



DATA
INFRASTRUCTURE

*Développer
la culture des données*

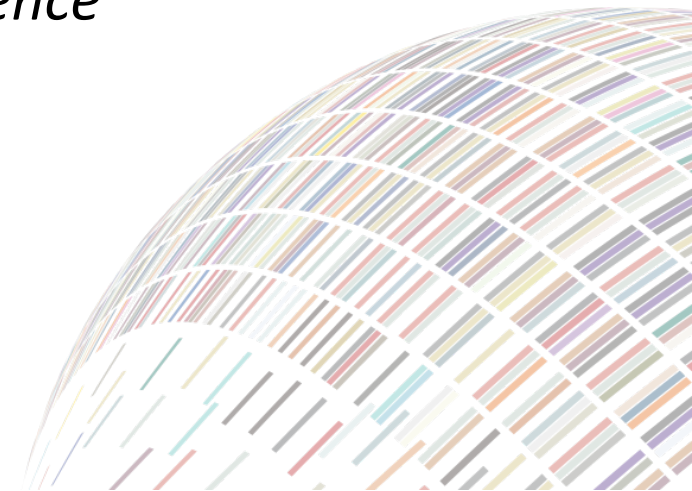


Providing data at the National Archive of Data from Official Statistics (ADISP)

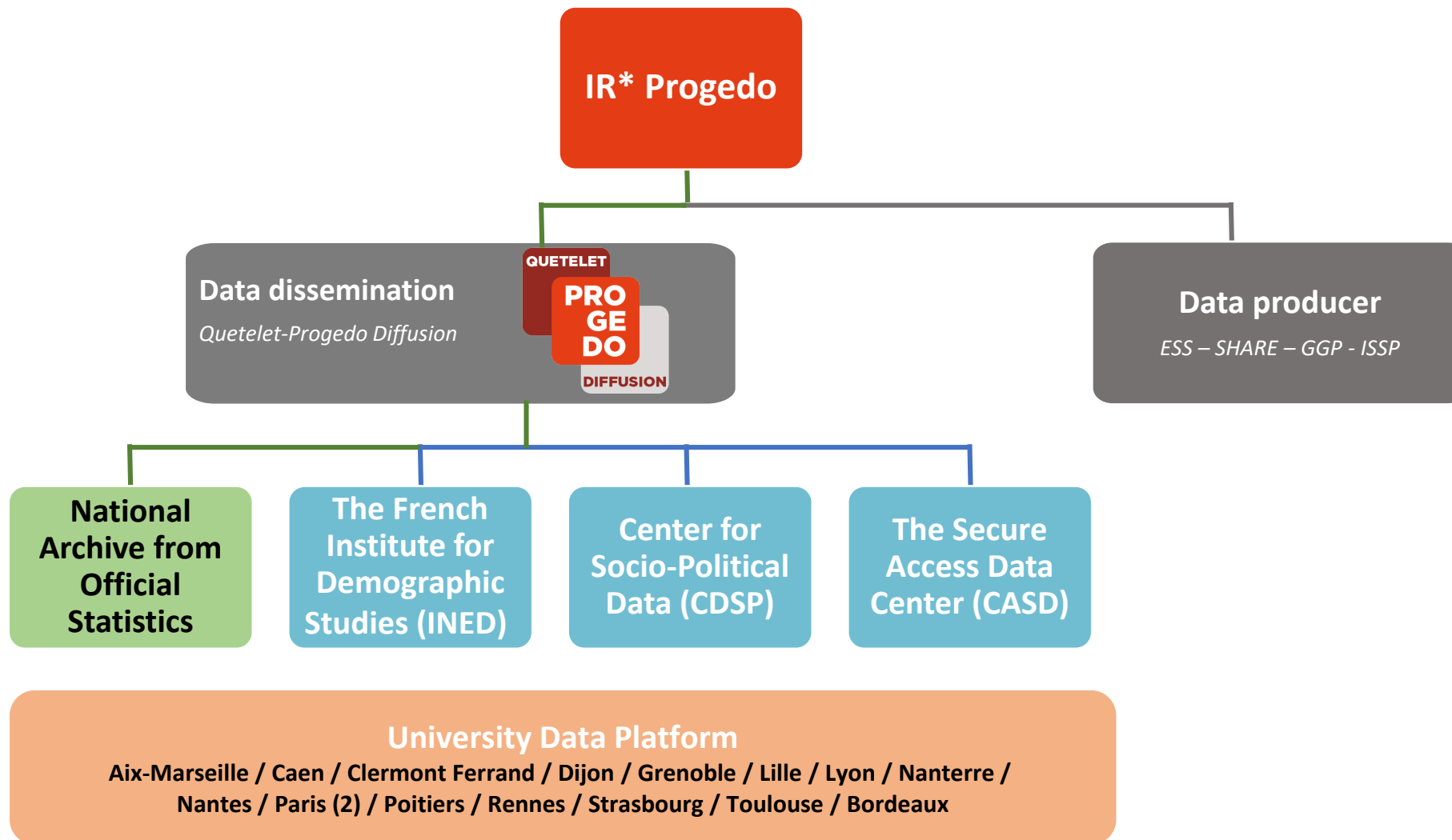
EDDI21 – 13th Annual European DDI User Conference

30 Novembre 2021, Paris

Elodie Pétorin, ADISP-Progedo



ADISP in Progedo universe



Adisp goal and missions

Goal : To develop the use of data in social sciences

4 daily missions :

1- **Collect and preserve data**

2- **Document data in DDI standard**

3- **Disseminate the data to the scientific community**

4- **Assist / advise users**

Access to the data free of charge

Researchers, Engineer, PhD
student, students

University and Institutional
organisation

The producers partners at ADISP



DEPS

Ministère de la Culture

DSED

interieur.gouv.fr
MINISTÈRE DE L'INTÉRIEUR



MEOS

MINISTÈRE
DES SPORTS

DEPP

MINISTÈRE
DE L'ÉDUCATION
NATIONALE



SIES

Ministère de
l'Enseignement Supérieur

SOes

Ministère de la Transition
écologique et solidaire



CEPREMAP

CENTRE POUR LA RECHERCHE ÉCONOMIQUE ET SES APPLICATIONS



ADISP catalog

Almost 1 500 references (Around 200 sources)

Only quantitative data

✓ Study and/or administrative data

✓ from 19th century to 2021

✓ metropolis and/or overseas departement

✓ Individual or aggregated data

L'**ADISP** - Archives de Données Issues de la Statistique Publique - diffuse des enquêtes et bases de données produites par l'INSEE, plusieurs services statistiques ministériels et des institutions de recherche publiques. Les données diffusées sont accessibles gratuitement à l'ensemble de la communauté scientifique. L'Adisp apporte également des conseils sur la recherche et l'utilisation des données.

[En savoir plus sur l'Adisp](#)

Nouveautés du catalogue

CEREMA - Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement

✓ Enquêtes Ménages Déplacements (EMD) : [Carcassonne 2015](#), [Dunkerque 2015](#), [Nimes 2015](#)

INSEE

- ✓ Base permanente des équipements : [2020](#)
- ✓ État civil - Fichier détail Décès : [2020](#)
- ✓ Cadre de Vie et Sécurité - Mayotte : [2020](#)
- ✓ Enquête mensuelle de conjoncture auprès des ménages : [2020](#)
- ✓ Etat civil mensuel communal - Naissances : [2020](#)
- ✓ Fichier des prénoms : [Edition 2021](#)
- ✓ Recensement de la population 2018 : [fichier détail individus localisés à la région](#) , [fichier détail logements](#)

Institut national de la Jeunesse et de l'Éducation Populaire (INJEP) et Agence de Service Civique (ASC)

✓ Enquête auprès des bénéficiaires du service civique : [2019](#)

Institut de recherche et de documentation en économie de la santé (IRDES)

✓ Enquête sur la protection sociale complémentaire d'entreprise (PSCE) : [2017](#)

Maurice Garden

✓ Mariages parisiens - [1885](#)

Ministère de l'Enseignement supérieur (SIES)

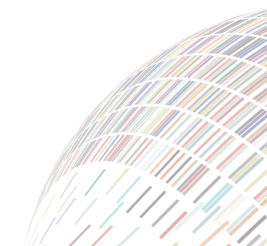
- ✓ Système d'information sur le suivi de l'étudiant - Universités : inscriptions : [2020-2021](#)
- ✓ Système d'information sur le suivi de l'étudiant - Universités : résultats : [2018-2019](#)

Ministère de l'Éducation (DEPP)

✓ Base centrale Scolarité : [2019](#), [2020](#)

What about the documentation ?

- ✓ New reference transmitted via secured transfer email (FileSender, etc..)
- ✓ New reference archived on our secured disk
- ✓ Data verification : missing data, inconsistencies, lack of documentation, etc...
- ✓ Documentation under Nesstar Publisher and production of an XML file



General structure of DDI (version 2.1)

1.0	Document description	<ul style="list-style-type: none">✓ 72 champs et sous-champs✓ Description de la phase (interne) de documentation
2.0	Study description	<ul style="list-style-type: none">✓ 101 champs et sous-champs✓ Éléments généraux sur l'enquête
3.0	File(s) description	<ul style="list-style-type: none">✓ 30 champs et sous-champs✓ Présentation générale du (des) fichier(s)
4.0	Variables description	<ul style="list-style-type: none">✓ 61 champs et sous-champs✓ Description de chaque variable du fichier
5.0	Other materials	<ul style="list-style-type: none">✓ 7 champs et sous-champs✓ Intégration des documents liés à l'enquête

270 field / subfield *precisely defined*

ADISP : we selected about **80 field**



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

Detailed description of data :

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

Xml file

Web page
creation

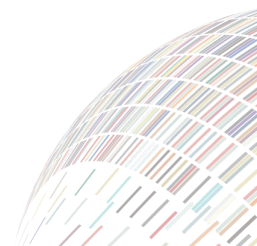
The screenshot displays a software interface for data documentation. It features a 'variables' table with columns for 'Number', 'Name', and 'Label'. The selected variable is 'v18' (Niveau d'études). To the right, there is a 'Category Hierarchy' tree showing levels like '1 - Oui', '2 - Non', and '3 - Conduite accompagnée et leçons de conduite'. Below the table, there are sections for 'Documentation' and 'Question', including fields for 'Texte de la question' and 'Univers'.

Number	Name	Label
v12	JOUR	Jour des déplacements
v13	P2	Sexe de la personne
v14	P3	Lien avec la personne de référence
v15	P4	Âge de la personne
v16	P5	Possession d'un téléphone portable
v17	P6	Possession d'une adresse de messagerie électronique
v18	P7	Possession du permis de conduire voiture
v19	P8	Niveau d'études
v20	P9	Proportionnalité de la personne

What DDI can do for you ?



Putting the pieces together, not an apple pie



The FAIR principles

F
indable



A
ccessible



I
nteroperable



R
eusable



Data should be easy to find

F
indable



- ✓ Rich metadata on every study or data set
- ✓ Every variable in the data set is included in the metadata
- ✓ Metadata are registered or indexed in a searchable resource (BdQ and ORDS).

RECHERCHER
DES DONNÉES

Pour rechercher des données au sein des catalogues de 3 des diffuseurs de Quetelet PROGEDO Diffusion (ADISP, CDSP et INED), deux moteurs de recherche sont mis à disposition :

ORDS

Outil de Recherche de Données en Sciences Sociales interroge les métadonnées des enquêtes sur les champs de la description générale (titre, référence, diffuseur, producteur, résumé, thèmes, etc.), à l'exclusion des questions, variables et modalités.

L'interrogation de la base documentaire se fait à l'aide de mots-clés en langage naturel et en français uniquement.

Faire une recherche

BASE DE QUESTIONS

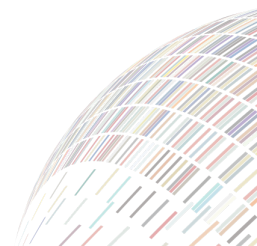
La Base de Questions permet d'explorer les données au niveau des questions, des modalités et des étiquettes de variable des données archivées.

Elle porte actuellement sur environ 40% du catalogue et correspond à plus de 300 000 questions ou variables.

La recherche peut s'effectuer par mot(s)-clé(s), requête avancée, producteur, série d'enquêtes et enquête.

Consultez le manuel d'utilisation de la base de questions

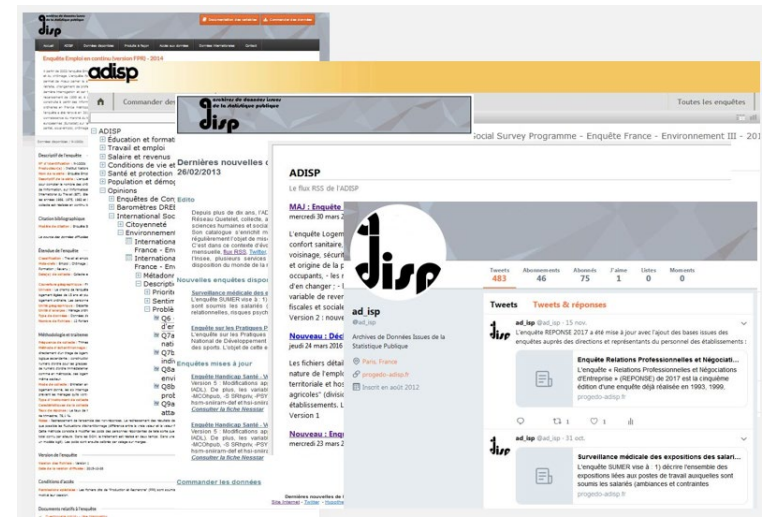
Faire une recherche



Publication and dissemination

Nesstar publication allows :

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



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

Recherche simple | Recherche avancée | Liste des enquêtes | Aide | Préférences | À propos | Votre sélection (0)

Réseau Quetelet

français | anglais | question | modalités | variable

Base de questions

La base de questions permet la recherche dans le texte des questions, les codes et étiquettes des modalités de réponse, les noms et les étiquettes de variable, d'une partie des enquêtes diffusées par le Réseau Quetelet.

Cette recherche peut s'effectuer par :

- mot(s)-clé(s),

Réseau Quetelet

Réseau français des centres de données pour les sciences sociales
French Data Archives for social sciences

Demandes | Messages ADISP | Demandeurs | **Enquêtes ADISP** | Mises à jour

Liste des références - Page 1 sur 73

Filtrer la liste des références [afficher]

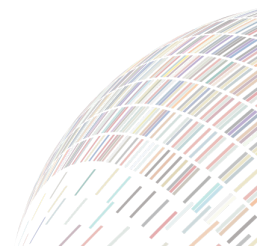
« Précédent | 1 | 2 | 3 | 4 | 5 | ... | 73 | Suivant » [Ajouter une référence]

Titre	Producteurs	Année	Référence	Type	Fichiers
Formation, qualification professionnelle (FQP) - 1964	• INSEE	1964	III-0001	FPR	4 voir éditer copier
Formation, qualification professionnelle (FQP) - 1970	• INSEE	1970	III-0002	FPR	4 voir éditer copier
Formation, qualification professionnelle (FQP) - 1977	• INSEE	1977	III-0003	FPR	4 voir éditer copier

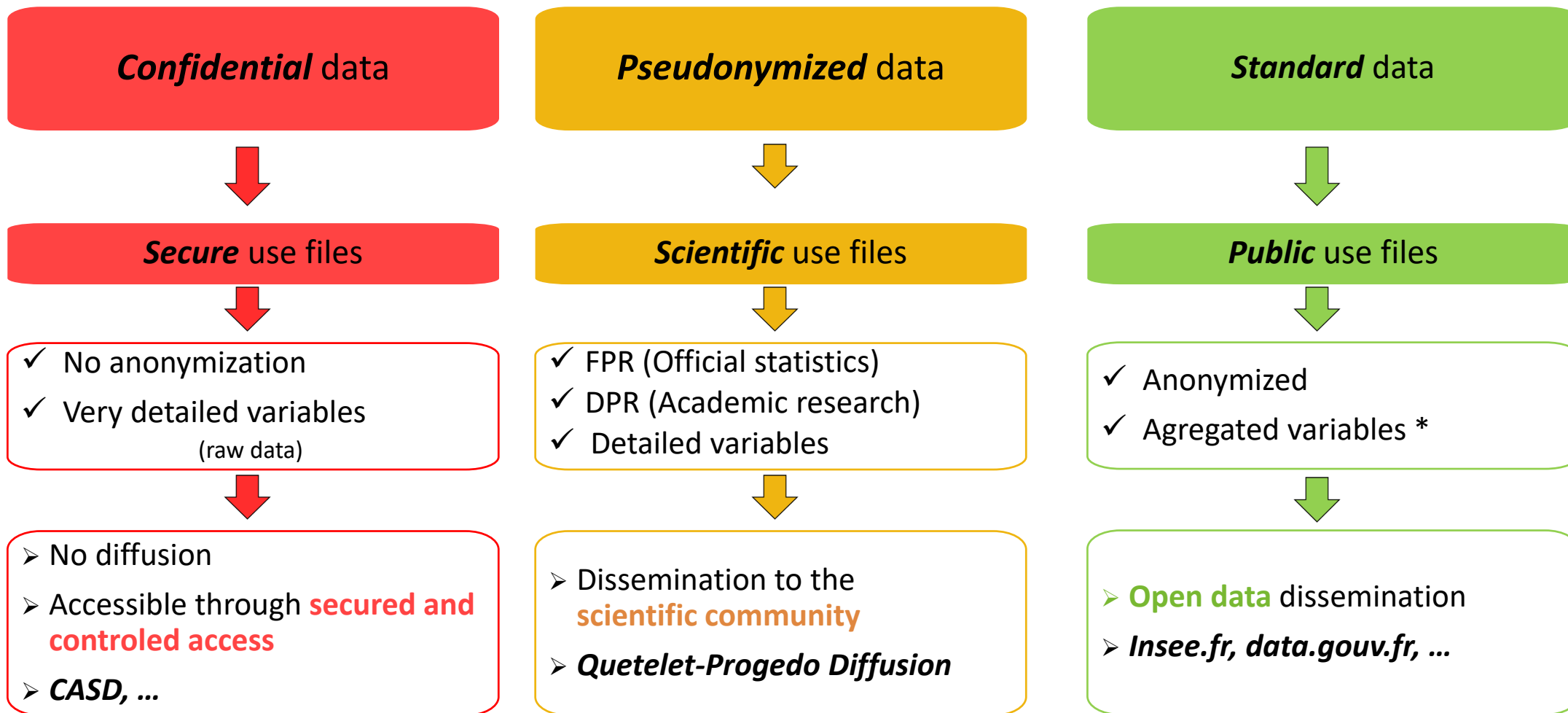
How can we provide access ?



- ✓ The protocol is open, free, and universally implementable
- ✓ Metadata are accessible, even when the data are no longer available
- ✓ Different level of access
- ✓ Work in progress : more user friendly portal



More detail on the file types and access



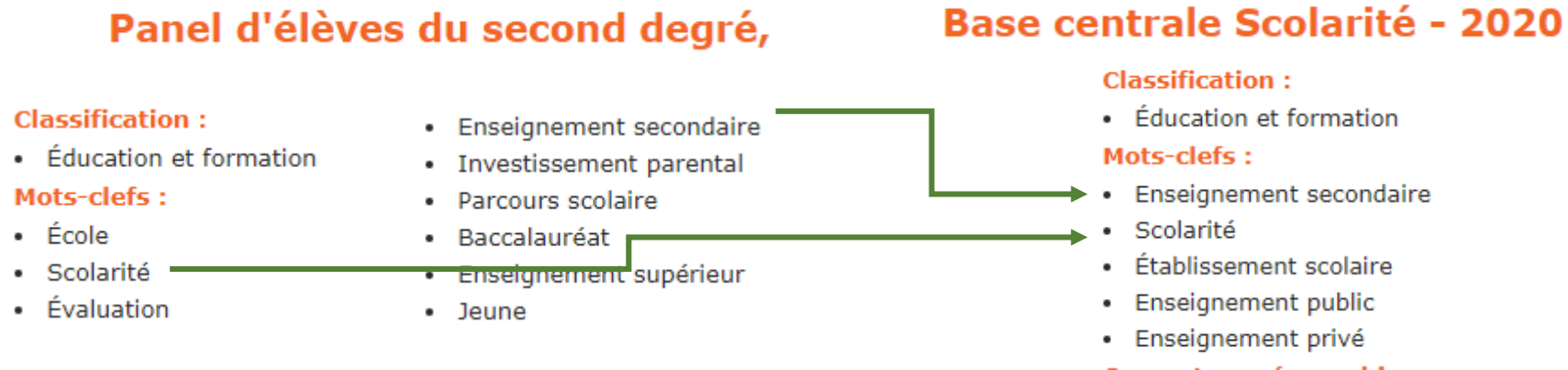
(*) Mostly on geographic variables, or profession



What about the interoperability ?

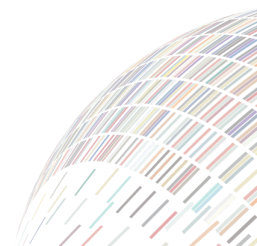


✓ Control vocabularies on each references



✓ Last version of DDI, difficult without IT support

✓ Work in progress : Use the european controlled vocabularies and translation of the metadata in english



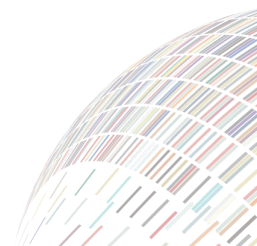
Reusing data



- ✓ Citation model for the use of data

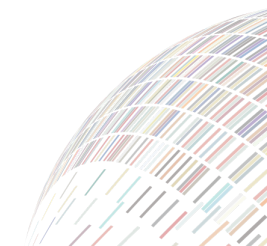
For example : lil-1416 : Budget des familles - 2017 (2017, INSEE)

- ✓ Metadata and data are well-described in DDI, they can be replicated and/or combined in different settings
- ✓ Internal persistent identifiers : *lil-1416*
- ✓ Work in progress : Digital Object Identifier



In conclusion

- ✓ Still a lot to do
- ✓ Provide FAIR data for almost 20 years
- ✓ Any idea to help us ?
- ✓ Any questions ?



Thanks for your attention

elodie.petorin@cnr.fr

DATA
INFRASTRUCTURE