



DATA
INFRASTRUCTURE

*Développer
la culture des données*



*Documenting is not enough : an example of
collaborative work in a French Research
Infrastructure*

Ada Chmilevschi PUDL

Elodie Pétorin ADISP/CNRS

License : [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

What's ProGeDo ?

- ✓ Large Research Infrastructure- government strategy. As a central player in national data policy in the human and social sciences
- ✓ The development of facilities for archiving, documenting data and making it available, including a system for secured remote access to confidential data (CASD box)
- ✓ The participation to the production of large multidisciplinary studies of national interest and selected by the European infrastructure roadmap (ESFRI)
- ✓ The animation of a national dynamics centered around the field of competence of the infrastructure.

At international level

✓ Supports the French parts of European and international surveys



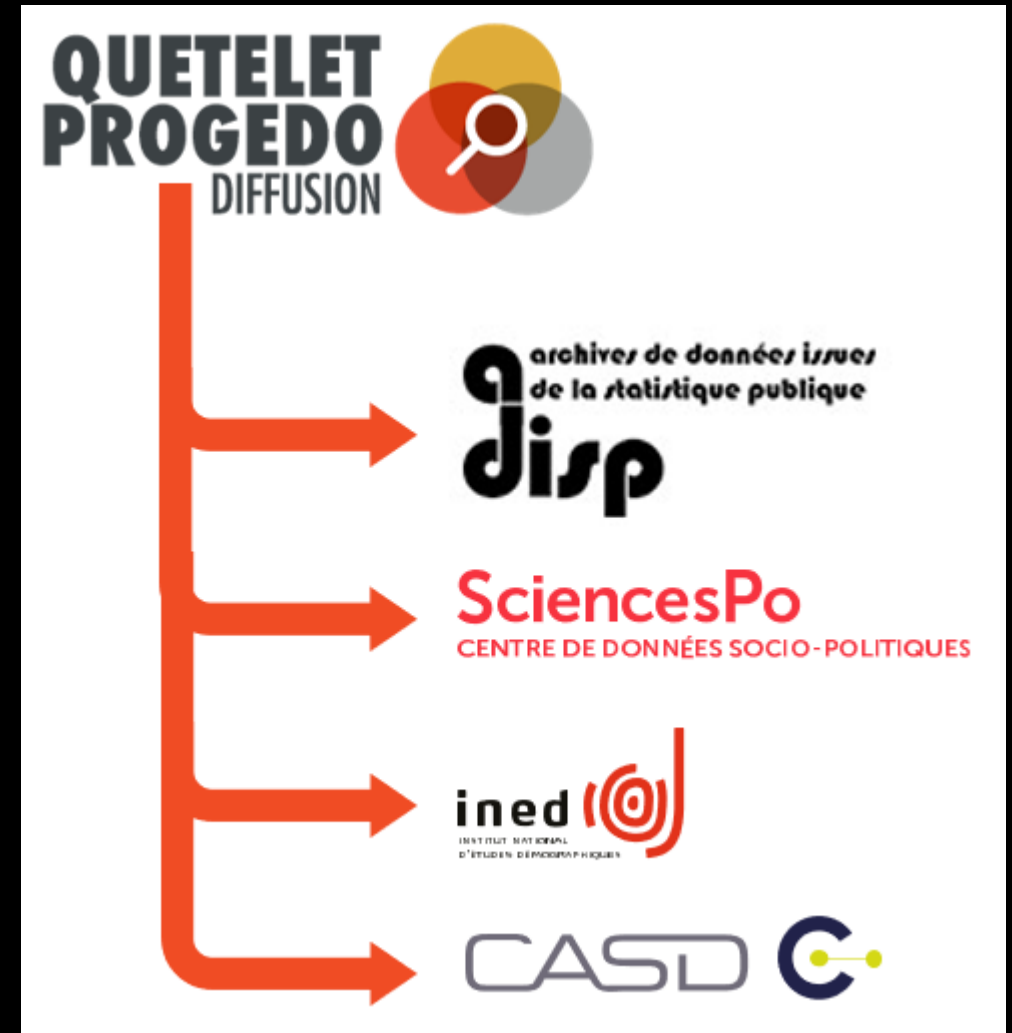
✓ Gives access to international database



✓ Member of CESSDA and, backed by CNRS, the French representative of the consortium

At national level:

- ✓ Organizes the dissemination of data through INED, CDSP and ADISP (data provider center)
- ✓ Manages the network of University Data Platforms



National Archive of Data from Official Statistics (ADISP)

- ✓ A service of the Large Research Infrastructure ProGeDo
- ✓ Collect and archive data
- ✓ Document data from Official Statistics (census data, localized data, data representative of results at a national level) by the use of professional standards for documentation and datasets (DDI standard)
- ✓ Disseminate data to the scientific community
- ✓ Assist and advise users

Documentation processes

- ✓ Around 1400 references
 - 80 to 100 new reference per year
 - 30 to 50 updates
- ✓ DDI 1.2 using Nesstar publisher



The screenshot shows the ADISP software interface. At the top, there are logos for 'disp', 'QUETELET PROGEDO', and 'DIFFUSION'. Below the logos, there are navigation tabs: 'ADISP', 'Commander des données', and 'Toutes les enquêtes'. The main area is divided into two panes. The left pane shows a list of data sources, including 'Base centrale Scolarité' for years 2004 to 2018, and a section for 'Métadonnées' (Metadata) with sub-items like 'Description des variables' and 'Table élèves'. The right pane shows a detailed view of a variable named 'NATIO_G: Nationalité' from the 'Jeu de données: Base centrale Scolarité - 2018'. It includes a table of values and categories with their respective counts and percentages, and a section for 'STATISTIQUES SOMMAIRES' (Summary Statistics).

Valeurs	Catégories	N	
1	Français	5566024	93,8%
2	Européen UE (sauf France)	104990	1,8%
3	Européen hors UE	28327	0,5%
4	Asie	57534	1,0%
5	Afrique	151144	2,5%
6	Amérique	24808	0,4%
7	Océanie	500	0,0%
9	Non défini, y compris apatride	285	0,0%

STATISTIQUES SOMMAIRES
Cas valides: 5933612
Cas manquants: 0
Cette variable est alphanumérique

Documentation processes on two levels

Descriptive instructions

General description of data :

- ✓ abstract
- ✓ producer
- ✓ universe, geography
- ✓ collection date and mode
- ✓ link to documents

CHAMP	EXEMPLES
Résumé de l'enquête	L'EMD de Nantes porte sur 154 communes ...
Thème(s)	Conditions de vie
Mots-clefs	Transport, véhicule, déplacement, stationnement, trajet, géolocalisation
Univers	Ménages résidant dans une commune du périmètre de l'EMD. Individus de 5 ans et plus. Déplacements effectués la veille de l'interview
Dates de collecte	Début : 2014-09-29 / Fin : 2015-03-15
Couverture géographique	154 communes (autres exemples : France métropolitaine, DOM, ...)
Unité géographique	Zonage en 123 secteurs (autres exemples : commune, département, région...)
Unité d'analyse	Ménages, individus, déplacements

Xml file

Web page creation

Detailed description of data :

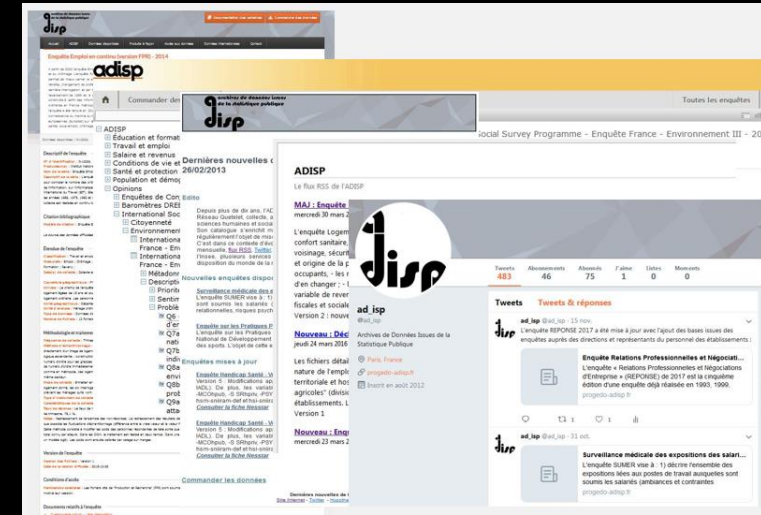
- ✓ question and investigators instructions
- ✓ label of variables and modalities

The screenshot shows a software interface for documenting variables. It features a table of variables with columns for Number, Name, and Label. The 'Documentation' tab is active, showing a tree view of documentation elements like 'Question', 'Texte de la question', 'Univers', and 'Instructions pour l'enquête'. The 'Texte de la question' field contains the text 'Possédez-vous le permis de conduire V1 (tourisme B) ?'. The 'Univers' field contains 'Les personnes âgées de 16 ans et plus (P4 > 15)'. On the right, there is a 'Category Hierarchy' tree and a 'Variable information' section with fields for 'Level Name', 'GeoMap LRI', 'Data Type' (set to 'Fixed String'), and 'Measure' (set to 'Nominal').

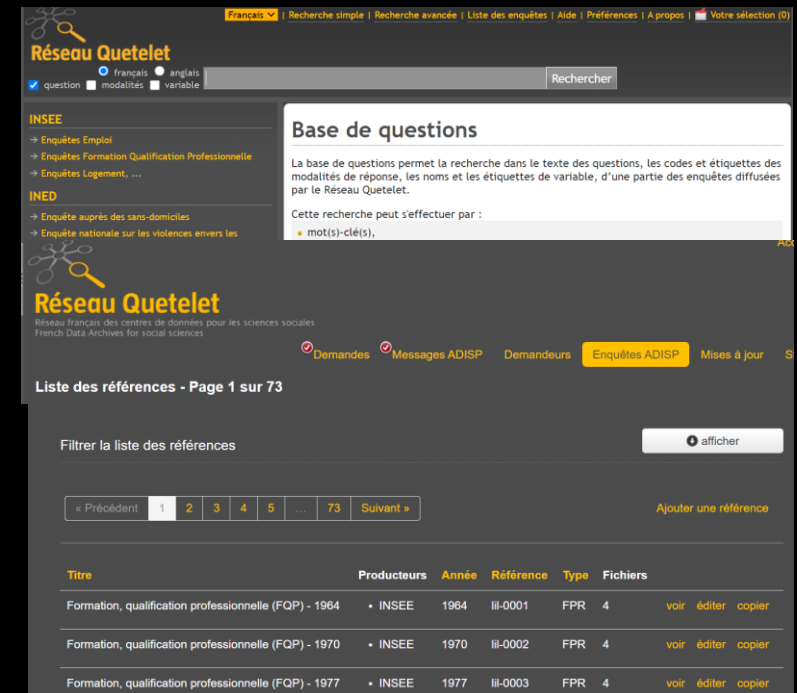
Publication and dissemination

Nesstar publication allows :

- ✓ Automatic feed of Adisp website and Adisp Nesstar platform
- ✓ Harvest in the social science browser Isidore

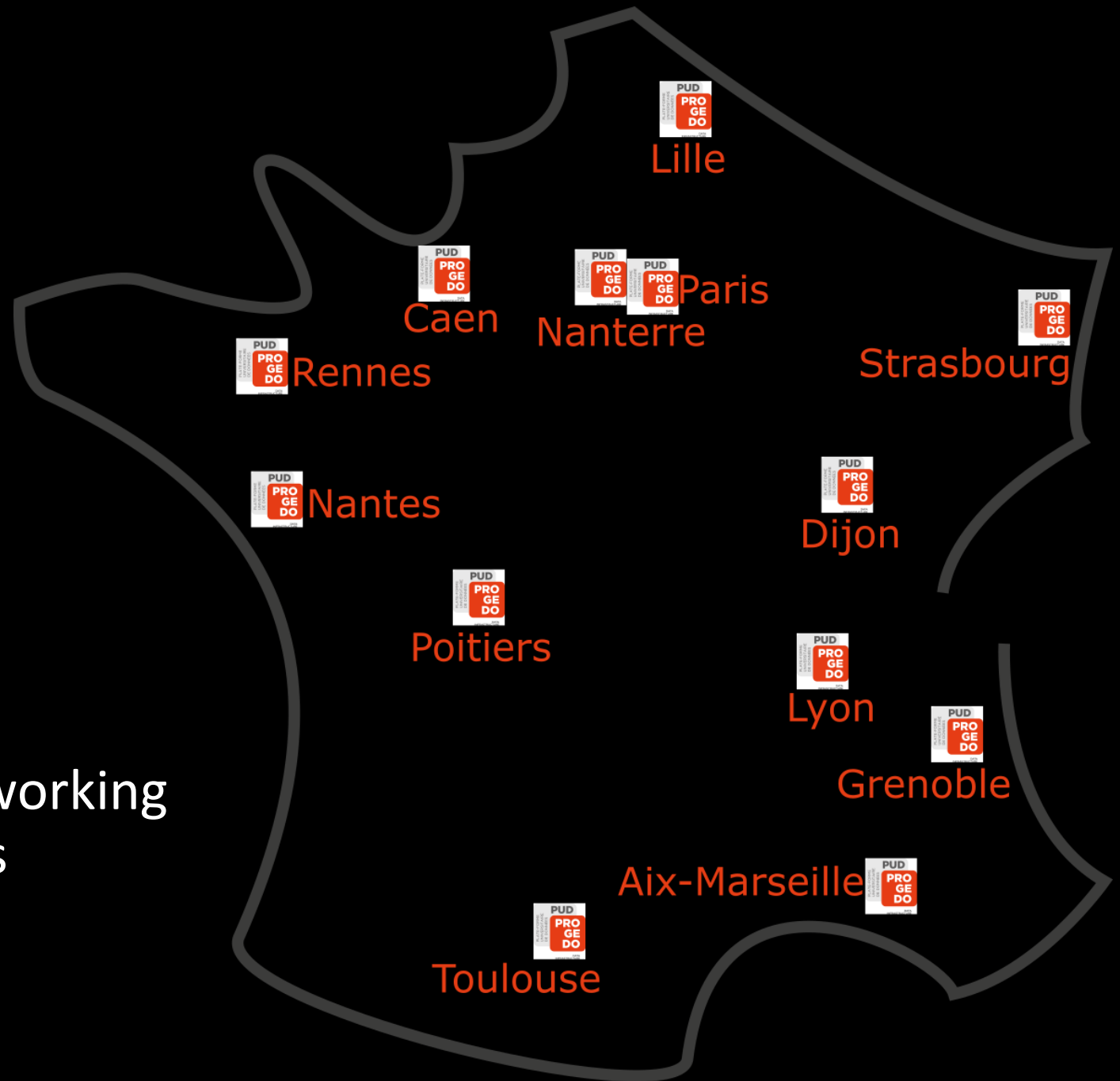


- ✓ Manually feeding the Question Bank (BdQ), Twitter account, Quetelet-ProGeDo diffusion command application using the xml notice



University Data Platforms (UDP)

Developing a data culture by working
closely with all research actors



University Data Platforms (UDP)

Working environment:

- ✓ Supported by public teaching and research institutions
 - ✓ Housed by Maisons des Sciences de l'Homme (national network of regional Institutes of human sciences)
 - ✓ Scientific coordination by the LRI ProGeDo
- ✓ **Service provider to :**
- ✓ Users involved in research (researchers, professors, students, research engineers)
 - ✓ Working with quantitative data in humanities and social sciences

ADISP and UDPs in the data cycle



ADISP and UDPs in the data life cycle



University Data Platforms

- ✓ Inform users on the most recent regulations and practices regarding data production :GDPR, DMP, FAIR principles
- ✓ Assist users in creating their DMP
- ✓ Direct them to the data protection officer and other necessary correspondents



- ✓ Data that reaches **ADISP** once produced will be more easily documented
- ✓ **ADISP** can assist users in finding the right data to use in their project

ADISP and UDPs in the data life cycle



University Data Platforms

- ✓ Organise workshops on methods
- ✓ At times advise individual users on which methods to use in their research
- ✓ Assist students and researchers in using the NESSTAR platform, through collective workshops or individually

ADISP

- ✓ Teaches UDP Engineers how to use Nesstar platform

ADISP and UDPs in the data life cycle



ADISP

- ✓ Receives the data and the documentation, verifies their consistency and uses them to create the metadata
- ✓ Data and metadata are disseminated using the Quetelet-Progedo-Diffusion application

ADISP

- ✓ Manages access to the data using this web portal
- ✓ Supplies access to the metadata, variables and primary statistics via a NESSTAR webview server

UDPs inform users on where and how they can access the data they are looking for

ADISP and UDPs in the data life cycle



ADISP

- ✓ Uses DDI standard for archiving and preserving the data and metadata via NESSTAR

UDPs

- ✓ Inform users on the latest requirements in data preservation and access

ADISP and UDPs in the data life cycle



ADISP

- ✓ The DDI standard that ADISP uses allows for multiple reuse of stored data over time

All the services supplied for the 'analyzing' stage by both parties are also necessary for re-using the data.

To conclude

A virtuous circle of mutual support

- ✓ ADISP documents and disseminates the data (while respecting FAIR principles)
- ✓ University data platforms (UDPs) inform users on the new regulations and principles of data production, which helps the latter create reusable data.

In every stage of the process, university data platforms and ADISP meet and communicate regularly as we are all part of PROGEDO

Thank you for your attention

Ada Chmilevschi (PUDL) pudl@meshs.fr

Elodie Pétorin (ADISP) elodie.petorin@cns.fr

