



Réseau Quetelet

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

The 2006 French Census

A New Collection, A New Dissemination.

What Place for a Data Archive?



Réseau Quetelet

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

- The situation of statistical data dissemination for academic researchers in France
- The case of the population census data dissemination before
- The renewed census
- The new situation for data dissemination



Réseau Quetelet

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

The French context (1)

A recent history

1946 Roper Institute, Univ. Connecticut

1962 ICPSR, Univ. Michigan

1967 UK Data Archive, Univ. Essex

but ...

1981 CIDSP, IEP Grenoble

1986 CES, then LASMAS, CNRS

1st convention between INSEE and LASMAS

and ...

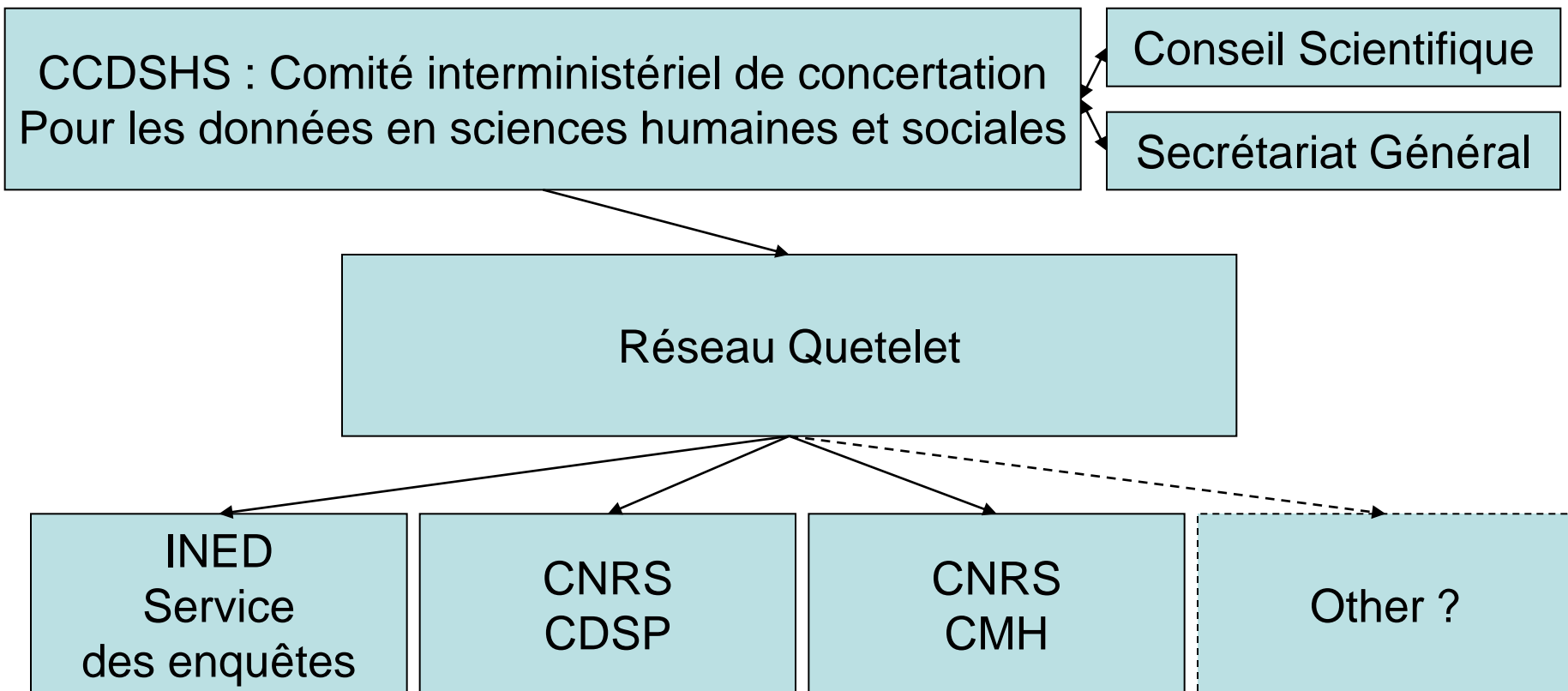
2001 CCDSHS

Réseau Quetelet

3rd convention between INSEE and Ministry of Education

The French context (2)

A complex structure



The French context (3)

INSEE, the national statistical institute

A service of the Ministry of Economy and Finance

A statistical institute

it manages national registers of individuals and enterprises

it carries out major surveys of households and enterprises

population censuses

it collects administrative files

A research institute

A statistical school (ENSAE ...)

most statisticians of statistical services of ministries come from

the INSEE



Réseau Quetelet

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

The French context (4)

The CNIL

(Commission Nationale de l'Informatique et des Libertés)

Created in 1978

Defines the rules for collecting, processing, storing and **accessing** to personal data

Provides more stringent requirements for sensitive data

Controls the dissemination of data by INSEE,
particularly of census data

- exhaustive file

- finely localized data

- sensitive variables

- symbolic role



Réseau Quetelet

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

The census dissemination before (1)

Some history

Until the 1982 census

- a golden age for researchers ...

- ... who could buy the census files and work on IBM big systems

The 1990 census

- the first restrictions (no more access to census tracks)

The 1999 census

- restrictions on dissemination of the microdata files

- new rules for using spatial units

 - a new elementary zoning (IRIS 2000)

- a new status for some variables (sensitive variables)

The census dissemination before (2)

The products

aggregate data

« Analyses » tables and « Profils » tables, by IRIS

« Mobilités » tables, by couples of communes

... and some other tables

microdata files

1/1 housing file

1/20 individual file (with rough sensitive variables)

1/4 individual file (with a rough zoning, without sensitive variables and without possibility of recoding new household variables)

The census dissemination before (3)

The CMH

disseminates

- the collections of standard aggregate tables

- new aggregate tables made at the request of researchers

- the microdata files (complete or extraction)

produces tabulations

- on microdata files conserved in the center

- on microdata files conserved at INSEE

 - who cannot be diffused outside of INSEE

 - who contains all the variables

 - (PSM : Produit Sur Mesure)

The census dissemination before (4)

The CMH

disseminates also auxiliary files

BDCOM *administrative and study zonings*

Correspondances *correspondences adresses-IRIS*

Contours *IRIS basemap*

produces tabulations on microdata files of former censuses

conserved in the center

conserved at INSEE



Réseau Quetelet

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

The renewed census (1)

Le recensement rénové de la population (RRP)

Why ?

To smoothe the costs

To reduce the interval

 between 2 censuses

 between the collection and the dissemination

To do each year a census year

... and some other reasons



Réseau Quetelet

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

The renewed census (2)

How ?

An annual survey

An annual combination of 5 annual surveys

The use of administrative registers and files

Different samplings according to the size of communes and the type of buildings and populations



Réseau Quetelet

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

The renewed census (3)

L'enquête annuelle de recensement (EAR)

each year a survey

but of a part of the population

In the communes under 10 000 inhabitants :

the communes are divided in 5 groups

each group is representative of the little communes of the region

each year

a survey of 1 group of communes

a survey of all housings

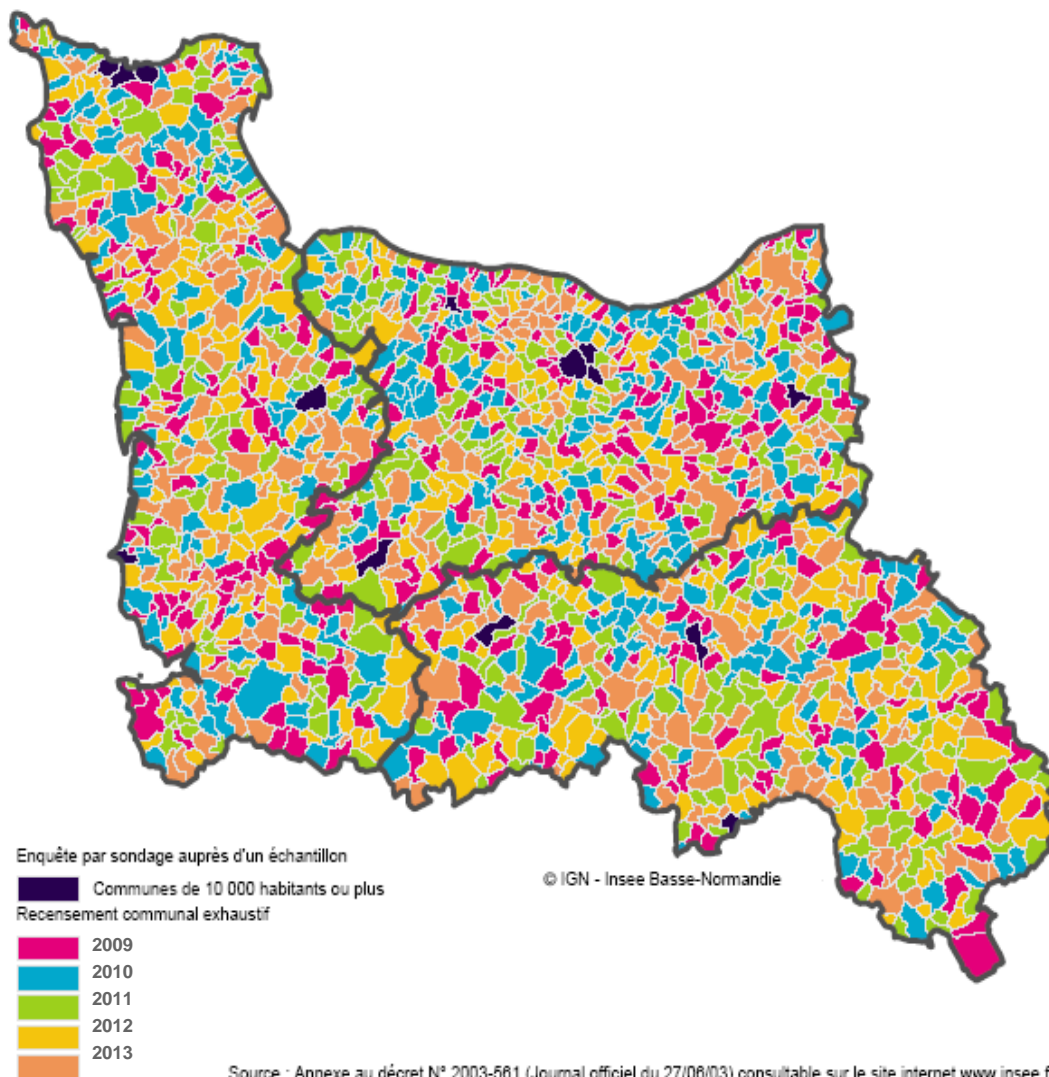
→ a 1/5 sampling

after 5 year (5 surveys for the 5 groups)

→ a 1/1 sampling



Les 5 groupes de rotation





Réseau Quetelet

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

The renewed census (4)

L'enquête annuelle de recensement (EAR)

In communes over 10 000 inhabitants :

- the housings of each commune are divided in 5 groups
- each group is representative of the IRIS of the commune
- each year

 - a survey of 1 group of IRIS

 - a survey of 8% of the housings of the IRIS

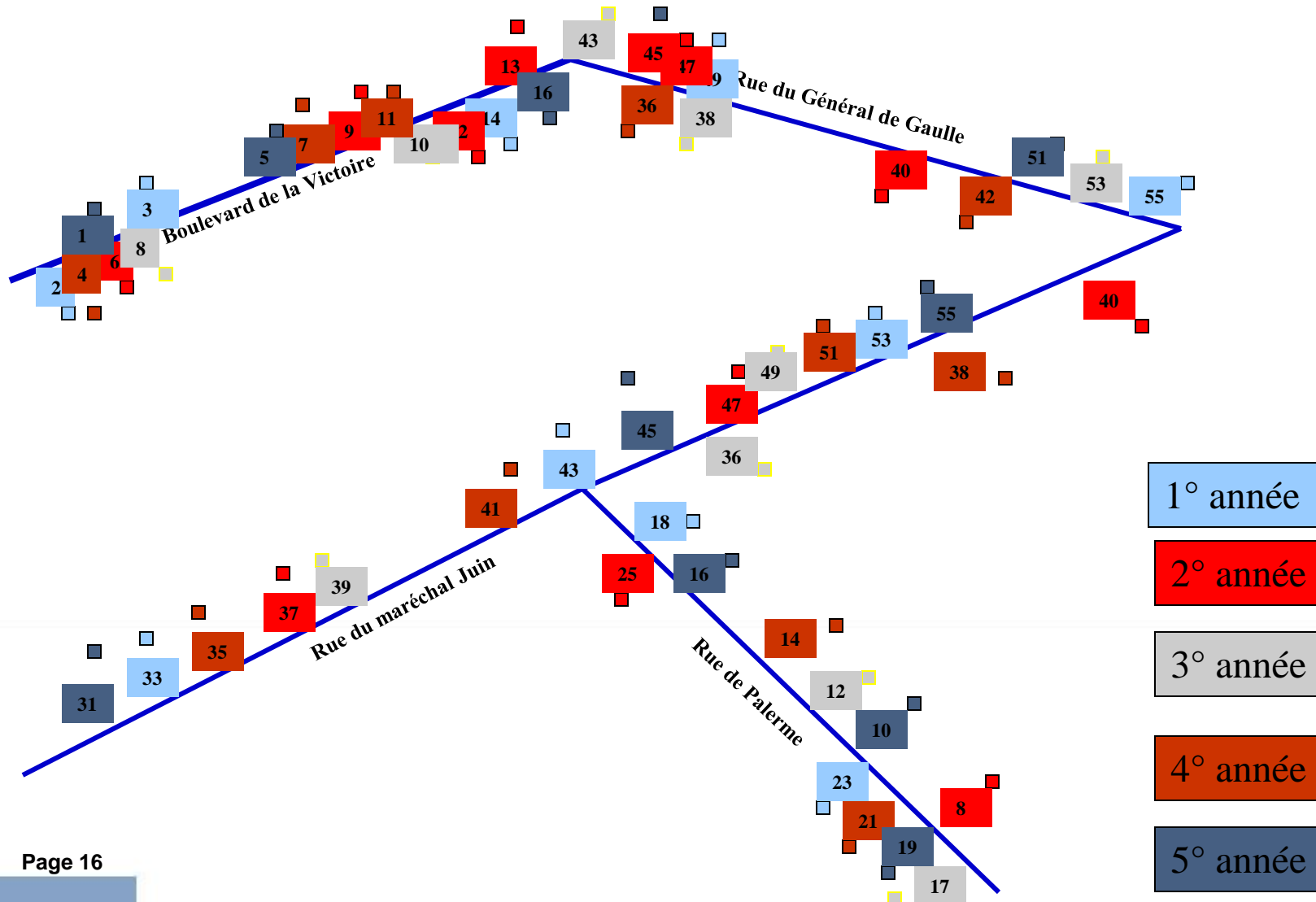
 - a 8 % sampling

- after 5 year (5 surveys for the 5 groups)

 - a 40 % sampling

Actually, it's more complicated (large housings, new housings ...)

Les 5 groupes de rotation dans les communes de plus de 10 000





Réseau Quetelet

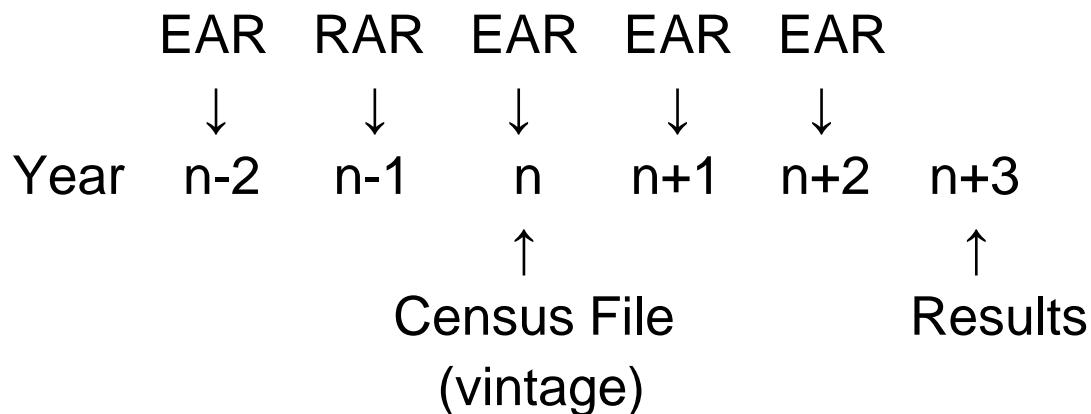
Réseau français des centres de données pour les sciences sociales
French Data Archives for social sciences

The renewed census (5)

Le fichier de recensement (Census file)

Each year

5 successive EAR are combined in 1 census file
with weights coming from administrative files





Réseau Quetelet

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

The renewed census (6)

Le fichier de recensement (Census file)

Actually, not 1 census file, but 2 census files

- a principal file

 - with only the original variables

 - with all the surveyed individuals

 - final sampling 65 %

- a complementary file

 - with all the variables (original and calculated)

 - with 25 % of the individuals of the little communes

 - final sampling 33 %

The renewed census (7)

The diffusion by INSEE

4 types of products for each census file

- aggregated tables at all the geographical levels

- detailed tables at geographical levels $\geq 2\ 000$ inhabitants

- aggregated detailed database at commune and IRIS levels

- 8 microdata files

 - 1 housing file

 - 2 individual files at the place of residence (region, canton)

 - 5 individual files for mobility (residence, work, study)

All the products can be downloaded from the INSEE website

Possibility of customized tabulations (PSM)



Réseau Quetelet

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

The renewed census (8)

The diffusion by INSEE

All is perfect ?

Constraints

geographical levels under IRIS

sensitive variables « variables à diffusion restreinte » (ethnicity)

... but it is the law

aggregated tables and microdata files with

not enough variables

nomenclatures not enough detailed for researchers

The role of CMH in dissemination

6 months ago

*All the census databases available on the INSEE website
availables on the CMH website (but in SAS, SPSS and Stata
formats)*

*All the census microdata files available on the INSEE website
availables on the CMH website (but in SAS, SPSS and Stata
formats)*

The role of CMH in dissemination (2)

Specific microdata files for researchers (FPR)

- More variables

- More detailed variables

- Signature of a specific licence

- Commitment to destroy the files after the study

2 kinds of FPR microdata files

- Census File (each year ... 1 year out of 5 ?)

- EAR File (1 year out of 2 ... 1 year out of 5 ?)

The role of CMH in dissemination (3)

PSM tabulations (Produit Sur Mesure)

- we pay the cost (250 € by tabulation)

- we write the program of tabulation (in SAS) with the user

- we contact INSEE and we receive the tabulation

The advantage of these customized tabulations

- to work on complete files

 - all the variables and the most detailed variables

- to get information you didn't find in standard tabulations and files



Réseau Quetelet

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

The role of CMH in dissemination (4)

CASD (Centres d'Accès Sécurisés à Distance)

secure remote access center

A new service

Organized by INSEE

Tested from 2009

With the collaboration of the Réseau Quetelet



Réseau Quetelet

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

The role of CMH in dissemination (5)

CASD (Centres d'Accès Sécurisés à Distance)

The architecture

- A server IN GENES (a part of INSEE)

- An SDbox on a computer of the user institution

- A secure link

The rules

- To work only on the requires files

 - the possible mergings are realised by INSEE

- The only outputs

 - anonymised tables

 - results of statistical analysis



Réseau Quetelet

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

The role of CMH in dissemination (6)

CASD (Centres d'Accès Sécurisés à Distance)

How to access ? You need to

- Have an SDbox in your academic institution

- Write a project

 - justifying your need for the required files and variables

- Present your project to the Comité du Secret

- Declare your project to the CNIL

- Contact the CASD ...

- ... and wait your turn !



Réseau Quetelet

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

The role of CMH in dissemination (7)

CASD and CMH are working together

To guide the users toward the good service

standard tables and files

FPR files

PSM tabulations

CASD

To look for the good files and documentation in the INSEE archives

To control the programs and the results

The role of CMH in dissemination (8)

6 months ago ... but today ?

Standard tables and files : OK

CASD : OK

PSM : new rules for confidentiality

not only for the last census, but also for previous censuses

all the demands are stopped since 5 months

always in discussion

FPR files

no file received

what files ?



Réseau Quetelet

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

Thank you
for your attention