



HAL
open science

DDI, Dataverse and Colectica: our data management combo

Alina Danciu, Geneviève Michaud

► **To cite this version:**

Alina Danciu, Geneviève Michaud. DDI, Dataverse and Colectica: our data management combo. 12th Annual European DDI User Conference (EDDI20), Nov 2020, En ligne, Paris, France. 10.5281/zenodo.4299083 . hal-03956404

HAL Id: hal-03956404

<https://sciencespo.hal.science/hal-03956404v1>

Submitted on 25 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.

DDI, Dataverse and Colectica: our data management combo

Alina DANCIU, Geneviève MICHAUD
Center for Socio-Political Data

EDDI 2020, 12th Annual European DDI Users Conference
Virtual event
December, 1, 2020

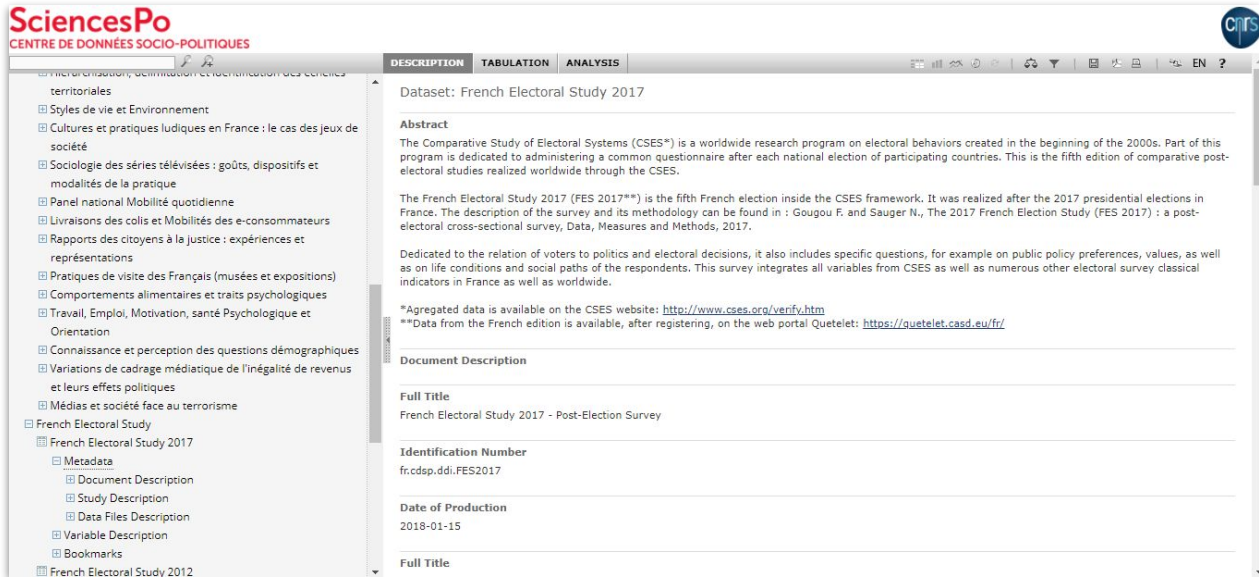
A quick look on...

**Going from Nesstar to Dataverse
(study-level metadata)**

Migrating DDI 1.2 to DDI 2.5

Using Colectica for variable management

Previous metadata & data management platforms



The screenshot displays the SciencesPo Centre de Données Socio-Politiques website. The left sidebar contains a navigation menu with categories such as 'Méthodologie, détermination et détermination des centres territoriales', 'Styles de vie et Environnement', 'Cultures et pratiques ludiques en France : le cas des jeux de société', 'Sociologie des séries télévisées : goûts, dispositifs et modalités de la pratique', 'Panel national Mobilité quotidienne', 'Livraisons des colis et Mobilités des e-consommateurs', 'Rapports des citoyens à la justice : expériences et représentations', 'Pratiques de visite des Français (musées et expositions)', 'Comportements alimentaires et traits psychologiques', 'Travail, Emploi, Motivation, santé Psychologique et Orientation', 'Connaissance et perception des questions démographiques', 'Variations de cadrage médiatique de l'inégalité de revenus et leurs effets politiques', 'Médias et société face au terrorisme', 'French Electoral Study', 'French Electoral Study 2017', 'Metadata', 'Document Description', 'Study Description', 'Data Files Description', 'Variable Description', 'Bookmarks', and 'French Electoral Study 2012'. The main content area is titled 'Dataset: French Electoral Study 2017' and includes an 'Abstract' section with the following text: 'The Comparative Study of Electoral Systems (CSES*) is a worldwide research program on electoral behaviors created in the beginning of the 2000s. Part of this program is dedicated to administering a common questionnaire after each national election of participating countries. This is the fifth edition of comparative post-electoral studies realized worldwide through the CSES. The French Electoral Study 2017 (FES 2017**) is the fifth French election inside the CSES framework. It was realized after the 2017 presidential elections in France. The description of the survey and its methodology can be found in : Gougou F. and Sauger N., The 2017 French Election Study (FES 2017) : a post-electoral cross-sectional survey, Data, Measures and Methods, 2017. Dedicated to the relation of voters to politics and electoral decisions, it also includes specific questions, for example on public policy preferences, values, as well as on life conditions and social paths of the respondents. This survey integrates all variables from CSES as well as numerous other electoral survey classical indicators in France as well as worldwide. *Agregated data is available on the CSES website: <http://www.cses.org/venfy.htm> **Data from the French edition is available, after registering, on the web portal Quetelet: <https://quetelet.csd.eu/fr/>

Below the abstract, there is a 'Document Description' section with the following details:

- Full Title:** French Electoral Study 2017 - Post-Election Survey
- Identification Number:** fr.cdsp.ddi.FES2017
- Date of Production:** 2018-01-15
- Full Title:** (This field is currently empty)

The reason why



DDI xml files migration

1st step - a R script to transform Nesstar DDI 1.2 files to DDI 2.5 files

The R script is available on CDSP's [Github](#).

2nd step - a Python custom program to:

- Import DDI 2.5 files into Dataverse (study level). During this step, a PID (DOI) was given to each study.
- The corresponding DOI was then inserted back in DDI 2.5 file.
- The DDI 2.5 file is uploaded into Dataverse.

Using the CMM controlled vocabularies

Dataverse property	Controlled vocabulary
Keywords	ELSST thesaurus
Kind_of_Data	DDI Alliance Controlled Vocabulary for General Data Format
Time_Method	DDI Alliance Controlled Vocabulary for Time Method
Analysis_Unit	DDI Alliance Controlled Vocabulary for Analysis Unit
Mode-Collection	DDI Alliance Controlled Vocabulary for Mode Of Collection
Sampling-Procedure	DDI Alliance Controlled Vocabulary for Sampling Procedure
Topic_Classification	CESSDA Topic Classification
Type_Instrument	DDI Alliance Controlled Vocabulary for Type of Instrument

1 tailor-made VC for institutions(using [ROR](#) and [ISNI](#))

Benefits of using Dataverse

Dataverse is a good solution for networked institutions:

- Single Sign On using institutional credentials
- Can harvest and be harvested
- General and Dataverse-level branding capabilities

Dataverse supports discoverability and persistence:

- Powerful search engine and filters
- Good Google referencing
- DOI

Pending issues with Dataverse

DDI import and export definitively are a hard route to follow.

Adding a language creates additional difficulties.

(see : Customizing a DDI compliant repository: a detailed course with Dataverse, session Managing DDI)



[Icon source](#)

We are getting there

SciencesPo

Ajouter des données ▾

Rechercher ▾

Guide d'utilisation

Support

Français ▾

S'inscrire

Se connecter

Classification des sujets ⓘ

SocietyAndCulture.SocialBehaviourAndAttitudes

Langue ⓘ

French

Producteur ⓘ

Sciences Po, Centre de données socio-politiques (CDSP), CNRS (Sciences Po, CNRS) (CDSP)

Endroit de production ⓘ

Sciences Po 27 rue Saint-Guillaume 75337 Paris cedex 07

Collaborateur ⓘ

Belton-Chevallier, Leslie, Institut français des sciences et technologies des transports, de l'aménagement et des réseaux, Laboratoire Dynamiques économiques et sociales des transports (DEST)
Blanquart, Corinne, Institut français des sciences et technologies des transports, de l'aménagement et des réseaux, Laboratoire Systèmes productifs, logistique, organisation des transports, et travail (SPLOTT)
Marei, Nora, Institut français des sciences et technologies des transports, de l'aménagement et des réseaux, Laboratoire Systèmes productifs, logistique, organisation des transports, et travail (SPLOTT)
Équipe Elipss, Sciences Po, Centre de données socio-politiques (CDSP), CNRS

Renseignements sur la subvention ⓘ

ANR-10-EQPX-19-01

Diffuseur ⓘ

Sciences Po, Centre de données socio-politiques (CDSP), CNRS (CDSP)

Date de disponibilité ⓘ

2020-02

Période couverte ⓘ

Start: 2016 ; End: 2016

Date de la collecte ⓘ

Start: 2016-06-23 ; End: 2016-06-30

Type de données ⓘ

Numeric; Text

6. Integrating Colectica in our toolbox



- For **variable-level metadata**, we continue to use Nesstar Publisher and will test R scripts
- We have also been using **Colectica** for a year now, for **longitudinal studies**
- We use Colectica Designer, Colectica Portal and Colectica Repository
- The pending question : “single” studies

Photo [source](#)

7. What's next

Improving data.sciencespo.fr

Testing **Dataverse** variable visualisation tools

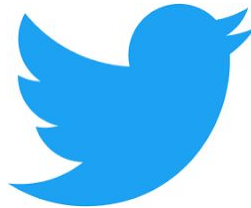
Setting up **new data and metadata management and publication workflows**



Get in touch!

alina.danciu@sciencespo.fr

genevieve.michaud@sciencespo.fr



@alina_danciu_

@GenevievMichaud

@CDSP_ScPo

Thank you

DOI: [10.5281/zenodo.4299083](https://doi.org/10.5281/zenodo.4299083)

