

Dataverse pour les entrepôts de données : Dataverse à SciencesPo

Geneviève Michaud

▶ **To cite this version:**

Geneviève Michaud. Dataverse pour les entrepôts de données : Dataverse à SciencesPo. Ateliers thématiques RDA France, RDA France, Sep 2019, Paris, France. 10.5281/zenodo.7499453 . hal-03919048

HAL Id: hal-03919048

https://sciencespo.hal.science/hal-03919048

Submitted on 2 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

SciencesPo
CENTRE DE DONNÉES SOCIO-POLITIQUES

dataverse à Sciences Po

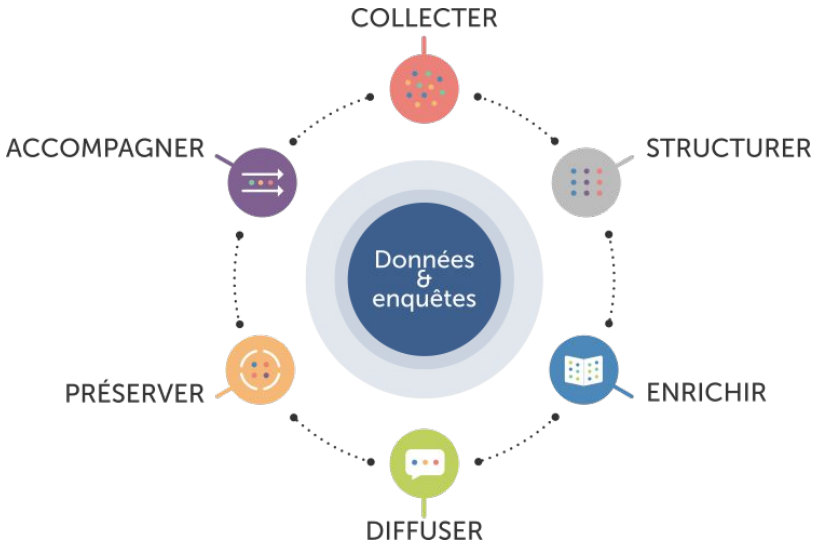
Atelier thématique "Dataverse pour les entrepôts de données", RDA France
12 Septembre 2019, CNRS rue Michel Ange, Paris

Geneviève Michaud, Centre de Données Socio-Politiques (UMS 828)

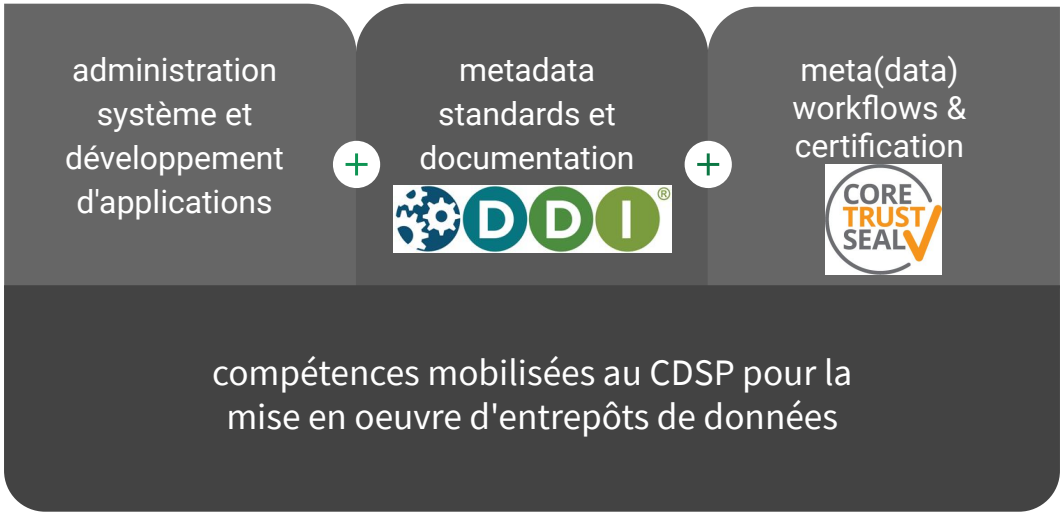
Creative Commons Attribution 4.0 International (CC BY 4.0)

le CDSP, centre de données et UMS

le CDSP et le cycle de vie des données de recherche



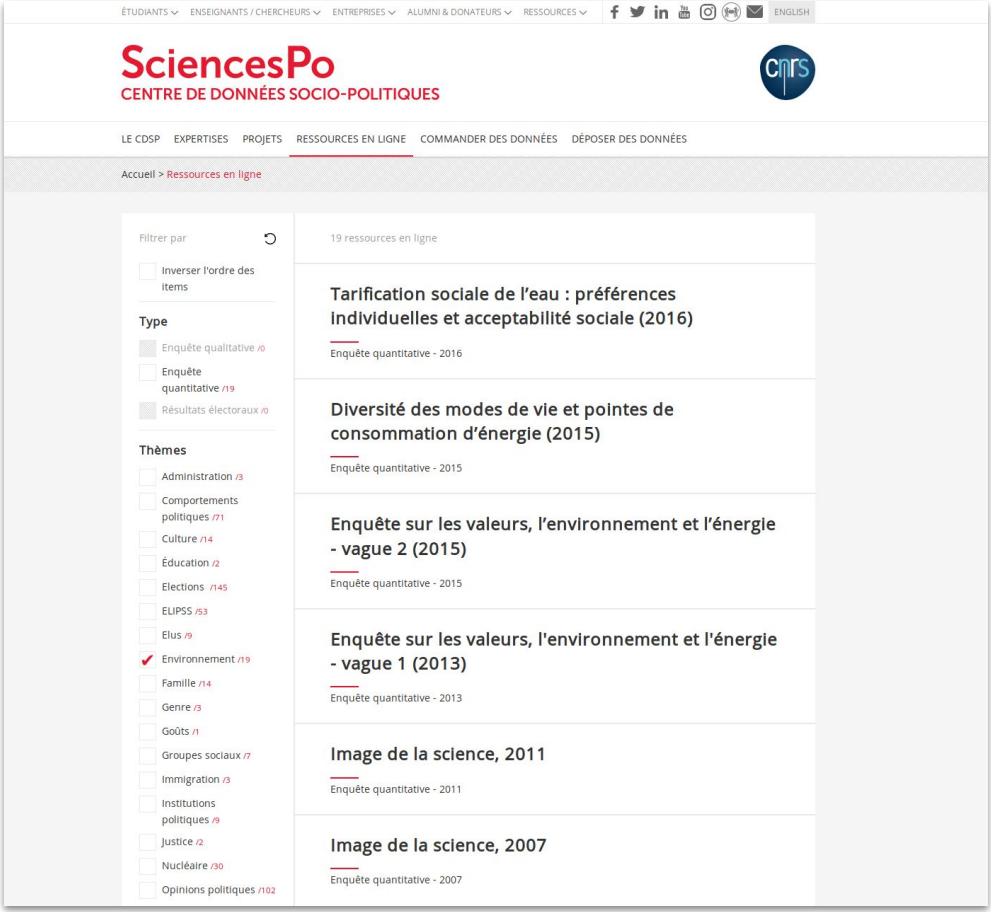
des compétences variées liées aux données de recherche



diversité des données

Données de recherche en Sciences Humaines et Sociales. Principalement données **d'enquêtes quantitatives** et **qualitatives**, ainsi que données électorales.

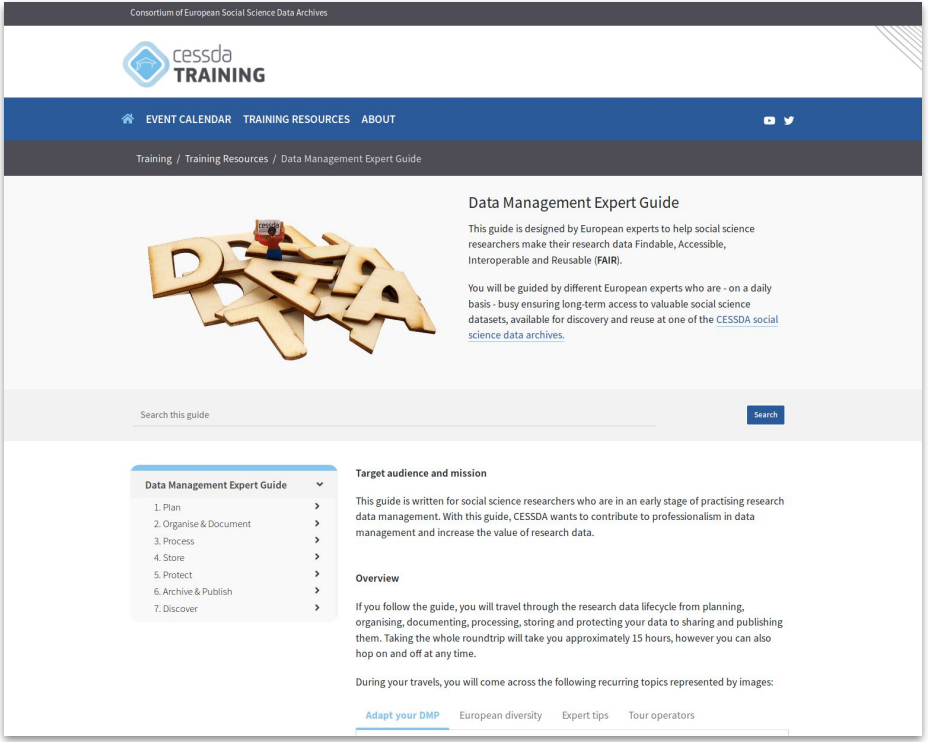
Définition d'un socle commun de documentation : le schéma de métadonnées DDI



un cadre structurant : CESSDA , ProGeDo

Le CDSP, partenaire de PROGEDO (TGIR) elle-même *service provider* français du CESSDA, (Consortium of European Social Science Data Archives).

Les recommandations du CESSDA constituent un **cadre** méthodologique **structurant**.



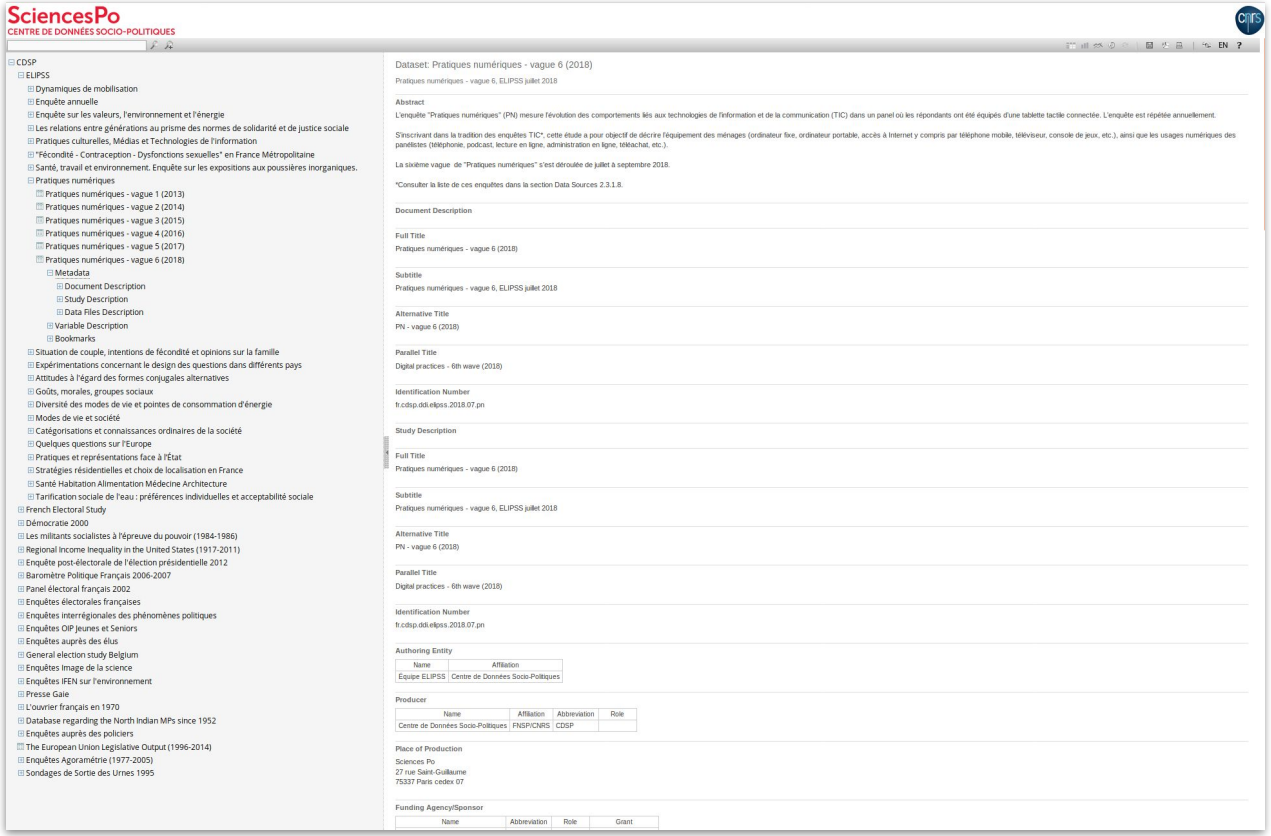
les entrepôts de données au CDSP

de Nesstar

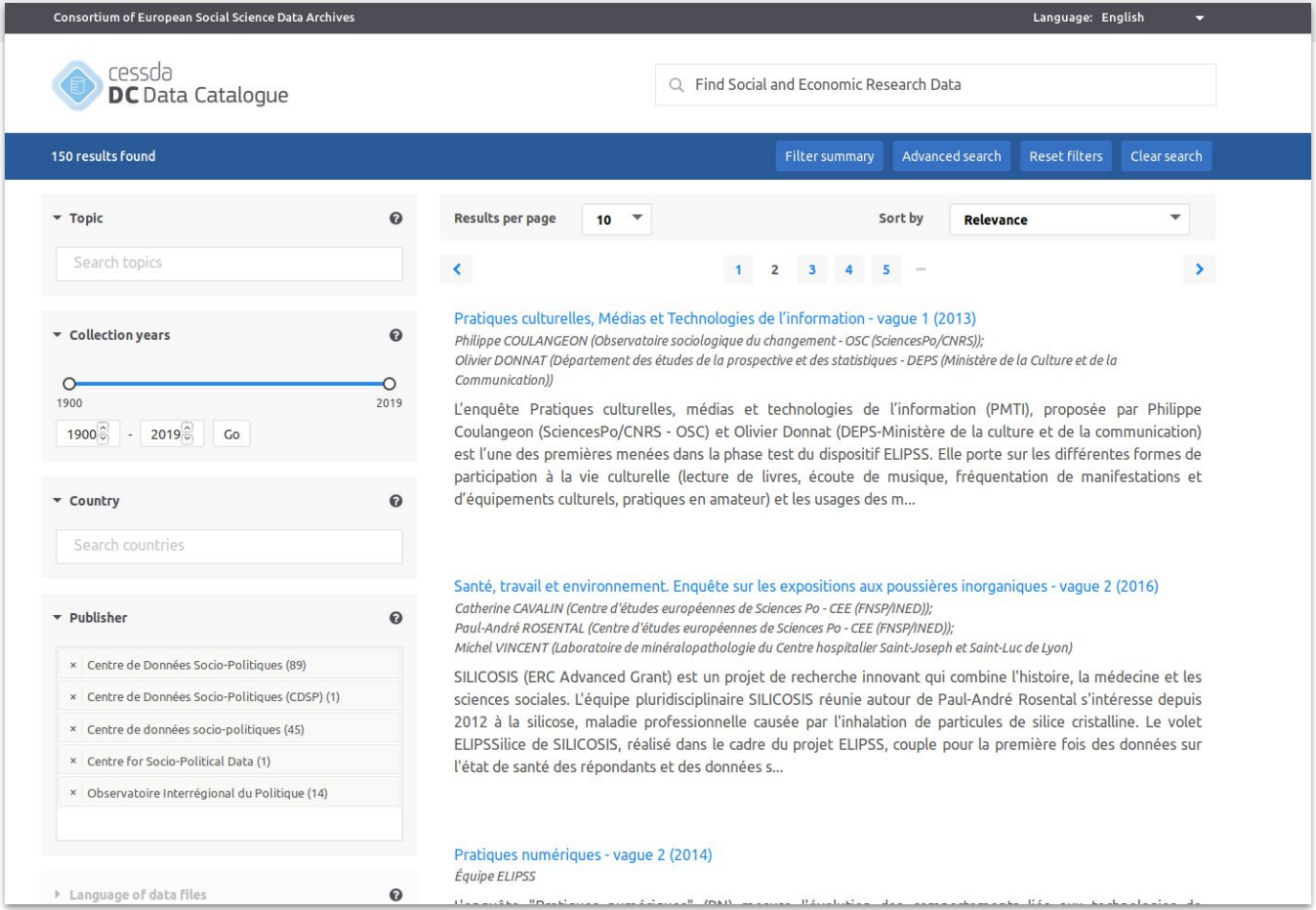
Depuis 2006 au CDSP.

- accessibilité
- inter-opérabilité
- standards (formats, protocoles)

- référencement
- PIDs
- modèle projet
- intégration



moissonnage du dataverse par CESSDA DC (protocole OAI-PMH)

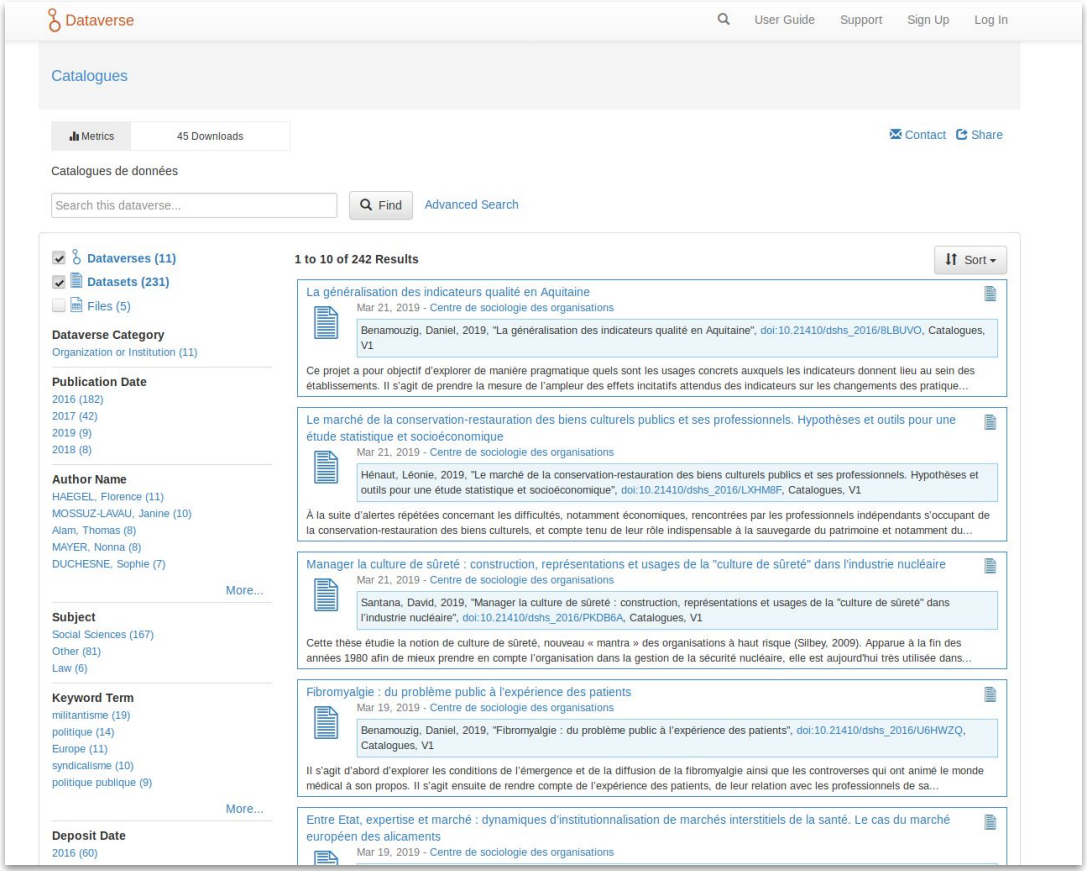


… à Dataverse

En 2016, lancement d'une première instance dataverse au CDSP.

Création d'une première collection pour le **signalement** d'enquêtes qualitatives (ANR ArchiPolis).

Développement des compétences de management et d'hébergement dataverse ; veille technologique.



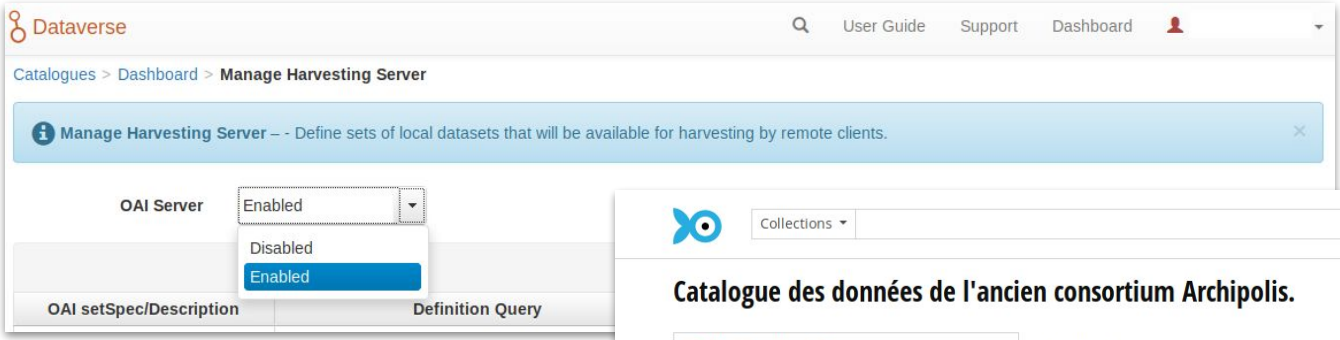
les atouts de dataverse

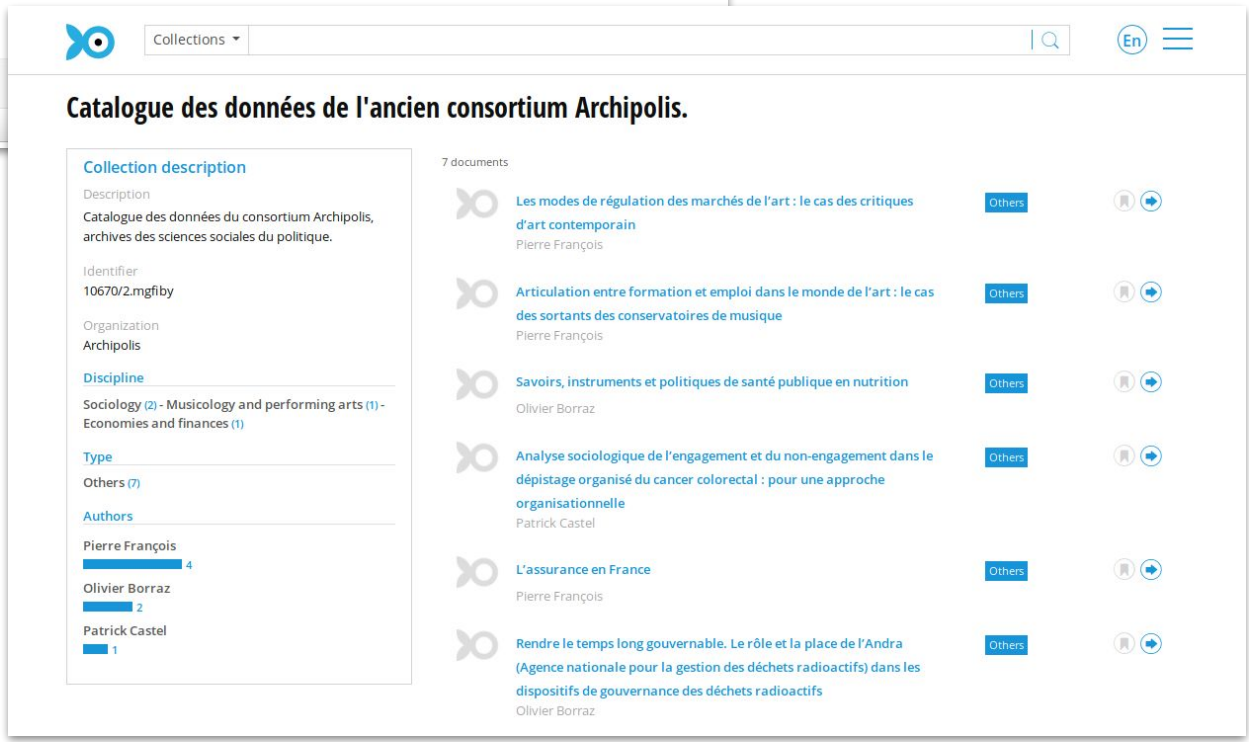


- cadre pour des données FAIR
  - PIDS, référencement
- intégration (API)
  - intégrable dans workflows de publication et l'ecosystème publications / données
- modèle open source
  - feuille de route publique, contributions communauté (ajout de fonctionnalités) : exemple DataverseEU



- granularité de la documentation





moissonnage du dataverse par Isidore

(protocole OAI-PMH)

un entrepôt de données institutionnel

For all today's technological affordances, data sharing remains fundamentally a social activity.

Data Communities
A New Model for Supporting STEM Data Sharing
Danielle Cooper, Rebecca Springer
https://doi.org/10.18665/sr.311396

un entrepôt à 3 collections

Une organisation conçue pour répondre aux **besoins des utilisateurs** des institutions et de la communauté de recherche (**chercheurs déposants** et **ré-utilisateurs**).

3 collections suivant déposant et le **niveau d'intervention** sur les données et métadonnées :

- 2 collections en **auto-dépôt :**
    - chercheurs Sciences Po
    - chercheurs extérieurs
- 1 collection pour les données documentées et diffusées par le CDSP





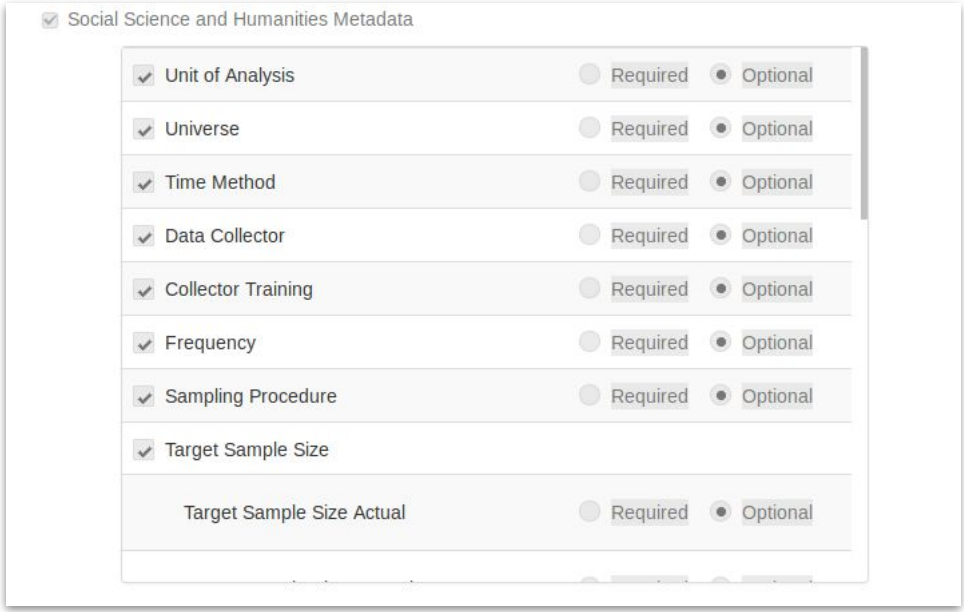
les métadonnées

sont modulaires :

- le *bloc* citation (obligatoire)
- le *bloc* "social sciences" (standard **DDI**)
- le *bloc* "geospatial" [...]

sont modulables : **vocabulaires contrôlés**

sont pré-paramétrables : par collection et/ou utilisation de **templates**

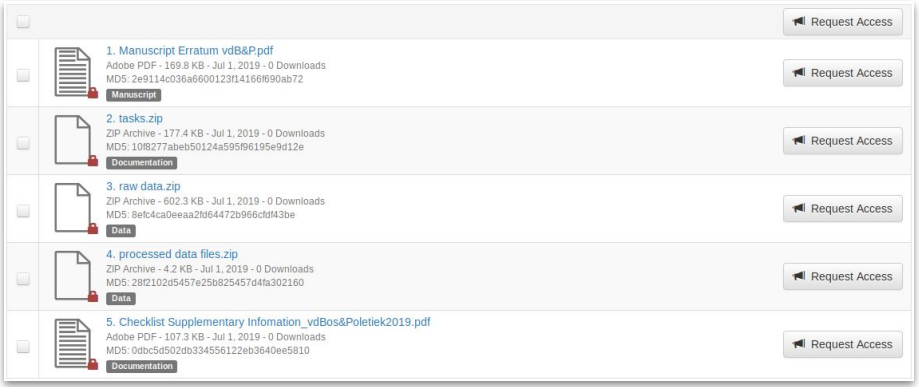


ouverture des données et modalités d'accès

Toutes les métadonnées, avec dataverse, sont publiques.

Toutes les données (notamment certaines données à caractère sensible) ne peuvent pas toujours être accessibles aussi librement.

Formalisation d'une **typologie** en 4 catégories, traduites en modalités d'accès.



[...] the realistic vision –and the challenge- of an institutional approach to RDM is probably a layered, component-based infrastructure with complementary support functions at various levels and with various types of data services, flexible and compliant with various situations, integrated in the national and international infrastructures, and embedded in an explicit and efficient organizational data strategy (policy), management (coordination) and follow-up (monitoring)

Joachim Schöpfel, Coline Ferrant, Francis Andre, Renaud Fabre. Research data management in the French National Research Center (CNRS). Data Technologies and Applications, 2018, ⟨10.1108/DTA-01-2017-0005⟩
https://dx.doi.org/10.1108/DTA-01-2017-0005