Ouvrir les données de la recherche.
Comment rédiger
un Plan de Gestion de Données (PGD) ?

Source de l'illustration : https://doranum.fr/plan-gestion-donnees-dmp/

Claudine DESFORGES
Chargée d'Etudes Documentaires
ENTPE - Direction de la Recherche
claudine.desforges@entpe.fr
Janvier 2022

UNIVERSITÉ DE LYON

ENTPE
L'école de l'aménagement durable des territoires

Le contexte : généraliser la science ouverte en France

A la suite du premier Plan national pour la science ouverte 2018-2021[1], le taux de publications scientifiques françaises en accès ouvert est passé de 41 % en 2017 à 56 % en 2019[2].

Le 2e Plan national pour la science ouverte 2021-2024[3] vise à **généraliser les pratiques de science ouverte**, en se fixant notamment comme objectifs d'ouvrir l'accès à 100 % des publications en 2030 ainsi qu'aux données de la recherche.

Que sont les données de la recherche ?

D'après l'OCDE[4], *« les données de la recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche. »*

Dans la pratique, ce peuvent être des données d'observation (capturées en temps réel, habituellement uniques, impossibles à reproduire, par ex., des mesures sismiques), d'expérimentation (obtenues à partir d'équipements de laboratoire, souvent reproductibles, parfois coûteuses (par ex., des résultats de réactions chimiques) ou computationnelles (générées par des modèles informatiques, par ex., une modélisation du changement climatique). Ce peuvent être des enregistrements (documents témoignant d'un phénomène ou d'une activité humaine), mais aussi des données compilées (issues du traitement de données brutes) ou des données de référence (validées par la communauté scientifique, réutilisables, par ex., le décodage du génome humain ou la cartographie de données astronomiques).

Les données de la recherche concernent l'ensemble du processus de recherche : elles peuvent être préparatoires (état des sources, base de la formulation de la question de recherche), brutes (collectées, avant leur mise en forme et analyse), traitées (après étalonnage ou correction), dérivées (ex : synthèse), publiées.

Leur cycle de vie peut être présenté de la manière suivante[5] :

créer les données → traiter les données → analyser les données → préserver les données → définir l'accès aux données → réutiliser les données.

---

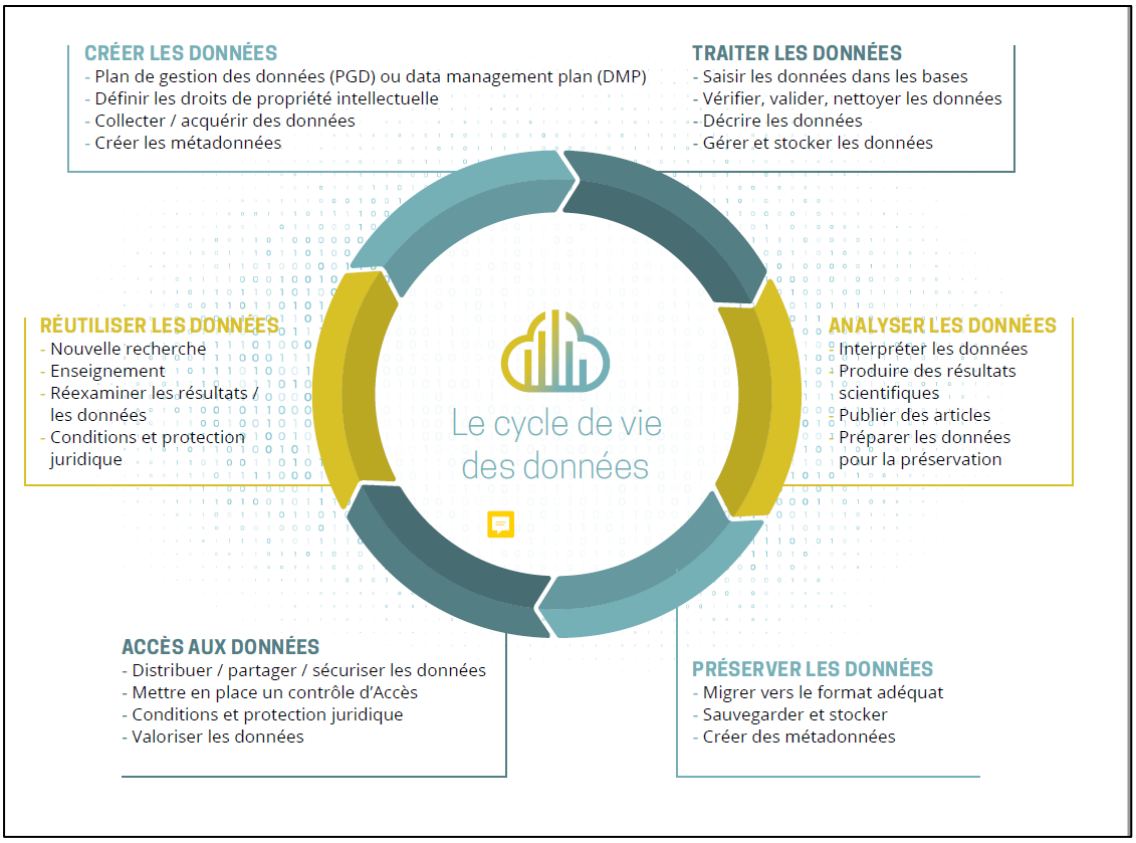
[1] Ministère de l'Enseignement supérieur, de la recherche et de l'innovation, Deuxième plan national pour la recherche ouverte, juillet 2018, https://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/67/2/PLAN_NATIONAL_SCIENCE_OUVERTE_978672.pdf

[2] D'après le MESRI, https://www.enseignementsup-recherche.gouv.fr/cid159131/plan-national-pour-la-science-ouverte-2021-2024-vers-une-generalisation-des-pratiques-de-science-ouverte-en-france.html

[3] Ministère de l'Enseignement supérieur, de la recherche et de l'innovation, Deuxième plan national pour la recherche ouverte, juillet 2021, https://cache.media.enseignementsup-recherche.gouv.fr/file/science_ouverte/20/9/MEN_brochure_PNSO_web_1415209.pdf

[4] Organisation de Coopération et de Développement Economique (OCDE), *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics,* Paris, 2007, 28 p. https://www.oecd.org/fr/science/inno/38500823.pdf

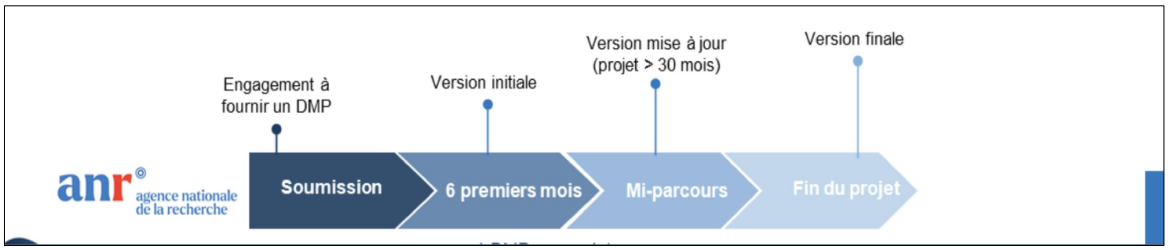
[5] D'après le *Guide pour la science ouverte* du Groupe INSA, p. 10, https://lalist.inist.fr/?p=51512



Pourquoi gérer et partager ses données ?

Un nombre croissant d'organismes de financement de la recherche demandent aux chercheurs de gérer leurs données de recherche de manière exemplaire, de les conserver et de les rendre accessibles dans la mesure du possible.

En France, c'est le cas de l'ANR qui, suivant le premier Plan national pour la science ouverte, demande l'élaboration d'un Plan de Gestion de Données (PGD)[6] pour les projets qu'elle finance depuis 2019. Un tel document constitue un livrable qui doit être fourni dans les 6 mois qui suivent le lancement scientifique du projet[7].



[6] Appelé aussi « *Data Management Plan* » (*DMP*).

[7] D'après Yvette Lafosse, « Plan de gestion de données. Bonnes pratiques de rédaction et outils », *Semaine Data SHS*, INIST, 6 décembre 2021.

Un PGD se présente sous la forme d'un document structuré de quelques pages dont l'objectif est de définir les actions à mettre en place pour gérer les données qui vont être collectées ou produites au cours d'un projet[8].

Cette démarche applique le principe « aussi ouvert que possible, aussi fermé que nécessaire » (respect de la vie privée, secret industriel et intérêts commerciaux légitimes, sécurité nationale …).

Sur le plan opérationnel, la gestion des données de la recherche, en permettant d'assurer la traçabilité du travail réalisé, a un impact sur la qualité des résultats de la recherche. Elle permet aussi aux chercheurs de gagner du temps en les réutilisant d'une recherche à l'autre.

Comment se préparer à rédiger un PGD ?

Le PGD vise à décrire le cycle de vie des données de la recherche, c'est-à-dire la façon dont elles seront produites, collectées, réutilisées, protégées et conservées à long terme. Il évolue tout au long du projet.

Il doit suivre les principes « FAIR » : les données doivent être Faciles à trouver (définir des identifiants), Accessibles (expliciter les protocoles d'interrogation), Interopérables (utiliser des termes normalisés), Réutilisables (bien décrire les métadonnées).

Il doit notamment spécifier qui pourra avoir accès aux données décrites, dans quel entrepôt elles seront déposées (institutionnel, disciplinaire ou généraliste) et ce, pendant quelle durée.

Quelques conseils pratiques pour réaliser un PGD complet[9] :

- Lister tous les matériaux de recherche qui seront manipulés : ceux réutilisés, collectés ou créés, outils et programmes élaborés…
- Faire le tour de tous les espaces de travail utilisés, de tous les sites internet consultés.
- Faire la même chose pour tous les outils utilisés : instruments de mesure, capteurs, smartphones, liseuses, dictaphones, appareils photo…).
- Catégoriser les données : types de données, stades de traitement, degré de sensibilité, propriétaires, destins de conservation ou de mise à disposition.
- Expliciter : donner un nom à chaque ensemble de données et rédiger une brève description des données incluses dans ces ensembles.
- Qualifier : choisir un modèle de PGD répondant aux questions : Qui ? Quoi ? Quand ? Où ? Comment ? Pourquoi ?

Une abondante documentation sur le sujet est disponible auprès de l'INIST-CNRS, sur la plateforme DoRANum (https://doranum.fr/tag/inist-cnrs/), qui a pour objectif d'accompagner la communauté scientifique dans la gestion et le partage des données de la recherche : actualités, ressources, FAQ… Contact : https://doranum.fr/contact/

[8] Voir le modèle proposé par l'ANR : https://anr.fr/fileadmin/documents/2019/ANR-modele-PGD.pdf

[9] D'après Justine Ancelin-Fabre, *Rédiger un plan de gestion pour ses données de recherche. De la pièce administrative au cadeau pour les générations futures !*, URFIST de Paris, 1er juin 2021.

Quelle est la structure d'un PGD ?

Un PGD doit globalement contenir les informations suivantes[10] :

RUBRIQUES	CONTENUS
Informations administratives	Nom du projet
	Description du projet
	Financeur(s)
	Responsable(s)
	Contact(s)
	Dates des différentes versions
	Politiques associées au projet
Description des données	Types et formats
	Méthodes de production ou de collecte
	Organisation, nommage, gestion des répertoires
	Informations nécessaires à l'interprétation des données
	Contrôle de la qualité des données
Exigences éthiques et légales	Droits de propriété intellectuelle
	Protection des données personnelles
	Sécurité du stockage et du transfert des données sensibles
	Accords conclus dans le cadre d'un partenariat
Stockage et conservation	Lieu(x) de stockage : espace de travail collaboratif, entrepôt…
	Système de nommage des fichiers
	Volumétrie des données
	Plan de sauvegarde : durée(s) de conservation
	Responsable(s) de la conservation
	Procédures de récupération
Modalités de partage	Modalités d'échanges de données pendant le projet
	Modalités d'accès, de diffusion et de partage des données après le projet
	Modalités de réutilisation des données

Comment réaliser un PGD en ligne ?

Afin de faciliter la mise en application des principes FAIR lors de la rédaction d'un PGD, l'INIST-CNRS a mis à disposition de la communauté scientifique un outil, OPIDoR[11] : https://dmp.opidor.fr/.

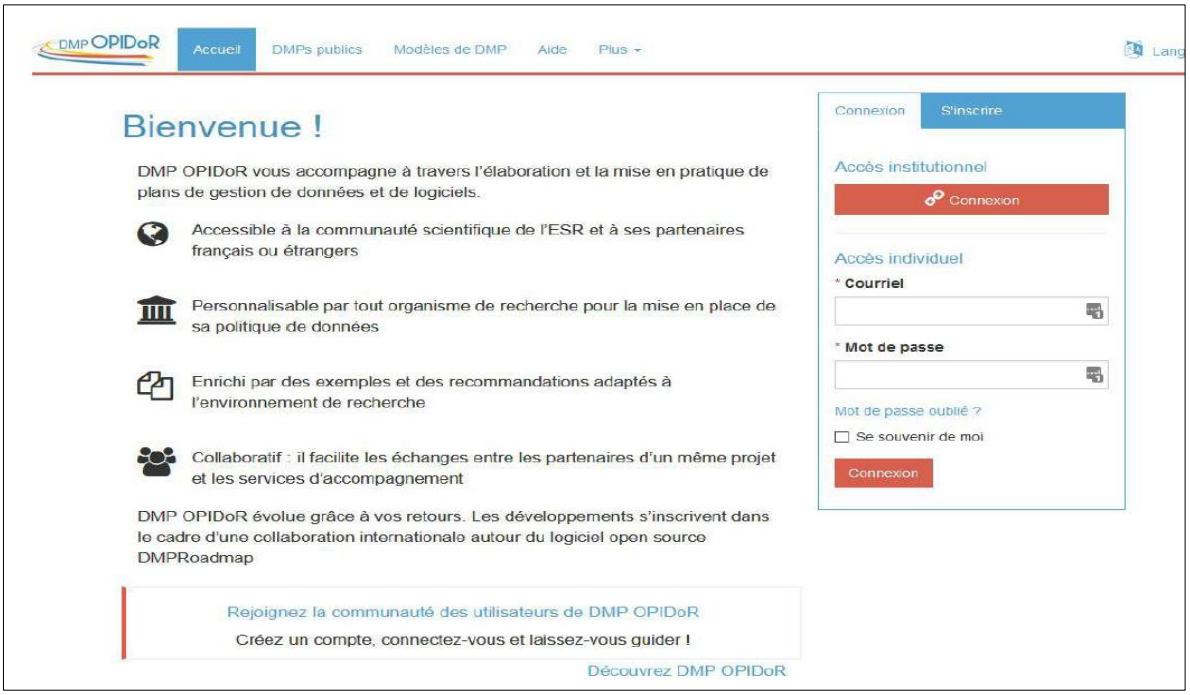
Cet outil, gratuit pour l'ESR, permet de rédiger des PGD en utilisant des modèles génériques préconisés par des institutions et des financeurs. Ces PGD, qu'ils soient en cours de rédaction ou finalisés, peuvent être partagés avec d'autres utilisateurs, exportés dans différents formats, et bénéficier d'aides en ligne. Contact : https://opidor.fr/contact/

[10] Cette présentation s'inspire des fiches didactiques de la délégation à l'information scientifique et technique du CIRAD : https://coop-ist.cirad.fr/gerer-des-donnees/rediger-un-pgd/1-qu-est-ce-qu-un-pgd

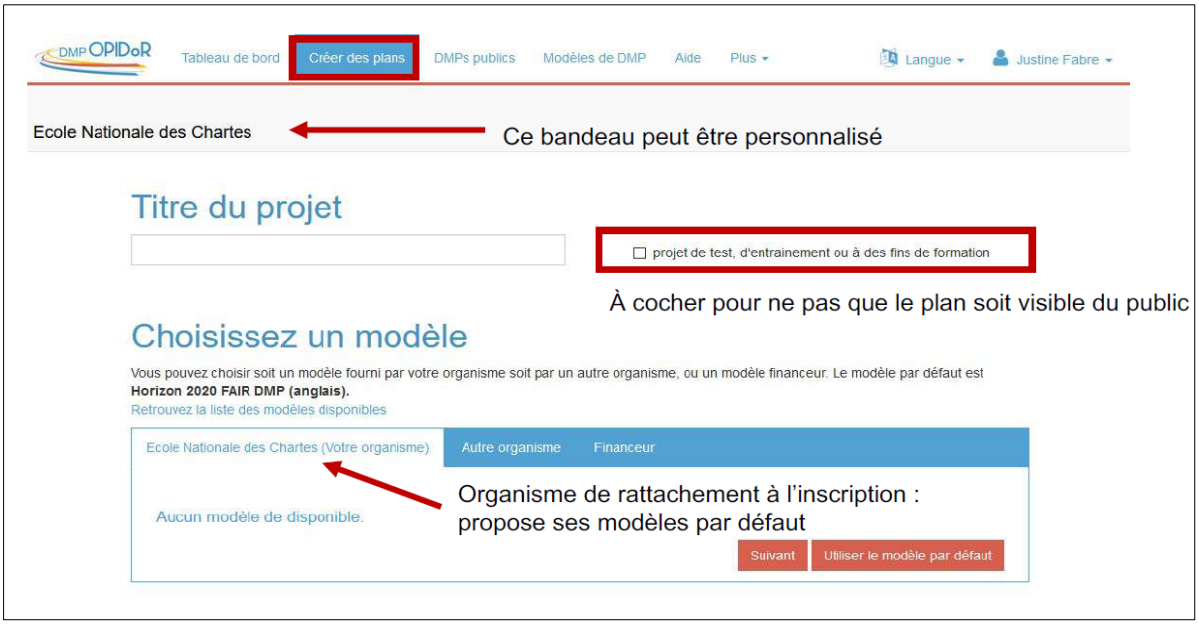
[11] Pour « Optimiser le Partage et l'Intéropérabilité des Données de la Recherche ».

Un exemple commenté[12] de rédaction d'un PGD :

Page d'accueil du PGD à renseigner → créer un compte puis se connecter. L'on pourra ensuite sélectionner le nom de son établissement dans une liste déroulante.

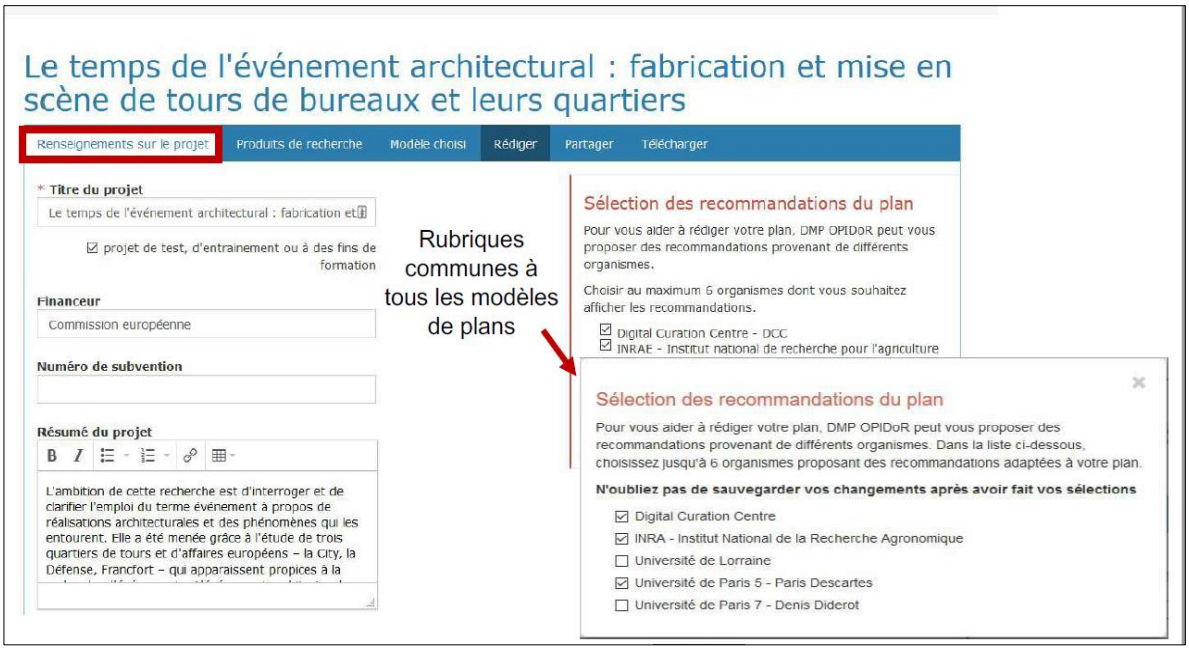


Saisir le titre du projet et choisir un modèle de PGD parmi ceux qui sont proposés :

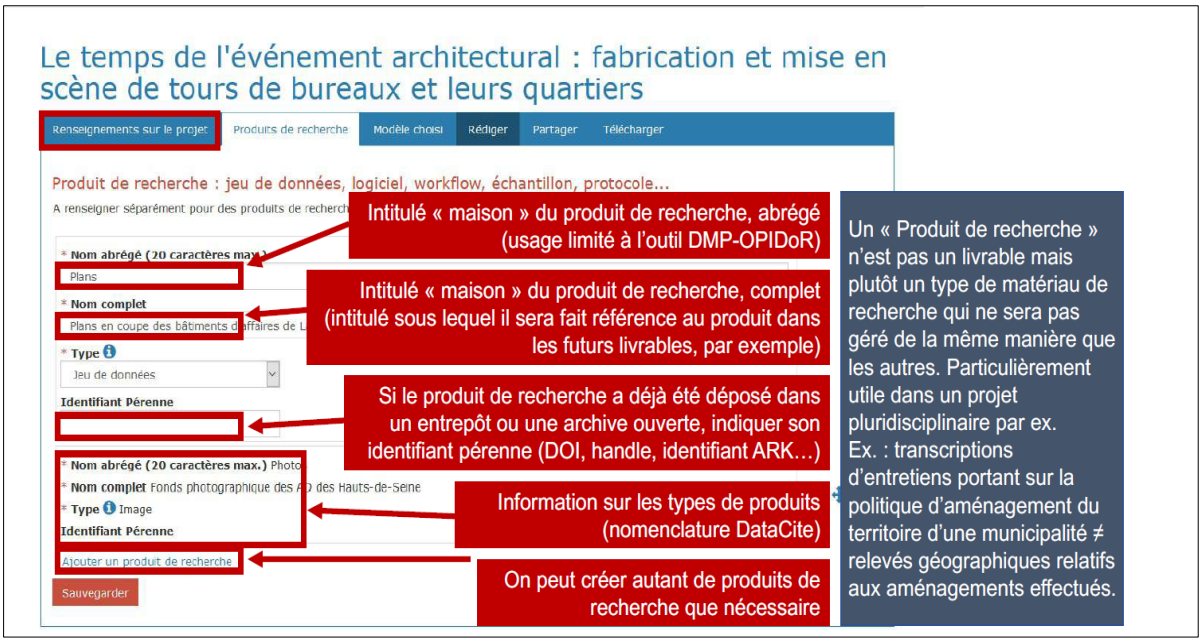


[12] Cet exemple est extrait de la formation proposée par Mariannig Le Béchec, *Plan de gestion de données pour les SHS*, URFIST de Lyon, 24 novembre 2021.

Résumer le projet, après avoir indiqué son financeur et son numéro de subvention :

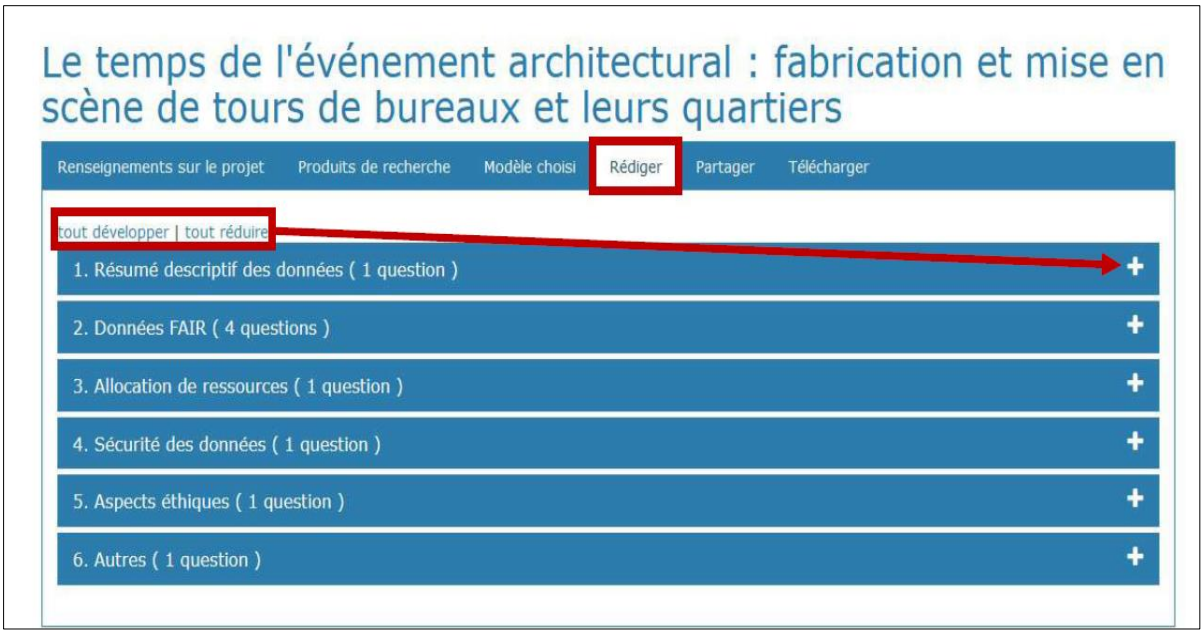


Décrire les produits de recherche associés, qui peuvent être des images, des sons, des corpus textuels, des logiciels, des protocoles…

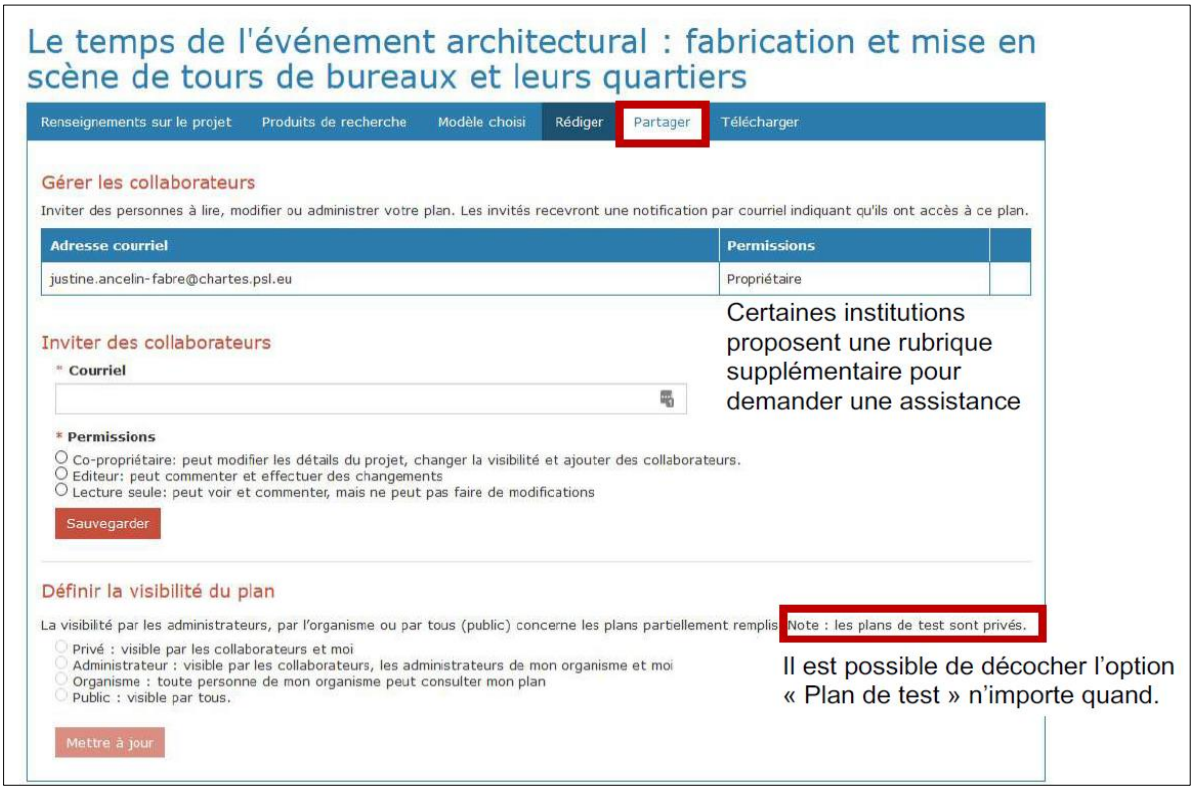


Il faut décrire chaque produit de la recherche séparément (objectif de la collecte ou de la génération des données, types et formats, estimation de leur volume, possibilités de réutilisation) en lui attribuant un identifiant pérenne, puis le joindre au formulaire par une action de glisser-déposer.

Par la suite, les différentes rubriques du PGD s'organisent ainsi :



Spécifier les modalités de partage de ces produits de recherche :



Pour terminer, indiquer les modalités de téléchargement du PGD (format, polices de caractère, etc.).