

**PRO
GE
DO**

DATA
INFRASTRUCTURE

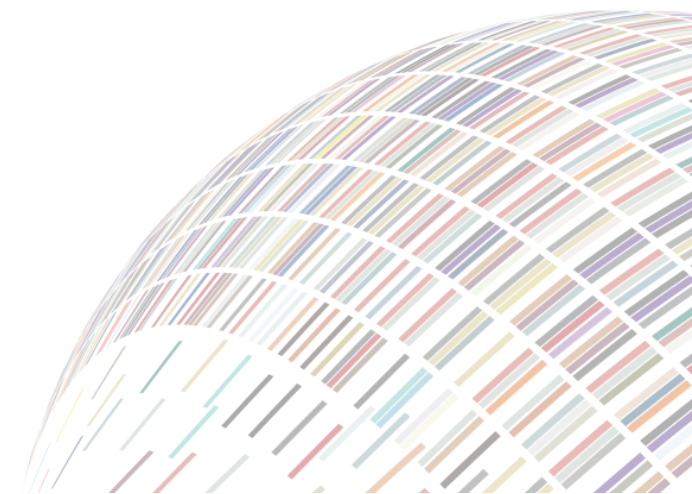
*Développer
la culture des données*

archives de données issues
de la statistique publique
disp

Quetelet-Progedo-Diffusion : un portail de stockage et de diffusion pour les grandes enquêtes en démographie historique.

Le 23 novembre 2023

Alena Lapatniova (Progedo-Adisp)



Grande expérience de la documentation et de la diffusion des données en SHS

Depuis plus de 20 ans l'Adisp documente et diffuse aux chercheurs SHS des données issues de la statistique publique.

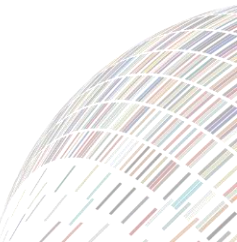
- Les ingénieurs de l'Adisp sont experts en documentation des enquêtes
- Le catalogue est connu par les chercheurs et les étudiants en SHS
- Expertise reconnue: PROGEDO a reçu le label « Centre de référence thématique » de Recherche.Data.Gouv



Les avantages à diffuser les données en démographie historique par PROGEDO

Pour que les données soient réutilisées, il faudrait qu'elles soient de préférence:

- Concentrées au même endroit pour être plus facilement trouvable;
- Bien décrites et documentées;
- Accessibles au même formats les unes et les autres.



Le catalogue Quetelet-Progedo Diffusion

The screenshot shows the website interface for the Quetelet-Progedo Diffusion data catalog. At the top, there is a navigation bar with the site logo and links for 'Recherche', 'Mode d'emploi', and 'Panier'. Below this, the site is identified as a member of CESSDA. The main header reads 'CATALOGUE DE DONNÉES EN SCIENCES HUMAINES ET SOCIALES POUR LA RECHERCHE'. Three key statistics are displayed: 1543 Jeux de données, 396 514 Variables, and 13 133 Utilisateurs. A search bar is present with a 'Rechercher' button. Below the search bar, there are filters for 'Séries (93)', 'Jeux de données (1543)', and 'Variables (396514)'. Two data items are shown as examples: 'Capacité des communes en hébergement touristique - 2022' and 'Table d'appartenance géographique des communes - 2022', both with 'Ajouter au panier' buttons. On the left side, there are filters for 'Années' (ranging from 2018 to 2022) and 'Thèmes' (including Données localisées, Population et démographie, Conditions de vie et société, Travail et emploi, and Salaire et revenus).

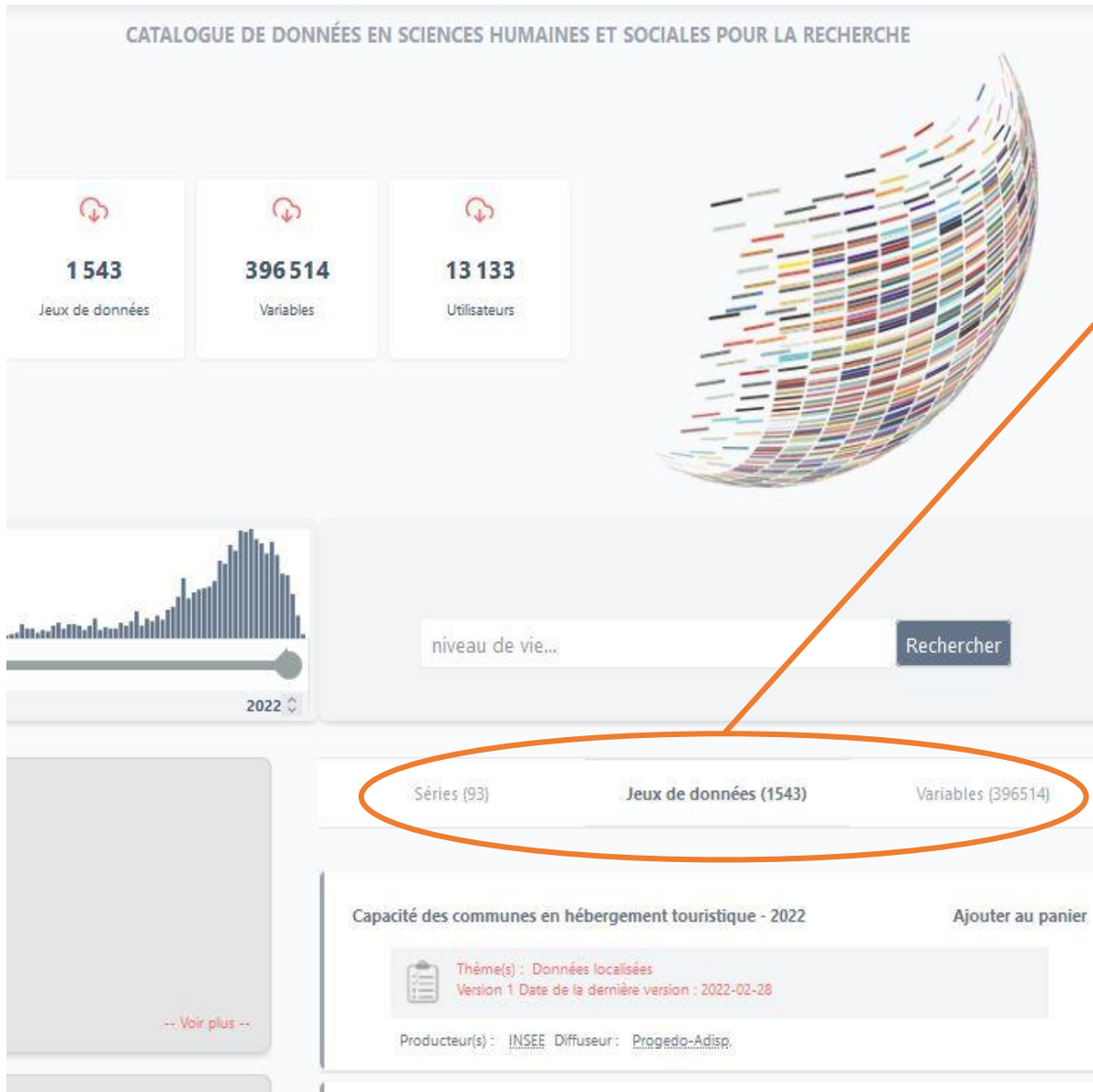
Le nouveau portail de données de Quetelet-Progedo-Diffusion propose :

- Une **interface** conçue pour trouver rapidement les données recherchées :
 - un **outil de recherche** intégré qui interroge dans tous les champs de description, aussi bien dans les titres des jeux de données que dans les résumés et la description des variables.

Le nouveau portail de données de Quetelet-Progedo-Diffusion, dispose aussi :

- des **filtres** pour affiner ses recherches :
- par exemple, par thèmes
 - ou par thème et producteur

La présentation des données



3 niveaux de présentation des données :

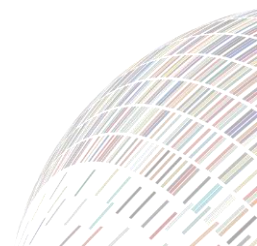
- Les **jeux de données**
- Les **séries** (ensemble de jeux de données)
- Les **variables**

Description fine des données (métadonnées) en utilisant le standard international DDI

Nous utilisons le schéma de description des données DDI qui permet :

- Décrire le projet de recherche dans le cadre duquel les données ont été collectées, préciser ses dimensions géographiques et temporelles et la population faisant l'objet de l'étude.
- Décrire les éléments méthodologiques tel que le plan et mode de collecte des données, la méthode d'échantillonnage, les sources primaires et secondaires utilisées, leur histoire et le lieu de leur conservation.
- Joindre les liens des publications apparentées (articles dans des revues, sur HAL, etc.), d'un site de valorisation etc.
- Décrire les fichiers de données et chaque variable présente, sa source par exemple, sa modification dans le temps, les données manquantes, etc. (ce que ne permet pas Dataverse par exemple).

Chaque jeu de données obtient un identifiant unique, un **DOI**, permettant de renvoyer vers les données dans un article, par exemple.



La description générale des données

6 onglets permettent de naviguer dans les **métadonnées** des enquêtes et dans la documentation qui leur est liée.

Base centrale Scolarité - 2021
Version 1 date : 2021-12-03

JEU DE DONNÉES

Identifiants :
lil-1508
doi:10.13144/lil-1508

Thème :
Éducation et formation

Séries : Base Centrale Scolarité (BCS)

Résumé Détails Variables Données Documentation Accès

Description générale
Les données extraites de la BCS couvrent 99% des élèves des établissements du second degré publics et privés (sous contrat et hors contrat) dépendant du Ministère de l'éducation nationale (France métropolitaine + DOM). Elles décrivent, d'une part, l'élève par ses caractéristiques sociodémographiques (âge, sexe, origine sociale, nationalité) et, d'autre part, son établissement

Résumé :

- Description générale
- Thème et mots-clefs
- Producteur / Diffuseur
- Date de production
- Version(s) des données
- Modèle de citation
- ...

Détails :

- Couverture géographique
- Univers
- Méthodologie
- Collecte
- Echantillonnage
- Pondération
- ...

Variables :

- Fichier d'appartenance
- Type de variable
- Statistiques descriptives

Données :

- Description des fichiers :
- Nombre d'observations
 - Nombre de variables

Documentation :

- Téléchargement libre du (des) ...
- Dictionnaire(s) des variables
 - Questionnaire(s)
 - Doc. méthodologiques
 - ...

Accès :

- Statut des données
- Modèle de citation
- Rappel des règles et conditions d'accès
- ...

EXEMPLE : Paroisses et Communes de France.

Loir-et-Cher – 1250-1990 (<https://data.progedo.fr/studies/doi/10.13144/lil-1506>)

Identifiants :
lil-1506
doi:10.13144/lil-1506

Thèmes :
Données historiques
Histoire
Démographie (Population, état civil, et recensement)

Série : [Paroisses et communes de France](#)

Couverture géographique :
Département du Loir et Cher et espace qui l'a formé avant la Révolution

Producteur :
[EHESS](#)

Diffuseur :
[Progedo-Adisp](#)

Résumé
Détails
Variables
Données
Documentation
Accès

Description générale
La base est constituée d'après les données compilées sur les 335 fiches publiées par C. Poitou, Dictionnaire d'histoire administrative et démographique, Loir-et-Cher, Paris, CNRS Éditions, collection Paroisses et communes de France, 1997, 591 p. Ce tableur documente les caractéristiques administratives d'Ancien Régime et de l'époque contemporaine des 334 paroisses et/ou communes et 101 variables démographiques, rangées par ordre chronologique.

Thèmes
Données historiques, Histoire, Démographie (Population, état civil, et recensement)

Mots-clés
Démographie, Administration, Population, Histoire administrative

Série
Paroisses et communes de France

Production

Producteur

Nom	Acronyme
École des Hautes Études en Sciences Sociales	EHESS

Auteur

Nom	Affiliation
Poitou, Christian	LDH-EHSS

Autres contributeurs et remerciements

Nom	Rôle	Affiliation
Motte, Claude	contributeur	LDH-EHSS
Carron, Diane	préparation des jeux de données, métadonnées	CRH-EHESS-CNRS

Date de production
1990

Période couverte par les données
1250 - 1990

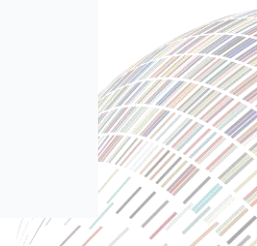
Diffuseur

Nom	Acronyme	Affiliation	URI
Archives de Données Issues de la Statistique Publique (ADISP)	Progedo-Adisp	Quetelet-Progedo Diffusion	http://www.progedo-adisp.fr

Version des données
Version 1

Date de la dernière version
2021-12-03

Citation bibliographique
doi:10.13144/lil-1506
Paroisses et communes de France - Loir-et-Cher - 1250-1990 - EHESS (producteur), Poitou, Christian (auteur), ADISP (diffuseur)



EXEMPLE : Paroisses et Communes de France.

Loir-et-Cher – 1250-1990 (<https://data.progedo.fr/studies/doi/10.13144/lil-1506>)

Résumé	Détails	Variables	Données	Documentation	Accès
<p>Série : Paroisses et communes de France</p> <p>Couverture géographique : Département du Loir et Cher et espace qui l'a formé avant la Révolution</p> <p>Producteur : EHESS</p> <p>Diffuseur : Progedo-Adise</p>					
<p>Couverture géographique Département du Loir et Cher et espace qui l'a formé avant la Révolution</p> <p>Unité géographique Paroisse ou commune.</p> <p>Unité d'analyse Espace politico-administratif. La paroisse fiscale (aussi dénommée collecte ou communauté) avant la Révolution puis la commune.</p> <p>Univers La population dénombrée et recensée par paroisse puis par commune avec les principaux rattachements administratifs et institutionnels.</p> <p>Type de données Données de recensement et données administratives agrégées</p> <p>Méthodologie</p> <p>Dimension temporelle Série temporelle : discrète.</p> <p>Date de collecte 1976 - 1996</p> <p>Fréquence de collecte Toutes les données existantes ont été collectées. La source d'origine est de fréquence irrégulière pour l'Ancien Régime, un pas de temps quinquennal est observé pour les années 1801 à 1946 puis irrégulièrement depuis la seconde guerre mondiale.</p> <p>Procédure d'échantillonnage La collecte effectuée par Christian Poitou a été aussi exhaustive que possible pour les des sources de l'Ancien Régime, sauf les rôles de taille qui ont été échantillonnés avec un pas de 5 ans. Le relevé des dénombrements et recensements généraux de l'époque contemporaine est complet jusqu'en 1990. Le relevé aussi exhaustif que possible des sources de l'Ancien Régime, cependant, il ne s'agit pas toujours de la population réelle, les données éparses issues des rôles de taille qui figuraient en annexe VIII dans le Dictionnaire Paroisses et communes ont été réinsérées dans ce tableau.</p> <p>Mode de collecte Compilation / synthèse. En ce qui concerne la compilation des diverses sources pour constituer la base, Christian Poitou a transcrit sur un tableur informatisé les données issues d'archives conservées dans dix-sept dépôts ainsi que celles des documents imprimés. Le CRH a vérifié la conformité entre les fichiers informatisés saisis dans les années 2000 et le dictionnaire édité en 1990. Les chiffres parfois aberrants ont été rectifiés par rapport aux données antérieures et postérieures.</p>					

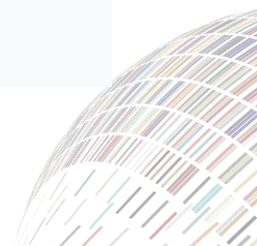
Caractéristiques de la collecte

La collecte a été faite en deux temps : le premier est celui de la compilation des données par C. Poitou pour préparer le dictionnaire Paroisses et communes dans les années 1990, la seconde est la remise en forme de ces données par D. Carron pour préparer le versement du jeu de données. Dans leur immense majorité, les sources de première main ont été établies par les institutions publiques voire par l'Église catholique avant la Révolution. Les données sont éparses avant le XVIIIe siècle (concernant soit un ressort particulier, soit une paroisse) et systématiques après la Révolution. Avant la Révolution la majeure partie des documents dénombrement les feux soumis à la taille, ce qui exclue par principe les personnes non assujetties à ce prélèvement (clergé, noble), elles peuvent être reportées mais sans régularité d'un document à l'autre. La valeur du feu dans cet espace est estimée à 4,5 habitants par feu. Le nombre de gabelants est fourni en nombre d'habitants à l'exclusion des enfants de moins de 8 ans, le chiffre reporté ne représente donc que 75 % environ de la population. Durant les XIXe et XXe siècles, les préconisations ont évolué en ce qui concerne le domicile de droit d'une part et le domicile de fait de l'autre pour les personnes présentes par intermittence en double compte (militaires, étudiants, travailleurs en foyers, hospitalisés, prisonniers) tout comme les modalités de comptage (comptage exhaustif, enquête au 1/8e). Le détail est précisé dans les sources. Le mode de collecte a varié d'une série à l'autre selon l'institution collectrice et la date considérée, les comparaisons entre séries devront donc être menées avec prudence et discernement en fonction des indications reportées dans les sources. Il est rare qu'une série renseigne toutes les paroisses d'Ancien Régime elle est souvent circonscrite à une élection ou un diocèse.

Traitements

Opération de contrôle

Une vérification des cotes d'archives a été opérée en 2021.



EXEMPLE : Paroisses et Communes de France.

Loir-et-Cher – 1250-1990 (<https://data.progedo.fr/studies/doi/10.13144/lil-1506>)

Résumé	Détails	Variables	Données	Documentation	Accès
--------	---------	-----------	---------	---------------	-------

V_1565_f **Nombre de feux en 1565** ^

Fichier : pcf_Loir_et_Cher
Type de variable : caracter
Univers : nombre de lieux renseignés : 24
Notes : Source: État des paroisses du pays de Berry avec le nombre de feux, BnF ms fr 3198 fol. 125-137.

Statistiques :

Nombre de réponses	-	334
Nombre total d'observations	-	334

V_1616_f **Nombre de feux en 1616** ^

Fichier : pcf_Loir_et_Cher
Type de variable : caracter
Univers : nombre de lieux renseignés : 63
Notes : Source: Rolle du nombre des feux qui sont es petites villes, bourgs et villages dépendant du duché d'Orléans, ADL, C suppl. 23. Publié dans C. Poitou, "Le peuplement de l'Orléanais en 1616", Bulletin de la SAHO, VIII, 1981, 61.

canton_1990 **Canton, subdivision de l'arrondissement, en 1990** ^

Fichier : pcf_Loir_et_Cher
Type de variable : caracter
Univers : nombre de lieux renseignés : 291
Descriptif de la variable : Situation du découpage cantonal coïncidant avec le recensement de la population le plus proche de la date de la parution du dictionnaire Paroisses et communes consacré à ce département.

Statistiques :

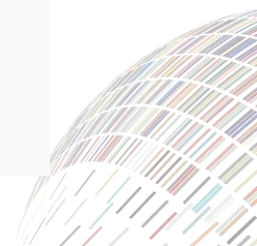
Nombre de réponses	-	334
Nombre total d'observations	-	334

coutume **Règles juridiques locales** ^

Fichier : pcf_Loir_et_Cher
Type de variable : caracter
Univers : nombre de lieux renseignés : 332
Descriptif de la variable : Rédigées dans le courant du XVI^e siècle, les coutumes ont consigné par écrit les règles et jurisprudences orales alors en vigueur dans douze provinces du nord du royaume de France. Il existait de nombreuses contestations quant à la coutume devant s'appliquer sur telle ou telle communauté d'habitants voire dans tel hameau et le ressort des juridictions compétentes pour trancher ces contentieux étant lui-même mouvant, c'est la situation existant à la Révolution qui est reportée dans les dictionnaires Paroisses et Communes comme dans ces tableurs. Les territoires qui ont formé le Loir-et-Cher étaient régis selon 12 coutumes.

Statistiques :

Nombre de réponses	-	334
Nombre total d'observations	-	334



EXEMPLE : Paroisses et Communes de France. Loir-et-Cher – 1250-1990 (<https://data.progedo.fr/studies/doi/10.13144/lil-1506>)



JEU DE DONNÉES

Paroisses et communes de France - Loir-et-Cher -
1250-1990

Version 1 date : 2021-12-03

[Ajouter au panier](#)

Résumé

Détails

Variables

Données

Documentation

Accès

Nom du fichier
pcf_Loir_et_Cher

Contenu du fichier
Le fichier présente les données sur les paroisses et communes du département du Loir-et-Cher. Abréviations utilisées : s_j = sans information : la paroisse ou la commune n'existe pas à la date considérée ; lac. = lacune dans la source utilisée ; n_c = non concernée : la source utilisée ne concerne qu'une partie de la circonscription et non l'ensemble des paroisses ou communes à cette date ; s_o = sans objet : la catégorie du libellé n'est pas opérante pour la paroisse ou la commune considérée. Lorsque la valeur fournie paraît fantaisiste, elle est suivie d'un signe "!".

Identifiants :
lil-1506
doi:10.13144/lil-1506

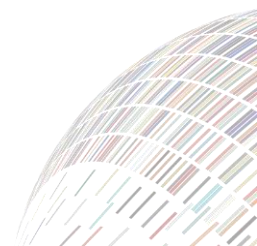
Thèmes :
Données historiques
Histoire
Démographie (Population, état civil, et recensement)

Série : [Paroisses et communes de France](#)


Couverture géographique :
Département du Loir et Cher et espace qui l'a formé avant la Révolution

Producteur :
[EHESS](#)

Diffuseur :
[Progedo-Adisp](#)



EXEMPLE : Paroisses et Communes de France. Loir-et-Cher – 1250-1990 (<https://data.progedo.fr/studies/doi/10.13144/lil-1506>)



JEU DE DONNÉES

Paroisses et communes de France - Loir-et-Cher -
1250-1990

Version 1 date : 2021-12-03

[Ajouter au panier](#)

Résumé


Détails

Variables


Données

Documentation

Accès



Dictionnaire des variables du département du Loir-et-Cher



Présentation Paroisses et communes

Identifiants :
lil-1506
doi:10.13144/lil-1506

Thèmes :
Données historiques
Histoire
Démographie (Population, état civil, et recensement)

Série : [Paroisses et communes de France](#)

Couverture géographique :
Département du Loir et Cher et espace qui l'a formé avant la Révolution

Producteur :
[EHES](#)

Diffuseur :
[Progedo-Adisp](#)

EXEMPLE : Paroisses et Communes de France. Loir-et-Cher – 1250-1990

(<https://data.progedo.fr/studies/doi/10.13144/lil-1506>)

Paroisses et communes de France - Loir-et-Cher - 1250-1990

Version 1 date : 2021-12-03

Ajouter au
panier

Résumé

Détails

Variabes

Données

Documentation

Accès

Localisation des données

Application de commande Quetelet-Progedo Diffusion

<https://commande.progedo.fr>

Centre d'archivage détenant des données

ADISP - Archives de Données Issues de la Statistique Publique

Type de diffusion

STD : Les fichiers Standard (STD) sont des données anonymisées accessibles aux demandeurs qui appartiennent à la communauté scientifique (cf. section "Restrictions").

Étendue de la collecte

La référence met à disposition 1 fichier de données et 1 dictionnaire des variables.

Nombre de fichiers de données

1 fichier de données

Notes

Les données sont disponibles aux formats SAS, SPSS, STATA et CSV.

Déclaration de confidentialité

Pour obtenir les données, l'utilisateur formule une demande sur le portail de commande de Quetelet-Progedo Diffusion ; acceptée sous condition de statut, présence d'une adresse e-mail institutionnelle et d'un projet de recherche.

Si la demande est acceptée, cela conduit à la signature d'un engagement individuel à destination de l'ADISP, et dans le cas des FPR, d'un engagement de confidentialité à destination du Comité du Secret Statistique. Ces engagements requièrent :

- D'utiliser les données exclusivement dans une finalité de recherche ;
- De ne pas céder ces données, sous quelque forme que ce soit, à une tierce personne, que ce soit à titre gratuit ou onéreux ;
- De traiter ces données conformément aux règles de l'art et du secret statistique ;
- De mentionner la source des données dans mes communications, publications... conformément au modèle de citation pour l'utilisation des données ;
- D'informer le diffuseur de ses communications, publications... et lui en faire parvenir les références ;

Restrictions d'usage

Les fichiers diffusés dans le cadre de Quetelet-Progedo Diffusion sont accessibles aux chercheurs français et étrangers, doctorants, post-doctorants, et étudiants de master à des fins de recherche, de production scientifique et dans certains cas d'enseignement. Toute utilisation commerciale est exclue.

Les critères d'une finalité de recherche sont la production ou reproduction, dans un but de validation de connaissances nouvelles de portée générale. Les résultats sont publics et libres de diffusion.

Personnes contact

Équipe de l'Adisp

email : diffusion.adisp@cnsr.fr

Texte à citer

doi:10.13144/lil-1506

Paroisses et communes de France - Loir-et-Cher - 1250-1990 - EHESS (producteur), Poitou, Christian (auteur), ADISP (diffuseur)

Requis du dépôt

Afin de fournir aux organismes producteurs l'information essentielle concernant l'utilisation des données, les utilisateurs de données sont tenus d'envoyer à l'ADISP (diffusion.adisp@cnsr.fr), si possible sous format électronique, les publications réalisées à partir de l'exploitation des données de l'enquête en envoyant la référence bibliographique et le résumé.

Conditions d'accès

L'utilisateur formule une demande sur le portail de commande de Quetelet-Progedo Diffusion, acceptée sous condition de statut, présence d'une adresse e-mail institutionnelle et d'un projet de recherche. Si la demande est acceptée, cela conduit à la signature d'un engagement individuel à destination de l'ADISP, et dans le cas des FPR, d'un engagement de confidentialité à destination du Comité du Secret Statistique. Si les données sont utilisées de manière collective, tous les membres du groupe de travail doivent faire une demande d'accès.

Clause de non responsabilité

Le producteur et le diffuseur des données ne sont pas responsables des interprétations ou déductions faites à partir de l'utilisation des données.

FORMATS DES DONNEES

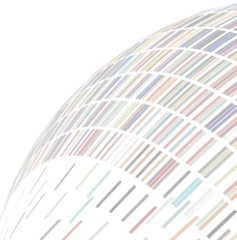
- Nous acceptons tous les formats, mais privilégions les formats ouverts.
- Nous transformons et diffusons toutes les enquêtes sous les 4 formats les plus utilisés actuellement dans les SHS:
 - CSV
 - STATA
 - SPSS
 - SAS

Comment déposer ses données

- Le QPD n'est pas un auto-dépôt, vous transmettez les données et la documentation et ce sont les ingénieurs de l'Adisp qui les déposent.
- Pour lancer la procédure de dépôt, ou pour toute question liée, contactez l'équipe de l'Adisp : diffusion.adisp@cnrs.fr
- Un rdv zoom pour échanger autour de vos données

Nous vous demanderons ensuite de préparer les données et une documentation suffisante pour les décrire :

- Consultez un document téléchargeable « Quelques conseils » en vue de dépôt des données pour vérifier le respect des règles de nommage des fichiers de données et des variables;
- Remplissez la notice téléchargeable en ligne qui renseigne sur votre recherche et vos fichiers de données ;
- Etablissez un dictionnaire des variables;
- Fournissez tout autre document susceptible d'apporter de l'information supplémentaire sur votre projet de recherche (liste complète de sources, une présentation du projet de recherche, publications etc)
- Envoyez vos données par un lien sécurisé (FileSender)
- Nous procéderons ensuite à la rédaction et la signature d'une convention de diffusion entre l'institution de l'équipe de recherche ou chercheur et le CNRS
- Nous intégrerons vos données et métadonnées sur nos sites et portail Quetelet-Progedo Diffusion et les ouvrirons à la diffusion. Nous faisons la promotion des nouvelles enquêtes diffusées via le Twitter et le carnet d'hypothèses.



Merci pour votre attention.

