



Data management plan ? Plan de gestion de données ? DMP OPIDoR vous guide!

Marie-Christine Jacquemot-Perbal, Inist-CNRS marie-christine.jacquemot@inist.fr

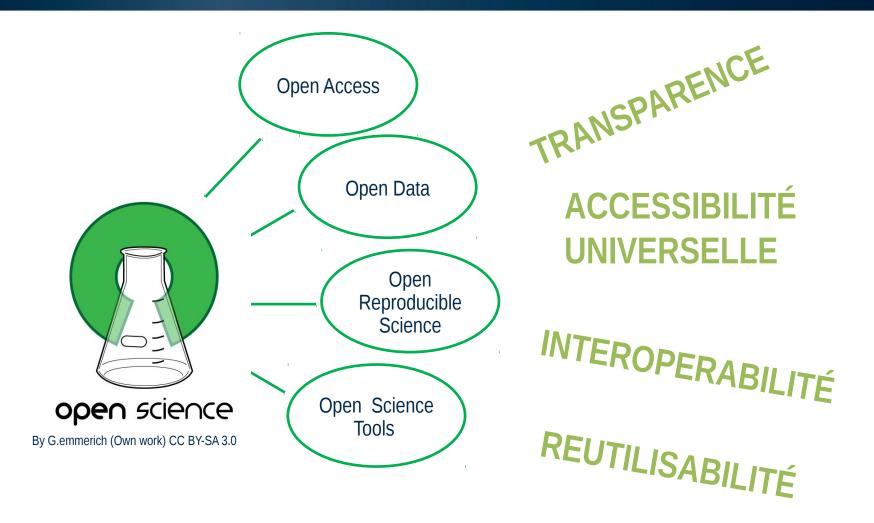
Laurent Rassinoux, Inist-CNRS laurent.rassinoux@inist.fr





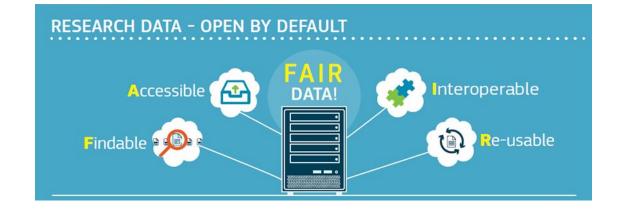
Séries Interopérables et Systèmes de Traitement

Open science



Adhésion aux principes directeurs FAIR







Open Data in Science in Europe

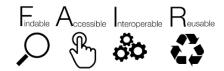
Statement by the European Members of the International Council for Science¹







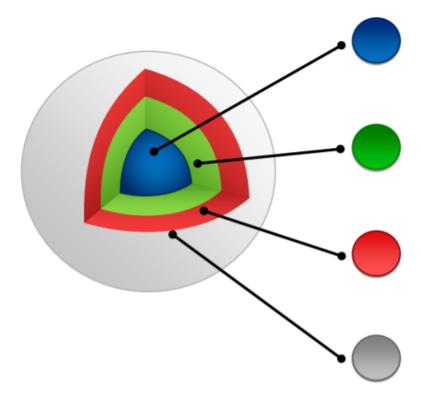
Les principes FAIR



- « Findable » : facile à trouver par l'utilisation d'identifiant global unique et pérenne, des métadonnées riches et le signalement dans un catalogue
- « Accessible »: grâce aux identifiants par des protocoles de communication ouverts et gratuits, +/- protocole d'authentification. Accès permanent aux métadonnées
- « Interoperable »: par l'application de standards pour les formats de représentation des données et métadonnées, l'utilisation de vocabulaires et ontologies, des références vers autres données/métadonnées
- « Reusable » : réutilisable grâce à une description riche, accompagnée si nécessaire d'informations sur la provenance, avec l'utilisation de standards communautaires et de licences

Wilkison et al. (2016) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/

« FAIR data object »



DATA

The core bits

At its most basic level, data is a bitstream or binary sequence. For data to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and code. These layers of meaning enrich the data and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)

Data should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Data should be represented in common and ideally open file formats. This enables others to reuse the data as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code use to process and analyse the data.

METADATA

Contextual documentation

In order for data to be assessable and reusable, it should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the data were created. To enable the broadest reuse, data should be accompanied by a 'plurality of relevant attributes' and a clear and accessible data usage license.

Figure 6: A model for FAIR Data Objects, noting the elements that need to be in place for data to be Findable, Accessible, Interoperable and Reusable.

Source : http://
doi.org/10.5281/zenodo.1285272

« 4 étapes pour réaliser les principes FAIRS »

« Turning FAIR data into reality: interim report from the European Commission Expert Group on FAIR data (Version Interim draft) »

Hodson, Simon, Jones, Sarah, Collins, Sandra, Genova, Françoise, Harrower, Natalie, Laaksonen, Leif, Mietchen, Daniel, Petrauskaité Rūta, Wittenburg, Peter. (2018) http://doi.org/10.5281/zenodo.1285272

- 1. FAIR + ouverture, « évaluabilité » des données, gestion à long terme
- Développer et soutenir un écosystème FAIR durable (politiques, DMP, identifiants, standards, entrepôts)
- 3. Garantir des données FAIR et des services certifiés
- 4. Ancrer une culture FAIR

Standards de données

Bonnes pratiques

- Organisation des répertoires
- Convention de nommage des fichiers
- Gestion des versions
- Formats des fichiers de préférence ouverts, non propriétaires ou standards internationaux
- Structuration des données
- Dictionnaire de données



Référentiel Général d'Interopérabilité

Standardiser, s'aligner et se focaliser pour échanger efficacement

<u>FACILE</u>: outil de validation du format d'archivage du CINES par analyse et expertise



Document 10

Data Management Principles Implementation Guidelines





Standards de métadonnées

Métadonnée : « Ensemble de données structurées décrivant des ressources physiques ou numériques, ou, sur un plan plus fonctionnel, de **l'information structurée** qui décrit, explique, localise la ressource et en facilite la recherche, l'usage et la gestion » (Morel-Pair, 2005)

- Métadonnées de découverte: ex. DataCite,
- Documentation ou mieux des métadonnées contextuelles riches : méthodes, instruments, qualité, provenance, références vers autres données, publications...: ex. EML, ISO19115, Darwin Core, DIF
- Annotation des métadonnées et données à l'aide de vocabulaires contrôlés et ontologies
- ⇒ Augmenter la qualité des données
- ⇒ Faciliter la réutilisation, l'interopérabilité des données au sein d'une discipline et entre les disciplines







Des identifiants uniques, pérennes et « résolvables »

Principaux systèmes d'identifiants uniques et globaux pour les données

Système Exemples

http://doi.org/10.5281/zenodo.267927

DOI http://doi.org/10.23728/b2share.fd8970a8bb59400c93f15618db8a00

e0

Handle https://www.repository.cam.ac.uk/handle/1810/253191

ARK http://n2t.net/ark:/b7272/q6td9v7j

PURL http://purl.oclc.org/fishpest

Identifier aussi les autres produits de la recherche

Workflow https://zenodo.org/record/46567#.WTfigmjyiF4

Logiciels/codes http://doi.org/10.5281/zenodo.1207385

Données de http://www.uniprot.org/uniprot/A0A1R2J2X8

références

Protocoles http://dx.doi.org/10.17504/protocols.io.hk3b4yn

Echantillon de sol https://app.geosamples.org/sample/igsn/IERCO00DC&he

ader=1

=> Accès, citation, reconnaissance, visibilité

Entrepôts de confiance - certification

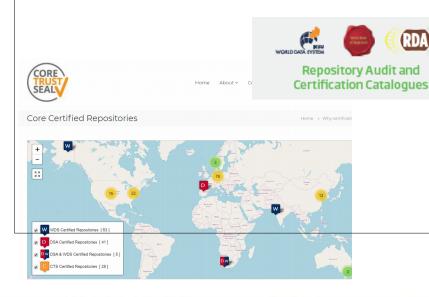
Définition d'un entrepôt de confiance https://smw-rda.esc.rzg.mpg.de/exports/tedt.html#Trusted_Repository "Trusted repositories are those repositories that are undergoing regular assessments according to a set of rules such as defined by Data Seal of Approval (DSA) or TRAC (ISO 16363)."

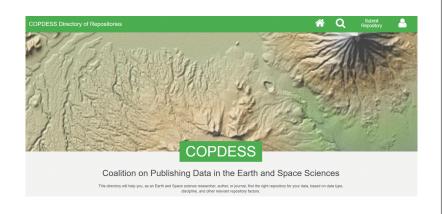
Comment trouver des entrepôts recommandés ?

Registre



Des entrepôts certifiés





Licences encouragées au niveau européen et international

- Licences couramment utilisées pour les données
 - <u>Licences de l'Open Knowledge Foundation</u>
 - ODC-by : Open Database Commons
 - ODC-ODbL : Open Database License (traduction française)
 - O PDDL: Public domain dedication and License
 - Licences Creative Commons version 4.0
 - Internationales
 - Droit des bases de données pris en compte
 - Modulables (4 options, 6 licences et déclaration CC0)



- Outils pour le choix de licences
 - Creative Commons : https://creativecommons.org/choose/
 - Public License Selector Institute of Formal and Applied Linguistics : http://ufal.github.io/public-license-selector/

Licences : décret d'application de la Loi pour une République numérique

Article D323-2-1



Créé par Décret n°2017-638 du 27 avril 2017 - art. 1

- I. L'administration peut soumettre la réutilisation à titre gratuit des informations publiques qu'elle détient aux licences suivantes :
- 1° La licence ouverte de réutilisation d'informations publiques ;
- 2° " L'Open Database License ".
- II. Lorsque ces informations publiques revêtent la forme d'un logiciel, l'administration peut soumettre leur réutilisation à titre gratuit aux licences suivantes :
- 1° Les licences dites " permissives " nommées " Berkeley Software Distribution License ", " Apache ", " CeCILL-B " et " Massachusetts Institute of Technology License " ;
- 2° Les licences " avec obligation de réciprocité " nommées " Mozilla Public License ", " GNU General Public License " et " CeCILL ".
- Les licences susmentionnées sont accessibles en ligne, dans leur version en vigueur, sur le site internet : http://www.data.gouv.fr 🗗.

Catalogues de données

Trouver et réutiliser les données :

Multidisciplinaires













Thématiques/disciplinaires







Politiques des données en France

LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (1)

NOR: ECFI1524250L

ELI: https://www.legifrance.gouv.fr/eli/loi/2016/10/7/ECFI1524250L/jo/texte Alias: https://www.legifrance.gouv.fr/eli/loi/2016/10/7/2016-1321/jo/texte

Directive INSPIRE

Décret relatif aux modalités de contribution obligatoire à l'inventaire du patrimoine naturel et modifiant le code de l'environnement





PRATIQUER UNE RECHERCHE
INTEGRE ET RESPONSABLE

UN GUIDE





Conseil scientifique de l'institut national des sciences de l'univers (INSU)

Recommandation

Sur le partage des données et des résultats de la recherche à l'INSU

Une bonne gestion pour éviter cela...



Copyright © 2017 by Heather Todd and Helen Morganhttp://library.ifla.org/1840/1/122-todd-en.pdf

Data management plan

- Elément clé d'une bonne gestion des données d'un projet de recherche
- Définit quelles données seront créées/collectées, comment elles seront structurées, documentées, partagées et conservées pendant et après le projet
- Document dynamique : initié en amont du projet (de préférence dès le montage du projet) et continuellement mis à jour tout au long du projet
- Incitation voire obligation par : agences de financement de la recherche (pilote H2020 Open Research Data par ex.), organismes de recherche
 - Recommandations, modèles

Planification, Anticipation, Coopération



DMP = opportunité pour établir un dialogue nécessaire pour la bonne conduite du projet

Scientifiques
Informaticiens
Personnels
d'accompagnement

Des services d'appui: informaticiens, professionnels de l'information/data librarians (documentalistes, bibliothécaires, archivistes)

=> répartition des rôles et responsabilités

Des outils: identifier les services liés à la gestion des données à différents stades du cycle de vie

https://cat.opidor.fr/





: un service ouvert pour l'ESR

DMP pour une Optimisation du Partage et de l'Interopérabilité des Données de la Recherche

- Hébergé à l'Inist-CNRS, sur un serveur sécurisé (pas une archive)
- Outil open source développé par une communauté internationale : code commun DMPRoadmap du Digital Curation Center (UK) et l'UC3 (USA)
- Mise à disposition d'une instance de production et d'une instance « bac à sable »
- Assistance conseil pour l'élaboration de DMPs et de modèles de DMPs

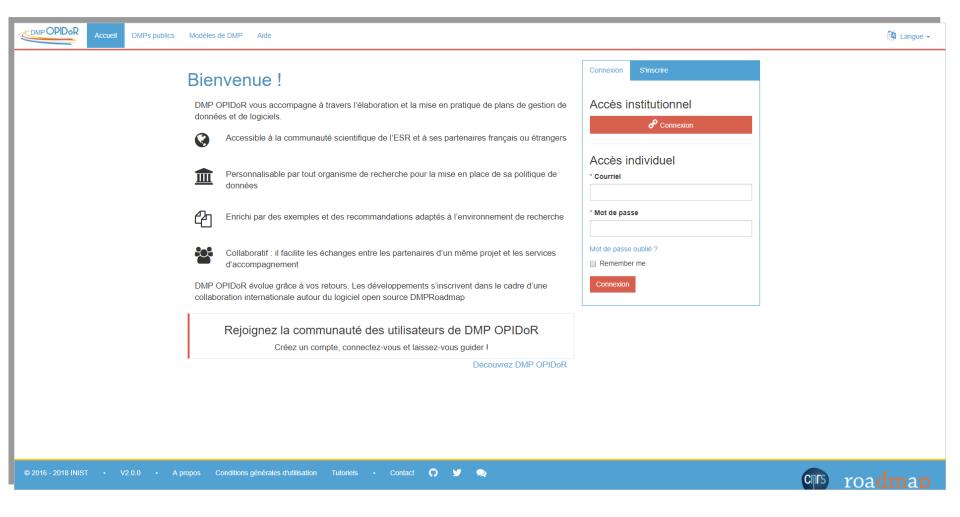
https://dmp.opidor.fr/	Instance de production
info-opidor@inist.fr	Assistance, conseils
dmpopidor@services.cnrs.fr	Liste de discussion
@OPIDoR_INIST	Twitter des services OPIDoR

DEMO

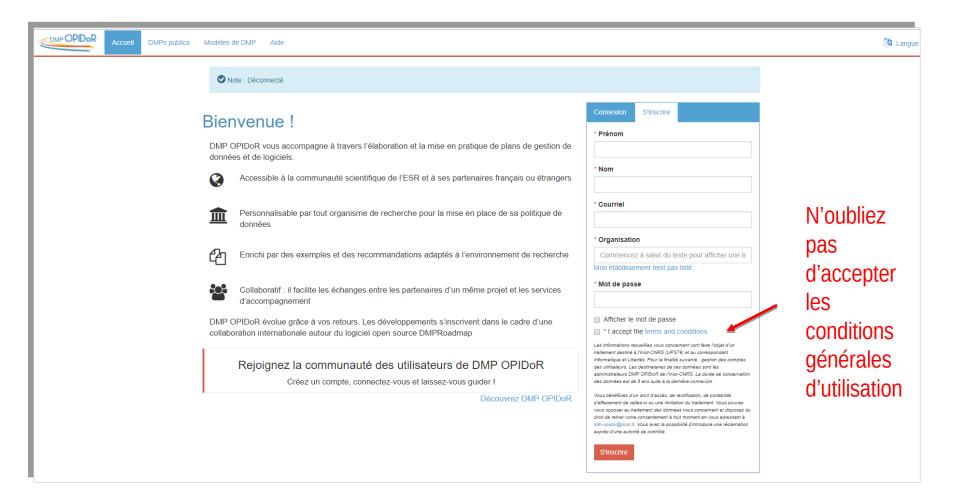


https://dmp.opidor.fr

ACCUEIL https://dmp.opidor.fr/

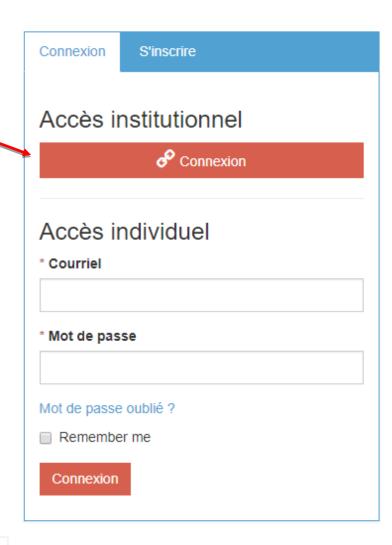


S'inscrire

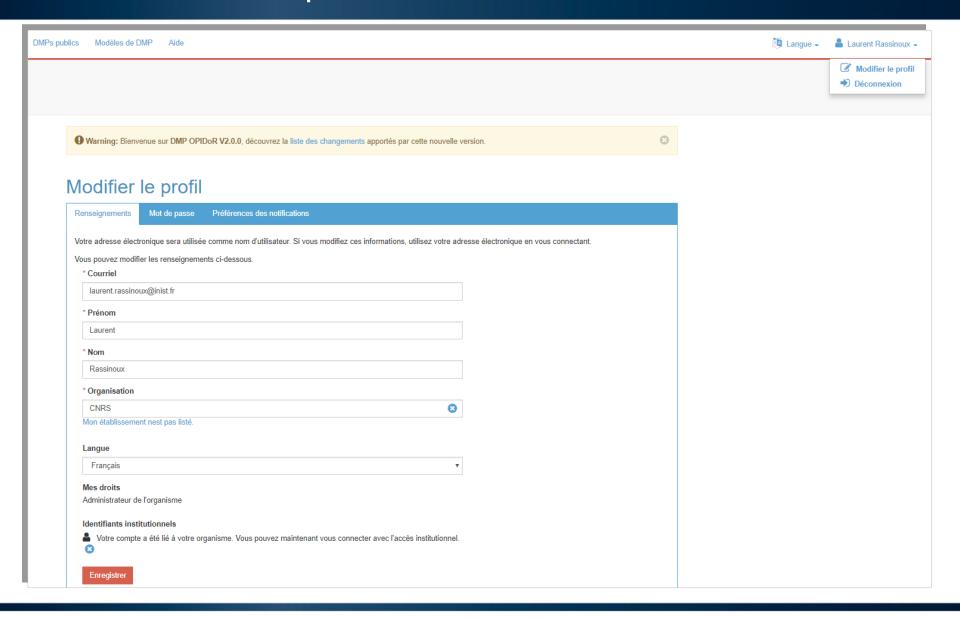


Se connecter

Si vous avez relié votre compte à votre authentification institutionnelle

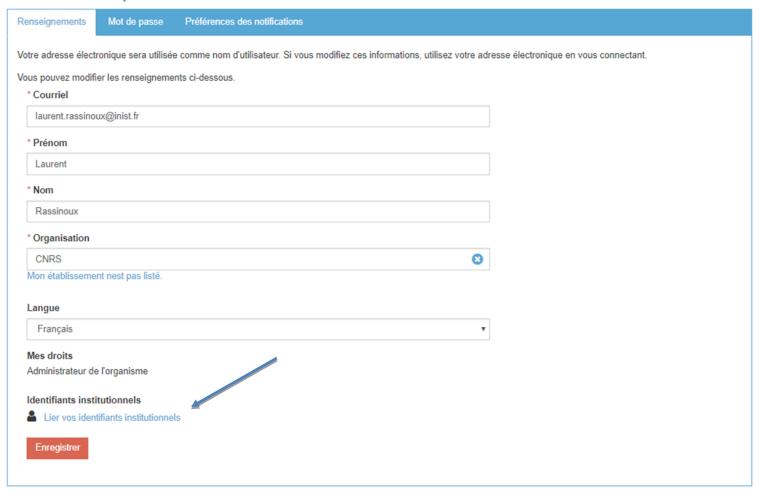


Lier votre compte à l'authentification institutionnelle



Lier votre compte à l'authentification institutionnelle

Modifier le profil



Authentification institutionnelle



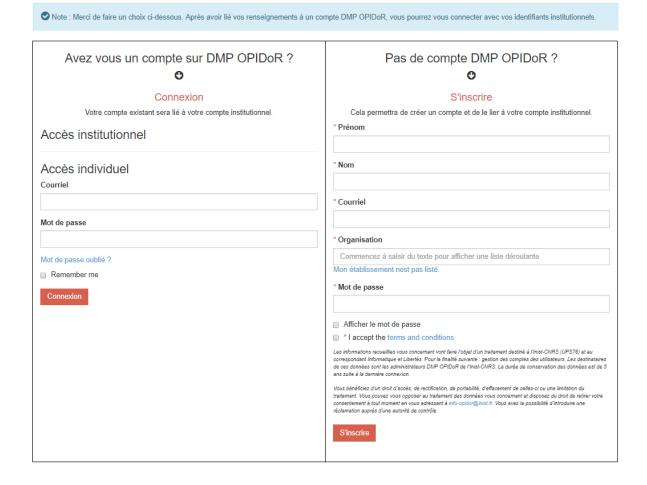
Accueil

MPs publics

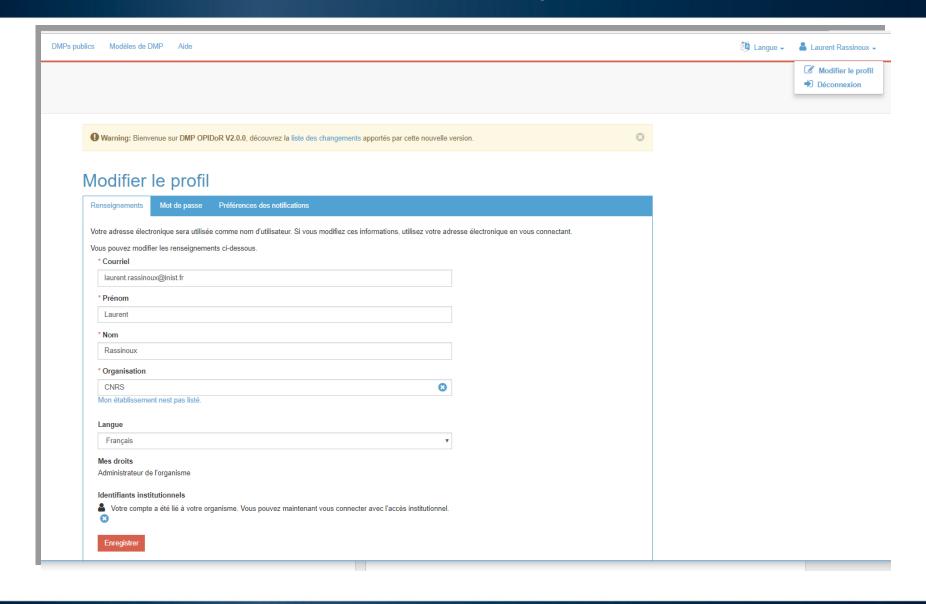
Modèles de DMP

Aide

Vous avez cliquez sur authentification institutionnelle mais vous n'avez pas lié votre compte ou vous n'avez pas encore de compte



Modifier votre profil



Gérer vos notifications

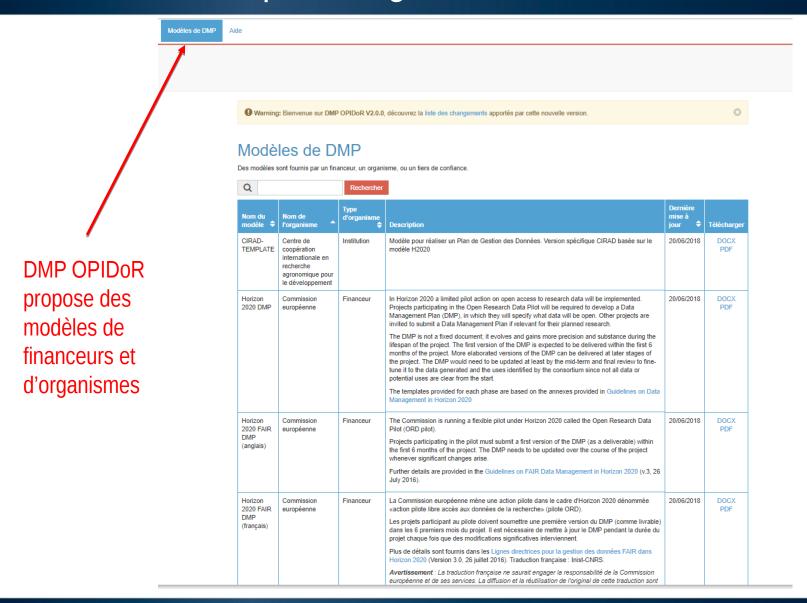
Modifier le profil



Créer un plan de gestion



Créer un plan de gestion : liste des modèles



Créer un plan de gestion : choix d'un modèle

Titre du projet

SIST18 Plan de test, d'entrainement ou créé en vue d'une formation

Choisissez un modèle

Différents type de modèles sont disponibles :

- · Modèles proposés par les organismes : Sélectionner l'organisme et cocher "Pas de financeur"
- · Modèles proposés par les financeurs : Cocher "Pas d'organisme" et sélectionner un financeur
- Modèles de financeurs personnalisés par les organismes : sélectionner l'organisme puis le financeur
- Modèle par défaut H2020 FAIR DMP (anglais): Cocher "Pas d'organisme" et cocher "Pas de financeur"

Plusieurs façons d'opérer suivant vos besoins

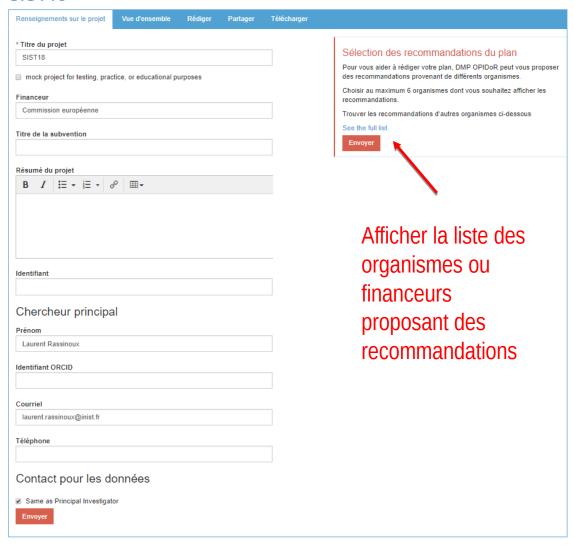
Retrouvez la liste des modèles disponibles

Selectionnez un organisme de recherche :

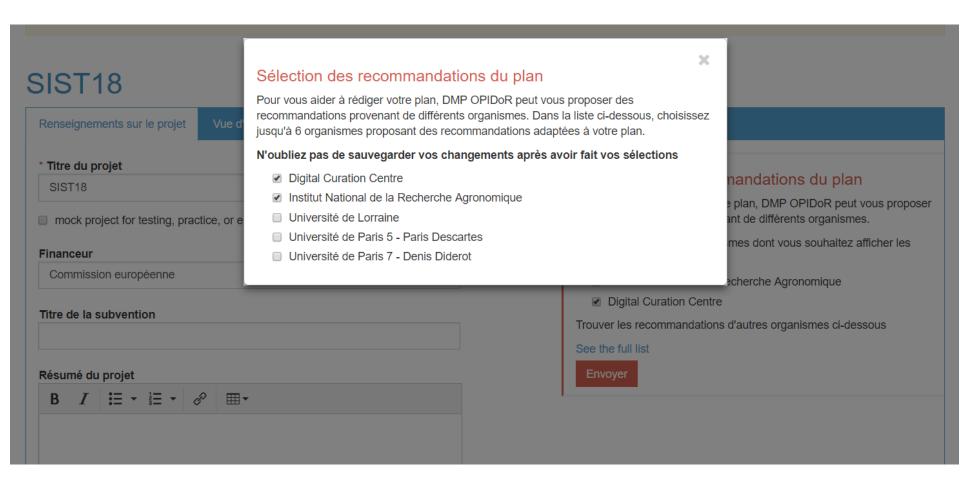
Selectionnez un organisme de recherche.		
Commencez à saisir du texte pour afficher une liste déroulante		- ou - ✓ Pas d'organisme de recherche
Selectionnez un financeur :		
Commission européenne	8	- ou - Pas de financeur associé à ce plan
Selectionnez un modèle :		
Horizon 2020 FAIR DMP (français)	٧	Nous avons trouvé plusieurs modèles correspondants à vos choix.
Horizon 2020 DMP Horizon 2020 FAIR DMP (anglais) Horizon 2020 FAIR DMP (français)		Certains financeurs ou organismes
		proposent plusieurs modèles

Créer un plan de gestion : choix de recommandations

SIST18



Créer un plan de gestion : choix de recommandations



Créer un plan de gestion

SIST18

Renseignements sur le projet

Vue d'ensemble

Rédiger Partager

Télécharger

Horizon 2020 FAIR DMP (français)

Ce plan est basé sur le modèle "Horizon 2020 FAIR DMP (français)" fourni par Commission européenne.

La Commission européenne mène une action pilote dans le cadre d'Horizon 2020 dénommée «action pilote libre accès aux données de la recherche» (pilote ORD).

Les projets participant au pilote doivent soumettre une première version du DMP (comme livrable) dans les 6 premiers mois du projet. Il est nécessaire de mettre à jour le DMP pendant la durée du projet chaque fois que des modifications significatives interviennent.

Plus de détails sont fournis dans les Lignes directrices pour la gestion des données FAIR dans Horizon 2020 (Version 3.0, 26 juillet 2016). Traduction française : Inist-CNRS.

Avertissement : La traduction française ne saurait engager la responsabilité de la Commission européenne et de ses services. La diffusion et la réutilisation de l'original de cette traduction sont régies par la décision de la Commission européenne du 12 décembre 2011 relative à l'utilisation des documents de la Commission (2011/833/UE)

Instructions

Rédiger

- 1. Résumé descriptif des données
 - 1. Fournir un résumé descriptif des données en abordant les points suivants
 - Indiquer l'objectif de la collecte/génération de données.
 - Expliquer le lien avec les objectifs du projet.
 - Préciser les types et formats de données générées/collectées.
 - Préciser si des données existantes sont réutilisées (le cas échéant).
 - Préciser l'origine des données.
 - Indiquer une estimation du volume des données (si possible).
 - Mettre en évidence le potentiel de réutilisation des données : à qui seront-elles utiles ?

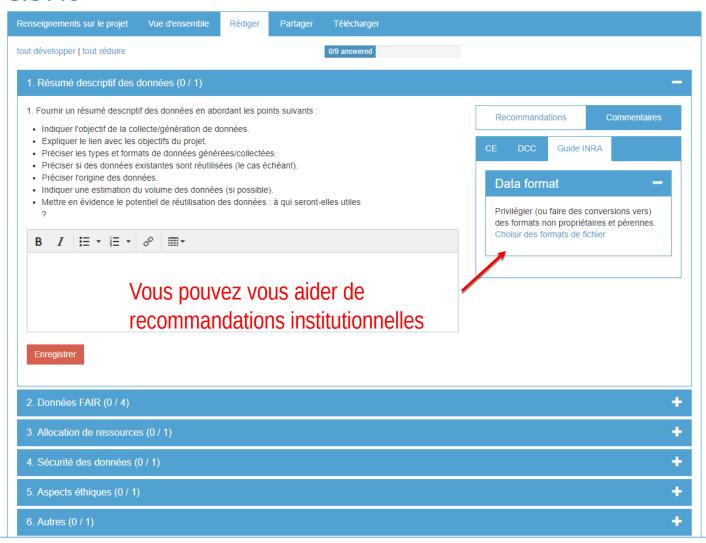
2. Données FAIR

- o 2.1. Rendre les données faciles à trouver, en incluant la mise à disposition des métadonnées
 - Indiquer comment les données seront facilement découvertes (description par des métadonnées).
 - Décrire la procédure d'identification des données et faire référence à un système standard d'identification. Utilisez-vous des identifiants pérennes et uniques tels que les DOL?
 - Définir les conventions de nommage utilisées.
 - Préciser les mots-clés pour faciliter la recherche.
 - Décrire comment seront gérées les versions.
 - Préciser les standards de création de métadonnées (le cas échéant). Si aucun standard n'existe dans votre discipline, décrire quelles métadonnées seront créées et comment.

Le plan est créé, vous arrivez sur un récapitulatif des questions

Créer un plan de gestion : les recommandations

SIST18



Créer un plan de gestion : les recommandations



Par exemple, la recommandation de l'INRA concernant les formats de données pointent vers son site INRA Datapartage

Gérer

Principes FAIR

Plan de gestion

DOI

Nommer et organiser ses fichiers de données

Documenter les données

Stocker

Choisir des formats de fichier

Archiver de manière pérenne

Partager / Publier

Réutiliser

Technologies

Documents de référence

gestion et partage des données scientifiques 8

- Accueil Gérer Choisir des formats de fichier
- Choisir des formats de fichier

Enjeux et critères de choix

Le principal enjeu d'un choix judicieux d'un format de fichier est la préservation de sa lisibilité sur le long terme, le principal risque...

> Lire la suite

Formats de fichiers par catégorie de données

Plusieurs sites maintiennent à jour des listes de formats pour différentes catégories de données, tout en mettant en avant les formats qui sont...

> Lire la suite

Boîte à outils



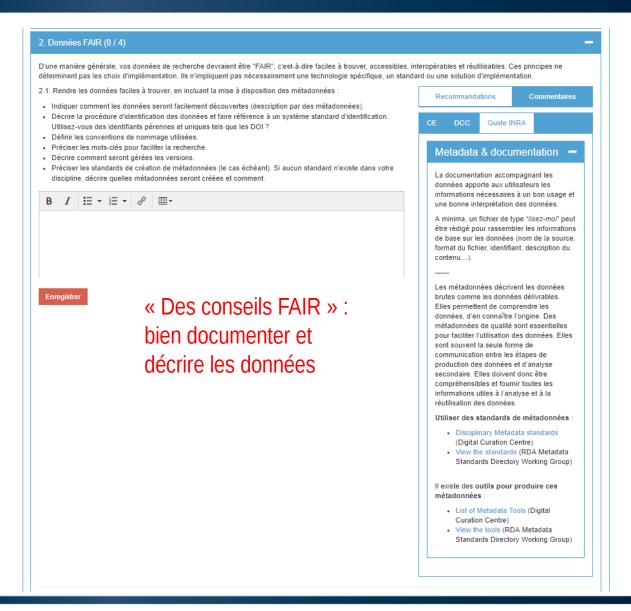
- DMP OPIDoR
- Puis-je diffuser mes données
- DOI Inra
- Choisir un entrepôt
- > Portail Data Inra
- Citer des données
- Publier un vocabulaire/une ontologie

On vous signale

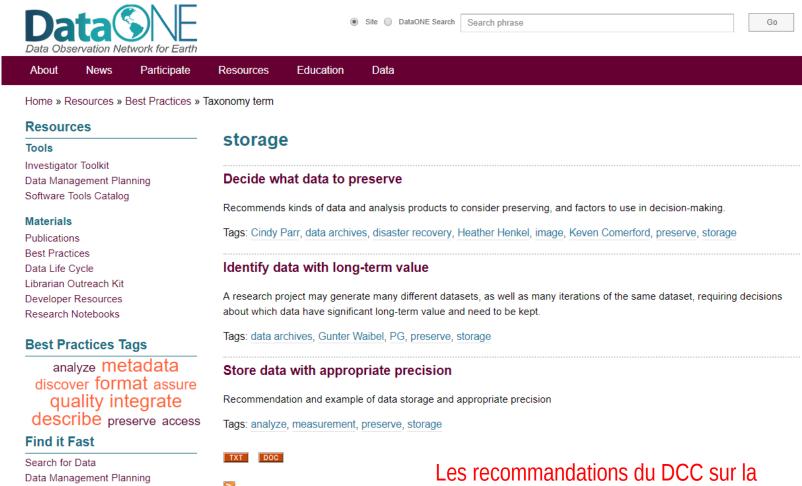


- Questions/Réponses
- Formations et tutoriels
- S'abonner aux actualités

Créer un plan de gestion : les recommandations







conservation et le stockage vous conduisent sur le site de DataOne

Best Practices

Software Tools Calendar

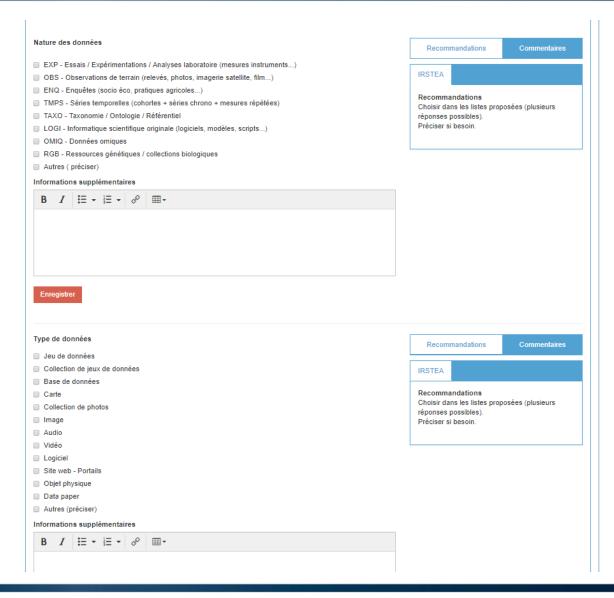
Ask DataONE

Une recommandation concernant le partage de données : cadre juridique, embargo

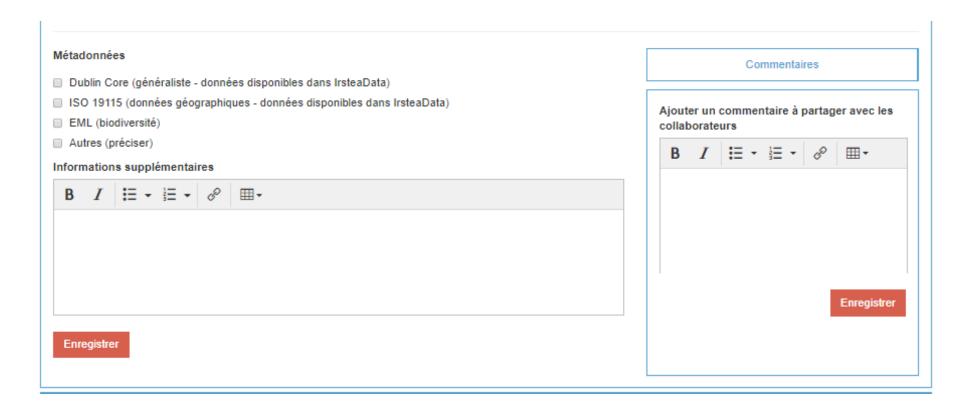


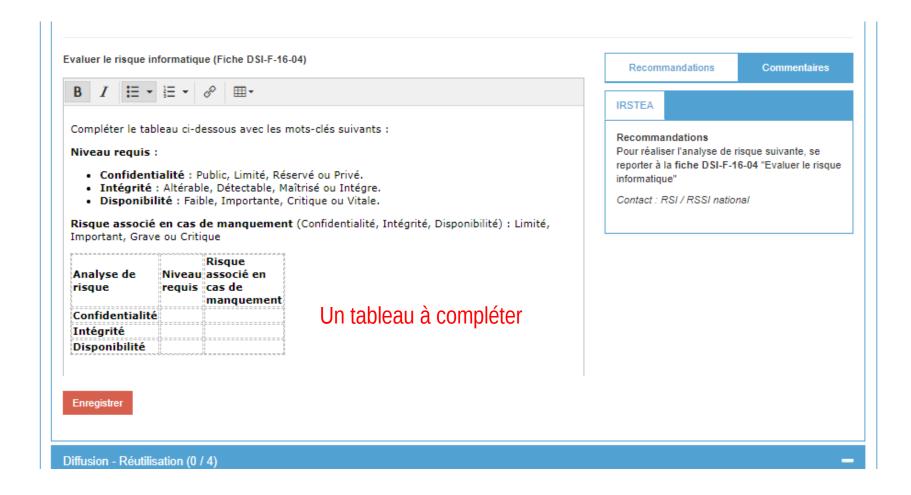
Un lien sur Datapartage (INRA) où vous retrouverez entre autres le guide "Ouverture des données de recherche. Guide d'analyse du cadre juridique en France"

L'IRSTEA propose une liste avec des valeurs à cocher pour décrire les données

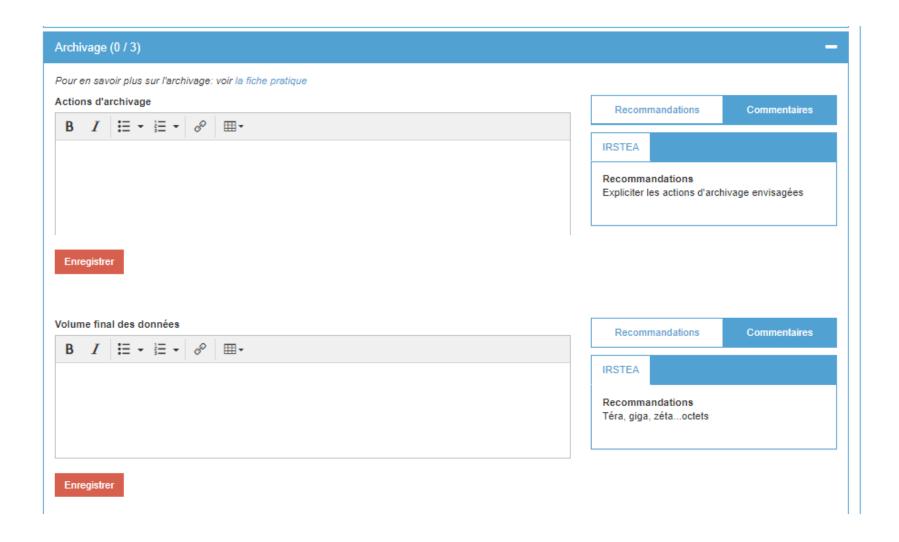


Ou bien aussi pour les standards de métadonnées utilisés

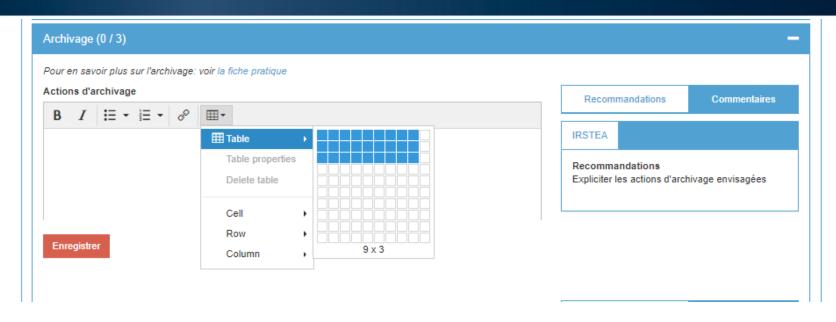


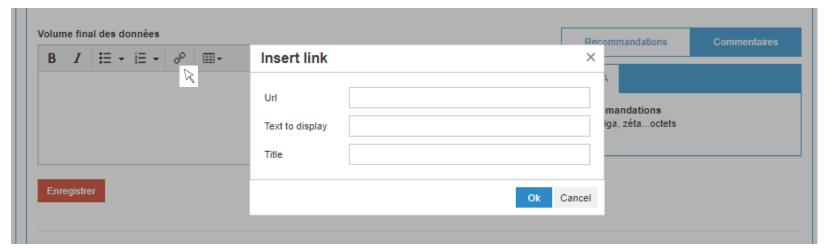


Créer un plan de gestion : l'éditeur



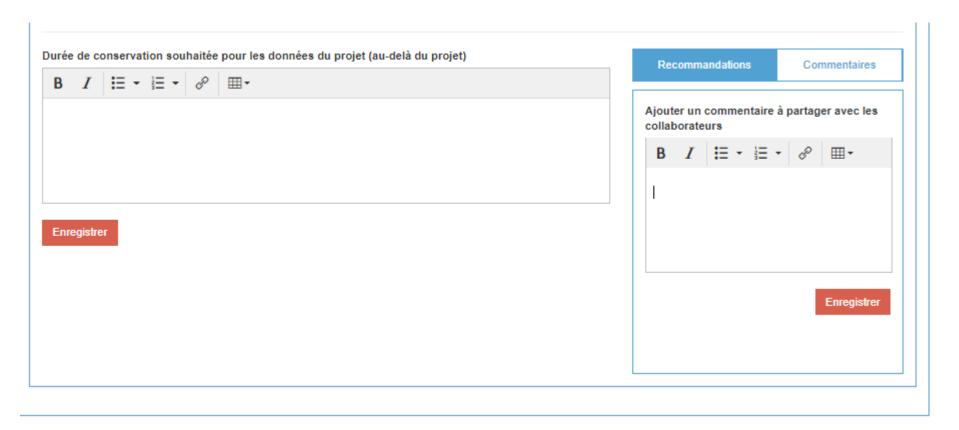
Rédiger un plan de gestion : l'éditeur



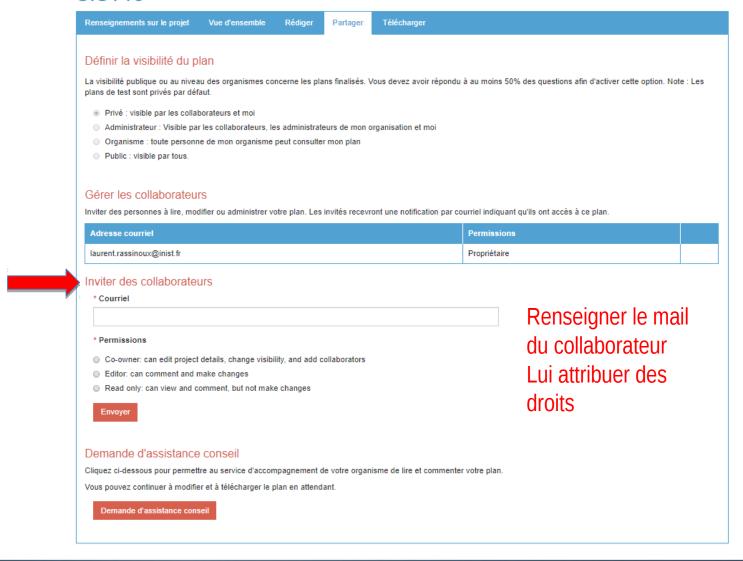


Outil collaboratif : laisser des commentaires

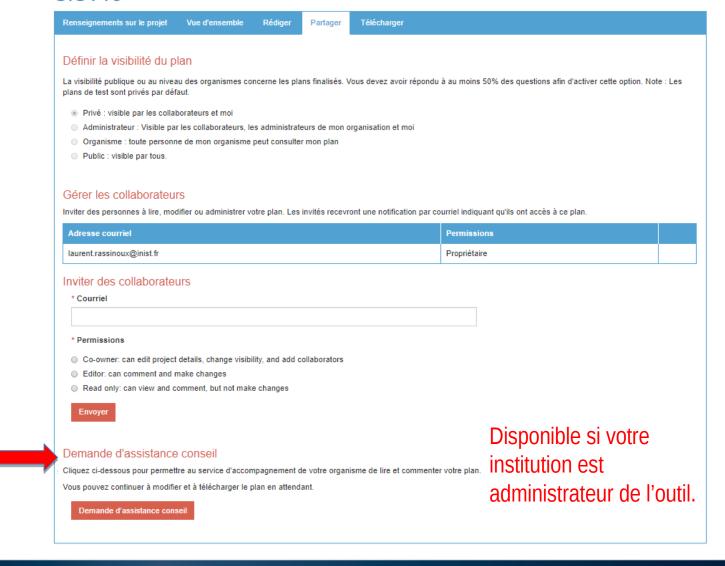
Vous pouvez ajouter des commentaires visibles par des collaborateurs



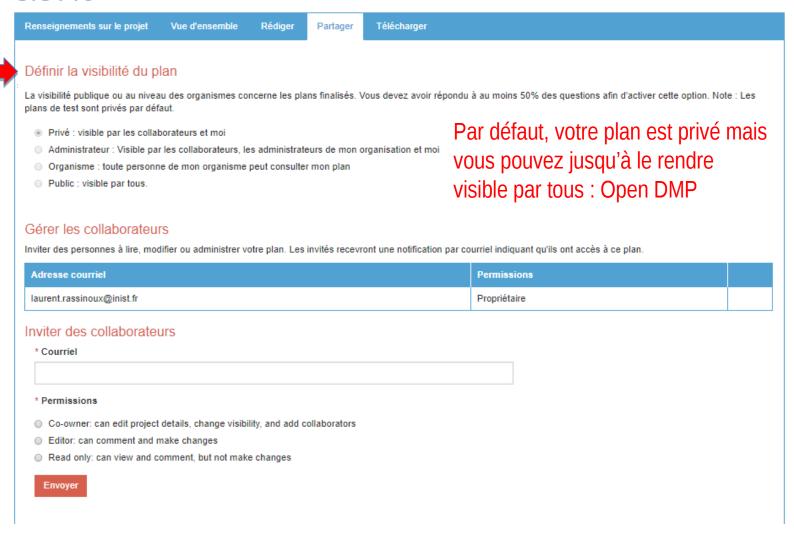
Outil collaboratif : partager votre plan de gestion



Outil collaboratif : demande d'assistance conseil



Outil collaboratif : visibilité de votre plan



Exporter votre plan de gestion de données

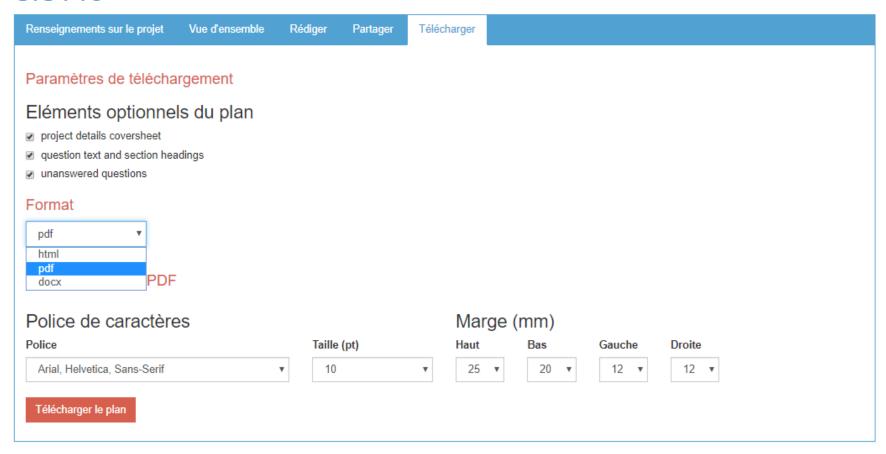


Tableau de bord : gérer ses plans



Tableau de bord

Créer un plan

DMPs publics

Modèles de DMP

Aide

Tableau de bord

Dans le tableau ci-dessous figurent les plans que vous avez créés, ainsi que ceux que vous partagez avec d'autres. Vous pouvez à tout moment les modifier, les partager, les télécharger, en faire une copie ou les supprimer.

Titre du projet	\$	Template	\$	Modifié →	Rôle	Test	Visibilité	Partagé	
SIST18		Horizon 2020 FAIR DMP (français)		21/06/2018	Propriétaire		Privé	Non	<u>Actions</u> ▼
									Modifier
Create plan									Partager
									Télécharger
									Créer une copie

Supprimer

Interface Administrateur

- DMP OPIDoR dispose d'une interface administrateur permettant à des institutions de :
- Créer leur(s) modèle(s) institutionnel(s) ou personnaliser un modèle de financeur
- Créer leur(s) recommandation(s)
- Personnaliser l'interface
- Offrir aux utilisateurs DMP OPIDoR affiliés aux institutions administratrices une possibilité de demander de l'assistance

Sur demande auprès de l'équipe DMP OPIDoR à info-opidor@inist.fr

AIDE ressources, exemples et guides

Aide

À propos de DMP OPIDoR

Ressources, exemples et conseils

Ressources utiles pour rédiger un Data ou un Software Management Plan

- Réaliser un plan de gestion de données 'FAIR': modèle [PDF, 33 pages]
 Reymonet N., Moysan M., Cartier A., Délémontez R. Université Paris Diderot et Université Paris Descartes. 2018. Ce document a pour vocation d'accompagner les chercheurs et chargés de projets dans la rédact.. 2018.
- <u>Découvrir des plans de gestion de données de la recherche, en 4 points [PDF, 6 pages]</u>
 Deboin M.C. 2014. Montpellier (FRA): CIRAD.
- Tutoriels « Le libre accès aux résultats de la recherche dans le cadre d'Horizon 2020 » [page web] Inist-CNRS 2015
- Modèle de Plan de Gestion de Logiciel de la Recherche V3.1, Projet PRESOFT, janvier 2018 [PDF, 18 pages] Gomez-Diaz T., Romier G. 2018.
- Checklist for a Software Management Plan. v0.1 [PDF, 25 pages]

The Software Sustainability Institute. 2016.

- Guidelines on FAIR Data Management in Horizon 2020 [PDF, 12 pages]
 European Commission. 2016, v3.0.
- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 [PDF, 11 pages]
 European Commission, 2017, v3.2.
- ERC Data Management Plan Template [page web]
- European Research Council. 2017.
- Checklist for a Data Mangement Plan [page web]
 Digital Curation Centre. 2013, v.4.0. Edinburgh (UK).

Une liste de 13 questions et de recommandations associées, qui présentent les principales problématiques abordées dans les DMPs.

 Adapt your Data Management Plan. A list of Data Management Questions based on the Expert Tour Guide on Data Management [PDF, 6 pages] CESSDA. 2017.

Exemples de Data Management Plans

Quelques exemples de DMP rédigés dans le cadre de projets Horizon 2020 :

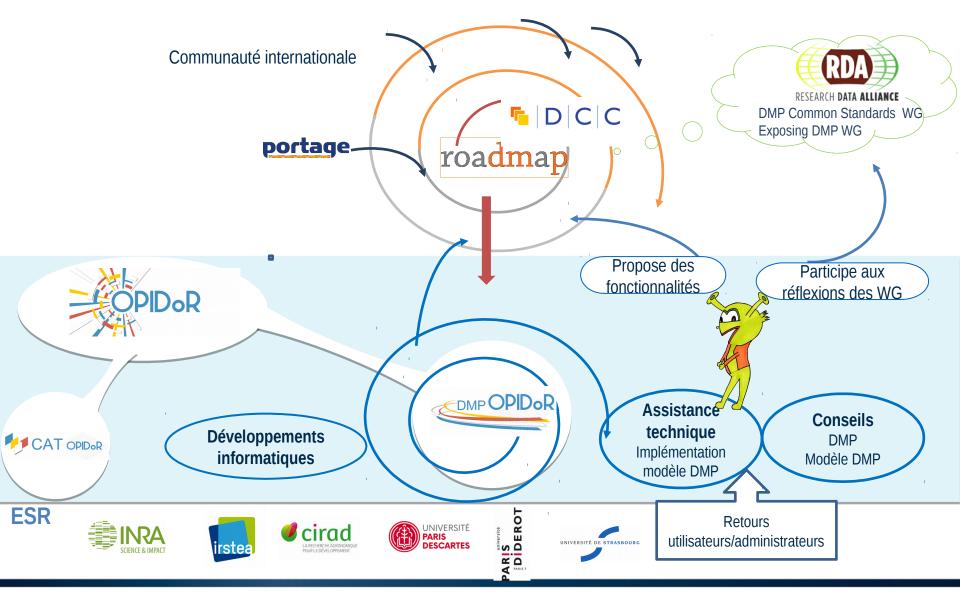
- AMECRYS Revolutionising Downstream Processing of Monoclonal Antibodies by Continuous Template-Assisted Membrane Crystallization. [PDF, 23 pages]
- PIE News Poverty, Income, and Employment News NEWS. [PDF, 29 pages]
- CompBiomed A Centre of Excellence in Computational Biomedicine. [PDF, 15 pages]
- AQUASCOM Network of Leading European AQUAtic MesoCOSM Facilities Connecting Mountains to Oceans from the Arctic to the Mediterranean; [PDF, 18 pages]

Autres exemples de DMPs

- Rio journal Data Management Plan collection [page web]
- Une sélection de plans de gestion de données publiés dans RIO (Research Ideas and Outcomes journal).
- Two social science DMPs [PDF, 7 pages]
- Exemples de plans réalisés par des chercheurs de l'Université de Leeds, partagés dans le cadre des supports de formation Leeds RoaDMaP.
- Psychology DMP IPDF, 11 pages1

Un DMP fictif très détaillé en psychologie élaboré par le projet pédagogique DMTpsych RDM, à partir de travaux expérimentaux précurseurs.

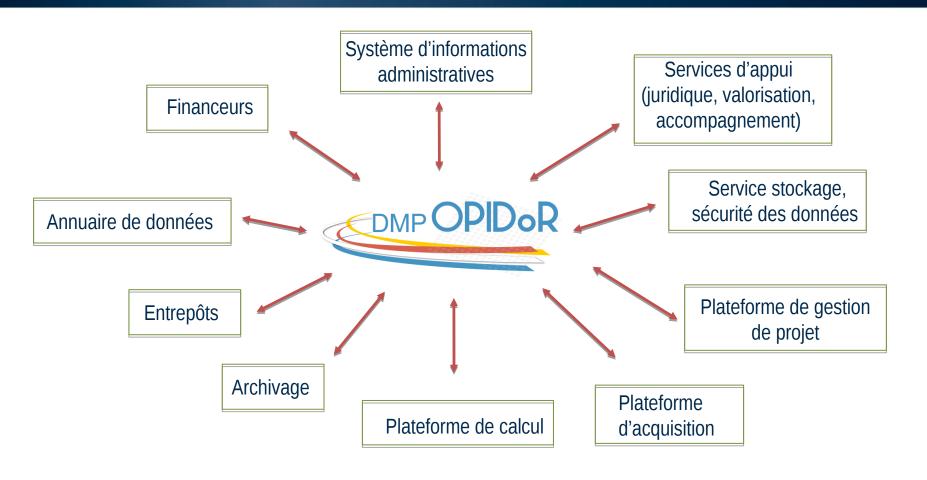
Evolution de DMP OPIDoR



Perspectives

Evolutions techniques	DMP Roadmap/DMP OPIDoR	Amélioration des fonctionnalités • API, duplication de la description des jeux de données
Standardisation de la structure du DMP Harmonisation par discipline	RESEARCH DATA ALLIANCE SCIENCE EUROPE Shaping the future of research	 DMP Common group Standards WG (RDA) Développer un modèle informationnel commun qui rendra le DMP exploitable par les machines et interopérables Domain Protocols for Research Data Management Modèles de DMP spécifiques à une discipline
Autres produits de la recherche à gérer	Software Sustainability Institute PRESOFT wellcome	 Logiciels/codes et autres matériels Software management plan Outputs management plan : données, logiciels, matériels (anticorps, lignées cellulaires, réactifs)

Intégration dans le workflow de recherche : une illustration



Coordination, alimentation semi-automatique, réutilisation des informations => Qualité des données, inventaire dynamique



Merci de votre attention

