



ined
INSTITUT
NATIONAL
D'ÉTUDES
DÉMOGRA
PHIQUES

La mise à disposition des enquêtes à l'Institut national d'études démographiques (Ined) : organisation et offre

Arianna Caporali, Amandine Morisset, Stéphane Legleye

Ined, Service des enquêtes et sondages

Journées de Méthodologie Statistique de l'Insee – 1^{er} avril 2015

Ceci est un résumé d'un article qui sera publié dans *Population* :

Caporali A., Morisset A., Legleye S. (à paraître 2015), « L'Ined et la mise à disposition des enquêtes quantitatives pour la recherche en sciences sociales », *Population*, 70(3).

- La mise à disposition des enquêtes est un atout pour les sciences sociales (ex. King 1995).
- Néanmoins, les activités des *survey data archives* sont parfois peu connues.

1. Historique et contexte actuel des *survey data archives* à l'international, en France et à l'Ined ; description restreinte aux archives d'enquêtes quantitatives.
2. Présentation du catalogue en ligne des enquêtes de l'Ined et du travail de préparation de données et métadonnées effectué par son équipe « mise à disposition ».

Historique des *survey data archives* – à l'international

- **Années 50 et 60** : création des premières *survey data archives* aux Etats-Unis et en Europe (Allemagne, Grand Bretagne, Belgique, etc.) (ex. Doorn and Tjalsma 2007; Silberman 1999).
- Contextes académiques favorisant la recherche empirique quantitative et les comparaisons internationales en sciences sociales.
- **Années 70** : création de **réseaux internationaux** pour établir des **normes internationales** pour la documentation des enquêtes (ex. Bisco 1966; Scheuch 2004).

La DDI (*Data Documentation Initiative*)



- Norme lancée en 2000, en langage XML, et adoptée par le **CESSDA** (*Consortium of European Social Science Data Archives*).
- Deux spécifications (Hansen et al. 2011; Kramer et al. 2011) :
 1. *DDI-Codebook* (DDI-C ou DDI 2) : description de l'étude empirique jusqu'au niveau variable,
 2. *DDI-Lifecycle* (DDI-L ou DDI 3) : documentation de l'ensemble du cycle de vie des données et des enquêtes longitudinales.
- **DDI-C** est largement utilisée en combinaison avec le logiciel **Nesstar** (*Network Social Science Tools and Resources*) qui permet l'exploration en ligne des données et métadonnées.

Historique des *survey data* archives – en France

- **Années 1980** : création des **premières *survey data*** archives (BDSP et LASMAS).
- 1999 : le rapport Silberman décrit les raisons du retard « français », une sociologie quantitative peu développée et un cadre juridique protectif vis-à-vis des données individuelles.
- **2001** : création du **Réseau Quetelet** pour harmoniser et coordonner l'accès aux enquêtes.
- 2004 : réforme de la loi « informatique et libertés » ouvrant à la réutilisation des données personnelles à des fins de recherche.
- 2008 : loi sur les archives ouvrant à l'utilisation de données très détaillées de la statistique publique à des fins de recherche.

Contexte actuel des *survey data archives*



**Accès aux « fichiers
production recherche »**

Centre Maurice Halbwachs - ADISP



**Accès aux
données très
détaillées**




- **Années 90** : la hausse des demandes d'accès génère le besoin d'organiser de façon formelle l'accès aux enquêtes.
- 2001 : le Services des enquêtes et sondages (SES) de l'Ined est co-fondateur du Réseau Quetelet.
- 2004 : premier fichiers accessibles selon une procédure définie et avec une documentation complète.
- **2012** : lancement du **catalogue Nesstar des enquêtes de l'Ined**.

Catalogue Nesstar des enquêtes de l'Ined



<http://nesstar.ined.fr>

ined  Catalogue Nesstar des enquêtes de l'INED [Commander les données](#) / [Retour au catalogue](#)

Catalogue des enquêtes de l'INED

- Enquêtes mises à disposition
 - Années 2000-2014
 - Années 1990-1999
 - Années 1980-1989
 - Années 1970-1979
 - Années 1960-1969
 - Années 1950-1959
- Enquêtes non mises à disposition
 - Années 2000-2014
 - Années 1990-1999
 - Années 1980-1989
 - Années 1970-1979
 - Années 1960-1969
 - Années 1950-1959
 - Années 1940-1949
 - Systèmes de suivi démographique en cours

Bienvenue sur le catalogue des enquêtes de l'INED

Le serveur Nesstar de l'INED permet un accès en ligne, direct et gratuit, aux enquêtes que l'Institut National d'Études Démographiques a réalisées depuis sa création en 1945 (soit un peu plus de 200 références).
Aucune inscription n'est nécessaire, et il est possible pour chaque enquête de :

- **Consulter la documentation liée à l'enquête**
- **Parcourir les données documentées**
- **Voir les tris à plat***
- **Réaliser des tris croisés ou analyses simples***
- **Télécharger l'ensemble de la documentation**
- **Télécharger les matériaux de l'enquête** (ex : questionnaire)

Pour télécharger des jeux de données, rendez vous sur [le portail des données du réseau Quetelet](#) ou cliquez à tout moment de votre navigation sur le lien "**Commander les données**" (en haut à droite de chaque page).

**Les données de certaines enquêtes doivent être pondérées avant de pouvoir être utilisées. Veuillez consulter la documentation associée à chaque jeu de données pour plus d'informations sur la pondération à appliquer.*

Aide à la navigation: Le serveur Nesstar de l'INED comporte deux cadres. Le cadre de gauche permet de parcourir les bases de données disponibles et le cadre de droite d'explorer et analyser les données.

Les enquêtes sont divisées en 2 catégories : "les Enquêtes mises à disposition" sont celles pour lesquelles il est possible de commander le(s) fichier(s) de données, tandis que les "Enquêtes non mises à disposition" sont celles où il n'est pas possible actuellement de commander le fichier de données (soit parce que celui-ci n'existe pas, soit parce qu'il n'est pas encore mis à disposition).

Pour rechercher un terme ou un sujet précis, utilisez les fonctions de recherche situées en haut à gauche de votre écran (sous le logo de l'INED).

Pour plus d'aide à la navigation, consultez [notre guide d'utilisation de Nesstar](#) ou cliquez à n'importe quel moment sur l'icône? située en haut à droite de votre écran.

Ce catalogue est géré par l'équipe "Mise à disposition" du service des enquêtes et sondages de l'INED.
Contact: enquetes@ined.fr

Catalogue Nesstar des enquêtes de l'Ined



<http://nesstar.ined.fr>

A screenshot of the Nesstar catalog website. The page title is 'Catalogue Nesstar des enquêtes de l'INED'. On the left, there is a navigation menu with categories like 'Catalogue des enquêtes de l'INED', 'Enquêtes mises à disposition', and 'Enquêtes non mises à disposition'. The main content area is mostly obscured by a large red oval that contains the following text:

Thèmes traités : fécondité, contraception, sexualité, couple, migrations, intégration, discrimination, genre, génération, inégalité, santé, vieillissement, habitat, emploi, etc.

Catalogue Nesstar des enquêtes de l'Ined



<http://nesstar.ined.fr>

The screenshot shows the 'Catalogue Nesstar des enquêtes de l'INED' website. On the left, there is a navigation menu with categories like 'Catalogue des enquêtes de l'INED', 'Enquêtes mises à disposition' (with sub-items for years 2000-2014, 1990-1999, 1980-1989, 1970-1979, 1960-1969, 1950-1959), and 'Enquêtes non mises à disposition' (with sub-items for years 2000-2014, 1990-1999, 1980-1989, 1970-1979, 1960-1969, 1950-1959, 1940-1949, and 'Systèmes de suivi démographique en cours').

A red box highlights the following text in the main content area:

+ 50 enquêtes mises à disposition : il est possible de commander le(s) fichier(s) de données.

Below this box, there is a list of actions:

- Consulter la documentation
- Parcourir les données
- Voir les tris à plat*
- Réaliser des tris croisés
- Télécharger l'ensemble
- Télécharger les matières

Below the list, it says 'Pour ou cliquez à' and '*Les données de certaines enquêtes'.

At the bottom of the page, there is an 'Aide à la navigation' section:

Aide à la navigation: Le serveur Nesstar de l'INED comporte deux cadres. Le cadre de gauche permet de parcourir les bases de données disponibles et le cadre de droite d'explorer et analyser les données.

Les enquêtes sont divisées en **2 catégories** : "les Enquêtes mises à disposition" sont celles pour lesquelles il est possible de commander le(s) fichier(s) de données, tandis que les "Enquêtes non mises à disposition" sont celles où il n'est pas possible actuellement de commander le fichier de données (soit parce que celui-ci n'existe pas, soit parce qu'il n'est pas encore mis à disposition).

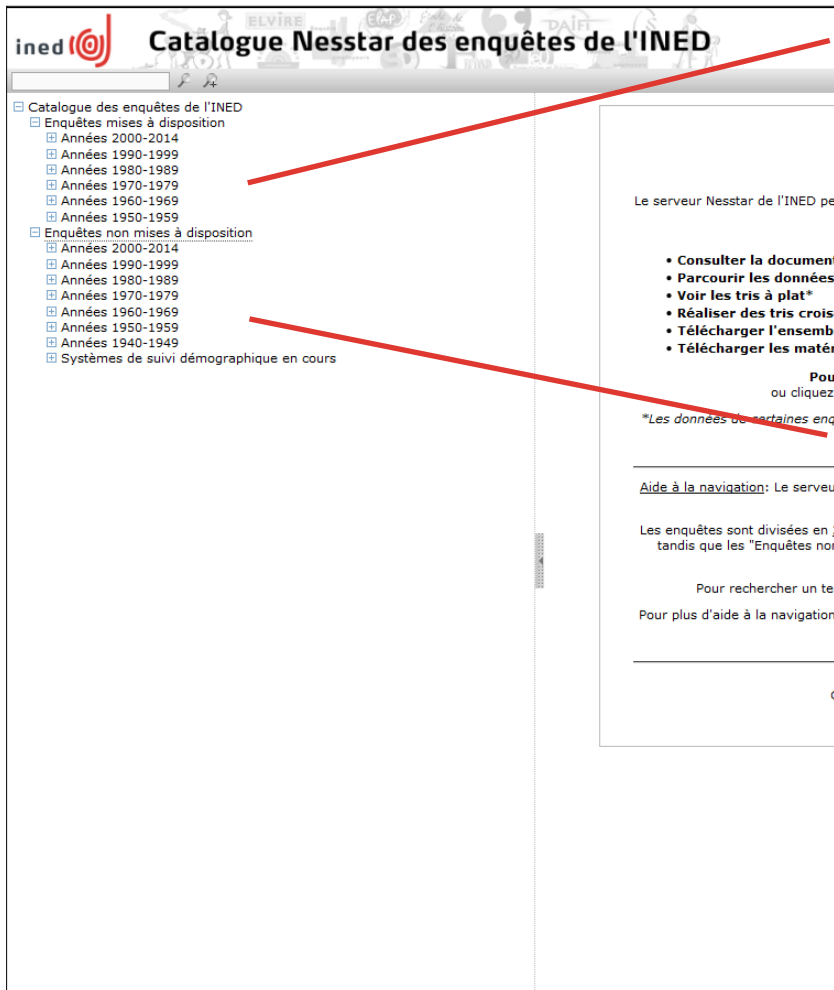
Pour rechercher un terme ou un sujet précis, utilisez les fonctions de recherche situées en haut à gauche de votre écran (sous le logo de l'INED).

Pour plus d'aide à la navigation, consultez [notre guide d'utilisation de Nesstar](#) ou cliquez à n'importe quel moment sur l'icône? située en haut à droite de votre écran.

Ce catalogue est géré par l'équipe "Mise à disposition" du service des enquêtes et sondages de l'INED.
Contact: enquetes@ined.fr

Catalogue Nesstar des enquêtes de l'Ined

<http://nesstar.ined.fr>



The screenshot shows the 'Catalogue Nesstar des enquêtes de l'INED' website. The left sidebar contains a tree view of survey categories:

- ☐ Catalogue des enquêtes de l'INED
 - ☐ Enquêtes mises à disposition
 - ☐ Années 2000-2014
 - ☐ Années 1990-1999
 - ☐ Années 1980-1989
 - ☐ Années 1970-1979
 - ☐ Années 1960-1969
 - ☐ Années 1950-1959
 - ☐ Enquêtes non mises à disposition
 - ☐ Années 2000-2014
 - ☐ Années 1990-1999
 - ☐ Années 1980-1989
 - ☐ Années 1970-1979
 - ☐ Années 1960-1969
 - ☐ Années 1950-1959
 - ☐ Années 1940-1949
 - ☐ Systèmes de suivi démographique en cours

The main content area displays a list of survey details, including a search bar and a list of actions:

- Consulter la documentation
- Parcourir les données
- Voir les tris à plat*
- Réaliser des tris croisés
- Télécharger l'ensemble
- Télécharger les matériels

Below the list, there is a section for navigation and search options.

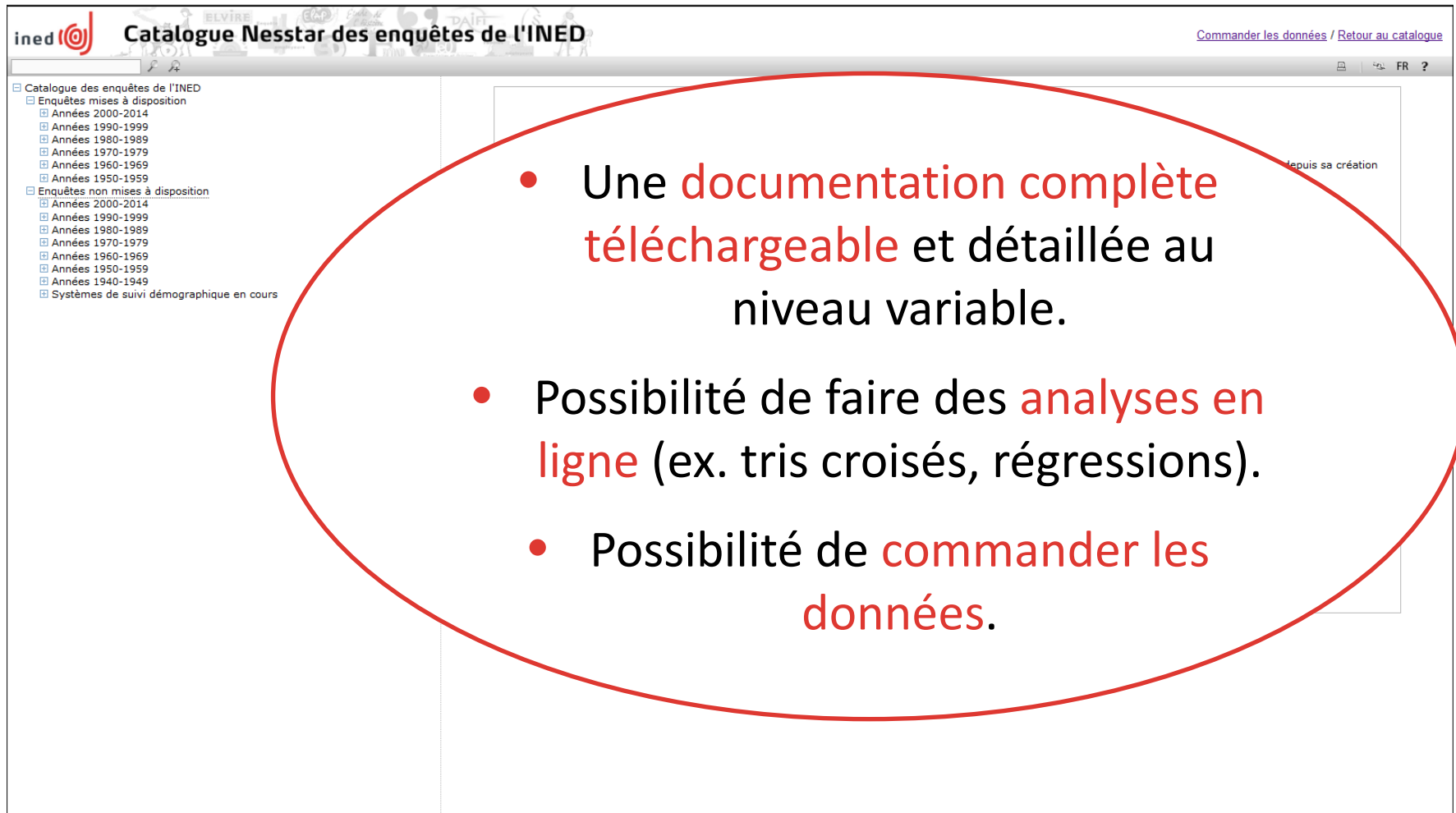
+ 50 enquêtes mises à disposition : il est possible de commander le(s) fichier(s) de données.

+ 190 enquêtes non mises à disposition : il n'est pas possible de commander le jeu de données (soit parce que celui-ci n'existe pas, soit parce qu'il n'est pas encore mis à disposition).

Catalogue Nesstar des enquêtes de l'Ined



<http://nesstar.ined.fr>

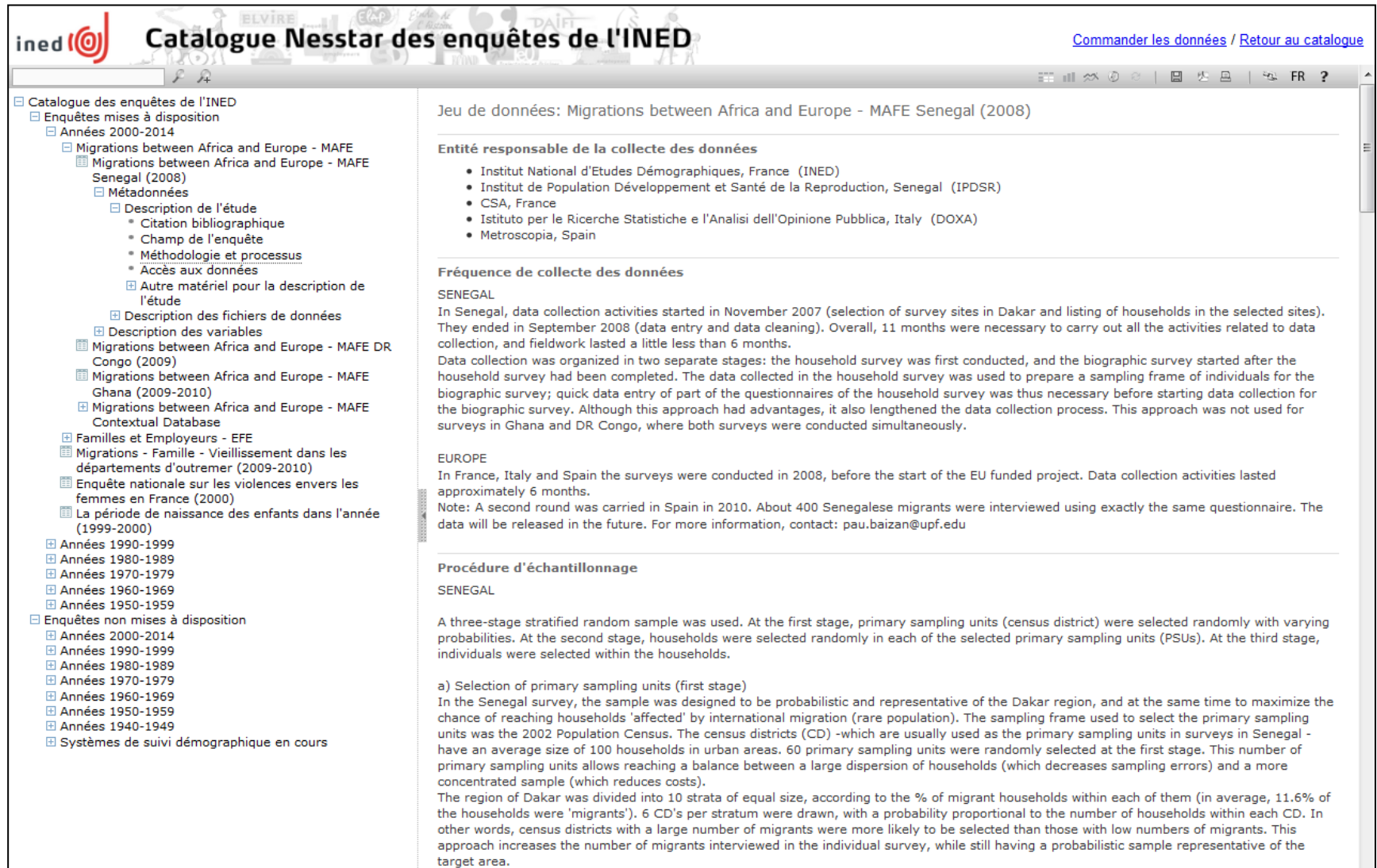
A screenshot of the Nesstar catalog website interface. The page title is 'Catalogue Nesstar des enquêtes de l'INED'. The left sidebar shows a tree view of survey categories: 'Catalogue des enquêtes de l'INED', 'Enquêtes mises à disposition' (with sub-items for years 2000-2014, 1990-1999, 1980-1989, 1970-1979, 1960-1969, 1950-1959), 'Enquêtes non mises à disposition' (with sub-items for years 2000-2014, 1990-1999, 1980-1989, 1970-1979, 1960-1969, 1950-1959, 1940-1949), and 'Systèmes de suivi démographique en cours'. The main content area is mostly obscured by a large red oval containing text. In the top right corner of the interface, there is a link 'Commander les données / Retour au catalogue'.

- Une **documentation complète téléchargeable** et détaillée au niveau variable.
- Possibilité de faire des **analyses en ligne** (ex. tris croisés, régressions).
- Possibilité de **commander les données**.

Préparation des données et des métadonnées

- **Documentation exhaustive** sur le projet et le(s) fichier(s) (ex. documents de travail, articles, questionnaires, dictionnaires des codes).
- **Création d'un fichier qui synthétise toutes les métadonnées** selon la norme DDI.
- Examen et **documentation** de toutes les **variables** (ex. labels, modalités, valeurs manquantes, questions, univers).
- Souvent, il faut compléter les informations manquantes ou choisir l'information correcte.

Affichage de métadonnées dans Nesstar



ined Catalogue Nesstar des enquêtes de l'INED [Commander les données / Retour au catalogue](#)

Jeu de données: Migrations between Africa and Europe - MAFE Senegal (2008)

Entité responsable de la collecte des données

- Institut National d'Etudes Démographiques, France (INED)
- Institut de Population Développement et Santé de la Reproduction, Senegal (IPDSR)
- CSA, France
- Istituto per le Ricerche Statistiche e l'Analisi dell'Opinione Pubblica, Italy (DOXA)
- Metroscopia, Spain

Fréquence de collecte des données

SENEGAL
In Senegal, data collection activities started in November 2007 (selection of survey sites in Dakar and listing of households in the selected sites). They ended in September 2008 (data entry and data cleaning). Overall, 11 months were necessary to carry out all the activities related to data collection, and fieldwork lasted a little less than 6 months.
Data collection was organized in two separate stages: the household survey was first conducted, and the biographic survey started after the household survey had been completed. The data collected in the household survey was used to prepare a sampling frame of individuals for the biographic survey; quick data entry of part of the questionnaires of the household survey was thus necessary before starting data collection for the biographic survey. Although this approach had advantages, it also lengthened the data collection process. This approach was not used for surveys in Ghana and DR Congo, where both surveys were conducted simultaneously.

EUROPE
In France, Italy and Spain the surveys were conducted in 2008, before the start of the EU funded project. Data collection activities lasted approximately 6 months.
Note: A second round was carried in Spain in 2010. About 400 Senegalese migrants were interviewed using exactly the same questionnaire. The data will be released in the future. For more information, contact: pau.baizan@upf.edu

Procédure d'échantillonnage

SENEGAL

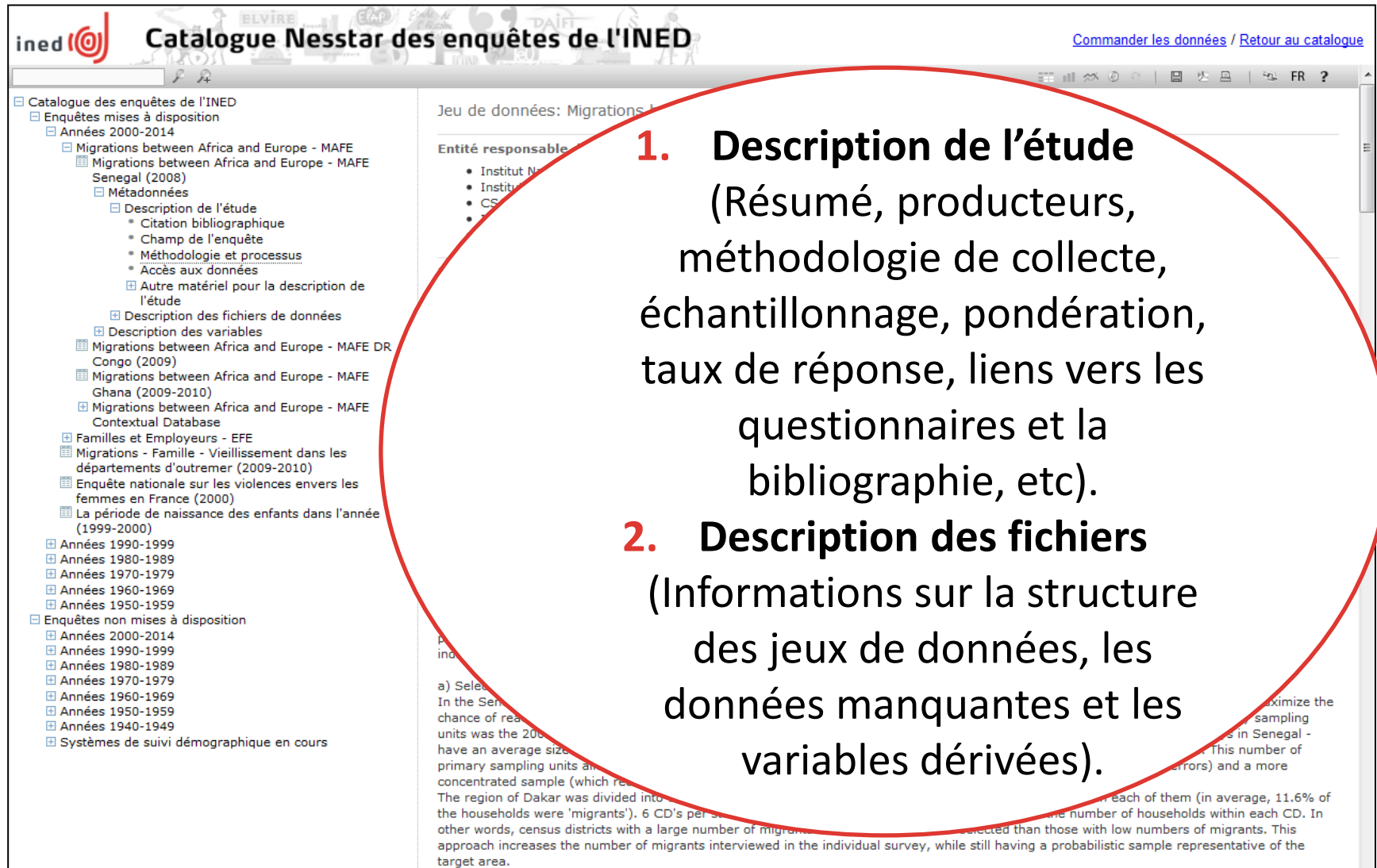
A three-stage stratified random sample was used. At the first stage, primary sampling units (census district) were selected randomly with varying probabilities. At the second stage, households were selected randomly in each of the selected primary sampling units (PSUs). At the third stage, individuals were selected within the households.

a) Selection of primary sampling units (first stage)
In the Senegal survey, the sample was designed to be probabilistic and representative of the Dakar region, and at the same time to maximize the chance of reaching households 'affected' by international migration (rare population). The sampling frame used to select the primary sampling units was the 2002 Population Census. The census districts (CD) - which are usually used as the primary sampling units in surveys in Senegal - have an average size of 100 households in urban areas. 60 primary sampling units were randomly selected at the first stage. This number of primary sampling units allows reaching a balance between a large dispersion of households (which decreases sampling errors) and a more concentrated sample (which reduces costs).
The region of Dakar was divided into 10 strata of equal size, according to the % of migrant households within each of them (in average, 11.6% of the households were 'migrants'). 6 CD's per stratum were drawn, with a probability proportional to the number of households within each CD. In other words, census districts with a large number of migrants were more likely to be selected than those with low numbers of migrants. This approach increases the number of migrants interviewed in the individual survey, while still having a probabilistic sample representative of the target area.

Catalogue des enquêtes de l'INED

- Enquêtes mises à disposition
 - Années 2000-2014
 - Migrations between Africa and Europe - MAFE
 - Migrations between Africa and Europe - MAFE Senegal (2008)
 - Métadonnées
 - Description de l'étude
 - Citation bibliographique
 - Champ de l'enquête
 - Méthodologie et processus
 - Accès aux données
 - Autre matériel pour la description de l'étude
 - Description des fichiers de données
 - Description des variables
 - Migrations between Africa and Europe - MAFE DR Congo (2009)
 - Migrations between Africa and Europe - MAFE Ghana (2009-2010)
 - Migrations between Africa and Europe - MAFE Contextual Database
 - Familles et Employeurs - EFE
 - Migrations - Famille - Vieillesse dans les départements d'outremer (2009-2010)
 - Enquête nationale sur les violences envers les femmes en France (2000)
 - La période de naissance des enfants dans l'année (1999-2000)
 - Années 1990-1999
 - Années 1980-1989
 - Années 1970-1979
 - Années 1960-1969
 - Années 1950-1959
 - Enquêtes non mises à disposition
 - Années 2000-2014
 - Années 1990-1999
 - Années 1980-1989
 - Années 1970-1979
 - Années 1960-1969
 - Années 1950-1959
 - Années 1940-1949
 - Systèmes de suivi démographique en cours

Affichage de métadonnées dans Nesstar



ined Catalogue Nesstar des enquêtes de l'INED [Commander les données / Retour au catalogue](#)

1. Description de l'étude
(Résumé, producteurs, méthodologie de collecte, échantillonnage, pondération, taux de réponse, liens vers les questionnaires et la bibliographie, etc).

2. Description des fichiers
(Informations sur la structure des jeux de données, les données manquantes et les variables dérivées).

Entité responsable

- Institut N...
- Institut...
- CS...

Jeu de données: Migrations 1

a) Selection of the sample
In the Senegal survey, the chance of reaching a household was the 200... units was the 200... have an average size... primary sampling units and... concentrated sample (which re... The region of Dakar was divided into... the households were 'migrants'). 6 CD's per... each of them (in average, 11.6% of... other words, census districts with a large number of migrants... collected than those with low numbers of migrants. This approach increases the number of migrants interviewed in the individual survey, while still having a probabilistic sample representative of the target area.

Description d'une variable dans Nesstar

ined Catalogue Nesstar des enquêtes de l'INED [Commander les données / Retour au catalogue](#)

DESCRIPTION TABLEAU ANALYSE

Jeu de données: Familles et Employeurs - Volet Familles (2004-2005)

Variable g_chrais: Raison principale du chômage

QUESTION LITTÉRALE
G7 - Pour quelle raison principale êtes-vous au chômage ?

Valeurs	Modalités	N
1	Vous avez démissionné de votre emploi précédent ou stoppé une activité indépendante	79 10,0%
2	Vous avez été licencié pour raisons économiques ou vous avez dû fermer votre entreprise	122 15,5%
3	Vous avez été licencié(e) pour d'autres raisons	101 12,8%
4	Votre CDD s'est terminé et n'a pas été renouvelé	337 42,8%
5	Vous n'avez jamais travaillé ou vous ne trouvez pas de travail	55 7,0%
6	Vous êtes travailleur intermittent et vous vous trouvez entre deux contrats	17 2,2%
7	Autre raison	77 9,8%

STATISTIQUES SOMMAIRES
Cas valides 788
Cas manquants 0
Cette variable est alphanumérique

INSTRUCTIONS POUR L'ENQUÊTEUR
Présenter carte 5
Une seule réponse possible

UNIVERS
Si F_SITUA en 4
Si F_TRAVAIL en 2

- Catalogue des enquêtes de l'INED
 - Enquêtes mises à disposition
 - Années 2000-2014
 - Migrations between Africa and Europe - MAFE
 - Familles et Employeurs - EFE
 - Familles et Employeurs - Volet Familles (2004-2005)
 - Métadonnées
 - Description des variables
 - TABLE DES HABITANTS DU LOGEMENT
 - TABLE INDIVIDU
 - Description de soi
 - Calendrier personnel
 - Débuts de vie de couple
 - Vos enfants
 - Emploi et conditions de travail
 - Recherche d'emploi et chômeurs
 - Durée de l'inactivité/interruption d'emploi
 - Proposition de travail reçue depuis la fin d'activité
 - 1ère raison du refus du travail proposé
 - 2ème raison du refus du travail proposé
 - 3ème raison du refus du travail proposé
 - 4ème raison du refus du travail proposé
 - 5ème raison du refus du travail proposé
 - Petits boulots depuis la fin d'activité
 - Formations suivies depuis la fin d'activité
 - Plus de temps consacré au conjoint depuis la fin d'activité
 - Plus de temps consacré aux enfants depuis la fin d'activité
 - Plus de temps consacré aux amis depuis la fin d'activité
 - Plus de temps pour des activités souhaitées depuis la fin d'activité
 - Plus de temps consacré pour d'autres activités depuis la fin d'activité
 - Raison principale du chômage

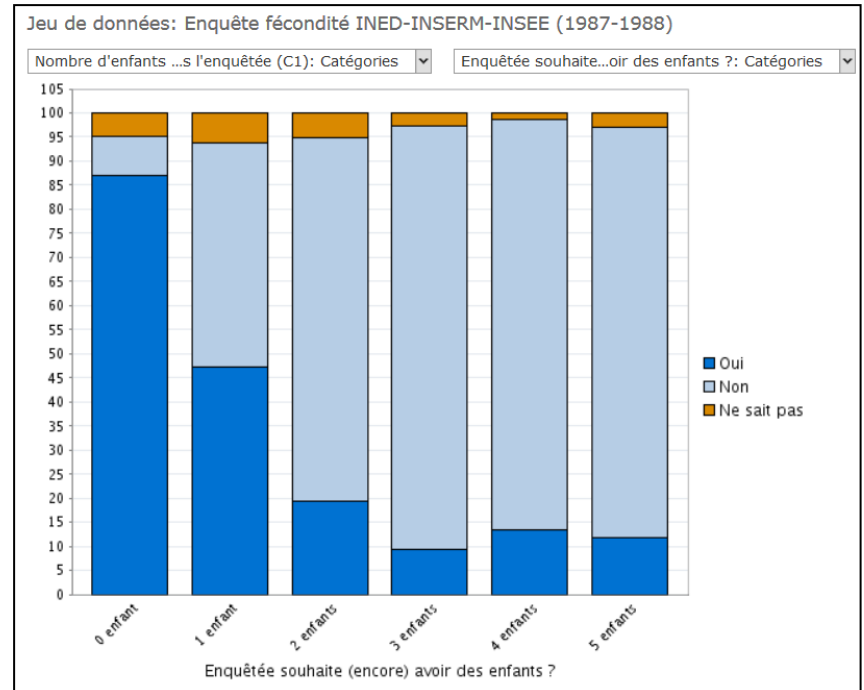
Exemple d'analyses dans Nesstar

Jeu de données: Enquête fécondité INED-INSERM-INSEE (1987-1988)

Nombre d'enfants ...s l'enquêtée (C1): Catégories Enquêtée souhaite...oir des enfants ?: Catégories

Enquêtée souhaite (encore) avoir des enfants ?	Oui	Non	Ne sait pas	Total	N=
Nombre d'enfants qu'a eus l'enquêtée (C1)					
0 enfant	87,0	8,1	4,9	100,0	3 704,0
1 enfant	47,3	46,6	6,1	100,0	2 109,2
2 enfants	19,5	75,4	5,1	100,0	2 958,7
3 enfants	9,5	87,8	2,7	100,0	1 276,6
4 enfants	13,5	85,1	1,4	100,0	313,9
5 enfants	11,9	85,1	3,0	100,0	122,9
Total	47,5	47,7	4,8	100,0	10 485,2

Filtre en place
 Pondération en place



Procédure d'accès aux données Ined

Nesstar =>



=> SES => Utilisateur

- Accès gratuit pour des fins de recherche.
- Procédure standardisée dans la plupart des cas.
- Formulaire d'engagement (ex. ne pas céder les données à un tiers, citer la source).
- Temps moyen pour obtenir l'accès : 1-2 semaines.
- Fichiers téléchargeables en SAS, SPSS, csv depuis le portail du Réseau Quetelet.

- Le catalogue en ligne des enquêtes de l'Ined contribue à leur diffusion (+8% des demandes d'accès).
- La mise à disposition des enquêtes devient de plus en plus importante avec la demande croissante d'accès aux données et l'augmentation du coût des enquêtes.
- L'expérience du SES suggère que, pour optimiser le processus de préparation des données et métadonnées, il est important de mieux collaborer avec les producteurs des données.