



Planifier la gestion des données de recherche

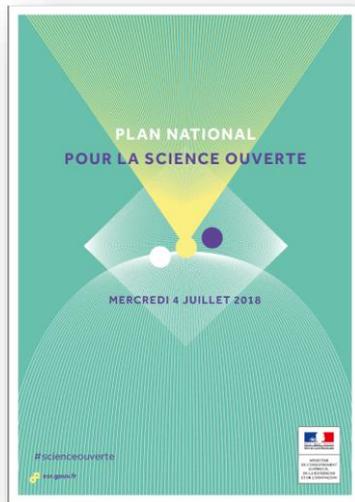
Le plan de gestion de données (DMP) un élément clé pour la mise en œuvre de bonnes pratiques de gestion et rendre les données FAIR.

07 février 2020

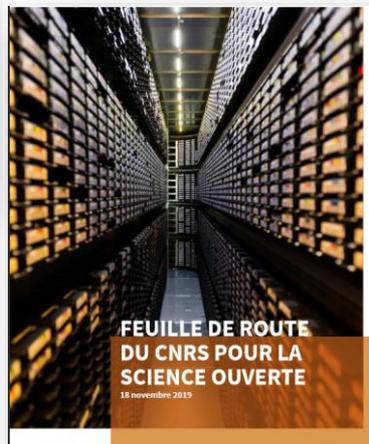
Journées “Préserver ses données numériques” - TGIR
Huma-Num - Centre de colloques du Campus Condorcet

POUR STRUCTURER ET OUVRIR LES DONNÉES...

« [...] les données produites par la recherche publique française soient progressivement structurées en conformité avec les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable), préservées et, quand cela est possible, ouvertes. »



DEUXIÈME AXE : **STRUCTURER ET OUVRIR LES DONNÉES DE LA RECHERCHE**

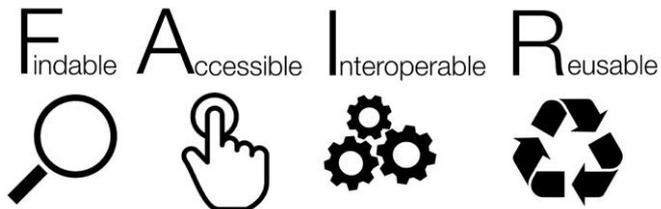


... développer une culture de la gestion/partage des données chez tous les acteurs du cycle de vie de la donnée ...

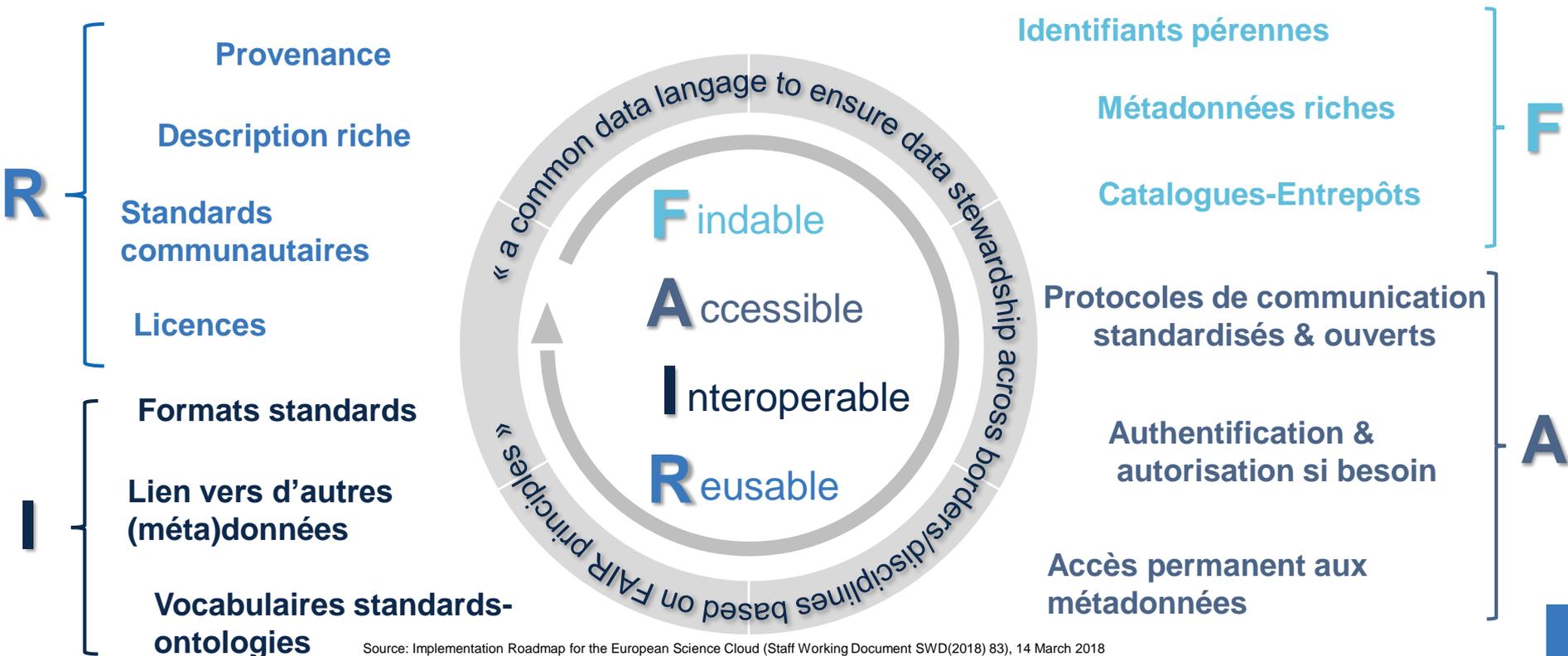
« LE MOUVEMENT DE PARTAGE DES DONNÉES DE LA RECHERCHE A ÉTÉ INITIÉ PAR LES CHERCHEURS EUX-MÊMES, QUI SE SONT ORGANISÉS, PAR DISCIPLINE SCIENTIFIQUE... »

LE TRAVAIL DE LA SCIENCE ET LE NUMÉRIQUE
- Données, publications, plateformes Une analyse systémique de la loi pour une République numérique

« Aussi ouvert que possible,
aussi fermé que nécessaire »



EN CONFORMITÉ AVEC LES PRINCIPES FAIR



Source: Implementation Roadmap for the European Science Cloud (Staff Working Document SWD(2018) 83), 14 March 2018

Des données FAIR plus faciles à partager et réutilisables par les hommes et par les machines



LE PLAN DE GESTION DE DONNÉES : UN ÉLÉMENT CLÉ POUR PRODUIRE DES DONNÉES FAIR

« Généraliser la mise en place de plans de gestion des données dans les appels à projets de recherche. »



Garantir des données fiables et bien gérées tout au long du projet, compréhensibles, disponibles et préservées sur le long terme pour une réutilisation future (démarche FAIR)

Favorise la mise en place de bonnes pratiques de gestion de données tout au long des projets

L'ANR met en place un plan de gestion des données pour les projets financés dès 2019.



MODÈLE
INTÉGRÉ
DANS DMP
OPIDOR

PLAN DE GESTION DES DONNÉES

IMPORTANCE SUR LE PLAN INTERNATIONAL

- **Echelle mondiale** : DMP de plus en plus recommandé ou exigé
- **Echelle européenne** : obligation pour Horizon 2020
- **En France** : obligation pour l'ANR
- **Echelle des organismes** : recommandations, politique d'établissements...
- **Echelle disciplinaire** : modèles de DMP par domaine

Les exigences des financeurs



Les plans de gestion de données - S. Cocard et D. L'Hostis, INRA, URFIST
Paris - 05 avril 2019

18



Le plan de gestion de données

UN ÉLÉMENT CLÉ POUR LA MISE EN ŒUVRE DE BONNES PRATIQUES DE GESTION ET RENDRE LES DONNÉES FAIR

DEFINITION

Data Management Plan - DMP

Document qui définit quelles données seront créées/collectées, comment elles seront structurées, documentées, partagées et conservées pendant et après le projet

Le cycle de vie des données

C'est l'ensemble des étapes de gestion, de conservation et de diffusion des données de recherche associées aux activités de recherche.



D'après Research data lifecycle – UK Data Service
<https://www.ukdataservice.ac.uk/manage-data/lifecycle>

PLAN DE GESTION DES DONNÉES

OUTIL DE GESTION DE PROJET



Document **évolutif**
(3 versions minimum)



Aide à bien organiser
les données



Description des données
selon le cycle de vie



Définit les
responsabilités



Aide à évaluer les
ressources nécessaires



Aide à obtenir des
données fiables

PLAN DE GESTION DES DONNÉES

ACTEURS ET CONTRIBUTEURS

Recommandations

- Structures institutionnelles
 - Financeurs
 - Organismes de recherche, universités
 - Infrastructures / centres de calculs
- Editeurs

Réseau d'aide



PLAN DE GESTION DES DONNÉES

EXEMPLE : MODÈLE SCIENCE EUROPE ANR

- Modèle composé de **6 grandes thématiques** illustrant les bonnes pratiques de gestion et de partage

Description des données et collecte ou réutilisation des données existantes

Documentation et qualité des données

Stockage et sauvegarde pendant le processus de recherche

Exigences légales et éthiques, codes de conduite

Partage des données et conservation à long terme

Responsabilités et ressources en matière de gestion des données

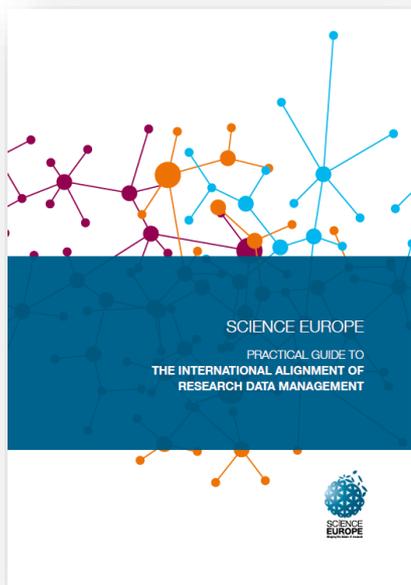


Modèle de Plan de gestion des données (PGD)



Science Europe. (2018). *Practical guide to the international alignment of research data management.*

PLAN DE GESTION DE DONNÉES



- Quelles données seront produites/utilisées au cours du projet ? (type, format, volume et accroissement...).
- Comment seront-elles produites ou transformées ?

Data description and collection or re-use of existing data



- Comment les données seront-elles identifiées, décrites ?
- Quels standards de métadonnées utilisera-t-on ?
- Comment seront générées les métadonnées ?

Documentation and data quality



- Comment, où, par qui, seront stockées, sauvegardées et sécurisées les données ?

Storage and backup



- Qui sera propriétaire des données produites ?
- Des données sensibles seront-elles produites ou utilisées ?
- Comment sera assurée leur anonymisation ?
- Des données externes seront-elles utilisées ?

Legal and ethical requirements



- Qui pourra accéder aux données ?
- Les données seront-elles publiées ?
- Comment ?
- Dans quel délai ?
- Sous quelle licence ?
- Quel est le plan d'archivage et de préservation à long terme ?

Data sharing, long term preservation



- En quoi consiste le projet ?
- Qui sont les partenaires ?
- Quelle est la politique de gestion des données ?
- Qui est responsable de la gestion des données ?
- Comment la gestion et le partage des données sont-ils financés, en particulier à long terme ?

Data management responsibilities and resources



INFRASTRUCTURES DE RECHERCHE EN SHS

CATÉGORIE	NOM	NOM COMPLET	ESFRI
TGIR	Huma-Num ¹	Humanités Numériques	DARIAH (2006) CLARIN (2006)
TGIR	Progedo	PROduction et GEstion de DONnées	ESS (2006) CESSDA (2006) SHARE (2006)
IR	ERIHS-FR ²	European Research Infrastructure for Heritage Science	ERIHS (2016)
IR	MÉTOPES ³	Données de recherche pour l'édition structurée	
IR	OpenEdition ⁴	Revue ouverte en Sciences humaines	
IR	RnMSH ⁵	Réseau national Maison des Sciences de l'Homme	

DONNEES DE RECHERCHE

SERVICES POUR LES DONNÉES (FAIR) – EXEMPLE DES SERVICES DU TGIR HUMA-NUM



TRAITER

Outils . Logiciels

Description des données et collecte ou réutilisation des données existantes

SIGNALER

Enrichissement sémantique
Accès unifié



isidore

Documentation et qualité des données



DIFFUSER

Machines virtuelles
Diffusion web

Partage des données et conservation à long terme



STOCKER

Entreposer . Organiser

Stockage et sauvegarde pendant le processus de recherche

Ex
ét

EXPOSER

Documenter . Partager



nakala

nakala

Responsabilités et ressources en matière de gestion des données

ARCHIVER

Préservation à long terme



DMP OPIDoR : aide à la création en ligne de DMPs



<https://dmp.opidor.fr>

Data Management Plan pour une Optimisation du Partage et de l'Interopérabilité des Données de la Recherche

Outil mis à disposition de l'ESR pour :

- Faciliter l'élaboration en ligne de DMP (et SMP)
- Favoriser la mise en place de bonnes pratiques de Gestion des données tout au long des projets de recherche
- Faciliter la mise en place des politiques des données des organismes de recherche

Outil open source développé par une communauté internationale :
code commun DMPRoadmap du Digital Curation Center (UK) et l'UC3 (USA)

- serveur sécurisé hébergé à l'Inist-CNRS, en France



MODÈLES DISPONIBLES – DMP OPIDoR

- Agences de financement
 - H2020 (FAIR DMP), ERC, ANR
- Organismes de recherche / Universités
 - Cirad, Inrae, Institut Pasteur, Sciences Po
 - Unistra (Université de Strasbourg), Université Paris Descartes & Paris Diderot
 - CC In2p3
 - MaSa - Mémoires des archéologues et des sites archéologiques - (à venir)
- Autres :
 - Projet PRESOFIT, SSI (UK)
 - Digital Curation Centre (UK), EPFL

Modèle par défaut (à venir) : Science Europe

NEW

ANR - Modèle de PGD (français)

1. DESCRIPTION DES DONNÉES ET COLLECTE OU RÉUTILISATION DE DONNÉES EXISTANTES

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

2. DOCUMENTATION ET QUALITÉ DES DONNÉES

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

4. EXIGENCES LÉGALES ET ÉTHIQUES, CODES DE CONDUITE

...

5. PARTAGE DES DONNÉES ET CONSERVATION À LONG TERME

...

6. RESPONSABILITÉS ET RESSOURCES EN MATIÈRE DE GESTION DES DONNÉES

...

Disponible sur https://dmp.opidor.fr/public_templates



maDMP : machine actionable DMP

***UN DMP PLUS ACTIF ET INTÉGRÉ
DANS LE WORKFLOW DE LA
RECHERCHE***

DES CONSTATS

Document textuel peu structuré (PDF, docx)

Limites de l'outil

Difficulté à décrire plusieurs jeux de données nécessitant une gestion différente

Pour les chercheurs

- Perçu comme un exercice administratif et non comme faisant partie intégrante de la pratique de la recherche
- Terminologie trop technique : métadonnées, ontologies (mapping), interopérabilité...
- Répétition de la saisie d'informations déjà saisies dans d'autres systèmes d'information
- Pas/peu d'expertise dans le choix du standard de métadonnées, entrepôt de données, sur les aspects juridiques...
- Absence/insuffisance de recommandations ou d'exemples par discipline

maDMP : MACHINE ACTIONABLE DMP

« Rendre les PGD exploitables par les machines afin d'améliorer l'expérience de toutes les personnes concernées en échangeant de l'information entre les outils et les systèmes de recherche et en intégrant les PGD dans les flux de travail existants »

maDMP



Funder



Ethics review



Legal expert



Researcher



Publisher



Repository operator



Infrastructure provider



Research support staff



Institutional administrator

Miksa T, Simms S, Mitchen D, Jones S (2019) Ten principles for machine-actionable data management plans. PLOS Computational Biology 15(3): e1006750. <https://doi.org/10.1371/journal.pcbi.1006750> <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006750>

maDMP : EXEMPLES DE BÉNÉFICES POUR L'UTILISATEUR

Je renseigne dans mon DMP

- le volume prévisionnel de mes données pour un centre de calcul... et je reçois un message m'informant de la création du compte projet et de la réservation de l'espace nécessaire
- le numéro de mon projet ANR... et les informations sur mon projet se complètent automatiquement
- un identifiant ORCID... et je récupère automatiquement toutes les informations auteur
- un nouveau membre de l'équipe projet... et il me suffira de le sélectionner pour récupérer toutes les informations lors de mon prochain DMP
- les informations de base de mon projet... pour avoir des suggestions d'entrepôt pour déposer mes données et de schémas de métadonnées pour les décrire



Des outils et services pour accompagner la gestion et le partage des données

DMP OPIDoR

Cat OPIDoR

PID OPIDoR

CONSEIL AU DEPÔT

EXPOSITION DES DONNÉES

Formation DoRANum

Cat OPIDoR : catalogue en ligne de services

Cat
OPIDoR

<https://cat.opidor.fr>

Pour des données de qualité, réutilisables et conformes aux principes FAIR

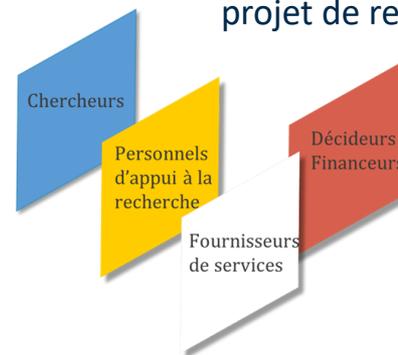
Cartographie des services français dédiés aux données scientifiques pour trouver :

- Les services utiles
- Pour chaque étape du projet
- Par domaine scientifique

Outil collaboratif : signalez les services que vous fournissez ou que vous utilisez



Pour tous les acteurs d'un projet de recherche



Conseil à la préparation des données et métadonnées



Accompagnement au dépôt dans un entrepôt existant :

- Préparation des données
- Enrichissement des métadonnées
- Reformatage des métadonnées
- Format des fichiers
- Choix de la licence
- etc.

En fonction des exigences de l'entrepôt choisi.



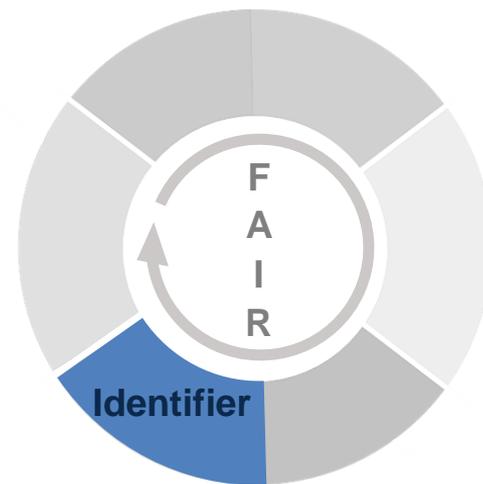
DataCite : attribution d'identifiants pérennes



Consortium international à but non lucratif créé en 2009, centré sur la valorisation des données de la recherche.

L'Inist :

- Opère ce service pour le compte du CNRS
- Est agence d'attribution de DOI (Digital Object Identifier) pour la France
- Offre un accompagnement personnalisé



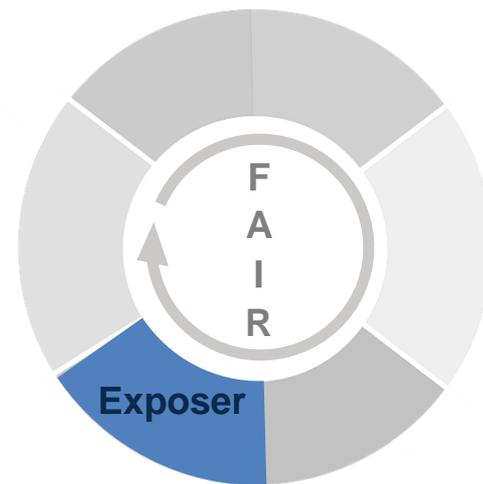
Omeka : exposition des données



Contact-
valorisationdonnees
@inist.fr

Aider à créer une plateforme web d'exposition des données :

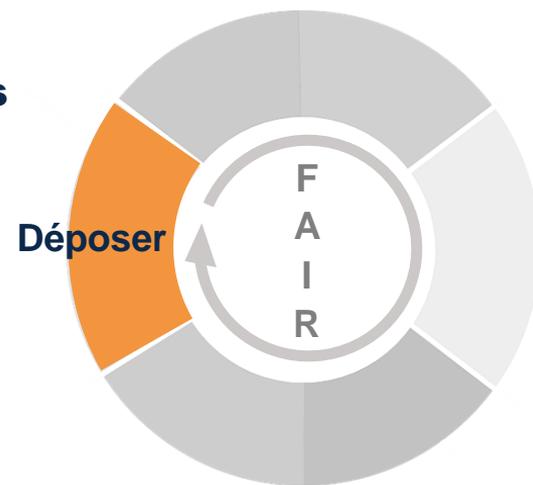
- Traitement documentaire des données
- Dématérialisation des documents
- Mise en ligne dans des applications web dédiées



Conseil au dépôt dans des entrepôts déjà existants



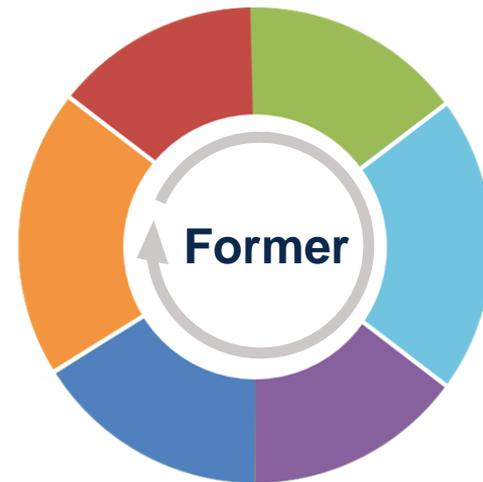
- Aide à l'expression des besoins de l'utilisateur
- Déploiement, démarche et grille d'analyse
- Conseil au choix de l'entrepôt
- Conseil au dépôt en accord avec les principes FAIR



DoRANum : plateforme e-learning sur la gestion, le partage, la valorisation des données



- Plus de 90 ressources pédagogiques
- Partenaires : Gis Réseau URFIST, Inra, ENPC
- 9 thématiques traitées :



Groupe projet CoSO



Enjeux et bénéfices



Aspects juridiques, éthiques, intégrité scientifique



Plan de gestion de données



Métadonnées



Identifiants pérennes



Dépôt & entrepôts



Stockage & archivage



Data papers & Data journals



Accès & visualisation

OPIDoR TOUR & FORMATIONS



10 octobre 2019 - Vandoeuvre



14 novembre 2019 - Grenoble



21 novembre - Paris



26 novembre - Poitiers



28 novembre - Orléans



7 janvier Villeurbanne



14 janvier - Strasbourg



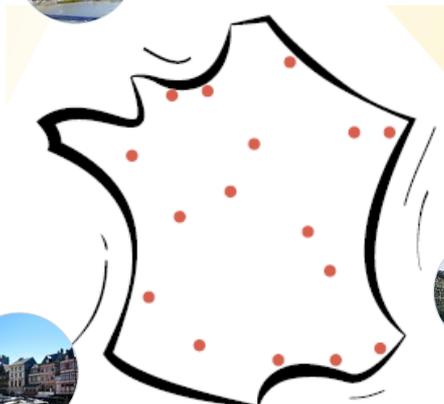
16 janvier - Rouen (en visio)



17 janvier - Caen (en visio)

20 janvier Guyancourt - LATMOS

28 janvier - Montpellier



4 février 20 - Marseille



6 février 2020 – Nice



2 mars - Nantes



3 mars 2020 - Rennes



5 mars 2020 - Toulouse

10 mars 2020 - Lille

12 mars 2020 - Talence

27 avril 2020 - Gif sur Yvette



Les équipes de l'Inist CNRS en lien avec les délégations régionales du Cnrs et les services des universités viennent à votre rencontre

Merci de votre attention !

Des questions ?

www.inist.fr

DMP OPIDoR : <https://dmp.opidor.fr/>

Cat OPIDoR : <https://cat.opidor.fr/>

PID OPIDoR : <https://opidor.fr/identifier/>

DoRANum : <https://dorandum.fr/>

www.cnrs.fr

