk</i> : projet de la Commission européenne pour conseiller les projets financés dans le cadre d'Horizon 2020, ce site comporte de nombreuses ressources sur les questions de propriété intellectuelle (documentation, formations, service support) : https://www.iprhelpdesk.eu/ ; <i>Guidance How to draw up your consortium agreement</i> : ce document donne des indications pour la rédaction de l'accord de consortium ; <i>DESCA 2020 Model Consortium Agreement</i> : ce site fournit un modèle pour la rédaction d'un accord de consortium : http://www.desca-2020.eu/.
Type et nature des données  FAIR Data	Ce champ précise les types de données qui seront produites ou utilisées dans le cadre du projet : données d'enquête, échantillons, code logiciel, corpus de texte, photographies, etc. Il fournit des premiers renseignements permettant d'orienter les besoins en stockage, en outils de traitement et de visualisation, les options de standards de métadonnées, d'outils de diffusion ainsi que les solutions de pérennisation à mettre en œuvre en vue de leur archivage. Cela permet également d'envisager la recherche d'autres types de données approchants, susceptibles d'être utilisés en matériaux complémentaires dans le projet.
Formats et outils de production, de traitement ou de conversion des données  FAIR Data	Parallèlement aux explications globales sur les modes de production des données, ce champ vient préciser les éventuels formats (outils de lecture associés) par lesquels transiteront les données tout au long de leur cycle de vie (de la production à la publication). Toutes les données et tous les formats n'ont pas vocation à être conservés (tout dépend de la nature et de la valeur des données au regard de la validation des résultats) mais il est important de pouvoir en retracer leur parcours. ⇒ « Soyez FAIR ! » : Afin de garantir l'accessibilité et la réutilisation des données dans le temps et par le plus grand nombre, il est fortement recommandé de privilégier des formats et des logiciels libres de droit (open source) notamment à des fins de diffusion ou d'archivage des données. A défaut, le recours à des formats et des outils standards ou largement documentés et partagés au sein de la communauté faciliteront le partage et la réutilisation des données entre chercheurs du domaine. ⇒ Personnes ressources : services informatiques ; équipes de recherche de la communauté ⇒ Liens et outils : pour que les données soient réutilisables sur le long terme, choisissez des formats pérennes : https://www.cines.fr/archivage/des-expertises/les-formats-de-fichier/ ; Pensez également à assurer la pérennité des logiciels utilisés pour produire les données, particulièrement pour des recherches menées sur le long terme. Il existe des <i>software management plans</i> pour vous y aider : https://www.software.ac.uk/resources/guides/software-management-plans Un exemple de projet de recherche sur le sujet : https://succes2017.sciencesconf.org/data/poster_SUCCES2017_PRESOFT.pdf

Formats et standards ou formats de métadonnées



FAIR Data

Les métadonnées sont l'ensemble des informations structurées permettant de décrire les données produites.

- ⇒ « **Soyez FAIR !** » : Afin de garantir l'interopérabilité et la réutilisation des données il est important de documenter avec soin les données produites, tout au long de leur cycle de vie (métadonnées descriptives, métadonnées de gestion, métadonnées d'archivage etc.). Plus les métadonnées (champs et contenus) seront structurées plus il sera aisé de les exploiter à des fins de recherche ou d'indexation. Il existe des formats très génériques de métadonnées pour qualifier des objets numériques (Dublin Core, DataCite Metadata Schema) mais de nombreuses disciplines ont progressivement développé des standards de métadonnées propres répondant davantage aux besoins spécifiques de leur communauté en matière de partage d'informations : *Climate and Forecast Metadata* pour les données climatiques, *Ecological Metadata Language (EML)* en écologie, *DICOM* pour l'imagerie médicale ou encore *l'International standard bibliographic description (ISBD)* pour les données bibliographiques. Un standard de métadonnées peut ensuite se traduire dans un ou plusieurs formats de données. La recherche de l'interopérabilité privilégiera des formats libres et ouverts à des formats propriétaires.
- ⇒ **Personnes ressources** : documentalistes, bibliothécaires, data curator
- ⇒ **Liens et outils** : le répertoire de métadonnées disciplinaires du *Digital Curation Center* britannique : <http://www.dcc.ac.uk/resources/metadata-standards> ; le répertoire de standards de métadonnées en science de la vie BioSharing : https://biosharing.org/standards/exchange_format

Section 5 : Stockage et partage des données au cours du projet

Nom du champ	Précisions et recommandation
Support des données	Un même type de données peut être conservé sur plusieurs supports (papier, serveur, disque dur, clé USB, etc.). Afin de garantir l'exploitation des données, leurs modalités de conservation et leur récupération complète en cas de perte, il est important d'avoir une vision globale sur l'ensemble des supports sur lesquels sont conservées les données tout au long du cycle de vie. L'obsolescence des supports peut être tout aussi problématique que l'obsolescence des formats (disparition des lecteurs de disquettes, fragilité des disques durs, disparition des magnétoscopes etc.). Garantir la conservation et la réutilisation des données sur le long terme peut donc nécessiter des transferts de données d'un support à un autre. Il convient d'anticiper l'achat du matériel adéquat. ⇒ Personnes ressources : archivistes (<i>digital forensics</i>) ; services informatiques
Modalités de stockage et d'hébergement FAIR Data	Ce champ décrit le matériel et le lieu physique de stockage des données : hébergement dans les <i>clouds</i>, serveurs locaux, disques durs, etc. Pour les données de santé, des recommandations spécifiques existent. ⇒ Personnes ressources : ingénieurs des laboratoires ; services informatiques ⇒ Liens et outils : la FAQ du site E-santé ;
Volumétrie prévisionnelle des données 	Estimez la volumétrie des données qui seront partagées par les équipes au cours du projet.
Accès et partage entre les partenaires  FAIR Data	De multiples acteurs parfois distants ou relevant de tutelles distinctes sont amenés à collaborer au sein du projet, il est donc important pour la réussite du projet autant que pour son bon déroulement d'anticiper la manière et les conditions dans lesquelles les différents partenaires pourront partager et s'échanger leurs données au cours du projet. Avant de se tourner vers d'éventuelles solutions commerciales gratuites ou payantes (Google Drive, Dropbox etc.) il est recommandé de s'interroger sur les outils à disposition au sein de son établissement. De plus en plus d'établissements proposent en effet des services de <i>cloud computing</i> institutionnels permettant notamment l'installation de machines virtuelles, d'espaces de stockage ou de partages de ressources. Il est par ailleurs indispensable que le contrôle des accès et les modalités de partage des données répondent aux nécessités de protection et d'éthique abordées lors du montage du projet. ⇒ Personnes ressources : services informatiques ; comités d'éthique ; CIL/Délégué à la protection des données ⇒ Liens et outils : la plateforme open source <i>Open Science Framework</i> : https://osf.io/

Règles de nommage et de classement  FAIR Data	Dans les projets collaboratifs, il est souvent nécessaire de mettre en place des espaces de partage de données ou de fichiers. En l'absence d'application de gestion électronique de document (GED) ou de base de données structurée et administrée, il est recommandé d'adopter des règles de nommage communes pour un partage intelligible des fichiers entre les partenaires. Afin de garantir l'intégrité et la compréhension des fichiers, le nommage doit respecter certaines règles formelles et idéalement comporter certaines mentions (date, objet, version etc.). Des métadonnées additionnelles peuvent également venir enrichir les propriétés des documents en complément des règles de nommage. ⇒ Personnes ressources : archivistes ; data curator ⇒ Liens et outils : ○ un outil de gestion des doublons : http://www.duplicatecleaner.com/; ○ des règles de nommage des fichiers de l'Université de Lausanne ; ○ les règles d'enregistrement des domaines Internet : https://www.afnic.fr/medias/documents/Cadre_legal/afnic-charte-de-nommage-2014-12-08.pdf ; ○ des outils libres de renommage des fichiers : ReNamer (http://renamer.fr.softonic.com/) et Ant Renamer (http://www.antp.be/software/renamer/fr). ○ exemple : « producteur_sujet_type de document_AAAAMMJJ_version.extension »
--	---

Section 6 : Risques, sécurité et éthique des données

Nom du champ	Précisions et recommandation
Risques et menaces concernant les données	Différents facteurs sont susceptibles de menacer l'intégrité, la disponibilité et la confidentialité des données produites au cours du projet. Les risques peuvent être de différentes natures : des risques naturels pesant sur les infrastructures (zones sismiques, inondables etc.), des risques techniques (corruptions ou pertes de données, problèmes de serveurs etc.), des risques de confidentialité (accès non autorisés, fuites de données sensibles, etc.) ⇒ Personnes ressources : services informatiques ; CIL/Délégué à la protection des données. ⇒ Liens et outils : ○ DRAMBORA est un outil du DCC permettant d'évaluer les risques pesant sur les entrepôts de données : http://www.dcc.ac.uk/resources/repository-audit-and-assessment/drambora ○ EBIOS est une méthode d'expression des besoins et identification des objectifs de sécurité de l'ANSSI : http://www.ssi.gouv.fr/uploads/2011/10/EBIOS-1-GuideMethodologique-2010-01-25.pdf

Mesures de protection des données  FAIR Data	L'application de mesures de sécurité particulières peut concerner une large gamme de données jugées sensibles au regard de la protection de la vie privée, d'intérêts commerciaux ou encore d'éléments relatifs aux activités militaires ou de défense. Ce champ vise à détailler la façon dont la protection des données sera mise en œuvre. Il doit lister les mesures prévues pour assurer la sécurité des données au vu des risques identifiés : • Au niveau de la disponibilité des données : garantir le fonctionnement des outils pour la continuité des accès par les équipes ; • Au niveau de l'intégrité des données : mettre à disposition des données de qualité dans les temps et espaces prévus. La production de données intègres est indispensable à l'obtention de résultats fiables. • Au niveau de la confidentialité : offrir un niveau satisfaisant d'accès et de préservation des données sensibles. ⇒ Personnes ressources : services informatiques ; CIL/Délégué à la protection des données ; fonctionnaire de sécurité défense ⇒ Liens et outils : ○ S'appuyer sur les recommandations du référentiel général de sécurité (RGS) : https://www.ssi.gouv.fr/administration/reglementation/confiance-numerique/le-referentiel-general-de-securite-rgs/ ○ La norme internationale de système de gestion de la sécurité de l'information ISO/CEI 27001 ○ La sécurité des données (CNIL) : https://www.cnil.fr/fr/garantir-la-securite-des-donnees ; ○ Les 7 principes clés de la protection des données personnelles (CIL-CNRS) : http://www.cil.cnrs.fr/CIL/spip.php?article1390 ○ Sur les changements liés au nouveau règlement européen sur la protection des données : https://www.cnil.fr/fr/reglement-europeen-sur-la-protection-des-donnees-ce-qui-change-pour-les-professionnels ○ Pour les recherches dans le domaine de la santé : https://www.cnil.fr/fr/recherches-dans-le-domaine-de-la-sante-le-nouveau-chapitre-ix-est-applicable
Ethique et gestion de la confidentialité 	Selon la norme ISO 7498-2 (ISO90), la confidentialité est la propriété d'une information qui n'est ni disponible ni divulguée aux personnes, entités ou processus non autorisés. Dans le droit européen, tous les responsables de traitement ou sous-traitants doivent veiller à préserver la confidentialité des données. Personnes ressources : CIL/Délégué à la protection des données ; comité d'éthique local

Section 7 : Dissémination et archivage des données

Nom du champ	Précisions et recommandation
Potentiel de réutilisation des données  FAIR Data	Dans une logique de retour sur investissement et d'encouragement à la réutilisation des données, l'identification des publics cibles est essentielle. Ils peuvent être de nature variée : grand public, communauté scientifique, secteur privé, projet de recherche identifié, etc. Ce champ peut aussi contenir des perspectives d'application ou de développement d'autres axes de recherche à partir des données du projet.
Principe général de diffusion FAIR Data	Le financeur peut indiquer dans son règlement la diffusion attendue qui s'applique à la majorité des données produites dans le cadre du projet, comme la Commission Européenne et l' <i>open research data</i> prévu par l'article 29.3 du <i>grant agreement</i> .
Modalités de diffusion des données envisagées  FAIR Data	Ce champ a pour vocation d'anticiper les choix et modalités de dissémination des données en fonction des objectifs d'impact du projet. En l'absence d'entrepôt ou de plateformes de données institutionnelles ou locales susceptibles de les accueillir, il est préférable de déposer les données dans un entrepôt idéalement ouvert, gratuit, référencé par les moteurs de recherche, proposant des URL et identifiants pérennes et présentant des garanties de préservation et d'archivage. Certains éditeurs, comme ceux listés ci-dessous, recommandent également des entrepôts ou partenaires pour le stockage et la diffusion des données : PLoS One: http://journals.plos.org/plosone/s/data-availability Nature Scientific Data : https://www.nature.com/sdata/policies/repositories ⇒ Personnes ressources : bibliothécaires, documentalistes ⇒ Liens et outils : ○ Pour trouver un entrepôt, consulter le Registry of Research data repository : http://www.re3data.org/ ○ EUDAT propose la base B2FIND : https://eudat.eu/services/b2find ○ ZENODO : entrepôt de l'UE pour les jeux de données orphelins issus de H2020 : http://zenodo.org/ ○ NAKALA est le service de stockage des données propose par le grand équipement HumNum pour les SHS : https://www.nakala.fr/ ○ Il est également possible de publier une description de données dans un <i>data journal</i> : http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList. ○ Améliorer l'exposition des données de la recherche : la publication de <i>data papers</i>²⁵ : https://archivesic.ccsd.cnrs.fr/sic_01427978

²⁵ Les *data papers* sont des articles décrivant des données de recherche, publiés dans des revues à comité de lecture.

Modalités d'archivage envisagées  FAIR Data	L'archivage doit être anticipé dès le début du projet : pérennité des formats choisis, tri des données, plateforme d'archivage sélectionnée. Pour en savoir plus sur le concept d'archivage pérenne : https://www.cines.fr/archivage/un-concept-des-problematiques/le-concept-darchivage-numerique-perenne/
---	--

PHASE 3 – DISSEMINATION ET ARCHIVAGE A L'ISSUE DU PROJET

Section 8 : Identification des jeux de données

Nom du champ	Précisions et recommandation
Nombre de jeux de données	Selon le projet, un jeu de données peut recouvrir des réalités différentes. En effet, un ou plusieurs jeu(x) de données peu(ven)t être lié(s) au projet de recherche, et désigner : <i>a)</i> un lot techniquement homogène, ou <i>b)</i> un lot intellectuellement cohérent même si celui-ci est composé de lots techniquement hétérogènes.
Liens et relations entre les jeux de données FAIR Data	(le cas échéant)

Section 9 : Protection – exception de diffusion

Justification de l'exception au principe de diffusion  FAIR Data	Le cas échéant, indiquer pour quelle raison vous ne pouvez pas diffuser le jeu de données décrit.
--	---

Sections 10 et 11 : Description et diffusion des jeux de données

Nom du champ	Précisions et recommandation
Intitulé du jeu de données	
Identifiant du jeu de données  FAIR Data	Attribuer un identifiant à des jeux de données facilite le suivi, la localisation, l'accès et la citation des données lors de leur publication ou à des fins de réutilisation. ⇒ « Soyez FAIR ! » : Afin de favoriser l'identification et l'accès aux données et jeux de données il est recommandé d'utiliser des systèmes d'identification informatique uniques permettant de désigner et de localiser les jeux de données de manière non ambiguë dans l'espace et le temps et de les intégrer dans des systèmes d'identification plus larges partagés sur le web. Nous recommandons en particulier le recours à des systèmes de type Handle, ARK, URI ou DOI. ⇒ Personnes ressources : documentalistes, bibliothécaires, services informatiques ⇒ Liens et outils : Un abonnement annuel de 180 euros pour l'attribution de 500 DOI est possible auprès de l'INIST : http://www.inist.fr/?Attribution-de-DOI
Auteurs et contributeurs 	Nom, prénom, affiliation.
Editeur (institution) 	L'éditeur correspond à la personne physique ou morale disposant de <i>la propriété intellectuelle sur les données</i> . Il implique souvent l'établissement d'affiliation ou de rattachement de l'auteur du jeu de données, ou le consortium de partenaires d'un projet collaboratif. Il s'agit de la personne physique ou morale assurant la responsabilité de la publication des données ou jeux de données.
Mots-clefs, vocabulaires contrôlés, ontologies associées  FAIR Data	Les champs de métadonnées (thésaurus, vocabulaires contrôlés etc.) doivent être normés autant que possible afin d'en faciliter l'exploitation. Ce champ indiquera les vocabulaires contrôlés et ontologies utilisés. Plusieurs vocabulaires contrôlés existent comme Rameau (http://rameau.bnf.fr/), Mesh (http://mesh.inserm.fr/mesh/), etc. Pour les ontologies biomédicales, il existe Bioportal (http://bioportal.bioontology.org/). Les mots-clefs servent quant à eux à caractériser les contenus et à en permettre la recherche.

Nature ou type de données  FAIR Data	Ce champ précise le type de données qui seront produites ou utilisées dans le cadre du projet : données d'enquête, échantillons, code logiciel, corpus de texte, photographies, etc. Dans le cas de création de logiciels, la loi « pour une République numérique » indique que les logiciels créés par des établissements publics doivent l'être par défaut en open source, sauf pour des raisons de sûreté de l'Etat, de sécurité publique, de secret de la défense nationale, de secret des relations extérieures, de sécurité des systèmes d'information, de secrets en matière industrielle et commerciale. Voir aussi le champ "Type de licence", page 28 de ce document, pour les licences d'utilisation.
Format et outils de lecture associés  FAIR Data	Le format des données est souvent un paramètre discriminant pour le choix de la plateforme de diffusion et de la licence associée. Le format et les outils de lecture associés sont également des mentions nécessaires dans le cadre de l'archivage et de la réutilisation des données pour garantir la capacité de lecture des futurs utilisateurs et la conservation des données dans le temps (anticiper l'obsolescence des formats). En effet, à des fins de traitement comme de conservation, les données peuvent être amenées à transiter par divers formats et/ou subir plusieurs opérations de <i>mapping</i> et de conversion au cours comme à l'issue du projet. Il est dans ce cas recommandé de documenter le plus précisément possible la nature de ces opérations et d'en indiquer idéalement l'objectif. Si le choix de découpage des jeux de données ne tient pas compte du cycle de vie des données (<i>raw data</i>, <i>derived data</i>, <i>published data</i>), il est important de préciser les formats et outils de production et de lecture associés à chaque étape de production des données du projet. Il indique enfin les outils, logiciels et matériels nécessaires à la lecture des données produites. ⇒ « Soyez FAIR ! » : Pour garantir une meilleure réutilisation et interopérabilité des données, il est recommandé de diffuser et d'archiver ses données sous des formats libres et ouverts et de prévoir en conséquence les opérations de conversion nécessaires depuis les formats propriétaires. Outre la barrière financière que représente l'acquisition d'un logiciel commercial, les formats propriétaires sont souvent moins bien documentés et soumis à une plus forte obsolescence que les formats et outils libres ou communautaires. ⇒ Personnes ressources : archivistes ; bibliothécaires ; services informatiques ⇒ Liens et outils : ○ Le référentiel général d'interopérabilité formule des recommandations en faveur de l'interopérabilité des formats dans les administrations publiques. ○ Les recommandations et l'outil FACILE du CINES concernant le contrôle et la validation des formats archivables : https://facile.cines.fr/ ○ Des outils de <i>digital forensics</i> comme BitCurator conçus pour les besoins des archivistes et des bibliothécaires en matière d'émulation, de lecture et de récupération de données : http://wiki.bitcurator.net/index.php?title=Main_Page.

Standards et formats de métadonnées  FAIR Data	Les métadonnées peuvent intervenir au niveau des données comme des jeux de données. Elles sont essentielles à la compréhension, à la gestion, au signalement comme à la conservation des jeux de données au cours ou à l'issue du projet. La production de métadonnées nécessite un travail de réflexion et de normalisation notamment lorsque cette réflexion intervient dans le cadre d'une création de bases de données. Si certaines métadonnées peuvent être automatiquement générées par les appareils ou les logiciels utilisés lors de l'acquisition ou du traitement des données (scanner, RMN, etc.), par des algorithmes spécifiques ou bien rapatriées par alignement, certains champs nécessitent généralement des enrichissements manuels ou des encapsulages au sein d'un format plus global. ⇒ « Soyez FAIR ! » : De nombreuses disciplines s'appuient désormais sur des standards et des formats de métadonnées spécifiques garantissant le partage et la transparence de l'information entre les différents partenaires, tout au long du projet. En l'absence de standard disciplinaire identifié pour la gestion et la description des données ou pour la diffusion des jeux de données, il reste possible : 1/ de recourir à des formats génériques de métadonnées utilisés pour la description d'objets numériques : DublinCore, METS, DataCite Metadata Schema, etc. 2/ de s'appuyer sur les métadonnées requises pour la diffusion des données et des jeux de données ayant vocation à être diffusés sur une plateforme externe (entrepôts de données, archives ouvertes etc.). Ces entrepôts imposent souvent un minimum de métadonnées obligatoires pour valider le dépôt. Selon l'orientation plus ou moins disciplinaire de la base, les métadonnées peuvent reprendre les standards disciplinaires évoqués plus haut. 3 / de créer son propre format de métadonnées. Dans ce cas ne pas oublier de le décrire et de le documenter avec précision en s'appuyant de préférence sur les langages XML ou RDF qui fournissent le meilleur cadre d'interopérabilité. ⇒ Personnes ressources : bibliothécaires, documentalistes, services informatiques ⇒ Outils et liens : ○ Le Digital Curation Center propose des ressources pour s'informer sur les formats disciplinaires : http://www.dcc.ac.uk/resources/metadata-standards. ○ La page GitHub du Research Data Alliance recense des standards de métadonnées, des outils et des études de cas sur ces questions : http://rd-alliance.github.io/metadata-directory/.
Modalités de diffusion et localisation du jeu de données  FAIR Data	Préciser ce qui a été effectivement réalisé, par rapport à ce qui avait été prévu dans le champ « Modalités de diffusion des données envisagées » de la Section 6 : dans quel entrepôt les données ont-elles été déposées ? quelle est l'URL du site créé (le cas échéant) ?, etc. ⇒ Outils et liens : répertoire d'entrepôts http://www.re3data.org/

Durée de l'embargo  FAIR Data	L'embargo est une protection temporaire qui permet de reporter à une date ultérieure la libre diffusion ou la réutilisation des données. L'embargo ne nécessite pas d'autorisation autre que celle du propriétaire. Les conditions de secret ou de discrétion au cours d'une période d'embargo sont moindres qu'une mesure de confidentialité (indispensable pour un projet de brevet ou de transfert de technologie). Dans le cadre des projets H2020, les délais d'embargo ne peuvent excéder 6 mois pour les sciences et techniques, de la vie et de la santé et 12 mois pour les sciences humaines et sociales. ⇒ « Soyez FAIR ! » La présence de données sensibles ou une ambition de publication ou d'exploitation différée des résultats n'excluent pas nécessairement toute diffusion des données. L'embargo se présente justement comme un compromis permettant de signaler l'existence des données (enjeux d'impact) sans compromettre leur exclusivité d'exploitation. Il permet de garantir à moyen terme leur accessibilité et leur réutilisation sans risque de perte entre la fin du projet et le moment de la diffusion. ⇒ Personnes ressources : bibliothécaires ; documentalistes ; services juridiques ; SATT
Type de licence  FAIR Data	La licence précise les conditions de partage et de réutilisation des données diffusées dans le cadre du projet, ainsi les éventuelles contreparties intellectuelles ou économiques qui y sont associées. Il est important de préciser qu'une diffusion en <i>libre accès</i> ne signifie pas nécessairement qu'une œuvre est <i>libre de droit</i>. La licence a notamment pour objectif de clarifier le statut juridique d'une œuvre et de préciser les conditions d'usage. De trop nombreux contenus ne sont pas réutilisés au maximum de leur potentialité en raison des ambiguïtés juridiques dues à l'absence de licences explicites. Il existe de nombreuses licences libres à utiliser selon les législations, les formats de données, les souhaits de protection des auteurs, les exigences des financeurs, etc. <u>Pour les jeux de données</u> (liste non exhaustive) • CC-by 4.0 : https://creativecommons.org/licenses/by-sa/4.0/deed.fr • CC0 : https://creativecommons.org/publicdomain/zero/1.0/legalcode.fr • La licence ETALAB : https://www.etalab.gouv.fr/licence-ouverte-open-licence <u>Pour les bases de données</u> (liste non exhaustive) • OKF – Open Database License (OdbL) : http://opendatacommons.org/licenses/odbl/1.0/ • OKF – ODC-By : http://opendatacommons.org/licenses/by/1.0/ • OKF – ODC-PDDL (Public domain) : http://opendatacommons.org/licenses/pddl/1.0/ <u>Pour les logiciels</u> (liste non exhaustive) • Open Licence Software (OPS) : https://opensource.org/licenses/OSL-3.0 • GNU-GPL : http://www.gnu.org/licenses/gpl.html

	Pour la documentation accompagnant les logiciels (liste non exhaustive) • GNU-FDL (Free Documentation Licence) : http://www.gnu.org/licenses/fdl-1.3.html ⇒ « Soyez FAIR ! » : Afin de garantir l'accès, l'interopérabilité et la réutilisation la plus large possible des données, il est recommandé de privilégier des licences libres largement partagées et documentées. ⇒ Personnes ressources : bibliothécaires ; documentalistes ; services juridiques ; services informatiques ; SATT ⇒ Liens utiles : ○ Le guide de sélection des licences du Digital Curation Center : http://www.dcc.ac.uk/resources/how-guides/license-research-data. ○ L'outil d'aide à la décision License selector (accessible par EUDAT B2SHARE) : https://ufal.github.io/public-license-selector/
--	---

Section 12 : Tri et archivage des données

Nom du champ	Précisions et recommandations
Sélection des données 	Toutes les données n'ont pas d'intérêt à être conservées à l'issue du projet. L'objectif de ce champ est d'indiquer l'intérêt des données sur le long terme (plus de 10 ans) et de déterminer ce que l'équipe souhaite en faire. Il permet d'indiquer comment procéder à la sélection des données. Ces dernières peuvent être conservées pour leur valeur informationnelle, leur potentiel de réutilisation, leur valeur de preuve ou leur valeur historique. Si les données n'ont pas toutes vocation à être archivées sur le long terme, le visa de l'archiviste reste néanmoins recommandé avant toute opération de destruction massive de données. Indiquez les jeux de données qui auront vocation à être conservés à moyen ou long terme en raison de leur valeur scientifique, juridique ou patrimoniale. En cas de doute, prenez contact avec l'archiviste de votre structure. ⇒ Personnes ressources : archivistes ⇒ Liens utiles : ○ Pour en savoir plus sur le concept d'archivage pérenne : https://www.cines.fr/archivage/un-concept-des-problematiques/le-concept-darchivage-numerique-perenne/ ○ Pour sélectionner les données à conserver :

	○ Instruction DAF-DPACI-RES-2007-002 du 15 janvier 2007 sur les archives du CNRS : http://www.archivesdefrance.culture.gouv.fr/static/871 ○ Un référentiel de conservation des archives de la recherche, produit par l'Association des archivistes français : http://www.archivistes.org/Referentiel-de-gestion-des-archives-de-la-recherche ○ Une autre possibilité est de s'appuyer sur la méthode et les recommandations du DCC : http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep
Durée de conservation préconisée	Cette durée varie fortement. Elle peut aller de quelques mois à l'éternité. Les durées de conservation doivent prendre en compte les exigences légales et réglementaires existantes. La durée de conservation permet également d'évaluer la durée d'utilisation de la plateforme d'archivage et donc les coûts induits.
Volume des données à archiver	Indiquer dans ce champ le volume final estimé. Une partie des coûts de la gestion des données est éligible au remboursement : le volume final des données est un critère d'évaluation des coûts de l'archivage.
Plateforme d'archivage  FAIR Data	Les plateformes d'archivage électronique choisies doivent être agréées par le Ministère de la Culture. Certaines universités disposent de conventions avec des plateformes : contactez le service d'archives de votre institution. Des spécificités existent concernant les données de santé. L'hébergement de données de santé à caractère personnel sur support informatique nécessite un agrément, accordé par le ministre de la santé (articles L.1111-8 et R.1111-9 à R.1111-15-1 du Code de la santé publique). ⇒ Personnes ressources : archivistes ⇒ Liens utiles : ○ La liste des plateformes d'archivage agréées par le ministère de la Culture est consultable sur ce site : https://francearchives.fr/fr/article/26287437 ○ Une liste d'hébergeurs agréés pour les données de santé est accessible à cette adresse : http://esante.gouv.fr/services/referentiels/securite/hebergeurs-agrees. ○ Le CINES propose un service de validation de formats « FACILE », afin de vérifier l'éligibilité de vos documents à un archivage sur sa plateforme : https://facile.cines.fr/