



HAL
open science

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

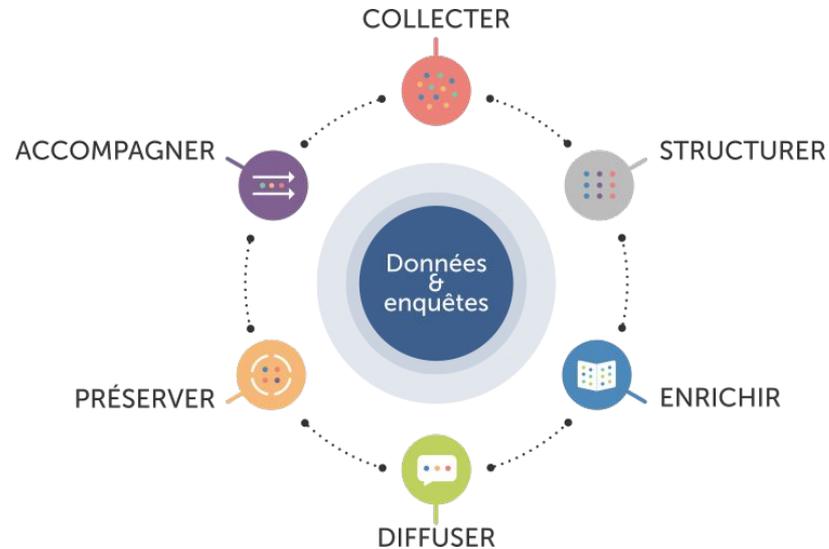
dataverse à Sciences Po

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)

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

administration
système et
développement
d'applications



metadata
standards et
documentation



meta(data)
workflows &
certification

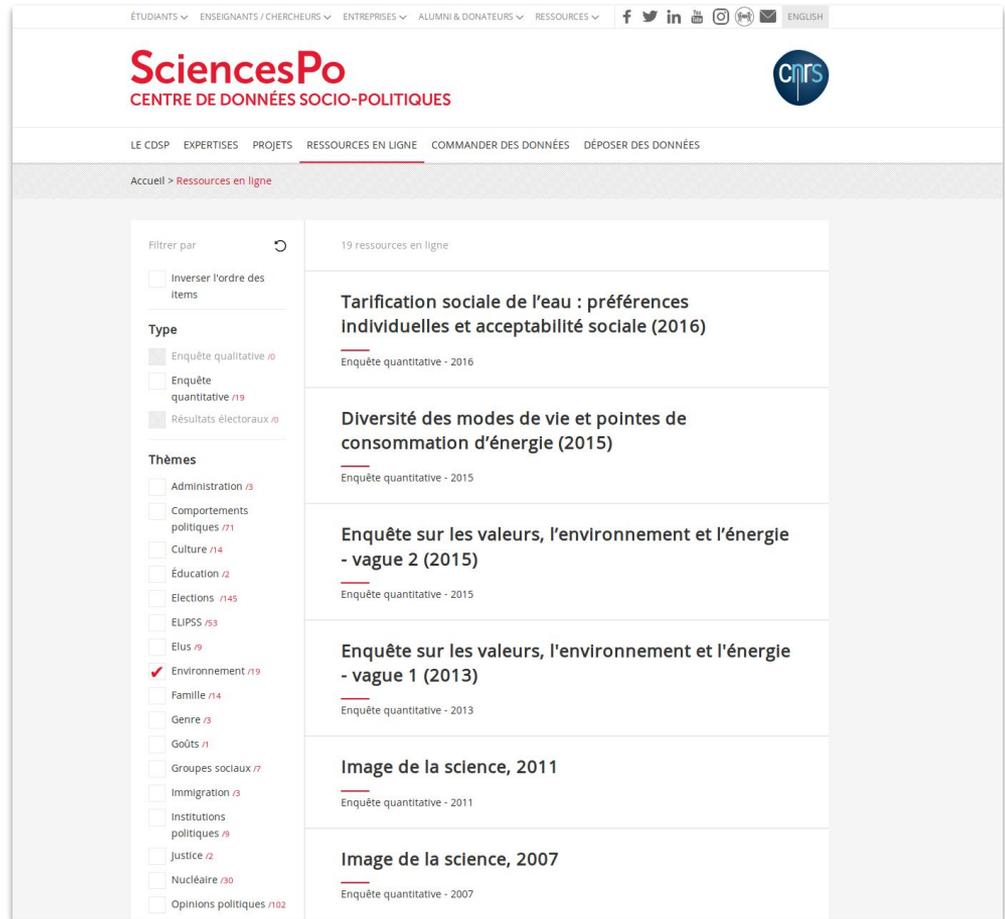


compétences mobilisées au CDSP pour la
mise en oeuvre d'entrepôts de données

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](#)



The screenshot displays the website interface for SciencesPo's Centre de Données Socio-Politiques. At the top, there is a navigation bar with menu items: ÉTUDIANTS, ENSEIGNANTS / CHERCHEURS, ENTREPRISES, ALUMNI & DONATEURS, RESSOURCES, and social media icons (Facebook, Twitter, LinkedIn, YouTube, Instagram, RSS, Email). The SciencesPo logo and the text 'CENTRE DE DONNÉES SOCIO-POLITQUES' are prominently displayed. Below this, a secondary navigation bar includes 'LE CDSP', 'EXPERTISES', 'PROJETS', 'RESSOURCES EN LIGNE' (which is highlighted), 'COMMANDER DES DONNÉES', and 'DÉPOSER DES DONNÉES'. The main content area shows 'Accueil > Ressources en ligne' and a list of 19 resources. A sidebar on the left allows filtering by 'Type' (Enquête qualitative /0, Enquête quantitative /19, Résultats électoraux /0) and 'Thèmes' (Administration /3, Comportements politiques /71, Culture /14, Éducation /2, Elections /145, ELIPSS /53, Elus /9, Environnement /19, Famille /14, Genre /3, Goûts /1, Groupes sociaux /7, Immigration /3, Institutions politiques /9, Justice /2, Nucléaire /30, Opinions politiques /102). The 'Environnement' theme is currently selected. The resource list includes: 'Tarification sociale de l'eau : préférences individuelles et acceptabilité sociale (2016)', 'Diversité des modes de vie et pointes de consommation d'énergie (2015)', 'Enquête sur les valeurs, l'environnement et l'énergie - vague 2 (2015)', 'Enquête sur les valeurs, l'environnement et l'énergie - vague 1 (2013)', 'Image de la science, 2011', and 'Image de la science, 2007'. Each entry is followed by 'Enquête quantitative' and the corresponding year.

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**.

Consortium of European Social Science Data Archives

CESSDA TRAINING

EVENT CALENDAR TRAINING RESOURCES ABOUT

Training / Training Resources / Data Management Expert Guide

Data Management Expert Guide

This guide is designed by European experts to help social science researchers make their research data Findable, Accessible, Interoperable and Reusable (FAIR).

You will be guided by different European experts who are - on a daily basis - busy ensuring long-term access to valuable social science datasets, available for discovery and reuse at one of the [CESSDA social science data archives](#).

Search this guide

Data Management Expert Guide ▾

- 1. Plan >
- 2. Organise & Document >
- 3. Process >
- 4. Store >
- 5. Protect >
- 6. Archive & Publish >
- 7. Discover >

Target audience and mission

This guide is written for social science researchers who are in an early stage of practising research data management. With this guide, CESSDA wants to contribute to professionalism in data management and increase the value of research data.

Overview

If you follow the guide, you will travel through the research data lifecycle from planning, organising, documenting, processing, storing and protecting your data to sharing and publishing them. Taking the whole roundtrip will take you approximately 15 hours, however you can also hop on and off at any time.

During your travels, you will come across the following recurring topics represented by images:

[Adapt your DMP](#) European diversity Expert tips Tour operators

les entrepôts de données au CDSP

de Nesstar

Depuis 2006 au CDSP.



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



- ◇ référencement
- ◇ PIDs
- ◇ modèle projet
- ◇ intégration

SciencesPo
CENTRE DE DONNÉES SOCIO-POLITIQUES

CDSP

- ELIPSS
 - Dynamiques de mobilisation
 - Enquête annuelle
 - Enquête sur les valeurs, l'environnement et l'énergie
 - Les relations entre générations au prisme des normes de solidarité et de justice sociale
 - Pratiques culturelles, Médias et Technologies de l'information
 - "Fécondité - Contraception - Dysfonctions sexuelles" en France Métropolitaine
 - Santé, travail et environnement. Enquête sur les expositions aux poussières inorganiques.
 - Pratiques numériques
 - Pratiques numériques - vague 1 (2013)
 - Pratiques numériques - vague 2 (2014)
 - Pratiques numériques - vague 3 (2015)
 - Pratiques numériques - vague 4 (2016)
 - Pratiques numériques - vague 5 (2017)
 - Pratiques numériques - vague 6 (2018)
 - Metadata
 - Document Description
 - Study Description
 - Data Files Description
 - Variable Description
 - Bookmarks
 - Situation de couple, intentions de fécondité et opinions sur la famille
 - Expérimentations concernant le design des questions dans différents pays
 - Attitudes à l'égard des formes conjugales alternatives
 - Goûts, morales, groupes sociaux
 - Diversité des modes de vie et points de consommation d'énergie
 - Modes de vie et société
 - Catégorisations et connaissances ordinaires de la société
 - Quelques questions sur l'Europe
 - Pratiques et représentations face à l'État
 - Stratégies résidentielles et choix de localisation en France
 - Santé Habitation Alimentation Médecine Architecture
 - Fertilisation sociale de l'eau : préférences individuelles et acceptabilité sociale
 - French Electoral Study
 - Démocratie 2000
 - Les militants socialistes à l'épreuve du pouvoir (1984-1986)
 - Regional Income Inequality in the United States (1917-2011)
 - Enquête post-électorale de l'élection présidentielle 2012
 - Baromètre Politique Français 2006-2007
 - Panel Electoral français 2002
 - Enquêtes électorales françaises
 - Enquêtes interrégionales des phénomènes politiques
 - Enquêtes OIP jeunes et Seniors
 - Enquêtes auprès des élus
 - General election study Belgium
 - Enquêtes Image de la science
 - Enquêtes IEN sur l'environnement
 - Presse Gaie
 - L'ouvrier français en 1970
 - Database regarding the North Indian MPs since 1952
 - Enquêtes auprès des policiers
 - The European Union Legislative Output (1996-2014)
 - Enquêtes Agoramétrie (1977-2005)
 - Sondages de Sortie des Urnes 1995

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

Consortium of European Social Science Data Archives Language: English

 Find Social and Economic Research Data

150 results found Filter summary Advanced search Reset filters Clear search

Topic ?

Collection years ?
1900 — 2019
 -

Country ?

Publisher ?

- × Centre de Données Socio-Politiques (89)
- × Centre de Données Socio-Politiques (CDSP) (1)
- × Centre de données socio-politiques (45)
- × Centre for Socio-Political Data (1)
- × Observatoire Interrégional du Politique (14)

▶ Language of data files ?

Results per page: 10 Sort by: Relevance

1 2 3 4 5 ...

Pratiques culturelles, Médias et Technologies de l'information - vague 1 (2013)
*Philippe COULANGEON (Observatoire sociologique du changement - OSC (SciencesPo/CNRS));
Olivier DONNAT (Département des études de la prospective et des statistiques - DEPS (Ministère de la Culture et de la Communication))*

L'enquête Pratiques culturelles, médias et technologies de l'information (PMTI), proposée par Philippe Coulangeon (SciencesPo/CNRS - OSC) et Olivier Donnat (DEPS-Ministère de la culture et de la communication) est l'une des premières menées dans la phase test du dispositif ELIPSS. Elle porte sur les différentes formes de participation à la vie culturelle (lecture de livres, écoute de musique, fréquentation de manifestations et d'équipements culturels, pratiques en amateur) et les usages des m...

Santé, travail et environnement. Enquête sur les expositions aux poussières inorganiques - vague 2 (2016)
*Catherine CAVALIN (Centre d'études européennes de Sciences Po - CEE (FNSP/INED));
Paul-André ROSENAL (Centre d'études européennes de Sciences Po - CEE (FNSP/INED));
Michel VINCENT (Laboratoire de minéralopathologie du Centre hospitalier Saint-Joseph et Saint-Luc de Lyon)*

SILICOSIS (ERC Advanced Grant) est un projet de recherche innovant qui combine l'histoire, la médecine et les sciences sociales. L'équipe pluridisciplinaire SILICOSIS réunie autour de Paul-André Rosental s'intéresse depuis 2012 à la silicose, maladie professionnelle causée par l'inhalation de particules de silice cristalline. Le volet ELIPSSilice de SILICOSIS, réalisé dans le cadre du projet ELIPSS, couple pour la première fois des données sur l'état de santé des répondants et des données s...

Pratiques numériques - vague 2 (2014)
Équipe ELIPSS

Enquête "Pratiques numériques" (PN) menée l'élaboration des comportements liés aux technologies de

... 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.

The screenshot displays the Dataverse website interface. At the top, the 'Dataverse' logo is on the left, and navigation links for 'User Guide', 'Support', 'Sign Up', and 'Log In' are on the right. Below the logo, the page title 'Catalogues' is shown. A 'Metrics' box indicates '45 Downloads'. There are links for 'Contact' and 'Share'. A search bar contains the text 'Search this dataverse...' with a 'Find' button and a link to 'Advanced Search'. On the left side, there are filters for 'Dataverses (11)', 'Datasets (231)', and 'Files (5)'. Below these are sections for 'Dataverse Category' (Organization or Institution (11)), 'Publication Date' (2016 (182), 2017 (42), 2019 (9), 2018 (8)), 'Author Name' (HAEGEL, Florence (11), MOSSUZ-LAWAU, Janine (10), Alam, Thomas (8), MAYER, Nonna (8), DUCHESNE, Sophie (7)), 'Subject' (Social Sciences (167), Other (81), Law (6)), 'Keyword Term' (militantisme (19), politique (14), Europe (11), syndicalisme (10), politique publique (9)), and 'Deposit Date' (2016 (60)). Each section has a 'More...' link. The main content area shows '1 to 10 of 242 Results' with a 'Sort' dropdown. The first result is 'La généralisation des indicateurs qualité en Aquitaine' by Benamouzig, Daniel, 2019. The second is 'Le marché de la conservation-restauration des biens culturels publics et ses professionnels. Hypothèses et outils pour une étude statistique et socioéconomique' by Hénaut, Léonie, 2019. The third is 'Manager la culture de sûreté : construction, représentations et usages de la "culture de sûreté" dans l'industrie nucléaire' by Santana, David, 2019. The fourth is 'Fibromyalgie : du problème public à l'expérience des patients' by Benamouzig, Daniel, 2019. The fifth is 'Entre Etat, expertise et marché : dynamiques d'institutionnalisation de marchés interstitiels de la santé. Le cas du marché européen des alicaments' by an author from the Centre de sociologie des organisations.

les atouts de dataverse

- ◇ cadre pour des données FAIR
PIDS, référencement



- ◇ intégration (API)
intégrable dans workflows de publication et l'écosystème publications / données

- ◇ modèle open source
[feuille de route publique](#), contributions communauté (ajout de fonctionnalités) : exemple DataverseEU



- ◇ granularité de la documentation

Dataverse User Guide Support Dashboard

Catalogues > Dashboard > **Manage Harvesting Server**

Manage Harvesting Server -- Define sets of local datasets that will be available for harvesting by remote clients.

OAI Server Enabled

Disabled

Enabled

OAI setSpec/Description Definition Query

moissonnage du dataverse
par Isidore

(protocole OAI-PMH)

Collections En

Catalogue des données de l'ancien consortium Archipolis.

7 documents

Collection description

Description
Catalogue des données du consortium Archipolis, archives des sciences sociales du politique.

Identifier
10670/2.mgfi by

Organization
Archipolis

Discipline
Sociology (2) - Musicology and performing arts (1) - Economics and finances (1)

Type
Others (7)

Authors

Pierre François 4

Olivier Borraz 2

Patrick Castel 1

- Les modes de régulation des marchés de l'art : le cas des critiques d'art contemporain** Others
Pierre François
- Articulation entre formation et emploi dans le monde de l'art : le cas des sortants des conservatoires de musique** Others
Pierre François
- Savoirs, instruments et politiques de santé publique en nutrition** Others
Olivier Borraz
- Analyse sociologique de l'engagement et du non-engagement dans le dépistage organisé du cancer colorectal : pour une approche organisationnelle** Others
Patrick Castel
- L'assurance en France** Others
Pierre François
- Rendre le temps long gouvernable. Le rôle et la place de l'Andra (Agence nationale pour la gestion des déchets radioactifs) dans les dispositifs de gouvernance des déchets radioactifs** Others
Olivier Borraz

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**

The image shows a configuration window titled "Social Science and Humanities Metadata". It contains a list of metadata elements, each with a checked checkbox and two radio buttons for "Required" and "Optional". The "Required" radio button is selected for all elements. The elements are: Unit of Analysis, Universe, Time Method, Data Collector, Collector Training, Frequency, Sampling Procedure, and Target Sample Size. Below this list is a section for "Target Sample Size Actual" with "Required" and "Optional" radio buttons, where "Required" is also selected.

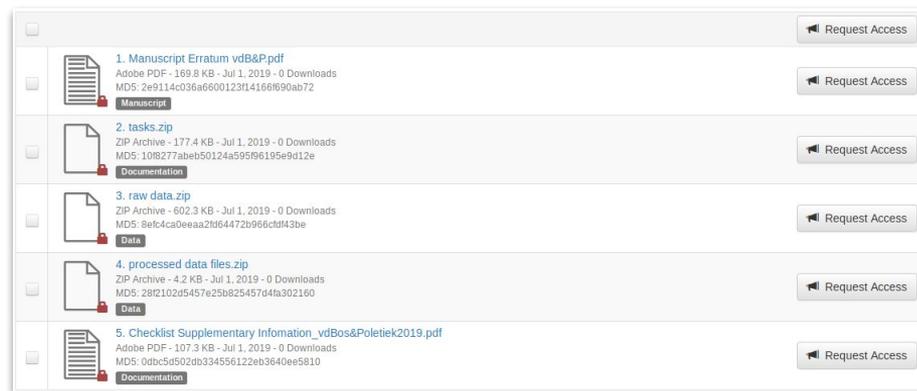
Metadata Element	Required	Optional
<input checked="" type="checkbox"/> Unit of Analysis	<input checked="" type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Universe	<input checked="" type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Time Method	<input checked="" type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Data Collector	<input checked="" type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Collector Training	<input checked="" type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Frequency	<input checked="" type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Sampling Procedure	<input checked="" type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Target Sample Size	<input checked="" type="radio"/>	<input type="radio"/>
Target Sample Size Actual	<input checked="" type="radio"/>	<input type="radio"/>

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>
