



HAL
open science

Métadonnées et standards de documentation

Lucie Marie

► **To cite this version:**

Lucie Marie. Métadonnées et standards de documentation. Semaine Data SHS : Traiter et analyser des données en sciences humaines et sociales, Sciences Po, Centre de données Socio-politiques (CDSP), CNRS; PUD des Grands Moulins de l'Université de Paris, Dec 2020, Paris, France. 10.5281/zenodo.4309815 . halshs-03169392

HAL Id: halshs-03169392

<https://shs.hal.science/halshs-03169392>

Submitted on 13 Dec 2022

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.

Métadonnées et standards de documentation

Lucie MARIE

Centre de données socio-politiques, Sciences Po, CNRS

Semaine DATA-SHS - Plateformes Universitaires de Données Paris
Evènement en ligne
7 décembre 2020

Un peu d'histoire !

Des bibliothèques d'Alexandrie en -280 av. J.C. au enjeux du BIG DATA :

Terme « metadata » par Philip Bagley, 1968 ;

Dictionnaire de Données Statistiques par Claude Poulain, 1983 ;

Différents types de métadonnées par Bo Sungren, 1995.

Recherche, Accessibilité, Interopérabilité, Réutilisation

FAIR DATA PRINCIPLES

AH!



FINDABLE



ACCESSIBLE



INTEROPERABLE



REUSABLE

Documentation à partir de standards internationaux :

Gestion des données ; interopérabilité des systèmes d'information ; science ouverte.

Data Documentation Initiative

« La Data Documentation Initiative (DDI) est une norme internationale permettant de décrire les données issues d'enquêtes et d'autres méthodes d'observation en sciences sociales, comportementales, économiques et de la santé. DDI peut documenter et gérer différentes étapes du cycle de vie des données de recherche, telles que la conception, la collecte, le traitement, la diffusion, la découverte et l'archivage. »

Source : DDI Alliance



Les utilisateurs DDI



Source : <https://ddialliance.org/community/join>



Centralising and Integrating Metadata from European Statistics



Statistics Canada



Statistique Canada



Cohort & Longitudinal Studies Enhancement Resources



A PARTNER IN SOCIAL SCIENCE RESEARCH

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH



ADA AUSTRALIAN DATA ARCHIVE



Le standard DDI

- Spécifications : plus de 120 balises pour construire un schéma de métadonnées
- Vocabulaires contrôlés : termes prédéfinis pour pouvoir naviguer plus facilement dans les métadonnées
- XML : langage structuré, balise les données et permet la portabilité des informations



Les versions DDI

- **Codebook** (DDI-C)
 - > Livre de code, enquêtes uniques (2000)
- **Lifecycle** (DDI-L)
 - > Référentiel de métadonnées, données longitudinales (2008)
- **Cross Domain Integration** (DDI-CDI)
 - > Sources hétérogènes de données (2020)



DDI-Codebook

- Dictionnaire de code : description d'une enquête à **posteriori**
- Donner une structure aux métadonnées pour rendre compte du contenu d'un seul fichier de données d'une enquête



DDI-C : deux niveaux de documentation



DDI-C : spécifications et métadonnées

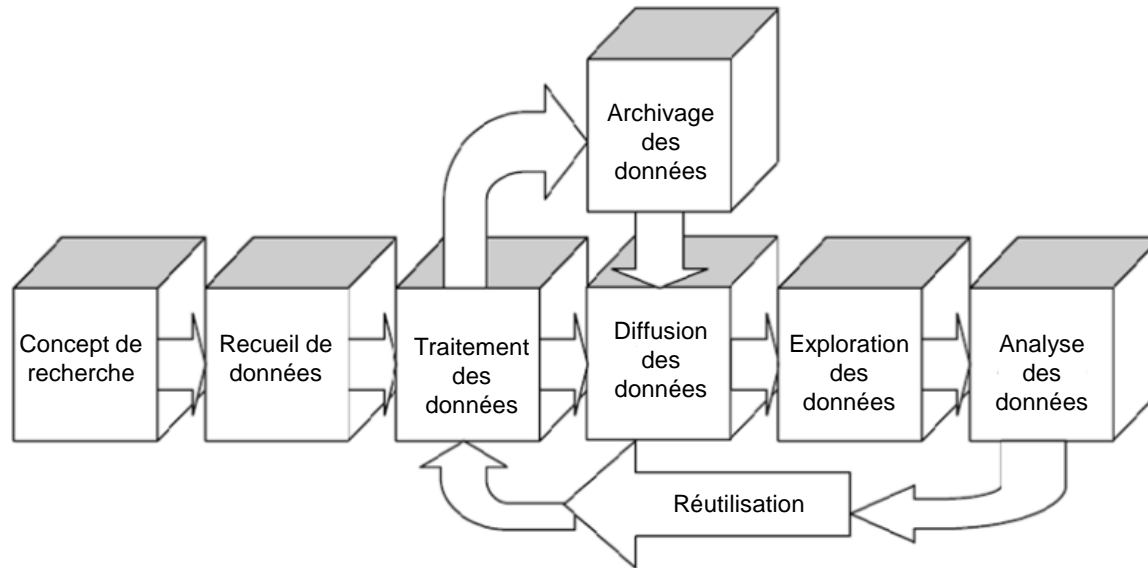


DDI-Lifecycle

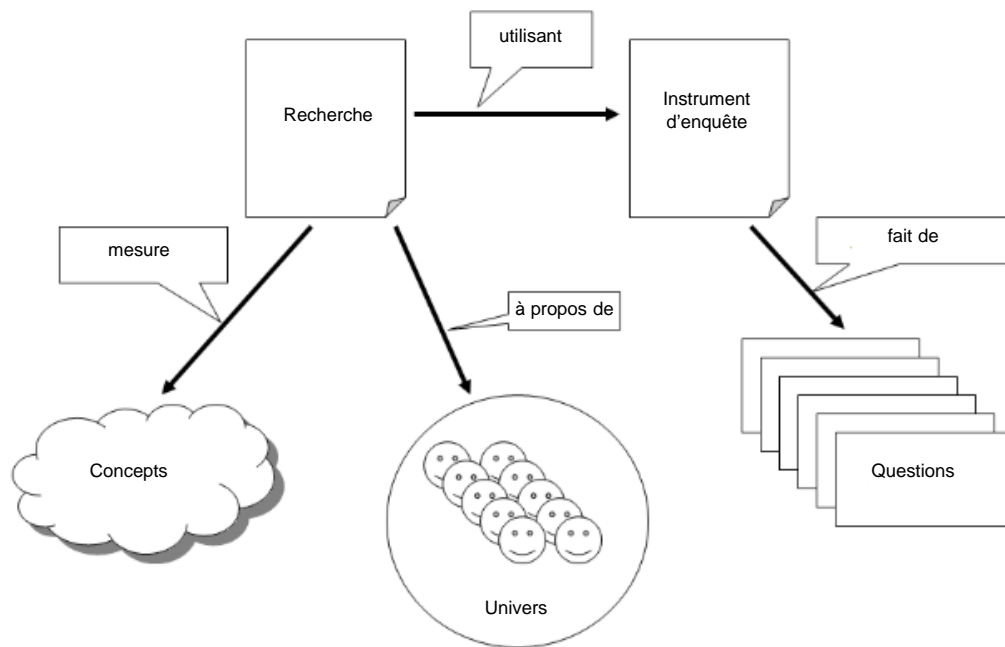
- Description du cycle de vie entier des données
- Principe de registre et de référentiel avec des métadonnées uniques
- Système de métadonnées centralisé : **saisir une fois réutiliser plusieurs fois**
- Principes clés : unicité, réutilisation et connexion des informations entre les différentes enquêtes



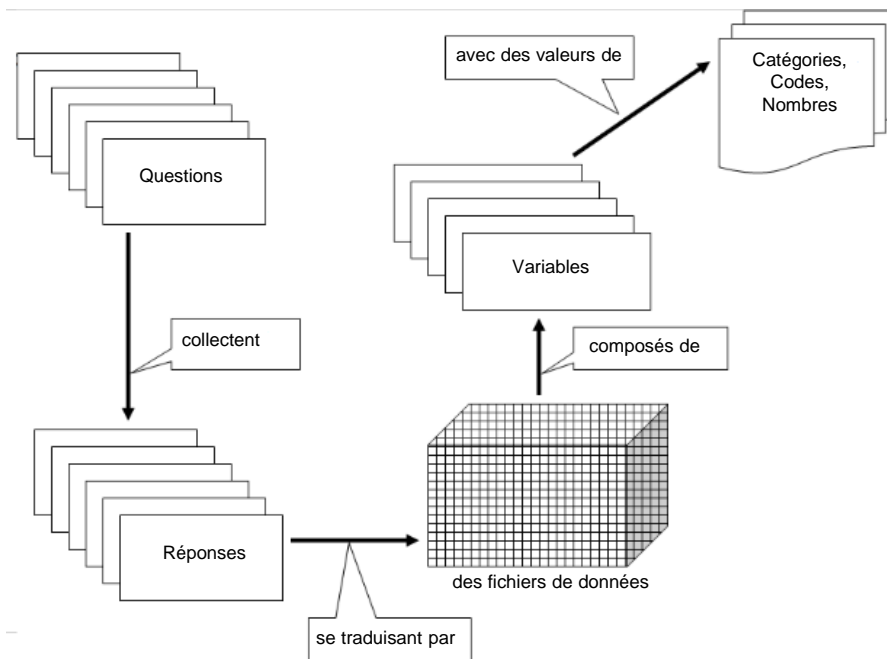
DDI-L : le cycle de vie des données



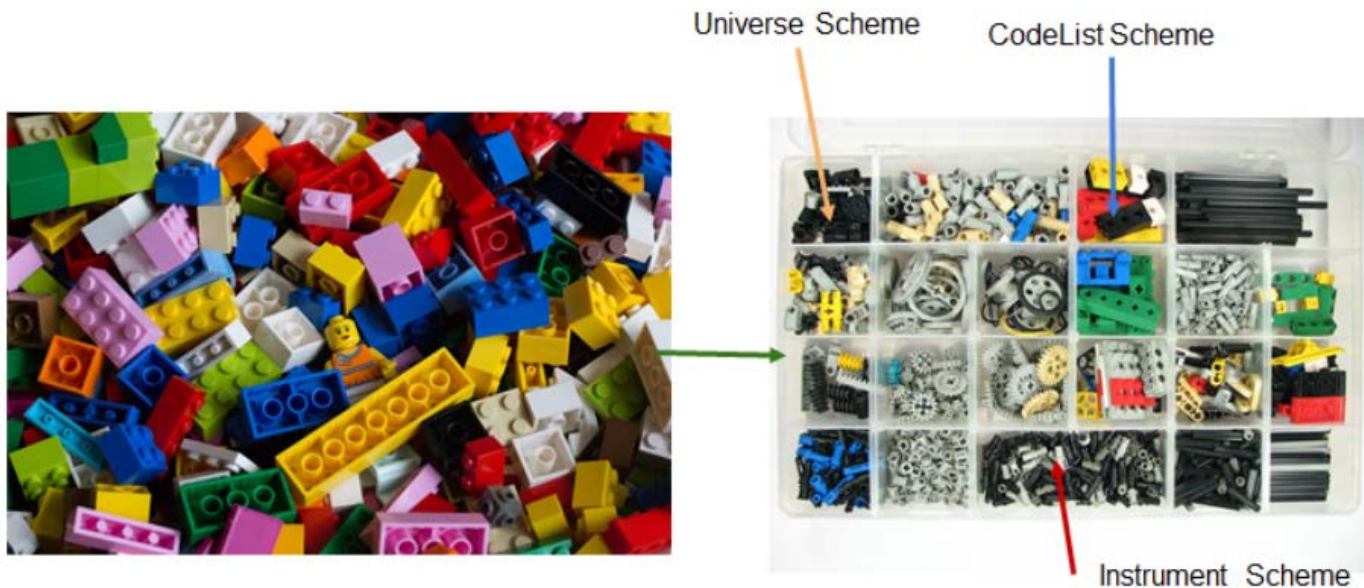
DDI-L en 60 secondes



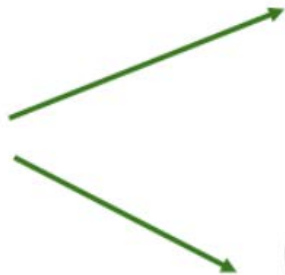
DDI-L en 60 secondes



DDI-L : spécifications et métadonnées



DDI-L : la réutilisation des métadonnées



DDI-C vs. DDI-L

DDI-C : Documentation d'enquêtes simples et uniques

Avantage :

- Documentation des variables contient la majorité des informations

Limites:

- Uniquement sur l'étape de traitement des données
- Métadonnées non-harmonisées

DDI-L : Documentation d'enquêtes longitudinales

Avantages :

- Cycle de vie entier des données
- Gestion cohérente et métadonnées harmonisées (référencement et réutilisation)
- Facilite l'exploration pour les utilisateurs
- Documentation multilingue

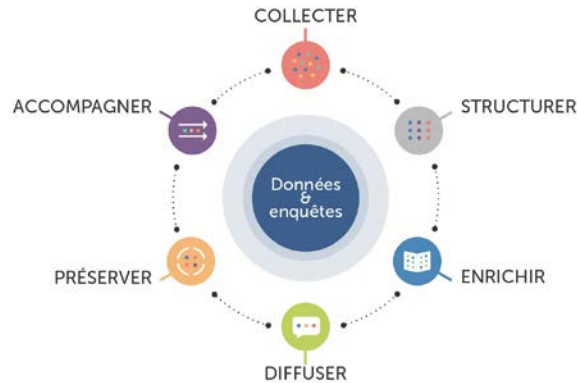
Limites :

- Ressources plus importantes

Les outils de diffusion de l'information du CDSP

Documentation et Diffusion & Projets Numériques

Mission : partager les (méta)données pour la communauté de la recherche - française et internationale



data.sciencespo & beQuali

Dataverse 



SciencesPo

Ajouter des données • Rechercher • Guide d'utilisation • Support • Français • S'inscrire • Se connecter

data.sciencespo

Statistiques 681 téléchargements

Contact Partager

L'entrepôt institutionnel de données de la recherche de Sciences Po

Chercher dans ce dataverse...

Trouver

Recherche avancée

Dataverses (14)

Jeux de données (314)

Fichiers (733)

Dataverse Category

Laboratoire (12)

Organization or Institution (1)

Research Project (1)

Publication Year

2020 (328)

Classification des sujets Terme

Elections (172)

Comportements et attitudes politiques (84)

Énergie et ressources naturelles (33)

Gouvernement, systèmes et organisations politiques (22)

Comportement social et attitudes (21)

Plus...

Subject

Social Sciences (319)

Mot-clé Terme

ATTITUDES POLITIQUES (95)

COMPORTEMENT POLITIQUE (66)

ELECTIONS PRESIDENTIELLES (43)

GOVERNEMENT RÉGIONAL (41)

1 à 10 de 328 résultats

Trier

Livraisons des colis et Mobilités des e-consommateurs : caractérisation des pratiques et des flux (2016)

5 mai 2020 - ELIPSS



Aguilera, Anne. 2020. "Livraisons des colis et Mobilités des e-consommateurs : caractérisation des pratiques et des flux (2016)". <https://doi.org/10.21410/7E4JOTU8V>, data.sciencespo, V2

L'enquête Livraisons des colis et Mobilités des e-consommateurs (LivMob), coordonnée par A. Aguilera, a pour objectif d'identifier les déterminants du choix d'un mode de livraison et d'analyser tout particulièrement le rôle des territoires et des mobilités. L'enquête approfondit...

Les rapports des citoyens à la justice : expériences et représentations (2018)

5 mai 2020 - ELIPSS



Vigour, Cécile. 2020. "Les rapports des citoyens à la justice : expériences et représentations (2018)". <http://doi.org/10.21410/7E4JHSF4LJH>, data.sciencespo, V2

L'enquête Rapports des citoyens à la justice : expériences et représentations (JustiRep) a pour objectif d'appréhender les représentations des citoyens à l'égard de la justice et du système judiciaire incluant les forces de l'ordre. D'une part, JustiRep rend compte de la pluralité...

Pratiques de visite des Français (musées et expositions) (2018)

Base de questions Quetelet

The screenshot displays the Réseau Quetelet search interface. At the top, there are navigation links for 'Français', 'Recherche simple', 'Recherche avancée', 'Liste des enquêtes', 'Aide', 'Préférences', 'A propos', and 'Votre sélection (2)'. The search bar contains the text 'élections régionales' and a 'Rechercher' button. Below the search bar, there are filters for 'question', 'modalités', and 'variable'. The main content area shows search results for 'élections régionales', with 115 results found. The results are sorted by 'score de pertinence'. The first result is a question: 'Pendant les campagnes électorales, les candidats essaient d'être en contact avec les citoyens et tentent de les informer par différents moyens. Lors de la campagne des élections régionales, avez-vous...'. This question has 8 modalities and a pertinence score of 5. The second result is a question: 'Beaucoup d'électeurs n'ont pas voté lors des élections régionales du 15 mars dernier. Vous-même, avez-vous voté lors des élections régionales du 15 mars dernier?'. This question has 6 modalities and a pertinence score of 5. The interface also shows a sidebar with 'Producteur' and 'Série d'enquêtes' filters, and a 'Pertinence' bar for each result.

Réseau Quetelet

français | anglais
question modalités variable

élections régionales Rechercher

Résultats 1 à 10 sur un total de 115 pour élections régionales
Trier par score de pertinence Afficher 5 modalités

Question Pendant les campagnes électorales, les candidats essaient d'être en contact avec les citoyens et tentent de les informer par différents moyens. Lors de la campagne des **élections régionales**, avez-vous...

(8) Regardé une émission sur les **élections** ou un débat entre les candidats à la télévision - [dn12_CPGNTV - Campagne électorale : visionnage d'émission débat à la télévision](#)

Lu des articles sur les **élections** dans un journal - [dn12_CPGNART - Campagne électorale : lecture d'articles dans les journaux sur les élections](#)

Parlé des **élections** avec vos amis ou votre famille - [dn12_CPGNDISC - Campagne électorale : discussion des election avec les amis, la famille](#)

Modalités (4)
Enquête [Dynamiques de mobilisation - vague 12 \(2015\)](#) - Equipe ELIPSS

Pertinence

Question Beaucoup d'électeurs n'ont pas voté lors des **élections régionales** du 15 mars dernier. Vous-même, avez-vous voté lors des **élections régionales** du 15 mars dernier ?

Modalités (6)
Variable [q37 - Participation aux élections régionales](#)

Enquête [Enquête interrégionale des phénomènes politiques 1998](#) - OIP

Pertinence

Question Pour ce que vous en savez, est-ce que les **élections régionales** de mars 1998 en [nom de la région] ont entraîné :

(2) un changement du président du conseil **régional** - [q13a1 - Notoriété du](#)

UpMet : une banque de questions en DDI-L

CDSP Search Explore Basket (0) Help

Religion du répondant

French Electoral Study French Electoral Study 2007 Religion

Appears Within

- Series
 - French Electoral Study
- Studies
 - French Electoral Study 2007
 - French Electoral Study 2012
- Variables
 - Religion
 - Religion du répondant

Represented Variable

Name religion.repondant
Label Religion du répondant
Conceptual Variable Religion
Category Set Socio-démographiques

Concordance

Statistics Code Comparison Correspondence Tree

French Electoral Study 2012 - French Electoral Study 2007

% of valid % of total


	French Electoral Study 2012 sd46	French Electoral Study 2007 r30
1 - Catholique	69.47%	65.95%
2 - Protestante	2.32%	1.70%
3 - Juive	0.55%	0.55%
4 - Musulmane	3.93%	1.65%
5 - Autre religion	1.36%	1.80%
6 - Sans religion	22.37%	27.79%

Dataset	Variable	Valid	Invalid	Min	First Quartile	Median	Third Quartile	Max	Mean	StdDev
French Electoral Study 2012	sd46	1985	29	1				6		
French Electoral Study 2007	r30	1997	3	1				7		

Des questions ?

Merci !

Pour me contacter :  lucie.marie2@sciencespo.fr

 [@luciemariejsph](https://twitter.com/luciemariejsph)

 [linkedin.com/in/lucie-marie-josephine](https://www.linkedin.com/in/lucie-marie-josephine)