Le plan de gestion de données et les identifiants DOI : composantes essentielles de l'écosystème FAIR

Claire François

16 novembre 2020

2ème JSO CNRS : « science ouverte 2020, où en sommes nous ? »

Inist | Institut de l'information scientifique et technique

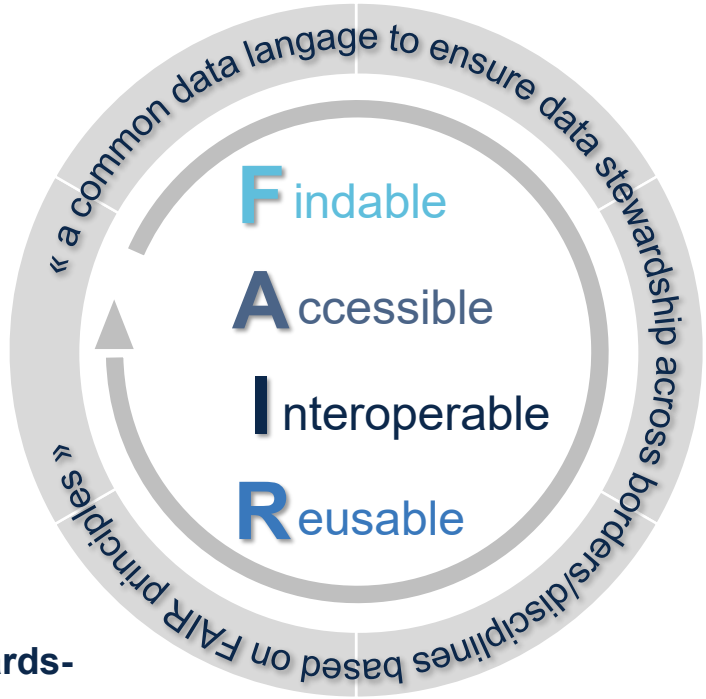
COMPOSANTES ESSENTIELLES DE L'ÉCOSYSTÈME FAIR

- Une politique
- **Des plans de gestion de données** (ou *PGD*) interconnectés
- **Des identifiants** associés à des ressources, métadonnées, thésaurus, …
- Des standards, ontologies
- Des plateformes collaboratives
- Des entrepôts CoreTrustSeal

Composantes essentielles de l'écosystème FAIR
le DMP et les identifiants pérennes

R
- Description riche
- Provenance
- Standards communautaires
- Licences

I
- Formats standards
- Lien vers d'autres (méta)données
- Vocabulaires standards-ontologies



F
- Identifiants pérennes
- Métadonnées riches
- Catalogues-Entrepôts

A
- Protocoles de 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

https://www6.inrae.fr/datapartage/Produire-des-donnees-FAIR

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

DMP OPIDoR

Outil d'aide à l'élaboration de plans de gestion de données

Un élément clé pour produire des données FAIR

LES MODÈLES DE DMPS DANS DMP OPIDoR

1 modèle générique (par défaut) - Science Europe

6 modèles de financeurs

ANR (2), ERC (1), Commission européenne(3)

19 modèles d'organismes de recherche / ESR

Cirad (2), Inrae (4), Institut Pasteur (2), Sciences Po (2), Univ Strasbourg (1), Université Paris Descartes* (1), Université Paris Diderot* (2), Université Paris Dauphine (1), ICM - Institut du Cerveau et de la Moelle épinière (1),

PACEA UMR 5199 CNRS - Université Bordeaux (1), MASA Consortium (1), cc-IN2P3 CNRS (1)

2 modèles SMP (software management plan)

PRESOFT projet/France Grille (2)

5 modèles hors France

DCC(2), EPFL (2), SSI (1)

33

Modèles de DMP

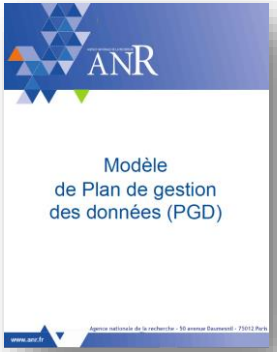
- Dont 5 modèles CNRS
- 16 en français
- 17 en anglais
- + 2 en préparation

Accompagnement à la rédaction de modèles

Paris Descartes, Paris Diderot et IPGP fusionnent pour devenir Universités de Paris au 01/01/20

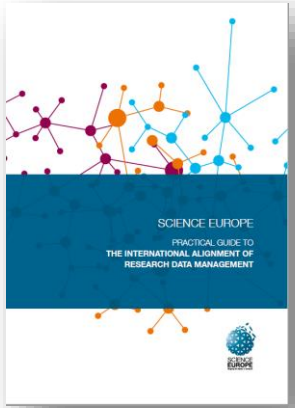
LA DYNAMIQUE DU PGD ANR

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



Déclaration conjointe du réseau des agences de financement françaises en faveur de la science ouverte.

Guide sur l'implémentation du modèle de DMP Science Europe dans les politiques science ouverte des financeurs en Europe.

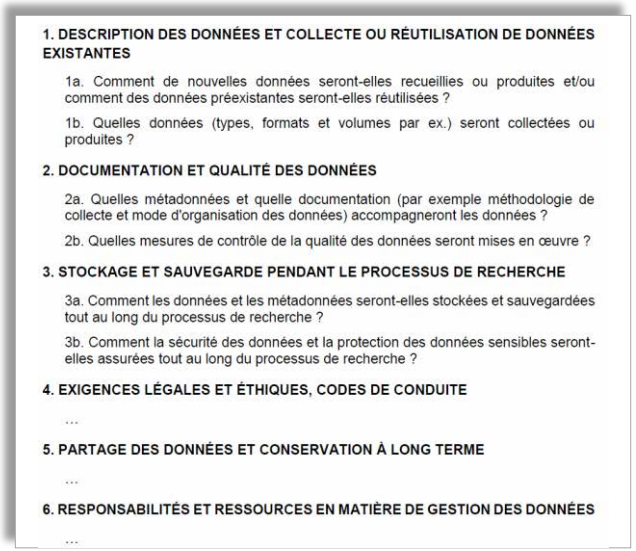


Science Europe. (2018). Practical guide to the international alignment of research data management.

Modèle intégré dans DMP OPIDoR

La dynamique du PGD ANR

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

OPIDoR Tour pour sensibiliser la communauté scientifique sur DMP OPIDoR et les services associés

Accompagnement à la rédaction de PGD

839

PGD ANR rédigés de septembre 2019 à octobre 2020

- *4 100 PGD rédigés*
  - *dont 613 CNRS*
- *24 PGD publics*

OPIDoR tour

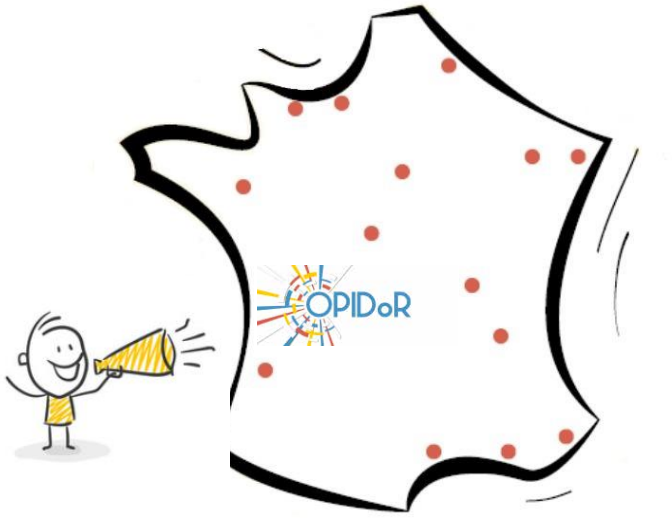
VALORISER LES DONNÉES DE LA RECHERCHE

19 dates dans tout l'hexagone

+ 700 participants

Typologie des participants

1. Chercheur / Enseignant chercheur : 31%
2. Documentaliste et professionnel de l'IST : 22%
3. Chargé de Valorisation/ Projet : 20%
4. Informaticien : 13%



« Cette formation au niveau local permet de prendre contact ('In Real Life') avec les collègues locaux, acteurs et/ou utilisateurs concernés par la gestion des données … »

Limites de l'outil actuel

- Document textuel peu structuré

Pour les chercheurs

- Perçu comme un exercice administratif et non comme faisant partie intégrante de la pratique de la recherche
- Terminologie trop technique : métadonnées, interopérabilité, ontologies…
- Répétition de la saisie d'informations déjà saisies dans d'autres systèmes d'information
- Pas/peu d'expertise dans le choix de 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 d'information et en intégrant les PGD dans les flux de travail existants »

doi: https://doi.org/10.1371/journal.pcbi.1006750.g001

Miksa T, Simms S, Mietchen D, Jones S (2019) Ten principles for machine-actionable data management plans. PLoS Comput Biol 15(3):e1006750. https://doi.org/10.1371/journal.pcbi.1006750

maDMP : bénéfices attendus

Je renseigne mon PGD :

- Avec le volume prévisionnel de mes données pour un centre de calcul, **je reçois automatiquement dans mon PGD** : l'accord de ma demande de création du compte projet avec l'espace alloué, le coût, le délai de réservation, ….
- Avec mon numéro de mon projet ANR **les informations sur mon projet se complètent automatiquement** dans mon PGD
- Avec un identifiant ORCID… je récupère toutes les **informations « auteur »** et participants au projet
- Avec les informations de base de mon projet … j'obtiens **des suggestions** d'entrepôts pour déposer mes données, de standards de métadonnées pour les décrire, de licence.

J'enregistre et partage les informations indispensables sur les données :

- pour la rédaction d'un Data Paper
- pour le versement ultérieur dans un entrepôts d'archivage pérenne

maDMP - CAS D'USAGES

FINANCEUR

ANR

En cours

Saisie automatique du DMP:
- Titre
- Description
- Participants, …

API sécurisées

Informations pour analyse :
- Volume de données coûts
- Ouverture, accessibilité, licences
- Entrepôts, archivage, …

DMP OPIDoR

Machine actionable DMP

Central hub of information on FAIR Digital Objects

Allocation de ressources:
- Volumes et processeurs attribués
- Métadonnées associées
- Coût estimé, …

API sécurisées

Demande de ressources:
- Volume, type, …
- Informations DMP
  - Research output,
  - Descriptions (metadatas)

CHERCHEUR PERSONNEL SOUTIEN

IFB

Demande de ressources:
- Volume,
- Délai,
- Type,
- Argumentaire issu du DMP

API sécurisées

Allocation de ressources:
- Volume attribué,
- Coût
- Délai de disponibilité accordé
- Type
- N° compte eDARI

GRAND EQUIPEMENT CALCUL INTENSIF

GENCI

En cours

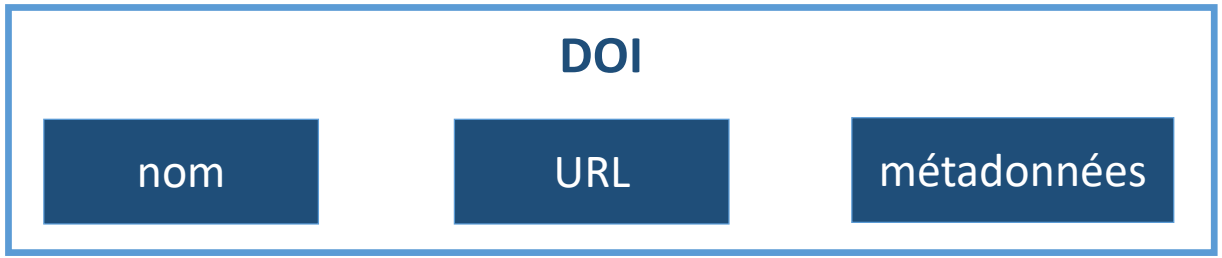
maDMP4LS machine-actionable DMP for Life Sciences **Projet ANR – Appel Flash science ouverte**

Identifiants pérennes DOI

Le service d’attribution de DOI de l’Inist CNRS

DOI (digital object identifier)

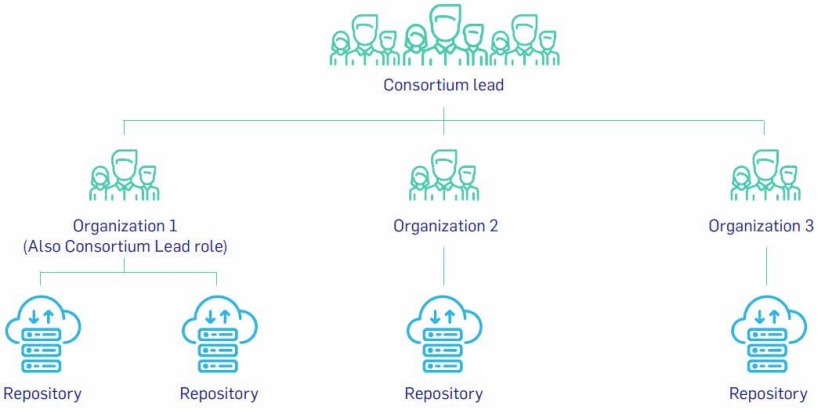
- Identifie les données et autres objets numériques issus de la recherche,
- est assigné à un objet de façon unique et permanente
- Facilite leur découverte, leur partage et leur réutilisation



Agence nationale d'attribution de DOI

DataCite est un consortium international à but non lucratif dédié à l'attribution d'identifiants pérennes (DOI) pour les données de la recherche.

Depuis 2010, **l'Inist-CNRS**, membre du consortium international, est **agence d'attribution (« consortium lead »)** pour la **France**.



14 000 €

BUDGET 2020

149 membres

Tarif annuel : 180 € par membre / an

LE SERVICE D'ATTRIBUTION DOI - INIST CNRS

Inist CNRS :

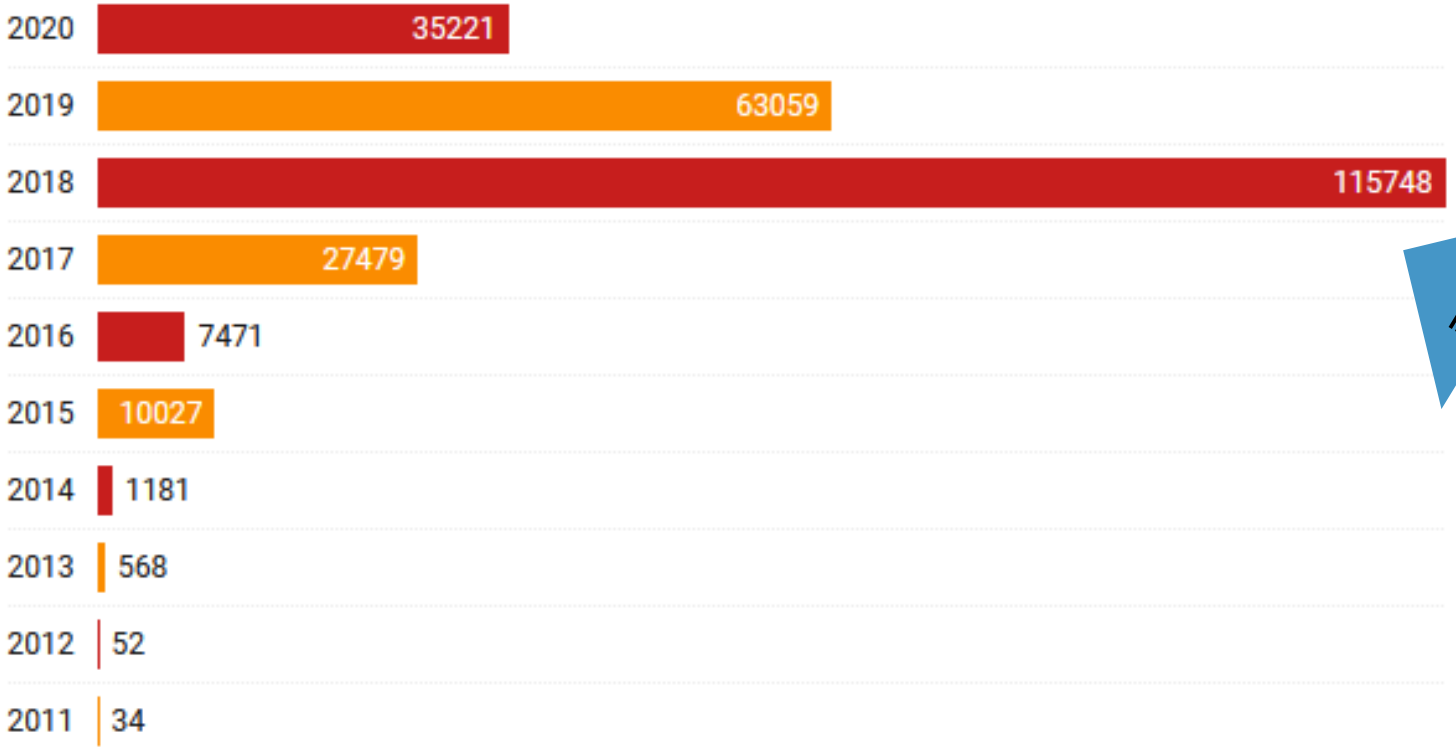
- Gestion des membres : contrats et comptes utilisateurs
- Accompagnement des membres :
  - Attribution de DOI aux collections/jeux de données
  - Enregistrement des métadonnées associées

Membre :

- Intégrité des métadonnées selon le schéma de métadonnées DataCite
- Stockage des données et maintenance : toutes les données enregistrées avec un DOI doivent être accessibles via un URL
- Persistance : veiller à la disponibilité des contenus enregistrés dans le temps

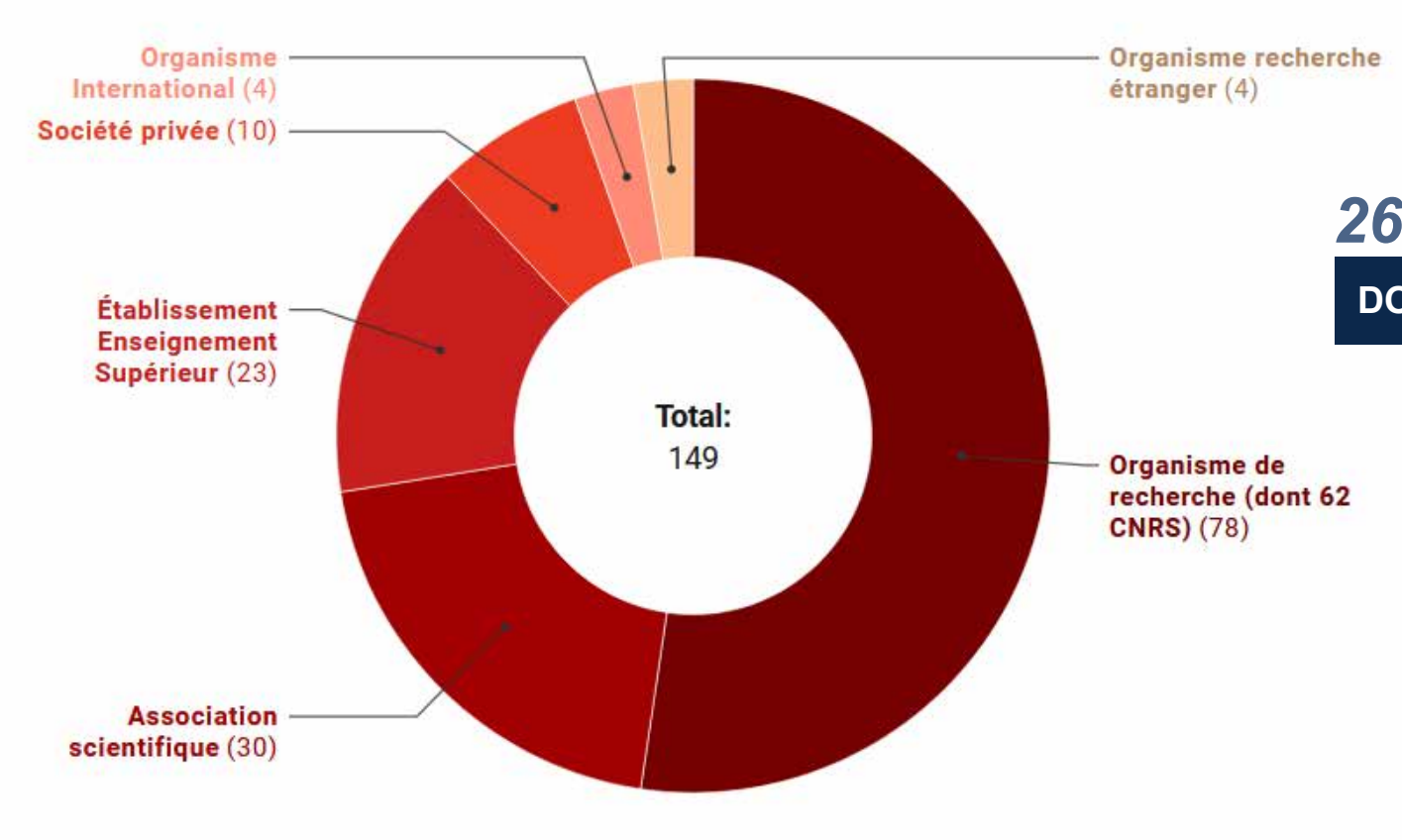
Évolution du service - nombre de DOI / an

31 10 2020



Mise en place de l'entrepôt Data Inrae

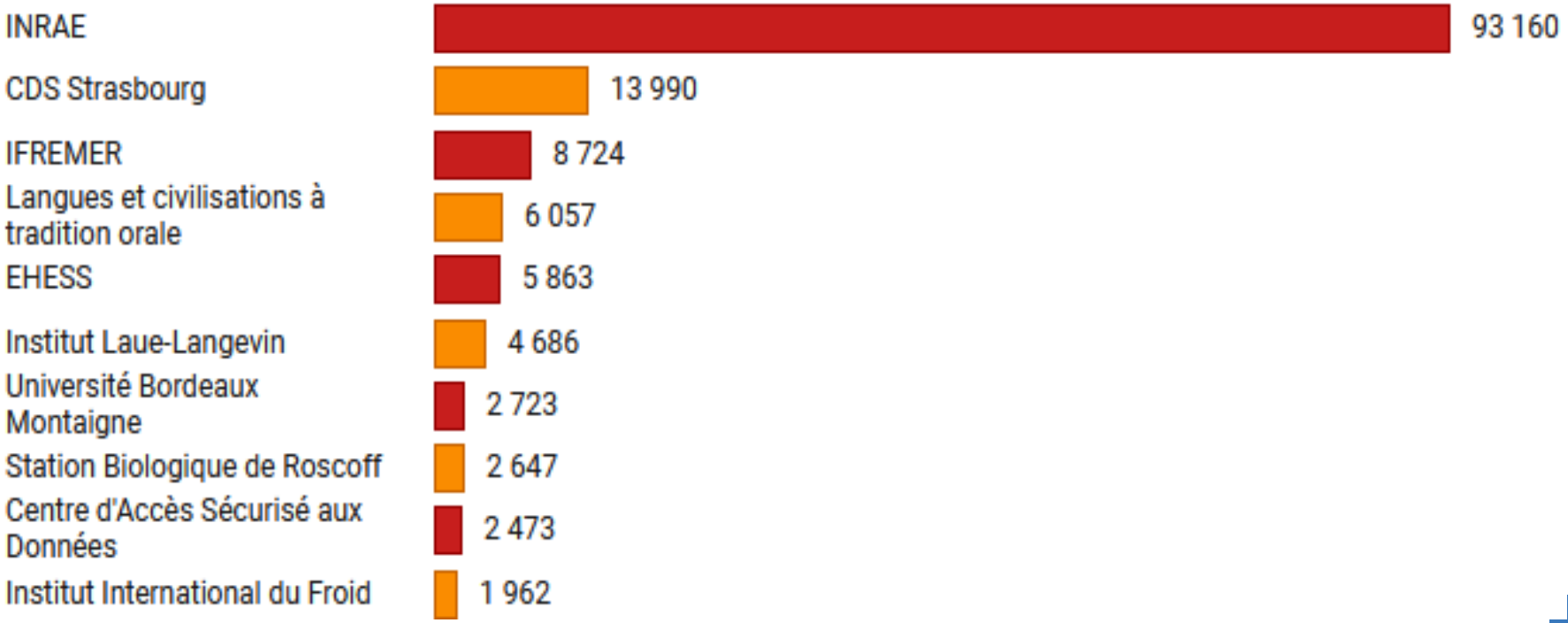
TYPOLOGIE DES MEMBRES (EN NOMBRE DE STRUCTURES)



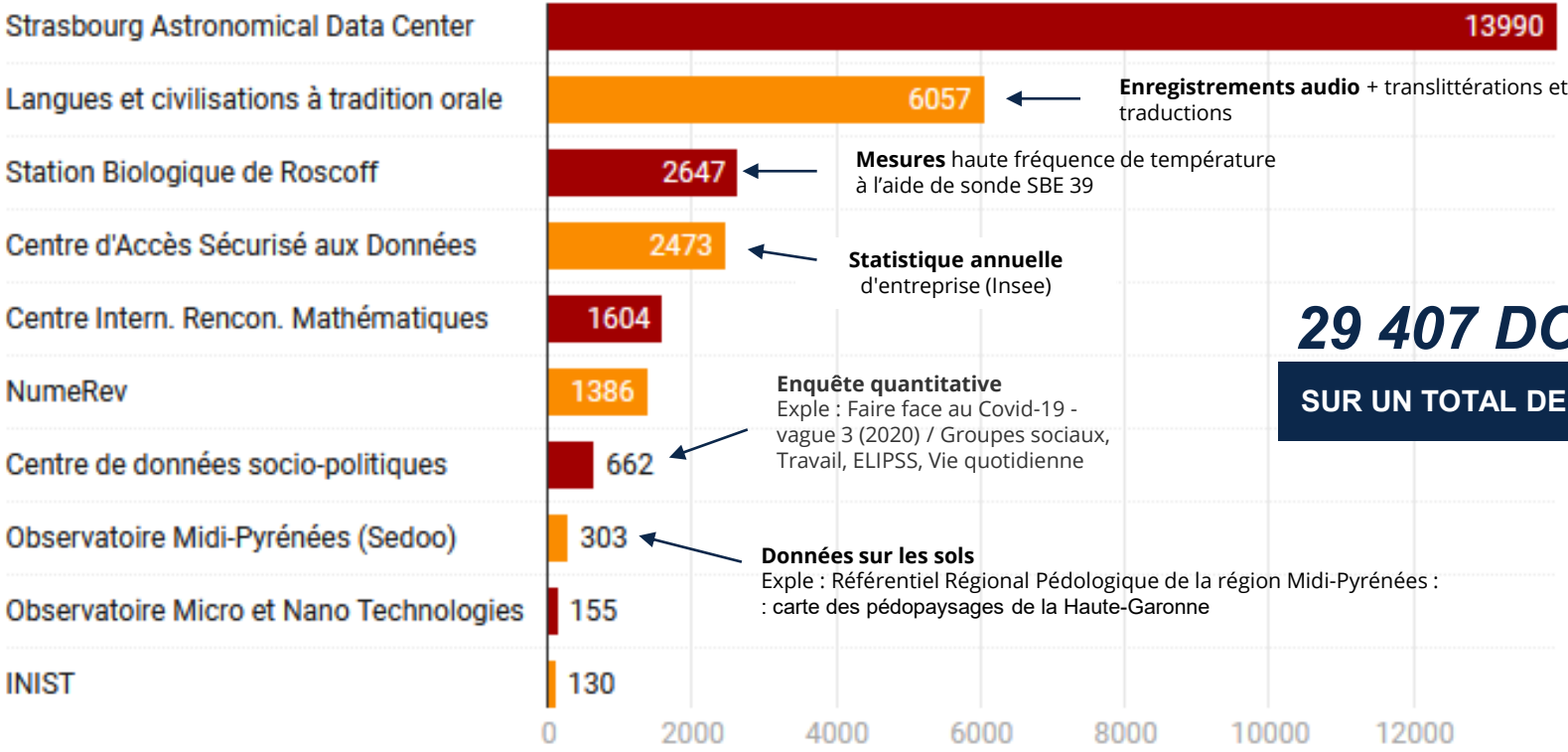
260 922

DOI attribués

LE TOP TEN DES MEMBRES (EN NOMBRE DE DOI, HORS ORGANISATIONS PRIVÉES)



TOP TEN DES MEMBRES CNRS (EN NOMBRE DE DOI)



Enregistrements audio + translittérations et traductions

Mesures haute fréquence de température à l'aide de sonde SBE 39

Statistique annuelle d'entreprise (Insee)

Enquête quantitative
Exple : Faire face au Covid-19 - vague 3 (2020) / Groupes sociaux, Travail, ELIPSS, Vie quotidienne

Données sur les sols
Exple : Référentiel Régional Pédologique de la région Midi-Pyrénées : : carte des pédopaysages de la Haute-Garonne

29 407 DOI

SUR UN TOTAL DE 29 991

Nouveau modèle économique DataCite - Structure du consortium Inist CNRS en 2021

24 organisations membres gérées par le « consortium lead » Inist CNRS en 2021

- 1 Cnrs qui regroupe les 49 laboratoires Cnrs (Labintel)
- 11 autres organismes de recherche et établissements d'enseignement supérieur (Universités, Grandes Écoles)
- 5 Associations scientifiques loi 1901
- 4 organismes internationaux (siège en France)
- 2 organismes étrangers (Universités)
- 1 organisation « Grandfather Bucket » (68 organismes qui attribuent peu de DOI)

26 000 €

Rappel en 2020 : 149 membres - budget de 14 000 euros

- Nouveau modèle d'adhésion à **DataCite** et Nouveau modèle économique
- Représentation française au sein des instances DataCite
- Gouvernance du consortium DataCite France
- Animation de la communauté des membres
- Services à développer

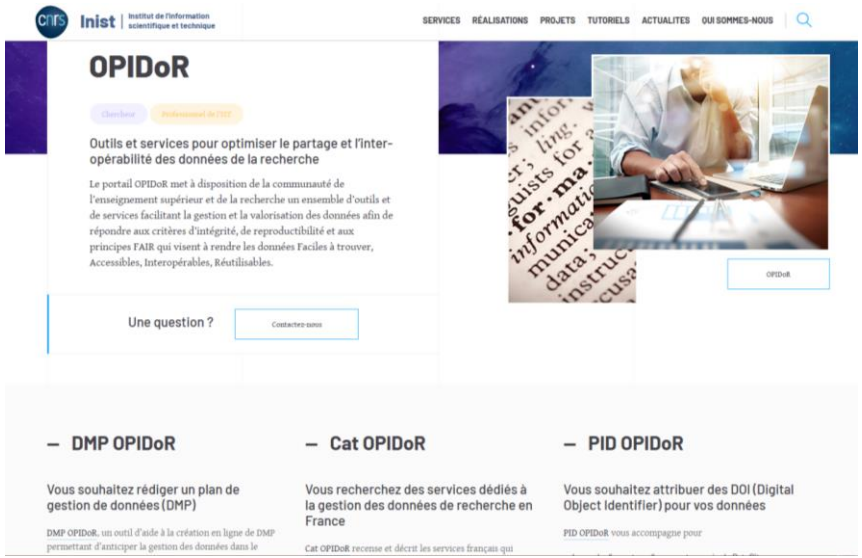
Objectifs

- Mettre à votre disposition un ensemble d'indicateurs sur le dépôt des données (via le DOI).
    - Exemples : Quelles sont les données enregistrées avec un DOI déposées par le CNRS ? Quels laboratoires ? Quelles thématiques ? Quels portails, entrepôts ? ...
- Mettre à disposition des corpus de métadonnées DataCite sur une thématique déterminée :
    - Exemple : Quelles sont les données déposées sur le thème de Cuba ? (exemple IRD)

Projet DIST Inist CNRS

C. Hadrossek (product owner), prestataire (scrum master), MC Jacquemot,

F. Tisserand, M. Yahia



MERCI !

Contact : claire.francois@inist.fr

https://www.opidor.fr

www.cnrs.fr https://www.inist.fr