

# Des outils et services pour accompagner la gestion des données de la recherche





#### **OPIDoR**

Optimiser le Partage et l'Interopérabilité des Données de la Recherche

Gif-sur-Yvette le 10 juin 2020 en visioconférence



Comment rédiger mon DMP ?



Qui sont les bons interlocuteurs?

Quelles sont les bonnes pratiques pour gérer et partager mes données ?

#### **Programme**



#### Durée : ~2h00 en fonction des questions

- Retour sur les sciences ouvertes et les principes FAIR
- Les services de l'Inist en soutien
- Les services du département DVDR
  - Le service formation
  - Le service exposition des données
  - Les services proposés par OPIDoR
    - DMP OPIDoR : planifier la gestion des données
    - Cat OPIDoR : repérer des services sur la gestion des données
    - PID OPIDoR : attribuer un identifiant pérenne aux données (DOI)













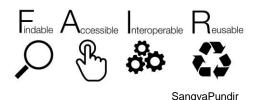
#### Structurer et ouvrir les données





DEUXIÈME AXE : STRUCTURER ET OUVRIR LES DONNÉES DE LA RECHERCHE « aussi ouvert que possible, aussi fermé que nécessaire »

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



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

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





#### Structurer et ouvrir les données

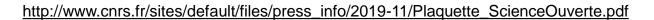




### 2 | LES DONNÉES DE LA RECHERCHE

Objectif: Les données (données brutes, textes et documents, codes sources et logiciels) produites par les chercheurs et les chercheuses CNRS ou avec des moyens mis en œuvre par le CNRS doivent être, dans la mesure du possible, rendues accessibles et ré-utilisables selon les principes FAIR pour une consolidation des connaissances essentielle au développement d'une science plus efficiente. « Les données doivent être aussi ouvertes que possibles, et fermées autant que nécessaire ».













Présentation par A. Petit, Président directeur général du CNRS - JNSO, 18/11/2019

Actions pour les données de recherche

- ▶ <u>Développer une culture de la gestion/partage des données</u> chez tous les acteurs du cycle de vie de la donnée
- Développer la publication des données (data papers), le dépôt conjoint publications/données et accompagner les chercheurs et les chercheuses dans l'utilisation des outils de gestion des données.
- ➤ <u>Soutenir et accompagner les infrastructures de recherche</u>, productrices de données, dans la définition et la mise en œuvre de politiques de données
- Soutenir et accompagner des Infrastructures de données Mettre en œuvre un service coordonné avec les instituts pour favoriser le dépôt des données pour tous les personnels des unités du CNRS
- Créer et afficher un répertoire des dépôts et des services de données dont le CNRS est responsable et auxquels il participe

https://webcast.in2p3.fr/video/feuille-de-route-du-cnrs-pour-la-science-ouverte





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



**Description riche** 

**Provenance** 

**Standards** communautaires

Licences

**Formats standards** 

Lien vers d'autres (méta)données

Vocabulaires standardsontologies

Findable

A ccessible Interoperable Reusable Reusable Pards-

Identifiants pérennes

Métadonnées riches

Catalogues - Entrepôts

Protocoles communication standardisés & ouverts

**Authentification &** autorisation si besoin

Accès permanent aux métadonnées

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

Traduction Inra https://www6.inra.fr/datapartage/Produire-de





### 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.





### Si vous voulez en savoir plus sur la science ouverte à l'ANR...



Webinaire « La politique science ouverte de l'ANR et le plan de gestion des données (PGD) » - 14 novembre 2019

Politique Science Ouverte ANR et le PGD
 https://anr.fr/fileadmin/documents/2019/Slides ANR webinaire science ouverte - PGD.pdf

=> Pour toute question : scienceouverte@anr.fr

 DMP OPIDoR, Planifier la gestion des données https://anr.fr/fileadmin/documents/2019/slides webinaire DMP OPIDoR.pdf













#### Des acteurs mobilisés pour vous accompagner



#### Les services du DVDR

Exposition des données (préparation traitement de données, métadonnées, ..): OMEKA



https://omeka.org/

DoRANum : plateforme e-learning sur la gestion des données (CoSo)



https://doranum.fr/





#### Des acteurs mobilisés pour vous accompagner

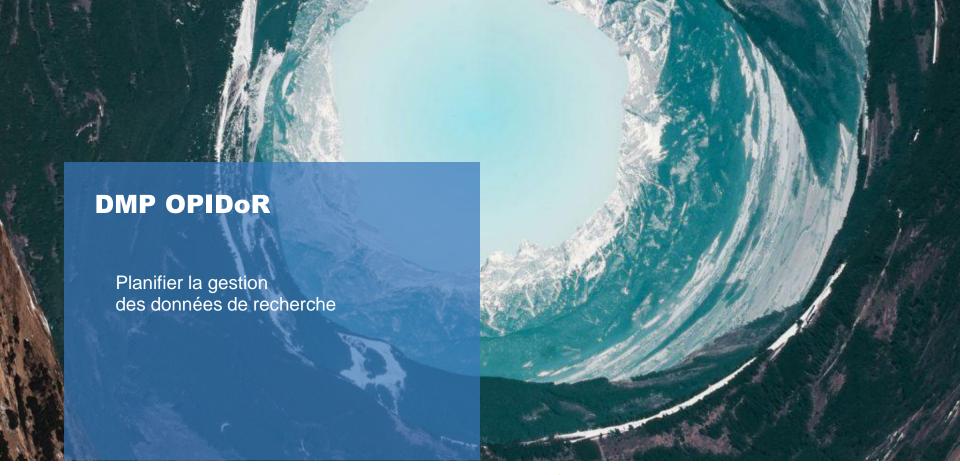


#### Les services d'OPIDoR

- Services en ligne, Twitter, liste de discussion DMP
- Soutien accompagnement, des équipes de recherche
  - Le DMP: DMP OPIDoR, création et rédaction de DMP
  - Les ressources et services en France disponibles sur les données de recherche : Cat OPIDoR
  - PID : DOI, assistance création, suivi, mise à jour
  - Conseil sur la préparation
    - Des métadonnées FAIRification
    - Le choix d'un entrepôt de données
  - Projets : maDMP













Plan de gestion de données

#### **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

- Prend en compte toutes les étapes du cycle de vie des données
- Initié très tôt, dès le montage du projet
- Document continuellement mis à jour
- Opportunité pour établir un dialogue entre les différents acteurs : effort collectif (scientifiques, data librarians, informaticiens, juristes, chargés de projet...)
- Incitation voire obligation par : organismes de recherche / universités et agences de financement de la recherche (ANR, Commission européenne/H2020...)



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



### Eléments de contenu du DMP

- Données administratives et description du projet
- Rôles et responsabilités (mise en œuvre du DMP)
- Informations sur les données de recherche : types, formats, volumes, qualité...
- Standards et métadonnées
- Politique d'accès, de partage et de réutilisation des données
- Stockage, sauvegarde et conservation (archivage)
- Ressources allouées

... tout en prenant en compte les principes FAIR



Les financeurs et organismes de recherche proposent des modèles de DMP



### Un outil d'aide à l'élaboration de DMP (1)

#### **DMP OPIDoR**

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

https://dmp.opidor.fr/

Outil basé sur le code commun DMPRoadmap (Digital Curation Center/UK et l'UC3/USA)

- adapté aux besoins de la communauté scientifique française
- serveur sécurisé hébergé à l'Inist-CNRS, en France
- mis à disposition de l'ESR



### Un outil d'aide à l'élaboration de DMP (2)



- Créer / rédiger un DMP à partir d'un modèle
- Partager son DMP avec un/des collaborateur(s)
- Ajouter des commentaires pour des collaborateurs dans un DMP
- Définir le niveau de visibilité d'un DMP
- Demander une assistance conseil auprès des services d'appui de son organisme de recherche (s'il existe)
- Télécharger son DMP sous différents formats (docx, pdf, ...)



### Modèles disponibles sur DMP OPIDoR

- Agences de financement
  - H2020 (FAIR DMP)
  - ERC
  - ANR
- Organismes de recherche / Universités
  - Cirad, Inra, Institut Pasteur, Irstea, Sciences Po
  - Unistra (Université de Strasbourg),
     Université Paris Descartes & Paris
     Diderot
- Autres :
  - Projet PRESOFT, SSI (UK)
  - Digital Curation Centre (UK), EPFL

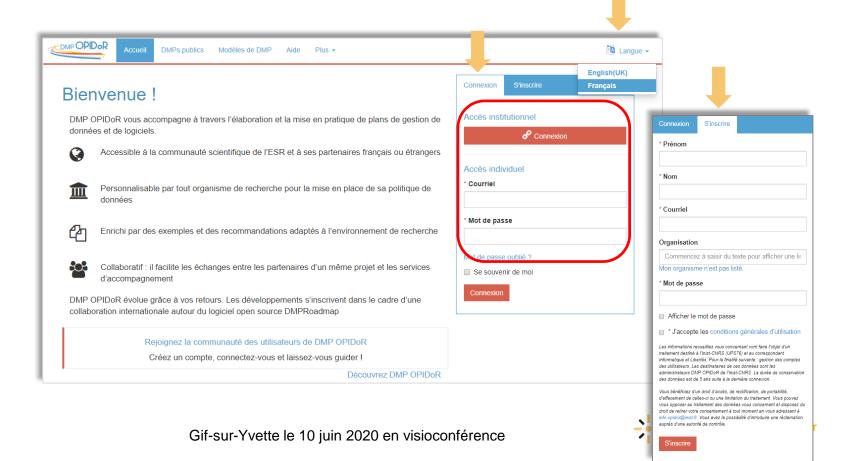
### ANR - Modèle de PGD (français)

- 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 serontelles 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

Gif-sur-Yvette le 10 juin 2020 en visioconférence

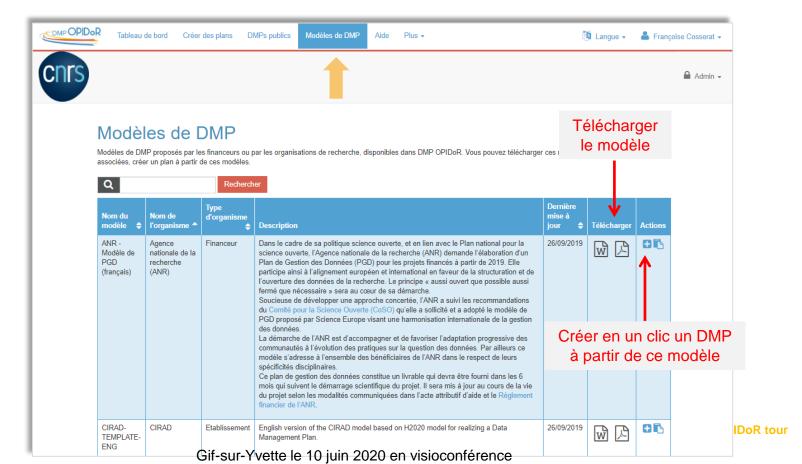
### 1. S'inscrire / Se connecter



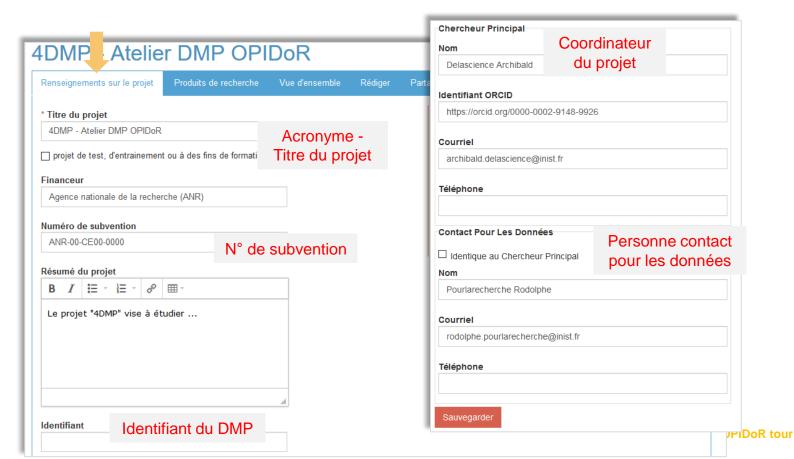
### 2. Création d'un plan de gestion (1)



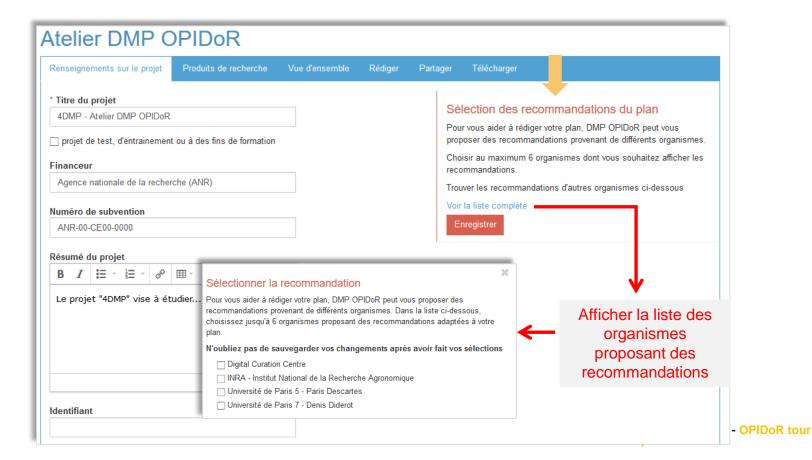
### 2. Création d'un plan de gestion (2)



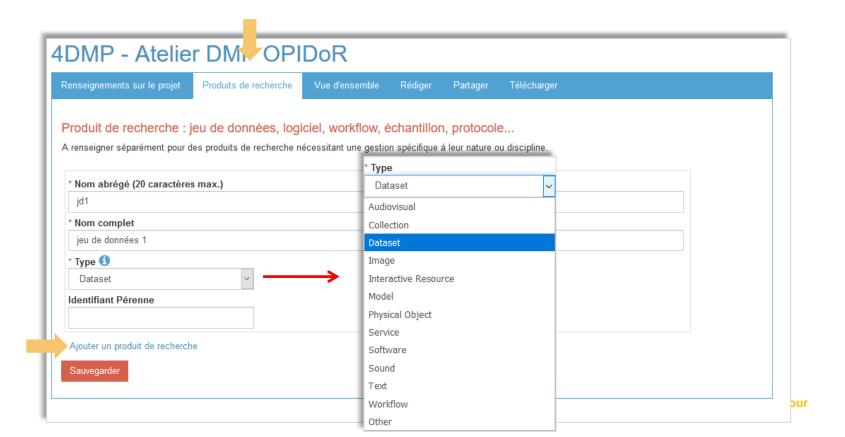
### 3. Renseignements sur le projet



### 4. Sélection des recommandations



### 5. Déclarer les produits de recherche



### 6. Rédiger le Plan de gestion



### 6. Rédaction - Recommandations

Ajouter des commentaires pour vos collaborateurs

Vous pouvez vous aider des recommandations

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



Recommandations

Commentaires

ANR

≀A DC

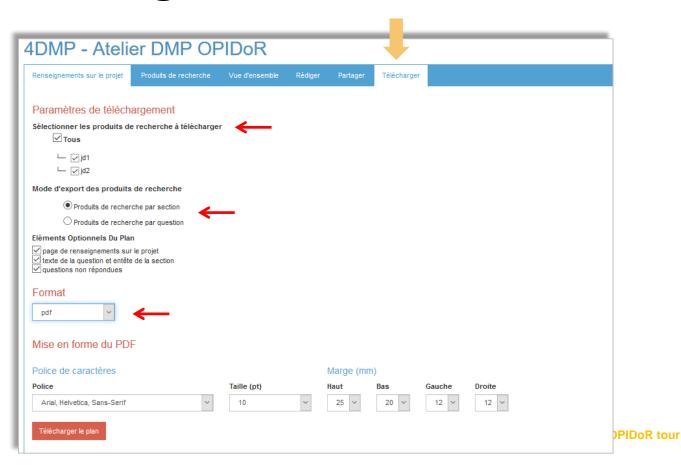
- Indiquer quelles métadonnées seront fournies pour aider à la recherche et à l'identification des données.
- Indiquer quels standards de métadonnées seront utilisés (par exemple DDI, TEI, EML, MARC, CMDI).
- Utiliser les standards de métadonnées des communautés scientifiques lorsque ceux-ci existent.



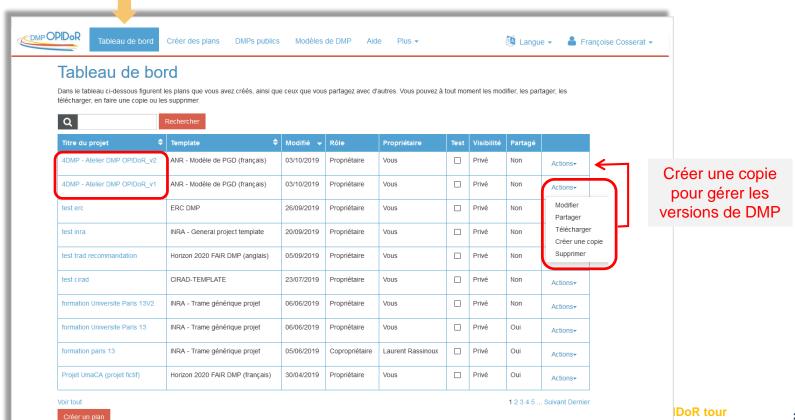
### 7. Partager



### 8. Télécharger



### Tableau de bord



### Pour les institutions : Interface Administrateur

- 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

### Les bonnes adresses



Contact: info-opidor@inist.fr

#### Equipe DMP OPIDoR - Inist CNRS

Anne Busin, Françoise Cosserat, *Benjamin Faure*, Marie-Christine Jacquemot, *Florence Maréchau*, *Florian Mazur*, Laurent Rassinoux

- DMP actuel
- DMP « machine actionable » en partenariat avec l'IFB: projet ANR Flash, de l'ANR et du GENCI

Liste de diffusion : <a href="https://listes.services.cnrs.fr/wws/info/dmpopidor">https://listes.services.cnrs.fr/wws/info/dmpopidor</a>

DMP OPIDoR : <a href="https://dmp.opidor.fr/">https://dmp.opidor.fr/</a>

V2,3,0: <a href="https://github.com/OPIDoR/DMPOPIDoR/releases">https://github.com/OPIDoR/DMPOPIDoR/releases</a>

Twitter: OPIDoR, DORANum, LaLIST











### Un catalogue en ligne de services

#### Cat OPIDoR

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

https://cat.opidor.fr/

- Recense et décrit les services français dédiés aux données scientifiques
- Proposé sous forme d'un wiki, cet outil collaboratif ouvert à tous permet de repérer les services utiles dans le cadre d'un projet de recherche

### Si vous cherchez ...

Où trouver une **aide personnalisée** pour la gestion des données de recherche ?

Quels sites web recensent des actualités sur ce thème ?

Existe-t-il un **entrepôt français** dans ma discipline de recherche ?

Où trouver des supports de formation en ligne ?

Quelles plateformes proposent des **ressources informatiques pour faire de l'analyse**, de la simulation ou de la modélisation ?





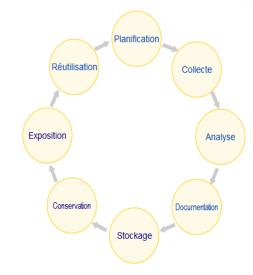
# Repérer les services dans le cadre d'un projet de recherche



Pour accompagner les chercheurs sur l'ensemble des étapes clés de la gestion, collecte, stockage, conservation

et ouverture des données

A quel stade du cycle de vie des données? [modifier]





# Repérer les services dans le cadre d'un projet de recherche



Par domaine scientifique

- Sciences Humaines & Sociales
- Sciences & Technologies
- Vie & Santé

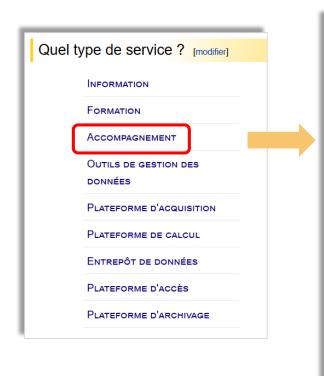
#### Des services

Information
FORMATION
ACCOMPAGNEMENT
OUTILS DE GESTION DES DONNÉES
PLATEFORME D'ACQUISITION
PLATEFORME DE CALCUL
Entrepôt de données
PLATEFORME D'ACCÈS
PLATEFORME D'ARCHIVAGE



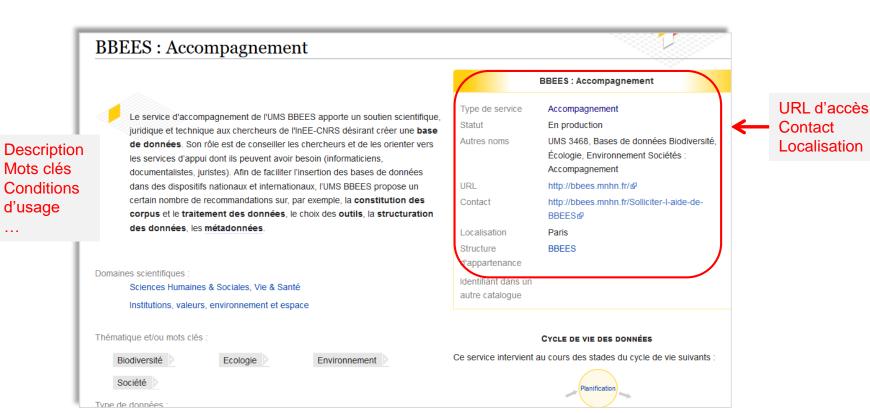
## Cat OPIDoR en pratique

Où trouver une aide personnalisée pour la gestion des données de recherche ?



Services	Domaine scientifique	Mots clés	Localisation	Stade du cycle de vie
BBEES : Accompagnement	Sciences Humaines & Sociales Vie & Santé	Biodiversité Ecologie Environnement Société	Paris	Planification Documentation
CeDONA	Sciences & Technologies Vie & Santé	-	Pessac	Collecte Planification Documentation Stockage Exposition Réutilisation
Centre Blaise Pascal Lyon	Sciences Humaines & Sociales Sciences & Technologies Vie & Santé	Formation Séminaire Soutien Projet scientifique Expertise calcul scientifique Calcul haute performance Modélisation numérique Codes de calcul Analyse de données	Lyon	Analyse Stockage
CeSAM	Sciences & Technologies	Astronomie Astrophysique	Marseille	Analyse Documentation Stockage Conservation Exposition Réutilisation

## **Un service dans Cat OPIDoR**



## D'autres modes de recherche

Par domaines et sous-domaines scientifiques

Par simple recherche



- SCIENCES HUMAINES & SOCIALES [Afficher]
  - Cultures et production culturelle
  - Individus, marchés et organisations
  - Institutions, valeurs, environnement et espace
  - L'esprit humain et sa complexité
  - L'étude du passé humain
  - Le monde social, diversité, population

Sciences & Technologies [Afficher]

VIE & SANTÉ [Afficher]

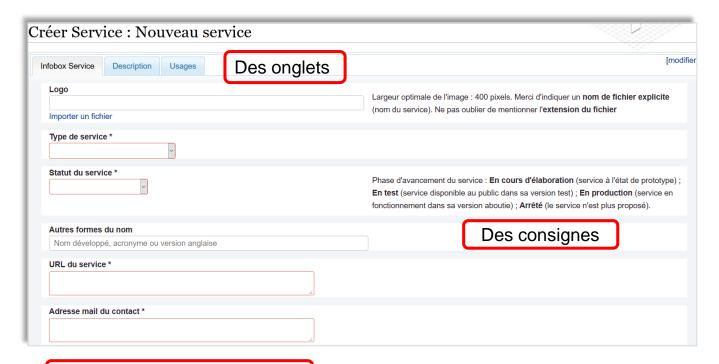
### Par localisation géographique





## **Contribuer à Cat OPIDoR**

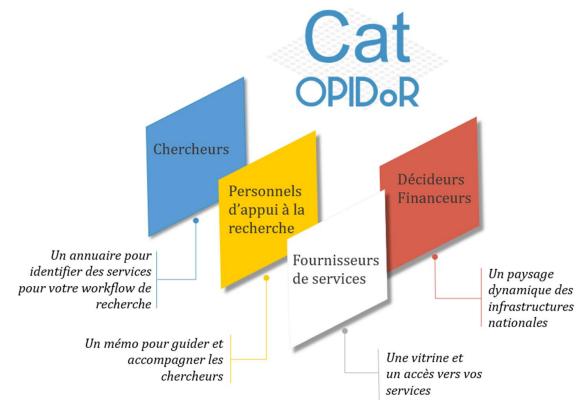




Quelques champs obligatoires



# A qui s'adresse ce catalogue?



## Les bonnes adresses



Equipe Cat OPIDoR – Inist CNRS

Ourida Aberkane, Anne Ciolek-Figiel, Benjamin Faure

Contact: infocatopidor@inist.fr

Cat OPIDoR : <a href="https://cat.opidor.fr/">https://cat.opidor.fr/</a>









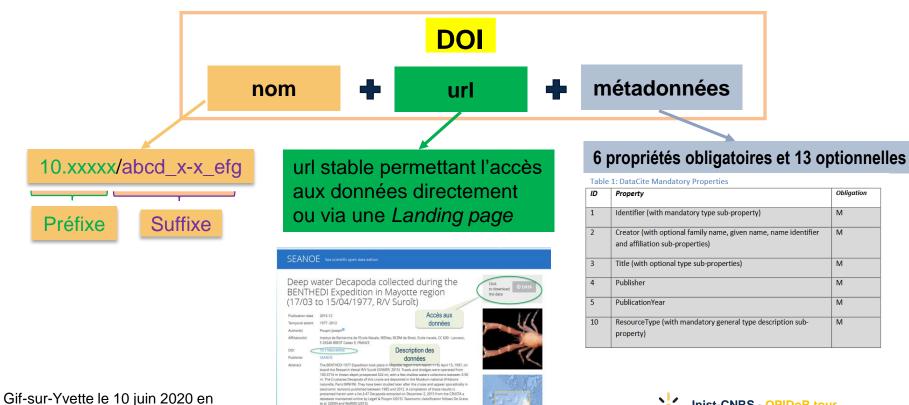
# Inist-CNRS: membre français du consortium DataCite pour l'attribution de DOI

- Sensibilisation et promotion des services DataCite
- Membre votant à l'Assemblée Générale : budget, modèle économique, développement du consortium...
- Groupes de travail DataCite :
  - Métadonnées : évolution du Schéma en fonctions des demandes des utilisateurs
  - Bonnes pratiques et stratégie...
- Fourniture des préfixes de DOI
- Assistance à la création et à la conversion de fichiers de métadonnées

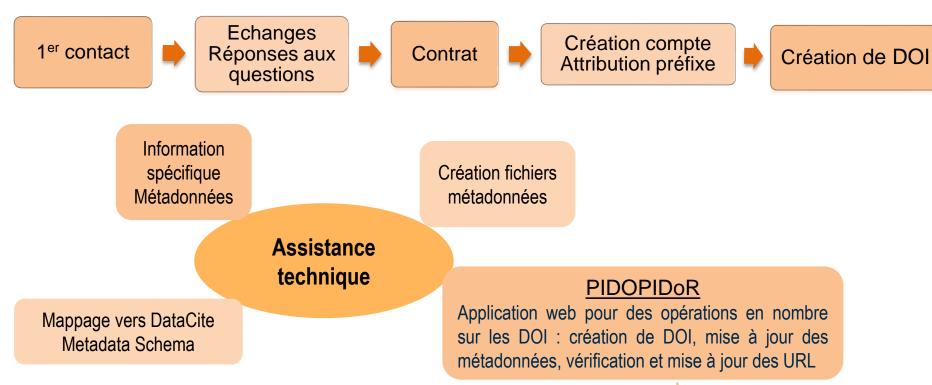
Gif-sur-Yvette le 10 juin 2020 en visioconférence



# **DOI: Triptyque**



# Workflow d'attribution de DOI par l'Inist-CNRS



## **Partenariat Inist-Institutions**

- Un devis pour les partenaires publics formalisé par l'Inist en relation entre le demandeur et sa tutelle (payeur)
- Coût annuel de 180 € HT pour un nombre illimité de DOI\*

\*modèle économique en cours d'évolution pour 2021



## Création de DOI

sur DataCite Fabrica



https://doi.datacite.org/

Automatique via l'API

#### Purpose of the DataCite REST API

The DataCite REST API allows any user to retrieve, query and browse DataCite DOI metadata records. In addition, DataCite Clients can register DOIs and DataCite Providers can manage Clients and prefixes via the API.

https://support.datacite.org/docs/api-create-dois

En grand nombre via 'PIDOPIDoR application web mise en ligne par l'Inist

Créer et mettre à jour aisément vos DOI

https://doidatacite.inist.fr/doidatacite/EzDoiMenu

Gif-sur-Yvette le 10 juin 2020 en visioconférence









## **Création de DOI**

PIDOPID<sub>o</sub>R

Créer et mettre à jour aisément vos DOI

Vous souhaitez ...

### PID OPIDoR







## Les bonnes adresses



Equipe PID OPIDoR – Inist CNRS

Nadine BUISSON, Mohamed S. YAHIA

Contact:datasets@inist.fr











## FAIR: les métadonnées

#### Les schémas

S'il convient d'utiliser des schémas de métadonnées, il est aussi important d'utiliser des vocabulaires communs homogènes en s'appuyant si possible sur les triples stores, le Linked Data et RDF ou OWL

#### Qu'est-ce qui distingue les schémas?

- Les domaines, problématiques des communautés scientifiques
- Certains formats de métadonnées génériques comme Dublin Core, DataCite (schéma.org) ne permettent pas de décrire de manière satisfaisante le contexte de production, les méthodes utilisées, ...
- Plus les métadonnées seront adaptées et spécifiques, meilleure sera leur compréhension : d'où la création de standards spécifiques: DDI (données enquêtes SHS), EML (écologie) et ISA-TAB (EBI)
- L'utilisation de standards bien définis permet également de faciliter les échanges, la recherche ou la comparaison des jeux de données (FAIR)

52

## FAIR: les métadonnées

Les schémas: https://doranum.fr/?s=metadonn%C3%A9es

#### Où se trouvent les informations relatives aux schémas?

- <u>FAIRsharing</u>: 'a curated, informative and educational resource on data and metadata standards, interrelated to databases and data policies'\_exemple <u>MIBBI</u>
- DCC: <a href="http://www.dcc.ac.uk/resources/metadata-standards">http://www.dcc.ac.uk/resources/metadata-standards</a>
- RDA: <a href="http://rd-alliance.github.io/metadata-directory/">http://rd-alliance.github.io/metadata-directory/</a>

#### Interopérabilité entre les schémas

- SIFR Bioportal: Portail d'ontologies en biomédecine et agronomie (An open platform to host French biomedical ontologies and terminologies based on the technology developed by the National Center for Biomedical Ontology)
   <a href="https://bioportal.bioontology.org/">https://bioportal.bioontology.org/</a>
- SIFR: Semantic Indexing of French biomedical Resources <a href="http://bioportal.lirmm.fr/">http://bioportal.lirmm.fr/</a>
- Loterre Linked open terminology resources <a href="https://www.loterre.fr/skosmos/fr/search?clang=fr&q=thrombose">https://www.loterre.fr/skosmos/fr/search?clang=fr&q=thrombose</a>

Gif-sur-Yvette le 10 juin 2020 en visioconférence

OBO Open Biological and Biomedical Ontology <u>www.obofoundry.org/</u>



# FAIR: les entrepôts

#### Quelle est pour vous la définition d'un entrepôt ?

Est-ce différent d'une banque de données, comme par exemple GenBank ou PDB (Protein Data Bank)?

- https://www.ncbi.nlm.nih.gov/genbank/eukaryotic genome submission annotation/
- https://www.rcsb.org/

#### Ce qui qualifie un entrepôt de données

Pouvoir déposer, conserver, partager, diffuser des jeux de données avec une description/métadonnée/documentation facilitant leur recherche et permettant leur <u>réutilisation</u> « aussi ouvert que possible aussi fermé que nécessaire »



# FAIR: les entrepôts

#### Où se trouvent les entrepôts?

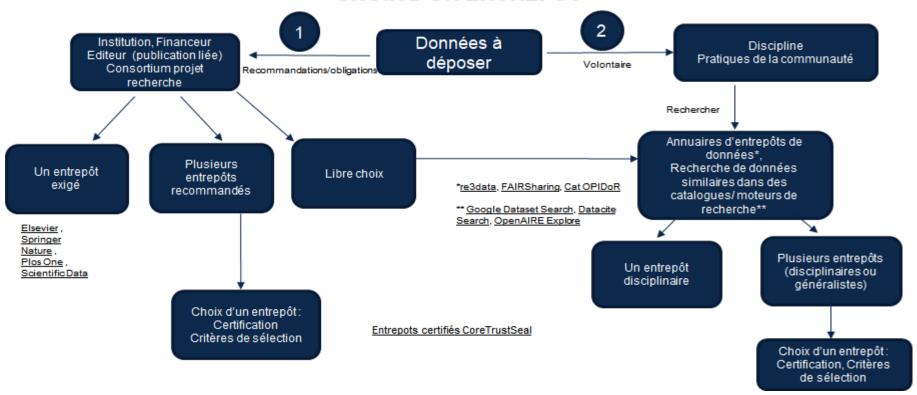
Si la question concerne la localisation géograhique des serveurs, cela peut être n'importe où quatre exemples :

- Zenodo en Suisse avec les serveurs du CERN : <a href="https://zenodo.org/search?page=1&size=20&q=segmentation%20AND%20MRI&type=dataset">https://zenodo.org/search?page=1&size=20&q=segmentation%20AND%20MRI&type=dataset</a>
- Pangaea en Allemagne : <a href="https://www.pangaea.de/?t=Chemistry">https://www.pangaea.de/?t=Chemistry</a>
- COD (Crystallography Open Database) en Lituanie : <a href="http://www.crystallography.net/cod/">http://www.crystallography.net/cod/</a>
- Dryad aux USA: https://datadryad.org/stash

Il peut sembler libre et ouvert mais peut appartenir à un éditeur scientifique, exemple Figshare et Springer <a href="https://figshare.com/">https://figshare.com/</a>

## FAIR : les entrepôts

#### **CHOIX D'UN ENTREPÔT**



# FAIR : les entrepôts

#### Critères de sélection

- type de données,
- · formats fichiers acceptés,
- volumétrie,
- métadonnées,
- licence,
- identifiant pérenne,
- facilite/aide dépôt,
- gestion des accès (embargo),
- gestion des versions,
- liens vers publications,
- statistiques d'usages,
- localisation serveur,
- visibilité,
- Conservation,
- sécurisée,
- coût...



## Les bonnes adresses



Equipe Métadonnées – Rntrepôts

Ourida ABERKAN, Françoise TISSERAND, Mohamed S. YAHIA

Contact :datasets@inist.fr









## Des acteurs mobilisés pour vous accompagner



#### Nouveaux services pour progresser dans la mise œuvre des principes FAIR

- Dans le DMP « machine actionable », l'intégrer dans le workflow des données de la recherche
- Dans l'implémentation des principes FAIR dans les « Data Repositories » entrepôts – bases de données, banques de données, …, dans l'environnement de travail du chercheur
- Par complémentarité avec les différents acteurs au sein des communautés, des institutions
- Par la prise en compte avec eux des différentes spécificités disciplinaires





#### **maDMP**



