



HAL
open science

Using Colectica to document socio-political surveys in DDI-Lifecycle: feedback and benefits

Lucie Marie

► **To cite this version:**

Lucie Marie. Using Colectica to document socio-political surveys in DDI-Lifecycle: feedback and benefits. 12th Annual European DDI User Conference (EDDI20), Nov 2020, Paris (virtual event), France. halshs-03219826

HAL Id: halshs-03219826

<https://shs.hal.science/halshs-03219826v1>

Submitted on 6 May 2021

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.

Using Colectica to document socio-political surveys in DDI-Lifecycle: feedback and benefits

[Lucie MARIE](#)

Center for Socio-Political Data (CDSP)

EDDI 2020, 12th Annual European DDI Users Conference

Virtual Event

December 1, 2020

Plan

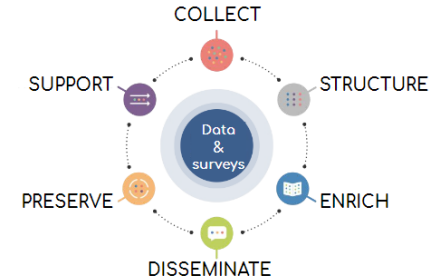
- CDSP missions & online tools
- UpMet project
- New variable documentation workflow
- Metadata visualisation on Colectica Portal
- Feedbacks & benefits
- Next steps

Center for socio-political data (CDSP) Sciences Po, CNRS

Collects and shares (meta)data for research community

Our sharing online tools

- Data: [Quetelet-Progedo platform](#)
- Metadata: from [Nesstar](#) to [data.sciencespo](#) & [Colectica](#)
- Question bank: [Quetelet-Progedo question bank](#)



UpMet project



“Upscaling metadata for increasing reuse in the social sciences”

Open science, FAIR principles, data stewardship & storage

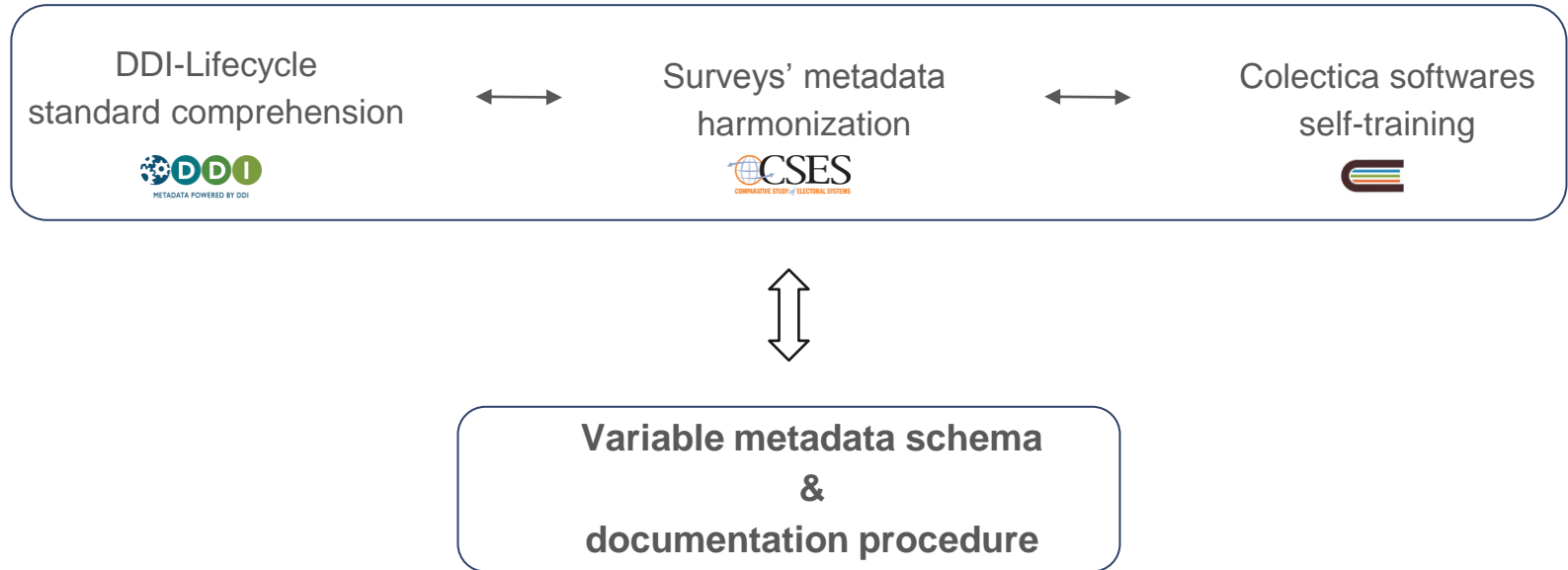
WP2: Question bank in DDI-L with Colectica

Objective: metadata harmonization

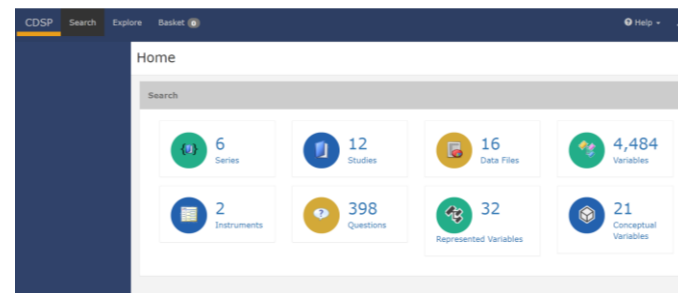
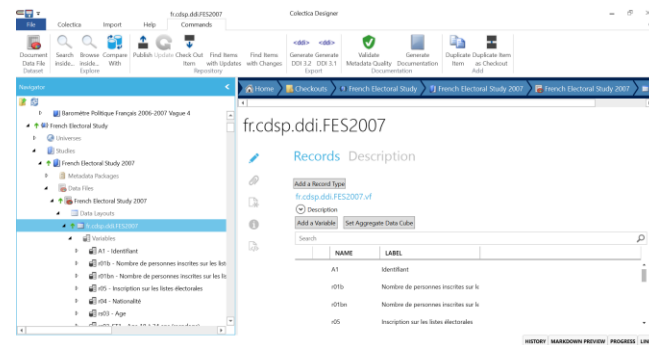
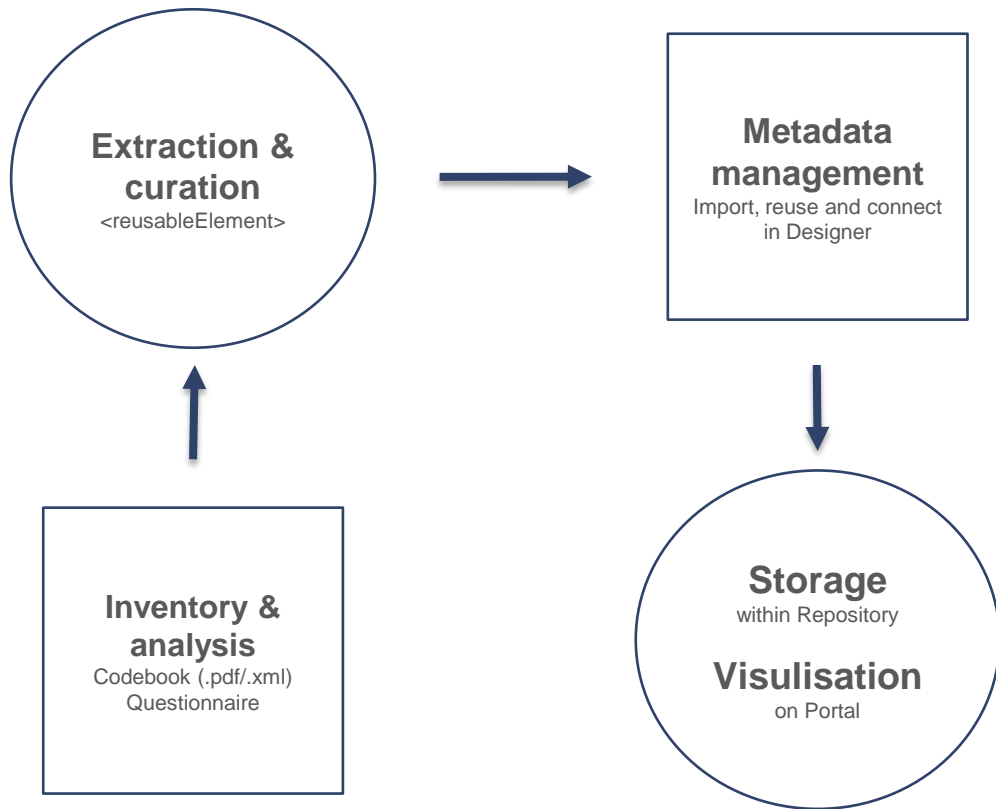
- Accessibility and search flexibility
- Semantic interoperability with CESSDA European Question Bank
- Documentation unicity

NEW VARIABLE DOCUMENTATION
shift from DDI-C to DDI-L

Experimentation phase



Workflow



Metadata inventory & curation



Identify and document reusable items, lineages and gaps across surveys

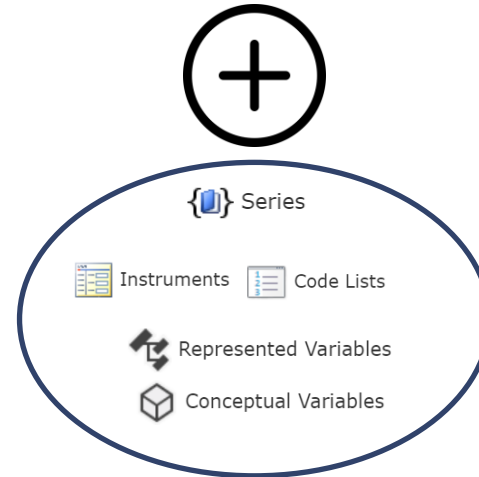
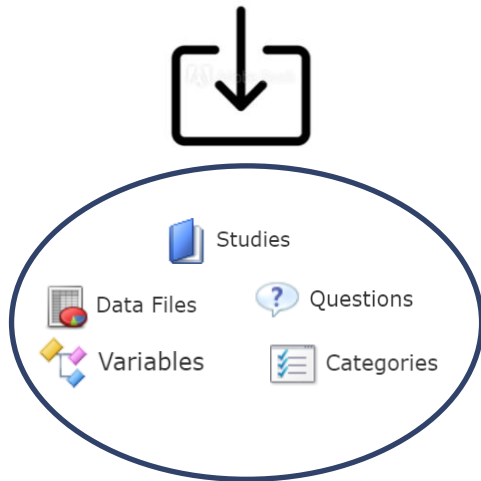


Extract unique and reusable items from DDI-C

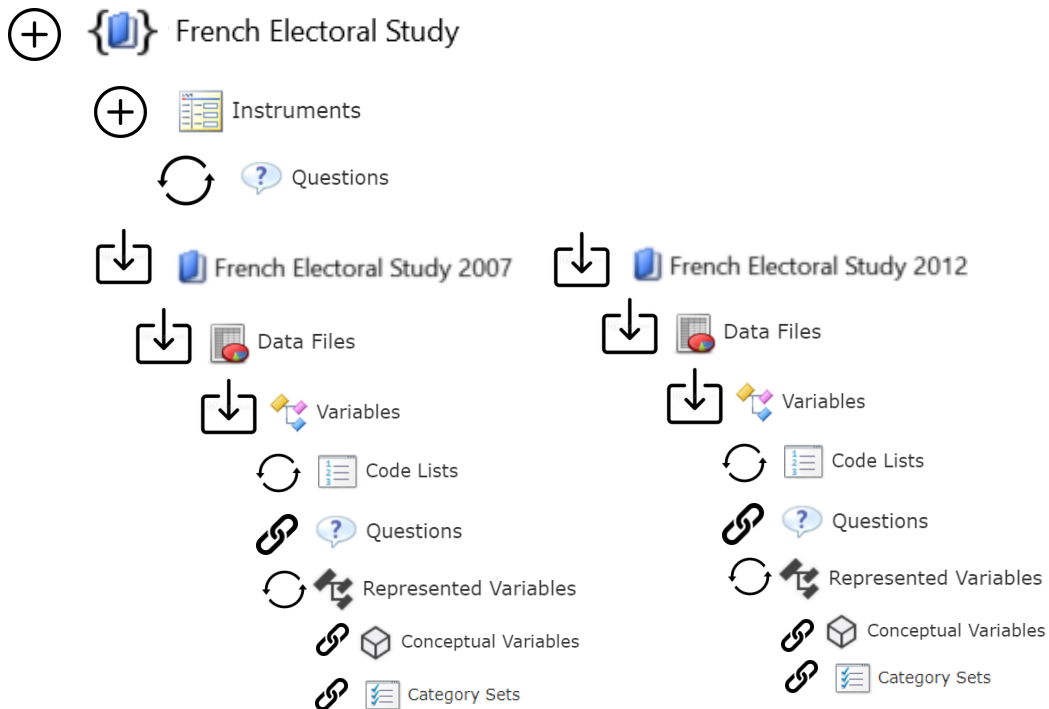
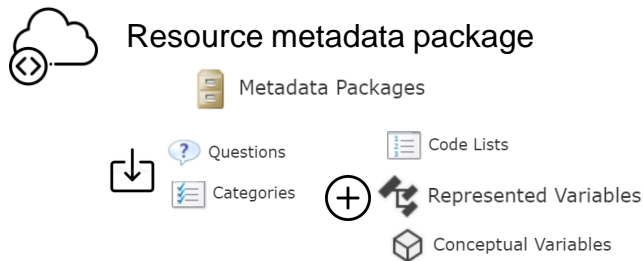
`<category>...</category>`

`<literalQuestion>... </literalQuestion>`

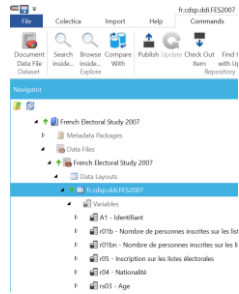
Import & create in Designer



Organise, reuse and connect in Designer



The metadata journey in the Colectica eco-system



Storage



```
<Fragment xmlns:r="ddi:reusable:3_2" xmlns="ddi:instance:3_2">  
<DataRelationship isUniversallyUnique="true" versionDate="2020-11-27T15:33:37.061722Z" xmlns="ddi:instance:3_2" version="59">  
<:URNurn:ddi:fr.cdsp:3a7a215d-2944-4a26-ac5c-a71a2cb55384/>  
<:Agencyfr.cdsp/>  
<:ID>3a7a215d-2944-4a26-ac5c-a71a2cb55384/</:ID>  
<:Version>59/</:Version>  
<DataRelationshipName>  
<:String xml:lang="fr-FR">fr.cdsp.ddi.FES2007/</:String>  
</DataRelationshipName>  
<LogicalRecord isUniversallyUnique="true">  
<:URNurn:ddi:fr.cdsp:ld9f4320-25a1-4215-84f0-68c00382033c:59/>  
<:Agency>fr.cdsp/</:Agency>  
</LogicalRecord>  
</DataRelationship>  
</Fragment>
```



Visualization



Publish metadata on Portal

Value	Label	Frequency	% of valid	% of all
1	Male	839	43.9%	43.9%
2	Female	1,149	56.1%	56.1%

Representation

Type: Categorical
Selection type: Selection
Codes: 1 Male, 2 Female
Role: input
Aggregation method: Unaggregated
Temporal: None
Geographic: None
Represented variable: Sexe
Missing values: Unlabeled Recodable Missing Value

METADATA VISUALISATION ***ON COLECTICA PORTAL***

Variable description and representation

The screenshot shows the CDSP interface with a dark blue header containing navigation links: CDSP, Search, Explore, Basket, and Help. A sidebar on the left lists 'Appears Within' categories: Series, French Electoral Study, Studies, French Electoral Study 2007, and Data Files. The main content area is titled 'Sexe' and shows the variable's metadata and two summary tables.

Variable Description

Name: r10
Label: Sexe
Dataset: fr.cdsp.ddi.FES2007

Value	Label	Frequency	% of valid	% of all
1	Male	831	41.55%	41.55%
2	Female	1,169	58.45%	58.45%

Valid	Invalid	Min	Max
2000	0	1	2

Representation

Type: Code List
Selection Style: SelectOne
Codes: sexe

- 1 Male
- 2 Female

Role: Input
Aggregation Method: Unspecified
Temporal: False
Geographic: False
Represented Variable: Sexe
Missing Values: Unnamed Reusable Missing Value

Source Questions and Instrument

The image displays a screenshot of the CDSP (Canadian Data System) interface, showing a detailed view of source questions and instrument details for R10.

Top Panel (CDSP): Includes navigation options: CDSP, Search, Explore, Basket (0), and Help.

Left Panel (Appears Within): A tree view showing the hierarchy of data sources:

- Series
 - French Electoral Study
- Studies
 - French Electoral Study 2007
- Data Files
 - French Electoral Study 2007

Source Questions

R10

Questionnaire CDSP Search Explore Basket (0)

R04

Appears Within

- Series
 - French Electoral Study
- Studies
 - French Electoral Study 2007
- Data Collections
 - French Electoral Study 2007

RS03

11 questions a

Lineage

Concorda

Statistics

Start

RS DEBUT

R05

Est-ce que vous êtes inscrit(e) sur les listes electorales pour pouvoir voter ?
UNE SEULE REPONSE POSSIBLE

R05==2 or 3 or 4

R04

Quelle est votre nationalité ?

SI R04==2

R10

Etes-vous...

RS03

Pouvez-vous me préciser votre âge ?
ENQUETEUR : REPONSE NUMERIQUE ENTRE 18 ET 99

QUESTIONNAIRE

ELECTION PRESIDENTIELLE

RS DE FIN

End

Concordance across surveys

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

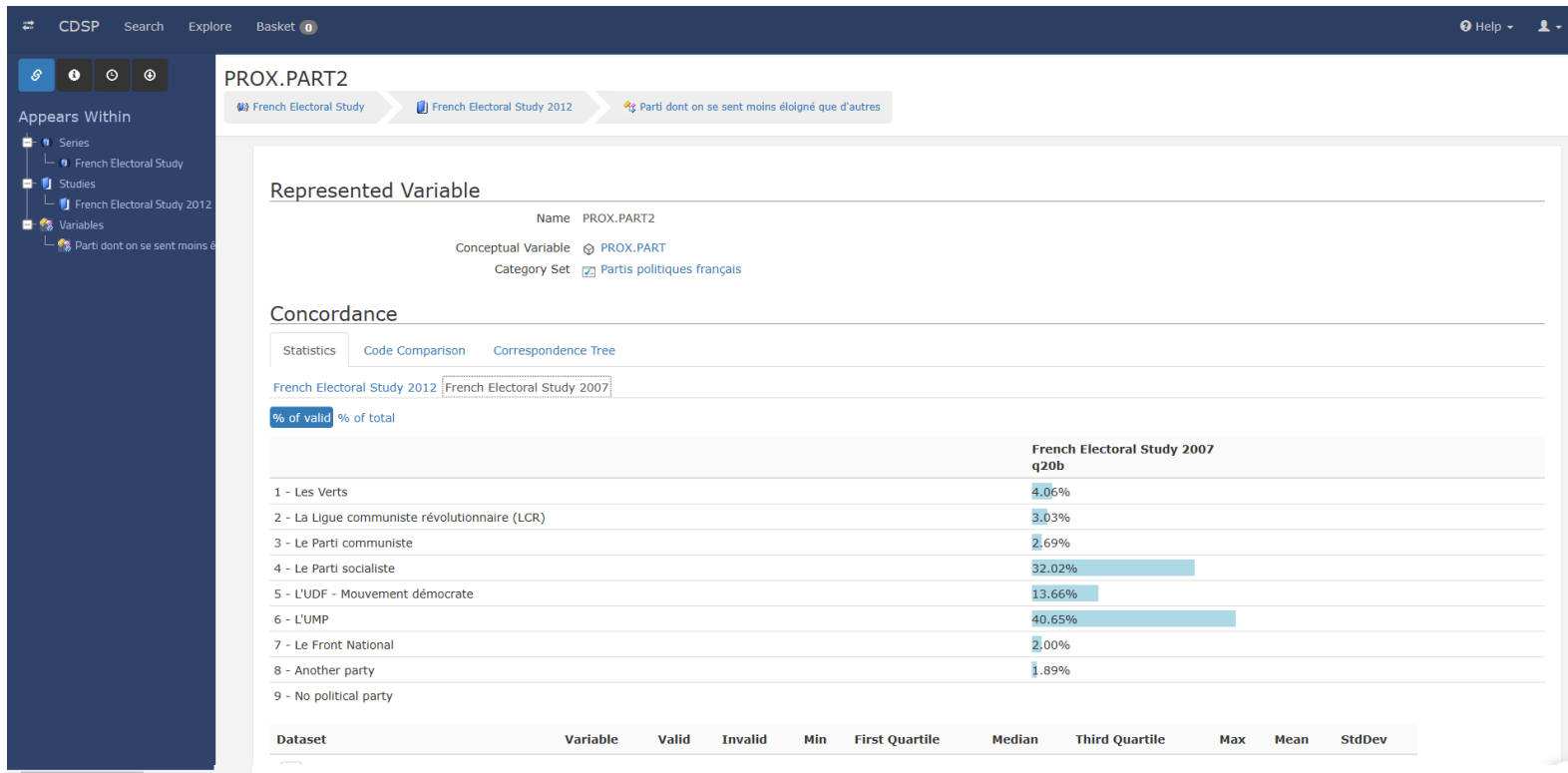
French Electoral Study 2012 - French Electoral Study 2007

% 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		

Concordance across surveys

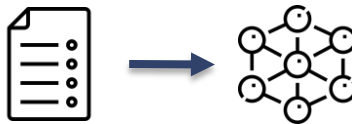




FEEDBACK & BENEFITS

Feedback: from DDI-Codebook to Lifecycle

A different documentation logic



Metadata schema correspondence in Designer



Disharmonize codebook documentation

Harmonisation: gaps in code lists

CDSP Search Explore Basket 0 Help

Geographic False
Represented Variable Sexe

Concordance

Statistics Code Comparison Correspondence Tree

BPF2007-R4 - BPF2007-R3

% of valid % of total

	BPF2007-R4 SEXE	French Electoral Study 2017 S2	BPF2007-R2 SEXE	BPF2007-R1 SEXE	French Electoral Study 2007 r10	French Electoral Study 2012 h1	BPF2007-R3 SEXE
1 - Male	47.87%		47.90%	47.86%	41.55%	44.89%	47.88%
2 - Female	52.13%		52.10%	52.14%	58.45%	55.11%	52.12%

Dataset	Variable	Valid	Invalid	Min	First Quartile	Median	Third Quartile	Max	Mean	StdDev
<input type="checkbox"/> BPF2007-R4	SEXE	5239	0	1				2		
<input type="checkbox"/> French Electoral Study 2017	S2									
<input type="checkbox"/> BPF2007-R2	SEXE	5647	0	1				2		
<input type="checkbox"/> BPF2007-R1	SEXE	5650	0	1				2		
<input type="checkbox"/> French Electoral Study 2007	r10	2000	0	1				2		
<input type="checkbox"/> French Electoral Study 2012	h1	2014	0	1				2		
<input type="checkbox"/> BPF2007-R3	SEXE	5240	0	1				2		

Appears Within

- Series
 - French Electoral Study
- Studies
 - French Electoral Study 2017
- Data Files
 - French Electoral Study 2017

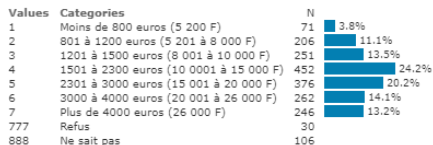
Harmonisation: semantic divergences

Dataset: French Electoral Study 2007

Variable r20: Revenus mensuels de foyer

LITERAL QUESTION

Si vous additionnez toutes les sources de revenus (salaires, allocations, autres), quels sont les revenus mensuels de votre foyer ?



SUMMARY STATISTICS

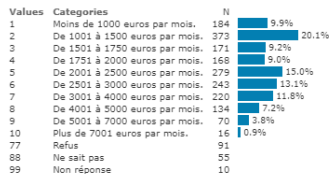
Valid cases 1864
Missing cases 136
Minimum 1,0
Maximum 7,0
This variable is numeric

Dataset: French Electoral Study 2012

Variable p3: Revenu net mensuel du foyer

LITERAL QUESTION

Si vous additionnez toutes les sources de revenus de votre foyer, quelle lettre correspond le mieux au revenu net mensuel de votre foyer ? Pour répondre, veuillez m'indiquer la lettre qui correspond à votre réponse ?



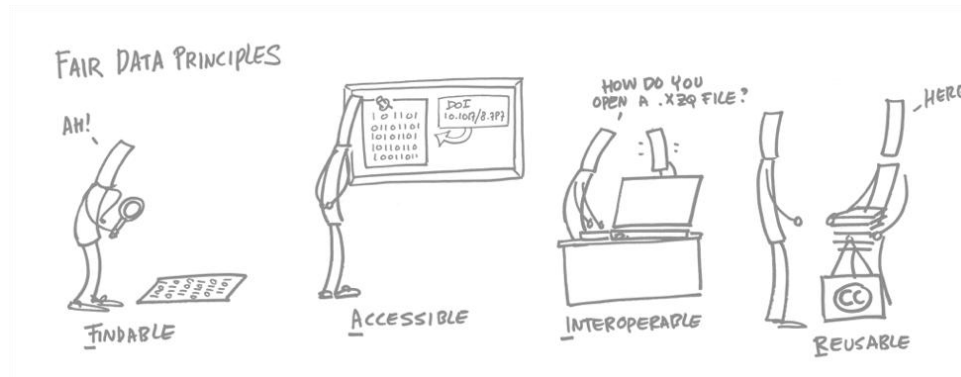
SUMMARY STATISTICS

Valid cases 1858
Missing cases 156
This variable is numeric

INTERVIEWER INSTRUCTIONS

TENDRE LISTE REPONSE ET LAISSER LE REPONDANT TOURNER LUI-MEME LA PREMIERE PAGE POUR ACCEDER A LA LISTE REPONSE DE FAÇON A CE QUE VOUS NE PUISSIEZ PAS VOIR LES ITEMS CORRESPONDANT AUX LETTRES. CODER LA SEULE REPONSE POSSIBLE

Benefits



- **FINDABLE** with one space of storage within repository
- **ACCESSIBLE** data and metadata via portal interface
- **INTEROPERABLE** xml ddi3.2 documentation compliant with the European Question Bank
- **REUSABLE** items for new surveys documentation

What's next?

- Improve skills on Colectica with trainings
- Publish surveys on public portal
- Harvest metadata by the CESSDA European Question Bank
- Automate part of the process
- Implement new surveys
- Report and share the migration strategy




Do you have any questions?

Thank you!

Let's keep in touch

 lucie.marie2@sciencespo.fr

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

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