

# Providing Access to Quantitative Surveys for Social Research: The Example of INED

**Arianna Caporali, Amandine Morisset, Stéphane Legleye**

IN **POPULATION** 2015/3 Vol. 70 , PAGES 537 TO 566

PUBLISHER **INED ÉDITIONS**

ISSN 1634-2941

ISBN 9782733210567

DOI 10.3917/popu.1503.0567

Uploaded: 01/21/2016

Article available online at

<https://shs.cairn.info/journal-population-2015-3-page-537?lang=en>



Discover the contents of this issue, follow the journal by email, subscribe...  
Scan this QR code to access the page for this issue on Cairn.info.



**Electronic distribution Cairn.info for Ined Éditions.**

You are authorized to reproduce this article within the limits of the terms of use of Cairn.info or, where applicable, the terms and conditions of the license subscribed to by your institution. Details and conditions can be found at [cairn.info/copyright](http://cairn.info/copyright).

Unless otherwise provided by law, the digital use of these resources for educational purposes is subject to authorization by the Publisher or, where applicable, by the collective management organization authorized for this purpose. This is particularly the case in France with the CFC, which is the approved organization in this area.



Arianna CAPORALI,\* Amandine MORISSET,\* Stéphane LEGLEYE\*

---

## Providing Access to Quantitative Surveys for Social Research: The Example of INED

*At a time of widespread debate on the future of open data, it is worth taking a closer look at the production, archiving, dissemination and sharing of survey datasets in order to pinpoint the challenges and potential benefits of broader access to quantitative social science data. After recounting the early days of data archiving and the creation of archival networks in various countries and pioneering institutions, Arianna CAPORALI, Amandine MORISSET and Stéphane LEGLEYE describe in detail the progressive implementation of a structured archiving policy at the French Institute for Demographic Studies (INED) and the dissemination of survey datasets via the Réseau Quetelet. A time-consuming, invisible and sometimes unrecognized task, the rational organization of survey data files and their accompanying documentation is key to ensuring that the data are made widely available for secondary analyses, and that the results obtained are scientifically valid.*

### *The importance of sharing survey data*

Activities to facilitate access to survey data are of key importance for the social sciences. It is thanks to such efforts that surveys can be shared between the teams involved in their design and with the broader research community. Sharing survey data is essential for the social sciences as it provides opportunities for secondary analyses, and for the testing and replication of studies. This enables researchers to understand, evaluate and build upon existing research, thereby contributing to the progress of their disciplines (King, 1995, 2006; ICPSR, 2012; Silberman, 1999). Sharing survey data also discourages scientific fraud and is useful for teaching analytical methods. It benefits survey producers by raising awareness of their work through citation, provides justification, through data re-use, of the high costs of surveys, and allows for further testing of data collection methodologies (Silberman, 1999).

---

\* French Institute for Demographic Studies, Paris.

Correspondence: Arianna Caporali, Institut national d'études démographiques, 133 boulevard Davout, 75980 Paris Cedex 20, email: arianna.caporali@ined.fr

Data sharing activities concern both data producers and survey data archives (ICPSR, 2012). Producers process the data for subsequent use and create coherent data files. They also prepare survey documentation, grant access to data files and archive them to assure their reusability. Survey data archives, i.e. archives that mainly deal with data at the individual (micro) level, serve the entire research community, and must provide accompanying documentation that is as exhaustive as possible. The role of data archives is to review data quality, create exhaustive metadata records, publish data and metadata files in online catalogues, manage users' requests for data access, provide assistance to users and liaise with data producers to report on survey data use (ICPSR, 2012).<sup>(1)</sup>

Data documentation is paramount to data sharing, because “without adequate documentation, scholars often have trouble replicating their own results months later” (King, 1995, p. 444). However, there is generally no specific budget for the preparation of metadata, even though produced data can be used by other research teams, avoiding the need to duplicate data collection operations. Moreover, researchers are reluctant to dedicate much time to this activity. The majority hastily assemble the documentation just before depositing the data in an archive. Often “researchers still wrestle with the idea that others may benefit from using their painstakingly gathered datasets and, perhaps even more important, they also fear that by making their datasets public, mistakes in the collection or processing of their data and their (to be) published results may be discovered” (De Moor and Van Zanden, 2008, p. 68). At the same time, in most western countries, the practice of depositing data has become a mandatory requirement of funding agencies, and a growing number of journals require authors to make available all data referenced in an article (De Moor and Van Zanden, 2008; Mochmann and Vardigan, 2011). Nonetheless, archivists who assemble the documentation and manage data access are often confronted with issues of metadata availability.

This article acknowledges the importance of data sharing and describes the archival activities carried out by the Surveys Department (SES)<sup>(2)</sup> of the French Institute for Demographic Studies (Institut national d'études démographiques, INED) to provide access to INED surveys. These surveys consist of quantitative individual (micro) level data produced for non-commercial purposes by researchers with public funds, often with the collaboration of other public bodies. It examines INED's activities in both the international and French contexts of access to quantitative social science survey data (hereafter

(1) Survey data archives differ from institutional repositories, such as depository and academic libraries, public record offices (which store administrative documents) and, in general, from memory institutions (which include photographic, sound and image archives, official statistics, etc.) (Doorn and Tjalsma, 2007). The role of these institutions is to preserve their collections. Data archives, on the other hand, aim at disseminating the data and at facilitating their immediate re-use (Silberman, 1999).

(2) The acronyms used in the article and their English translations, where applicable, are listed in Appendix 1.

referred to as survey data). It does not cover other types of social science research data that fall beyond its scope, such as qualitative survey data, electronic texts, linguistic corpora, historical and archeological data, administrative data, or data produced for commercial purposes.

The article begins by tracing the origins of survey data archives. It then illustrates how international standards for documenting survey data were established and describes the development of survey data archives and the provisions regulating access to social science data in France. Last, it reviews the development and the current organization of the activities that provide access to survey data at INED. It ends with some reflections on future developments in archiving and sharing of survey data in the social sciences.

## I. Origins and development of survey data archives for the social sciences

The development of survey data archives started after the Second World War on the initiative of researchers in the political sciences.<sup>(3)</sup> The post-war geo-political context encouraged international comparative studies in this field and revealed a need for greater sharing of survey data (Bisco, 1966; Doorn and Tjalsma, 2007; Silberman, 1999). These archives allowed “the institutionalization of data sharing” (Silberman, 1999, p. 26), because they gave structure to an activity that was previously performed on a mainly informal basis.

UNESCO played a key role in this development. It promoted a debate on the costs and benefits of archiving and encouraged the creation of data archives (Scheuch, 2003; Silberman, 1999).<sup>(4)</sup> Most of the early archives originated in the academic world. They were developed in contexts where researchers could carry out large-scale surveys with state financial support, and where official statistics offered insufficient data for social research and/or were difficult to access (Silberman, 1999).<sup>(5)</sup> With a remit to provide access to anonymized<sup>(6)</sup> survey data produced by researchers, they also took on the role of managing

(3) The first survey data archive, named Elmo Roper's Public Opinion Research Center, was created in the USA, at Williams College (Massachusetts) in 1947 (Bisco, 1966; Doorn and Tjalsma, 2007; Hastings, 1964; Silberman, 1999). In Europe, the Zentralarchiv was founded at the University of Cologne in 1960 to gather data from research institutes in the Federal Republic of Germany (Bisco, 1966; Scheuch, 2003).

(4) The International Social Science Council (ISSC) of UNESCO held international conferences on research data archiving. The first three conferences were held in the 1960s and focused mainly on archiving survey data. Subsequent meetings organized in the 1970s focused on the need to develop international networks of data archives (Section II). After the 1970s, UNESCO focused on the legal aspects of data archiving (Scheuch, 2003).

(5) The relationship between the development of survey data archives and official statistics is complex and varies across countries. Silberman (1999) offers an overview for Canada, France, Germany, Great Britain, and USA.

(6) Anonymized surveys consist of data files where direct identifiers (e.g. name and address) are removed and indirect identifiers (e.g. geographic location, detailed occupation) are provided in aggregated forms (ICPSR, 2012).

surveys from national statistical institutes (Silberman, 1999). In France, as we shall see (Section III), the development of these archives was delayed for institutional and legal reasons.

Since the 1980s, progress in information technology has led to the creation of internet portals that make it easier to find distributed resources (Doorn and Tjalsma, 2007). To facilitate replication of studies, some journals have adopted a “data availability policy” (DAP), which requires authors to deposit the data used in their articles (De Moor and Van Zanden, 2008; King, 2006). Archives have also increasingly acquired qualitative data (Corti, 2000; Duchesne and Garcia, 2014), and developed systems to allow for secure access to very detailed data<sup>(7)</sup> in compliance with the principle of statistical confidentiality (Silberman, 2011; Le Gléau and Royer, 2011).

## II. International networks of survey data archives and data documentation standards

International networks were rapidly set up to coordinate the first survey data archives. The conferences organized by UNESCO in the 1960s and 1970s (Footnote 4) played a role in their development (Bisco, 1966; Rokkan and Scheuch, 1963; Rokkan, 1966; Scheuch, 2003). For example, the Council of European Social Science Data Archives (CESSDA)<sup>(8)</sup> was created in 1976 with the aim of exchanging data and technologies and developing comparative research (Doorn and Tjalsma, 2007; Marker, 2013; Scheuch, 2003; Silberman, 1999). In 2013 it became a European research infrastructure with a legal status.<sup>(9)</sup> The International Association for Social Science Information Services and Technology (IASSIST), an organization of individual data archivists, has held annual conferences since the mid-1970s<sup>(10)</sup> (O’Neill Adams, 2006).

The international conferences organized by these networks led to the definition of international standards for documenting metadata (Scheuch, 2003; Silberman, 1999) that are essential for developing data access beyond national boundaries (Blank and Rasmussen, 2004; Rasmussen and Blank, 2007). While standardized documentation was an issue from the outset, major

(7) We define “very detailed data” as disaggregated data (e.g. geographic locations or nationalities) that, if matched, may identify respondents.

(8) [www.cessda.net/](http://www.cessda.net/) The websites cited in this article are listed in Appendix 2.

(9) CESSDA brings together the main survey data archives in Europe, and provides an online catalogue of the data available in these archives. It was put on the European Strategy Forum on Research Infrastructure (ESFRI) Roadmap, and was identified as a candidate to form a European Research Infrastructure Consortium (ERIC) (Marker, 2013). In 2013, CESSDA was established as a limited company under Norwegian law (CESSDA AS).

(10) Among other international networks of survey data archives, in the USA the Council of Social Science Data Archives (CSSDA) was active from 1962 until 1970 as a confederation of institutions aimed at coordinating and disseminating the activities of its members (Bisco, 1966; O’Neill Adams, 2006). At the international level, the International Federation of Data Organizations (IFDO) was founded in 1977 to coordinate worldwide data services (Scheuch, 2003; Silberman, 1999).

progress was made thanks to innovations in information technology. North American and European representatives of survey research and data archives created the Data Documentation Initiative (DDI) in 1995. The DDI project led to “an international XML-based standard for the compilation, presentation, and exchange of documentation for datasets in the social and behavioral sciences” (Vardigan et al., 2008, p. 108). This standard, called DDI, is an updated follow-on of the Standard Study Description (SSD) agreed in Copenhagen in 1980 by CESSDA, and of the OSIRIS standard, developed at the University of Michigan in the 1970s (Marker, 2013). Its primary object was to replace paper codebooks with metadata in electronic human-readable format (Box 1). Today DDI is widely used; “if DDI did not exist, it would be more difficult for systems to ‘talk to’ each other, data analysis would have less metadata available to aid in interpreting data, and there would be continuous ‘reinvention of the wheel’” (Wackerow and Vardigan, 2013, p. 163). DDI is supported by IASSIST members and it is recommended by CESSDA. Its usage has been facilitated by the creation of the user-friendly software Nesstar (Networked Social Science Tools and Resources, Box 2) for online publication of data and metadata (Vardigan et al., 2008), a requirement for inclusion in the CESSDA catalogue.

### Box 1. Data Documentation Initiative (DDI)

The Data Documentation Initiative (DDI, [www.ddialliance.org/](http://www.ddialliance.org/)) is an international standard for documenting social science surveys. It is based on eXtensible Markup Language (XML) schemas, and it comprises a group of items for the general description of empirical studies down to the level of each single variable of a dataset. The XML format presents the advantage of being a machine-actionable format. It is robust, uniform and extensible, and is supported by numerous applications (Mochmann and Vardigan, 2011; Leighton, 2002; Vardigan et al., 2008). Moreover DDI is compatible with other metadata standards such as Dublin Core (a bibliographic standard). DDI version 1.0 was launched in 2000. Today, there are two lines of development of the DDI specification:

- DDI-Codebook (DDI-C or DDI 2), at version 2.5. as of November 2014, was introduced in 2002 to document simple survey data. It is document-centric and focused on the elements of a traditional codebook. This specification is widely used, especially because it can be implemented with Nesstar user-friendly software (Box 2).
- DDI-Lifecycle (DDI-L or DDI 3), at version 3.2. as of November 2014, was introduced in 2008 and designed to document surveys across their entire life cycle. It can be used from the beginning of a survey project to document any of its phases. It is particularly suited for longitudinal studies because it contains features that allow explicit comparisons between items of different waves (Hansen et al., 2011; Kramer et al., 2011). Furthermore, a dialogue has been initiated between the developers of DDI and SDMX (Statistical Data and Metadata Exchange; Gregory and Heus, 2007), the standard used to document aggregated data by the international statistical institutes of Eurostat and the OECD among others. These developers are collaborating to develop compatibility between the two standards (Data without Boundaries, 2013).

### Box 2. Nesstar (Networked Social Science Tools and Resources)

Developed in the second half of the 1990s in the context of CESSDA expert seminars and of European Union (EU) funded projects (Marker, 2013), Nesstar is currently managed by Norwegian Social Science Data Services (NSD). It is user-friendly software for managing documentation in DDI-C format without having to know the XML language. It can also be used to publish data and metadata on the Internet when the software is linked to a server tool. Nesstar provides a convenient and intuitive means for users to search, browse and explore metadata, and to visualize and analyse the corresponding data online. It is possible to access documentation at the level of each variable (including a frequency table), create personalized cross-tabulations and graphs, and even to produce simple analyses (correlations, regressions). A data weighting tool is also available to ensure correct interpretation of results and analyses. Another attractive feature of Nesstar is its search tool that can be used for simple or advanced searches in the documentation and variables. Finally, survey documentation and analysis results can be exported in various formats. A complete explanation of these functionalities is available on the Nesstar website ([www.nesstar.com/](http://www.nesstar.com/)). Although other software supporting the DDI standard has been developed (for a complete list see DDI website), Nesstar is the most widely used among European survey data archives.

## III. The French context: making up for lost time

### *Survey data archives and regulations on data access up to the twentieth century*

Although some French researchers took part in early discussions about sharing survey data, no survey data archives were created in France before the 1980s (Silberman, 1999). There were two main reasons for this. First, there was little institutional and structural support in France for the large-scale university surveys which, as discussed above (Section I), prompted the development of survey data archives in other countries. Second, the surveys of the National Institute of Statistics and Economic Studies (Institut national de la statistique et des études économiques, INSEE) covered a large spectrum of topics, and French researchers could use the aggregated data published by the Institute. It was also possible to access its individual-level data files,<sup>(11)</sup> but this was difficult due to a legal framework which placed strong emphasis on personal data protection.

Individual-level data were protected under two separate laws. First, the Act of 7 June 1951 prohibited the communication of individual data collected

(11) INSEE also took part in international academic projects, such as the 1966 Time Use Survey, and granted access to its surveys (Chenu, 2011; Szalai, 1972). This access was managed by the INSEE departments or, at that time, by INSEE's regional economic observatories (B. Riandey, personal communication, 13 November 2014). INSEE's surveys were archived internally and transmitted to the French national archives (R. Silberman, personal communication, 26 November, 2014).

by official statistics (INSEE, and Ministerial Statistical Departments).<sup>(12)</sup> These data could be disclosed only in aggregated forms and in anonymized files. From 1984, some exceptions were granted, but exclusively for data on companies, with the creation of a committee for confidentiality of corporate statistics (Comité du secret statistique concernant les entreprises)<sup>(13)</sup> (Gaeremynck, 2009; Silberman, 2011). Second, the Data Protection Act of 6 January 1978 established that, whenever identification was possible, personal data could be collected and processed only within a limited time, after notification of the French Data Protection Authority (Commission nationale de l'informatique et des libertés, CNIL),<sup>(14)</sup> and for a specific purpose. This Act, which grew in importance with the development of computer technologies,<sup>(15)</sup> made it difficult to re-use personal data for purposes other than those that justified their initial collection (Riandey, 2000; Silberman 1999, 2011).

Against this legal background,<sup>(16)</sup> data archives were first created to facilitate the sharing of anonymized individual data files (Silberman, 1999). A socio-political database (Banque de données sociopolitiques, BDSP) was established at the Institute of Political Studies in Grenoble and incorporated into a centre for computerization of socio-political data (Centre d'informatisation des données sociopolitiques, CIDSP).<sup>(17)</sup> This archive contained socio-political data produced by private and public statistical offices, as well as data produced by academic researchers. Another archive, run by a research unit of the National Centre for Scientific Research (Centre national de la recherche scientifique, CNRS), the Laboratory for Secondary Analysis and Applied Methods in Sociology

(12) This Act confirmed the role of INSEE as coordinator of official data collection, and reiterated the obligation for interviewees to provide accurate answers and for statisticians to safeguard data confidentiality, already established under French law (Lang, 2008). Statistical confidentiality underpinned the bond of trust between the interviewer and interviewee, and penal sanctions applied in case of violation.

(13) Anonymization made it difficult to analyse companies for which certain types of information (e.g. size) were essential. The decree of 17 July 1984 created the Committee for confidentiality of corporate statistics as part of the French Statistical Advisory Committee (Conseil national de l'information statistique, CNIS), with the task of evaluating requests to access data on companies. It considered the scientific relevance of data requests, the applicants' motives, and the reliability of their institution (Gaeremynck, 2009; Silberman, 2011).

(14) A declaration to the CNIL was sufficient in cases of personal data allowing for direct or indirect identification. In cases of sensitive data, however (i.e. ethnic origin, political and religious opinions, health, sexual behaviours), authorization was required.

(15) This Act was drafted in response to the government's SAFARI project for administrative file automation and registration of individuals that was designed to link up all files kept by the French administrative authorities (Riandey, 2000; Silberman, 2011).

(16) Three further legislative texts deserve to be mentioned: 1) the Act of 17 July 1978, which created a committee for access to administrative documents (Commission d'accès aux documents administratifs, CADA), responsible for safeguarding rights of access to these documents, 2) the Archives Act of 3 January 1979, which authorized free access to survey data after 100 years for facts or acts of a private nature, and after 30 years for information of an economic or financial nature, and 3) the manual of tax procedures (Livre des procédures fiscales) which made no provision for access to fiscal documents for research purposes (Silberman, 2011).

(17) The Centre for Socio-Political Data (Centre de données sociopolitiques, CDSP, for more information: <http://cdsp.sciences-po.fr/>) of Science Po Paris took over from the BDSP in 2005 (Chenu, 2011).

(Laboratoire d'analyse secondaire et de méthodes appliquées à la sociologie, LASMAS), which has since become the National Archive of Data from Official Statistics, (Archive de données issues de la statistique publique, ADISP) of the Centre Maurice-Halbwachs (CMH),<sup>(18)</sup> was created in Paris. One of its missions was to provide access to survey data gathered by public statistical bodies. These archives were established on the initiative of individual specialists rather than as a common institutional effort<sup>(19)</sup> (Silberman, 1999).

Access to stored data by researchers was regulated by a number of agreements (Rhein, 2002; Silberman, 1999), such as the one signed between INSEE and the CNRS in 1986 that gave LASMAS the right to disseminate a few anonymized INSEE surveys to all CNRS researchers. CNRS also signed similar conventions with other public data producers, such as the Ministry of Education. French research institutions, such as INED and the French National Institute for Agricultural Research (Institut national de la recherche agronomique, INRA), signed analogous agreements with the public data producers, in order to gain access to their data. Agreements for data produced by researchers and archived at CIDSP were less formal and established on a more individual basis, but licenses for dissemination to researchers were similar. The general principles regulating access to data produced with public funding were clarified in 1994, in the so-called “Balladur Circular”. While access to these data was free of charge, institutions wishing to acquire them had to pay the cost of the service.

In the late 1990s, Claude Allègre, minister of education, research and technology, noted a growing need among social science researchers for access to survey data, and commissioned Roxane Silberman, the director of LASMAS, to conduct a study on French social sciences and their data (Allègre, 1999). Based on a survey among CNRS and university laboratories, Silberman (1999) identified three main groups of problems. The first group concerned data access. Data files in France were not regularly upgraded to the latest computer innovations. Furthermore, some surveys were not fully documented or not documented at all, which made their re-use impossible. Above all, there were no provisions in the agreements between CNRS and INSEE entitling university researchers to access data produced by public data producers, so many scholars had to rely on personal contacts with INSEE or other administrations. There was also uncertainty about university researchers' intellectual property rights and their obligations to provide access to the data they produced. In the absence of any regulation, surveys co-produced by public institutions and universities were not made available by LASMAS. Furthermore, to protect citizens' personal data, an increasing level of anonymization, especially for census micro data files, prevented French social researchers from making detailed analyses. Some

(18) For more information: [www.cmh.ens.fr/greco/adisp.php](http://www.cmh.ens.fr/greco/adisp.php), (Chenu, 2011).

(19) The creation of the BDSP was encouraged by Frédéric Bon. LASMAS was created on the initiative of Alain Degenne and the specialists who had previously worked with Jacqueline Frisch at the Centre d'études sociologiques (centre for sociological research) headed by Raymond Boudon.

of the BDSP studies were protected and could not be accessed. The second group of problems concerned data use. Compared with other countries, French quantitative sociology was less developed and less well-equipped in terms of computers and software for data processing. The third group related to data production. Large-scale surveys produced entirely by researchers in the academic world were still rare in France.

In response to these problems, Silberman (1999) called for the creation of a “purpose-built archival structure” (p.47) and for a major reform of French policy on access to social science data, including legal provision for access to data for research purposes, and measures to make all survey data produced with public funds available for re-use. Experience in other western countries had shown that archives played a key role in stimulating the development of quantitative sociology and in involving researchers in survey production. A large archival structure was needed to harmonize survey documentation and to enable France to play an active role at CESSDA. The Silberman report laid the foundations for the creation of the Centre Quetelet (Chenu, 2011) which today manages access to most of the social science survey data produced in France and to some international surveys (see Box 3).

### *The Réseau Quetelet: creation, objectives and data access*

The Centre Quetelet was set up by decree on 12 February 2001 (Chenu, 2003). Together with the university data platforms (Plateformes universitaires de données, PUD), it implements the policy of the coordinating committee for humanities and social science data (Comité de concertation pour les données en sciences humaines et sociales, CCDSHS). This committee is responsible for defining national public policy on data in the social sciences and humanities, with three main purposes: 1) to facilitate access to data that are useful for research, 2) to develop the use of these data, and 3) to provide support for the production of large-scale surveys in the social sciences. The university data platforms provide local assistance to users and support for survey implementation. The Centre Quetelet is responsible for collecting, managing and archiving social science data and for providing user training in the technical and scientific innovations in this field.

The Centre Quetelet was founded in December 2001 as a CNRS institution in partnership with three other institutions: École des hautes études en sciences sociales (EHESS), INED, and the University of Caen (Arduin, 2004; Chenu, 2003; Riandey, 2003). It had three founding members, BDSP (current CDSP, Footnote 17), which provides socio-political surveys; LASMAS (current ADISP), responsible, in particular, for data from public data producers; and INED through its Surveys Department, which provides the socio-demographic surveys produced by INED since its creation in 1945 (Box 4 and Section IV). In 2005, the Centre Quetelet was transformed into a network with the name of Réseau Quetelet (Chenu, 2011). It is a member of CESSDA and, since 2013, has been

### Box 3. Examples of options for accessing French quantitative surveys in the social sciences

As of November 2014, there are four main options for accessing quantitative survey data in the French social sciences. They provide information with varying levels of detail:

- The INSEE website ([www.insee.fr/](http://www.insee.fr/)) offers highly anonymized individual data files that can be downloaded without making a formal request. It is also possible to ask for custom data tabulations that include only specific (anonymized) variables.
- If more detailed data are needed, researchers can make a formal request to the Réseau Quetelet for access to surveys for scientific use (Section III). More than 1,100 references are available, including socio-political surveys from CDSP (e.g. post-electoral surveys), socio-demographic surveys by INED (Section IV), and official statistics from CMH-ADISP (e.g. data from INSEE, and the statistical departments of French ministries). ADISP may also prepare custom data tabulations in cooperation with INSEE.
- Another partner of the Réseau Quetelet, CASD/GENES (Section III), manages access to very detailed data from INSEE and French ministries in particular. On request, CASD can also prepare files linking different sources.
- For researchers wishing to study France in international surveys, such as the Survey of Health, Ageing and Retirement in Europe (SHARE), the European Social Survey (ESS), the Generations and Gender Surveys (GGS), access may be sought through the research infrastructures and the institutions that manage these data.

This system of organization means that a single survey (e.g. the Labour Force Survey) can be obtained from different access points. Of course, these options do not cover all possible modes of access to French quantitative surveys (Silberman, 2011). The French Longitudinal Study of Children (Étude Longitudinale Française depuis l'Enfance, ELFE) follows a specific protocol (for more information: <https://pandora.vjf.inserm.fr/public/>).

the French service provider within the new CESSDA set up as a European research infrastructure with a legal status (see Footnote 9). It is also a member of PROGEDO, the French social science research infrastructure which brings together the main actors involved in quantitative surveys.

The Réseau Quetelet has three main missions: data documentation, data access and data dissemination (Arduin, 2004). Its partners are responsible for gathering all the information available about each archived survey and of restructuring it in accordance with international data documentation standards. The Réseau Quetelet prepares and sends the datasets to users in compliance with deontological rules. It also advises researchers on how to use datasets and informs them about data availability and the latest technological innovations in software and methods. It maintains ties between users and data producers at both national and international levels.

### Box 4. INED Surveys Department (SES)

INED has built up a specialized Surveys Department which oversees data collection for projects conducted by INED researchers and, increasingly, in partnership with other institutions (INSEE, INSERM, etc.). It has three main missions. Its first mission centres on quantitative and qualitative data collection, ranging from questionnaire design to final data presentation. The Surveys Department offers expert know-how at all stages of the data collection process, and is closely involved in many large-scale projects both in France and abroad. Second, the department is actively involved in methodological research and in assessing innovative survey techniques and protocols in the social sciences. Its third mission is to make INED quantitative surveys data available to users.

In order to accomplish its missions, the Réseau Quetelet runs a website<sup>(20)</sup> that catalogues and documents all the data provided by its partners. In line with CESSDA recommendations, each of its partners has adopted DDI as a standard to document the data, implemented through the Nesstar software. Efforts are being made to provide survey metadata both in French and English. Some international surveys, such as the INED survey Migrations between Africa and Europe (MAFE), are documented in English only (Box 5).

Access to data managed by the Réseau Quetelet is regulated through two types of conventions: one with data providers (its partners) and the other one with data users. The general principle is that access is provided free of charge and exclusively for non-commercial research purposes. Access procedures are available both in French and in English. They may vary depending on the status of users, their institutional affiliation, and the nature of the requested data files. While procedures are open to users of all nationalities, more detailed information (e.g. an extended description of research projects) may be required from users who do not belong to a French university or public research institution. Data requests can be submitted via the Réseau Quetelet website which was overhauled in March 2014. Several datasets from different data producers can now be requested in a single application. Users undertake to comply with a set of rules.<sup>(21)</sup> The mean waiting time for access to data may vary from a couple of days to several weeks.

Soon after the creation of the Réseau Quetelet, the legislation on data protection was modified. Increasing requirements for data anonymization were making it more difficult to analyse French statistics. Spatial variables and sensitive variables (especially on nationality and country of birth) were being provided in increasingly aggregated forms (Riandey, 2000; Silberman, 2011). First concerning personal data, the 1978 Data Protection Act was reformed in

(20) [www.reseau-quetelet.cnrs.fr/](http://www.reseau-quetelet.cnrs.fr/)

(21) The main rules are: use data exclusively for research purposes; ensure that the data files are not damaged; refrain from giving these files to third parties; process data in accordance with scientific methods; present the results of analyses so as to prevent any identification of respondents; use the received data as fully as possible; mention the source of the data in publications.

### Box 5. INED surveys and their main subjects

INED surveys carried out since the 1980s include:

- Couple formation, 1983-1984
- Family situations, 1985 (with INSEE)
- 3B bis survey - Family, work and migration event histories, 1988-1989 (with the Catholic University of Louvain)
- Local family circle, 1990
- Geographical mobility and social integration (MGIS), 1992 (with INSEE)
- Analysis of sexual behaviour in France, 1992 (with INSERM)
- Transition to adulthood, 1993-1994
- Survey on homeless people in Paris, 1994-1995
- Family situation and employment, 1994 (with INSEE)
- Survey on outcomes of children born outside marriage, 1996-1997
- National survey on violence against women in France, 2000
- Cystic fibrosis observatory in France, 2000-2007
- Handicap, disability and dependence in prisons (HID-prison), 2001 (with INSEE)
- Fertility intentions (3 waves), 1998, 2001, 2003 (with INSEE)
- Adoption survey in 10 *départements*, 2003-2004
- Families and employers, 2004-2005 (with INSEE)
- Generations and Gender Survey (GGS), an international survey whose French component has been entrusted to INED (with INSEE): Survey of family and intergenerational relationships (ERFI), 2005, 2008, 2011
- Context of sexuality in France, 2006 (with INSERM)
- Migration between Africa and Europe (MAFE), a major research initiative bringing together 10 European and African research centres, 2008-2010
- Migrations - Family - Ageing in the overseas *départements*, 2009-2010 (with INSEE)

INED has also developed demographic surveillance systems in West Africa, notably in Senegal since 1970 (Bandafassi, Mlomp), and in Mali since 1989 (Tominian Circle).

2004 (Act no. 2004-801 of 6 August), following the European Parliament and Council Directive 95/46/EC of 24 October 1995 – effective in member states since 1998 – on the protection of individuals with regard to the processing of personal data and on the free movement of such data. It introduced the possibility of using personal data collected for other purposes to carry out statistical or historical research (Silberman, 2011).<sup>(22)</sup> The re-use of personal data that might lead to direct or indirect identification and/or that was of a sensitive nature had to be submitted to the CNIL (Footnote 14). Second, with regard to official statistics,

(22) The European Council's 1981 Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data made provision for access to data for research purposes. France opted not to implement it, however. In 1994, access was authorized for medical research, subject to approval by the committee for personal data protection in medical research (Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé, CCTIRS) (Silberman, 1999; Silberman, 2011).

the 1951 Act (Footnote 12) was reformed by the Archives Act of 2008 to permit the re-use of very detailed personal data collected by official statistics for statistical, historical and research purposes (Gaeremynck, 2009).<sup>(23)</sup> It is also worth mentioning the Act of 22 July 2013 on higher education and research which opened access to fiscal data for scientific research purposes. As already established for data on companies (Footnote 13), and for data on public bodies (Order no. 2004-280 of 25 March), access to very detailed personal data was subject to the approval of the committee for statistical confidentiality (Comité du secret statistique – the reference to corporate statistics having been dropped from its title).

These developments opened the path for organizing secure access to very detailed data from official statistics. The Secure Data Access Centre (Centre d'accès sécurisé aux données, CASD) was created in 2010 by the Group of National Economics and Statistics Schools (Groupe des écoles nationales d'économie et statistique, GENES), at that time an INSEE department, and it became a partner of the Réseau Quetelet (Le Gléau and Royer, 2011; Silberman, 2011). It also became the French data provider in the European Data Without Boundaries (DWB) project, which supports equal and easy access to official micro-data in Europe (Silberman, 2013). Contrary to the access provided by the other partners of the Réseau Quetelet, access to the CASD is not free of charge. Moreover, the procedure involves obtaining approval from the committee for statistical confidentiality, the data producers and, in the case of personal data, the CNIL.<sup>(24)</sup> It can take several weeks, depending on the number of authorizations required.

#### IV. The case of INED: the triple role of data producer, user and provider

##### *From archiving to providing access to surveys*

Since its creation in 1945, INED has carried out numerous socio-demographic surveys, sometimes in collaboration with other public bodies (Box 5). For many years, access to these surveys was managed in an informal manner.<sup>(25)</sup> From the 1970s, surveys and their documentation were archived and transmitted to the French national archives (Archives nationales) thanks to the work of Suzanne Helgoual'ch and Henri Bastide (Comité d'archivage de l'Ined, 2001a).

(23) The 2008 Archives Act also shortened the period of restricted access to survey data from 100 to 75 years for facts and behaviour of a private nature, and from 30 to 25 years for information of an economic and financial nature.

(24) The approval of other bodies may also be necessary. A detailed explanation of the procedure required by the CASD is provided by Le Gléau and Royer (2011) and on the CASD website (<https://casd.eu/>). The CASD has made it possible to satisfy researchers' needs while continuing to respect the principle of statistical confidentiality. The data are stored on a server, cannot be copied, and are accessible for a limited period using a personal password. Anonymization of analysis results is obligatory.

(25) Data requests could come from French as well as foreign researchers, and the procedures were managed either by research units or by the INED Surveys Department (B. Riandey, personal communication, 13 November 2014).

However, these activities were not sufficient to fully accomplish INED's mission of disseminating knowledge across the research community, to the public authorities and the public at large, as laid down in its statutes of 1945 and 1986 (amended in 2001) and in its strategic orientations (INED, 2002).

Initial discussions as to whether and how INED should engage in activities for providing access to its surveys began in the early 1990s. Surveys were becoming more costly and requests to access the data for secondary analyses were increasing (Bozon, 1995). Jacques Magaud, director of INED at that time, commissioned Michel Bozon, then head of the Surveys Department, to draw up a report on the management of access to INED surveys. Bozon consulted various INED researchers and recorded a variety of different, sometimes divergent, proposals, as well as certain misgivings (Bozon, 1995). He concluded that certain principles of data access, previously tacitly understood, needed to be formalized and that INED surveys needed to be made accessible for re-use, accompanied by full documentation.

Defining rules for accessing INED surveys was one of the objectives laid down in INED's strategic orientations for the years 2002-2005. It was important to respect the intellectual property rights of survey producers, whether INED researchers or not, but also to factor in the costs of data access. Priority access was to be reserved for partners who had contributed to survey funding and design. INED, unlike other data archives, only disseminated and provided access to surveys conducted by INED alone or in partnership with other institutions. Except in some special cases, INED was not required to provide access to data files produced by others without its collaboration.

Against this background, and spearheaded by François Héran, INED's director at that time, the Institute was a co-founding partner in the creation of the Réseau Quetelet in December 2001, in which it occupied an intermediate position between a producer of official statistics and an academic research body; INED researchers not only produce surveys but also use them. Specifically, the Surveys Department (Box 4) was given the task of applying the institute's policy on documenting and providing access to its surveys, in close liaison with the Réseau Quetelet (INED, 2002). This policy did not take practical effect until 2004.<sup>(26)</sup> Anonymized data files, along with their documentation, could be requested and accessed by INED and non-INED researchers, provided that they accepted the accompanying terms and conditions and agreed to comply with the rules governing use of the data<sup>(27)</sup> (Service des enquêtes et des sondages de l'Ined, 2004). The data were then sent by post on a CD-ROM.

The activities aimed at providing access to surveys were progressively ramped up (INED, 2006) and, from 2006, under the impetus of François Héran

(26) However, from 2001, some survey files were made available to university demographers for teaching purposes (Comité d'archivage de l'Ined, 2001b).

(27) The conditions are detailed in an internal document (Service des enquêtes et des sondages de l'Ined, 2004) and are similar to those established by the Réseau Quetelet (Footnote 21).

and Cécile Lefèvre, head of the Surveys Department at that time, additional resources were deployed (Comité d'archivage de l'Ined, 2006). These activities included producing documentation and re-formatting data files, and extended beyond the tasks of data archiving per se. As a result of this policy, in 2008, some recent surveys became available before being sent to the National Archives.

In 2009, the Surveys Department was also given the task of providing access to the Generations and Gender Surveys (GGS) that form part of the Generations and Gender Programme (GGP).<sup>(28)</sup> Furthermore, work was undertaken to determine the most appropriate ways of promoting the use of INED surveys, to better understand current and potential users, and hence to identify the surveys that should take priority in terms of user access (Comité d'archivage de l'Ined, 2009). This was made possible by adopting the DDI standard (Box 1) and posting the entire catalogue of INED surveys online using Nesstar software (Box 2).

### *Implementation of DDI and the Nesstar software*

After considering the option of using INSEE's Statistical Data Dictionary (Dictionnaire de données statistiques, DDS) to archive and document surveys, it was decided to adopt the DDI standard in 2003. The DDI rules were screened and tested, along with the functionalities of the Nesstar software liable to facilitate implementation of the DDI-C standard (Comité d'archivage de l'Ined, 2003, 2004). It was not until the end of 2008, however, that work began to transfer the INED survey collection to Nesstar.

With the Nesstar software, also adopted by the other members of CESSDA and the Réseau Quetelet,<sup>(29)</sup> an online catalogue has been generated containing all survey records. The INED-Nesstar Catalogue (Catalogue Nesstar des enquêtes de l'Ined),<sup>(30)</sup> was officially launched in June 2012. It replaces the previous dissemination tool. Since it came online, more than 5,800 users (from 93 countries) have viewed more than 67,700 pages of the catalogue. In 2014, the number of users grew by 160%.

The GGS surveys are managed through another catalogue, called GGP Online Codebook & Analysis, which was launched in 2010, and is available at the GGP homepage.<sup>(31)</sup> This catalogue has been visited by more than 5,100 users (from 74 countries), who viewed about 19.4 pages on average per session. The number of users increased by about 180% in 2014.

(28) The GGP is a pan-European research infrastructure aimed at providing internationally comparable individual-level data on demographic behaviours, along with contextual information on demographic, social, economic, and political macro-conditions. To this end, the GGP complements individual level-data collected through the GGS, a panel survey with a wave every three years, to the contextual data in its database (Vikat et al., 2007; Caporali et al., 2014). For further information: [www.ggp-i.org/](http://www.ggp-i.org/).

(29) The Nesstar software license was acquired by the Réseau Quetelet for all its partners.

(30) <http://nesstar.ined.fr/webview>

(31) [www.ggp-i.org/online-data-analysis.html](http://www.ggp-i.org/online-data-analysis.html)

### *Surveys made available and data access procedures*

As of November 2014, the INED Nesstar catalogue contains 248 references. They cover a large spectrum of socio-demographic themes such as fertility, contraception, sexuality, marriage, migration, immigrants' integration, discrimination, gender, generation, inequality, health, ageing, housing, employment, etc. The surveys in the catalogue are divided into two categories (each organized by decade and by year of production):

- Available surveys (55 references), i.e. whose data file(s) can be supplied to users;
- Unavailable surveys (193 references), i.e. whose data file(s) cannot be supplied to users (either because they do not exist, or because access has not yet been authorized).

For the available surveys, the catalogue offers complete, downloadable data documentation, and enables users to perform and export some basic analyses. The catalogue can be consulted in both French and English. However, for most surveys the documentation is only available in French. Requests for access to data can be submitted via the Réseau Quetelet portal, and access is granted in accordance with its guiding principles. Thanks to the new online catalogue, the number of data access requests is increasing (+8% over the last year).

The GGP Online Data Codebook & Analysis catalogue is entirely in English. As of November 2014, it offers data for two waves of the survey and for 17 countries (Australia, Austria, Belgium, Bulgaria, Czech Republic, Estonia, France, Georgia, Germany, Hungary, Italy, Lithuania, Netherlands, Norway, Poland, Romania, Russian Federation). The data files are regularly updated to the most recent versions of the surveys. Access to the GGP surveys is restricted to researchers and is managed via an online platform by the United Nations Economic Commission for Europe (UNECE). The mean waiting time for access is five days. The conditions are laid down in agreements signed by UNECE and GGP participant countries (see the GGP webpage for more information). This catalogue has also contributed to GGP data dissemination. The number of registered users of GGS micro-data files has grown by 33% over the last year (Generations and Gender Programme, 2014).

### *The work of data preparation and data documentation*

Before making a survey available for re-use, the sometimes “invisible” task of preparing and documenting the data must be performed. This involves gathering as much documentation as possible about both the survey itself and its data file(s). This documentation is often scattered: some documents may be stored in electronic files or in boxes of archives kept at INED, others may come from the research teams or take the form of articles and working papers based on the survey data. There are various possible file formats which do not represent the same workload. For the more recent surveys, documentation

files are in “current” formats, i.e. Word or pdf documents from which information can be extracted and copied, and data files are in SAS, SPSS or Stata formats that do not need to be converted. In the case of old surveys, it is not rare to have only hand- or type-written paper documents, or files in text format, for example (or even no file at all<sup>(32)</sup>). Documents are sometimes incomplete, and variable names may be missing.

Sometimes it is also necessary to select the correct information. There may be a lack of documents or, on the contrary, some information may be duplicated. At this stage, the collaboration of research teams involved in data collection is of crucial importance, both to optimize the gathering and selecting of information and to synthesize it. If they supply exhaustive documents and clean data files, the work of data preparation and documentation can be simplified and surveys can be made available much faster.

Once all the data files and related documentation are ready, they are imported into Nesstar. The aim is to make all the information clear and understandable for all users. Metadata are documented according to the DDI fields chosen by the Surveys Department. These fields are classified in three groups:

- Document description: information on the Nesstar file (survey concerned, author, etc.);
- Study description: information on the survey (summary, researchers, producers, funding agencies, dates of collection, field, methodology, sampling procedure, etc.) and links to the questionnaire, other associated or connected surveys, and bibliography;<sup>(33)</sup>
- Description of data files: information on the datasets themselves (how they are constructed, missing data, notes on which variables are replaced with derived ones, etc.).

In Nesstar, the variable labels and categories are also entered, so datasets that are already documented represent a significant time saving. In addition, every variable is examined in great detail: the texts of the question (text before the question, actual question wording, text after the question, instructions to interviewers), the universe (i.e. the persons to whom the question was asked) and who answered the question (interviewer or respondent). Other information may be added, such as the questionnaire from which the variable is taken (when there are several questionnaires) or, in case of derived variables, the variables and the software programs used in the calculation. To ensure maximum precision, each dataset is entirely reviewed, and variables are reorganized to follow the order of the questionnaire.

(32) In the past, data from questionnaires were recorded on perforated cards. They were then transferred to computer files, but these repeat operations led to loss of information. As a consequence, data from certain surveys will never become accessible.

(33) The online bibliography is a part of the survey metadata. It consists of a bibliographic reference list, updated in collaboration with INED Documentation department.

This work of documentation, verification and harmonization is very time-consuming. A good codebook, and clean and anonymized files speed up the process considerably. A non-anonymized file requires additional work to detect potentially identifying variables which must be recoded or deleted before dissemination.<sup>(34)</sup>

The GGP surveys are documented using the same DDI fields as for INED surveys. The datasets are already anonymized and the variable labels and categories are already specified. Some of the survey metadata are common across all GGP surveys (e.g. how the study should be cited, data content keywords and summary). However, even for these surveys, the preparation of country-specific metadata (i.e. information that varies across harmonized surveys, such as data collection methods and processing, and specific variables) may prove to be time-consuming.

## Conclusion

This article reviewed the development of activities to provide access to quantitative social science surveys, detailing their development and current practices at INED. First, it gave a brief history of data sharing and data archives in social science. The need to organize access to survey data for secondary analysts was first felt in the 1950s in the field of political science. The development of European and international comparative research led to the establishment of national data archives. The article went on to examine the origins of international data archiving networks that fostered the creation of harmonized international standards for documenting metadata. Developed in parallel (and thanks) to the IT revolution, these standards paved the way for online data access beyond national borders via the Internet. The DDI international standards are today recommended by CESSDA and widely used in European archives.

The article then gave an account of the development of social science survey data archives in France, where the establishment of a national data archival structure is more recent. The main factors that explain this delay are a weak academic tradition in large-scale (international) surveys and strong laws to protect personal data confidentiality. The Réseau Quetelet was created in the early 2000s as a structured organization to manage previously dispersed data archives. Today, it centralizes access to most French social science surveys. Moreover, the French legal setting has progressively taken into account researchers' needs by authorizing re-use of personal and very detailed data for

(34) Identifying variables are kept in a separate file which will be archived. To sum up, for a single survey, there may be four main types of data files: raw data files (after data collection); research data files (cleaned and weighted data files not necessarily completely anonymized) for use by the survey teams; dissemination data files (fully anonymized data files); and data files sent to the national archives (which most often correspond to the research data files).

research purposes. In particular, this was made possible thanks to the 2004 reform of the Data Protection Act and the 2008 reform of the Archives Act.

The article also reviewed the development of activities at INED to provide access to surveys. Despite initial discussions in the 1990s about the creation of formal rules for granting access to surveys for research purposes, it was not until 2000 that the institute initiated these activities. The INED Surveys Department, co-founder of the Réseau Quetelet, today offers its expertise in these activities to international projects, such as the GGP. Like all Réseau Quetelet partners, and in line with CESSDA recommendations, INED has implemented the Nesstar software for publishing and exploring data and metadata online. The INED and GGP Nesstar catalogues have contributed to the dissemination of their surveys. Last, the article describes the Surveys Department's activities to prepare the surveys and their documentation before providing access to them. When survey metadata are not structured to the DDI standard, the work of data and metadata preparation may be a time-consuming process. For future surveys, access to data could be speeded up through closer collaboration with data producers in preparing survey metadata based on DDI requirements (ICPSR, 2012; Vardigan et al., 2008).

In the future, the trend towards open data in the social sciences (Silberman, 2013), made possible by advances in information technology, will increase the importance of activities to provide access to social science survey data. Data access has become a transnational issue and action is being taken at the European level to integrate and strengthen infrastructures which provide access to data (notably CESSDA and the Data Without Boundaries project; Silberman, 2013). These developments will require constant monitoring of the necessary technological upgrades and the development of resources to ensure wider knowledge of available surveys and facilitate their dissemination.

**Acknowledgements:** We are grateful to Jacques Véron for his precious comments on a previous version of this article and Benoît Riandey for giving us useful additional information on the French context. We wish to specially thank Roxane Silberman for reading the article and for her time and availability in providing us with further valuable information. We are also grateful to the editorial committee and to the anonymous reviewers of *Population* whose comments contributed much to the final version of this article.



## APPENDICES

## Appendix 1: List of acronyms

*Note:* For entities with no official English title, a translation is given in brackets for information.

ADISP	Archives de données issues de la statistique publique / National Archive of Data from Official Statistics
BDSP	Banque de données sociopolitiques (socio-political database)
CADA	Commission d'accès aux documents administratifs (committee for access to administrative documents)
CASD	Centre d'accès sécurisé aux données / Secure Data Access Centre
CCDSHS	Comité de concertation pour les données en sciences humaines et sociales (coordinating committee for humanities and social science data)
CCTIRS	Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé (committee for personal data protection in medical research)
CDSP	Centre de données sociopolitiques / Centre for Socio-Political Data
CESSDA	Consortium of European Social Science Data Archives
CIDSP	Centre d'informatisation des données sociopolitiques (centre for computerization of socio-political data)
CMH	Centre Maurice-Halbwachs
CNIL	Commission nationale de l'informatique et des libertés / French Data Protection Authority
CNIS	Conseil national de l'information statistique / French Statistical Advisory Committee
CNRS	Centre national de la recherche scientifique / French National Centre for Scientific Research
CSSDA	Council of Social Science Data Archives
DDI	Data Documentation Initiative
DDS	Dictionnaire de données statistiques / Statistical Data Dictionary
DWB	Data without Boundaries
EHESS	École des hautes études en sciences sociales
ELFE	Etude longitudinale française depuis l'enfance / French Longitudinal Study of Children

ERFI	Étude des relations familiales et intergénérationnelles (French Generation and Gender survey)
ESS	European Social Survey
EU	European Union
GENES	Groupe des écoles nationales d'économie et statistique / Group of National Economics and Statistics Schools
GGP	Generations and Gender Programme
GGS	Generations and Gender Surveys
HID-prison	Handicap, Incapacités, Dépendances en prison (survey on disability and dependence in prison)
IASSIST	International Association for Social Science Information Services and Technology
ICPSR	Inter-University Consortium for Political and Social Research
IFDO	International Federation of Data Organizations
INED	Institut national d'études démographiques / French Institute for Demographic Studies
INRA	Institut national de la recherche agronomique / French National Institute for Agricultural Research
INSEE	Institut national de la statistique et des études économiques / National Institute of Statistics and Economic Studies
INSERM	Institut national de la santé et de la recherche médicale / National Institute for Health and Medical Research
ISSC	International Social Science Council
LASMAS	Laboratoire d'analyse secondaire et de méthodes appliquées à la sociologie / Laboratory for Secondary Analysis and Applied Methods in Sociology
MAFE	Migrations entre l'Afrique et l'Europe / Migrations between Africa and Europe
MGIS	Mobilité géographique et insertion sociale (geographical mobility and social integration)
NESSTAR	Networked Social Science Tools and Resources
NSD	Norwegian Social Science Data Services
OECD	Organization for Economic Cooperation and Development

PROGEDO	Production et gestion des données en sciences humaines et sociales (production and management of humanities and social science data)
PUD	Plateformes universitaires de données (university data platforms)
SAFARI	Système automatisé pour les fichiers administratifs et le répertoire des individus (system for administrative file automation and registration of individuals)
SDMX	Statistical Data and Metadata Exchange
SES	Service des enquêtes et des sondages / INED Surveys Department
SHARE	Survey of Health, Ageing and Retirement in Europe
SSD	Standard Study Description
UNECE	United Nations Economic Commission for Europe
UNESCO	United Nations Educational, Scientific and Cultural Organization
XML	eXtensible Markup Language

## Appendix 2: List of websites as of November 2015

ADISP	<a href="http://www.cmh.ens.fr/greco/adisp_en.php">http://www.cmh.ens.fr/greco/adisp_en.php</a>
CASD	<a href="https://casd.eu/en">https://casd.eu/en</a>
CDSP	<a href="http://cdsp.sciences-po.fr/index.php?lang=ANG&amp;idTexte=accueil">http://cdsp.sciences-po.fr/index.php?lang=ANG&amp;idTexte=accueil</a>
CESSDA	<a href="http://www.cessda.net">www.cessda.net</a>
DDI	<a href="http://www.ddialliance.org">www.ddialliance.org</a>
ELFE – Data platform	<a href="https://pandora.vjf.inserm.fr/public/">https://pandora.vjf.inserm.fr/public/</a>
GGP	<a href="http://www.ggp-i.org/">http://www.ggp-i.org/</a>
GGP Online Codebook & Analysis	<a href="http://www.ggp-i.org/online-data-analysis.html">www.ggp-i.org/online-data-analysis.html</a>
INED-Nesstar Catalogue	<a href="http://nesstar.ined.fr/webview/">http://nesstar.ined.fr/webview/</a>
INSEE	<a href="http://www.insee.fr/en/default.asp">http://www.insee.fr/en/default.asp</a>
Nesstar	<a href="http://www.nesstar.com">www.nesstar.com</a>
Réseau Quetelet	<a href="http://www.reseau-quetelet.cnrs.fr/spip/?lang=en">http://www.reseau-quetelet.cnrs.fr/spip/?lang=en</a>

## REFERENCES

- ALLÈGRE Claude, 1999, “La lettre de mission”, in Silberman Roxane, *Les sciences sociales et leurs données*, rapport du ministère de l'Éducation nationale, de la recherche et de la technologie, pp. 5-6, [www.ladocumentationfrancaise.fr/var/storage/rapports-publics/004000935/0000.pdf](http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/004000935/0000.pdf)
- ARDUIN Pascal, 2004, “Présentation du Centre Quetelet : description des missions, activités, partenaires et projets de mise à disposition des données”, Journée d'information sur l'archivage historique et la mise à disposition des données d'enquêtes en SHS, Institut national d'études démographiques (INED), Paris, France.
- BLANK Grant, RASMUSSEN Karsten Boye, 2004, “The Data Documentation Initiative: The value and significance of a world-wide standard”, *Social Science Computer Review*, 22(3), pp. 307-318.
- BISCO Ralph L., 1966, “Social science data archives: A review of developments”, *American Political Science Review*, 60(1), pp. 93-109.
- BOZON Michel, 1995, “L'accès aux enquêtes de l'ined. Principes généraux, règles pratiques, problèmes en suspens”, INED internal report, 11 p.
- CAPORALI Arianna, KLÜSENER Sebastian, NEYER Gerda, KRAPF Sandra, GRIGORIEVA Olga, 2014, “The contextual database of the Generations and Gender Programme: Concept, content and research examples”, Stockholm University, Stockholm Research Report in Demography, 6.
- CHENU Alain, 2003, “Une infrastructure pour les données en sciences humaines et sociales”, *Courrier des statistiques*, 107, pp. 29-31.
- CHENU Alain, 2011, “Introduction”, in Chenu Alain, Lesnard Laurent (eds.), *La France dans les comparaisons internationales. Guide d'accès aux grandes enquêtes statistiques en sciences sociales*, Paris, Presses de Science Po, pp. 9-17.
- COMITÉ D'ARCHIVAGE DE L'INED, 2001a, “Compte rendu et relevé de décisions de la séance du 18 janvier 2001”, INED internal document.
- COMITÉ D'ARCHIVAGE DE L'INED, 2001b, “Compte rendu et relevé de décisions de la séance du 11 décembre 2001”, INED internal document.
- COMITÉ D'ARCHIVAGE DE L'INED, 2003, “Compte rendu et relevé de décisions de la séance du 16 janvier 2003”, INED internal document.
- COMITÉ D'ARCHIVAGE DE L'INED, 2004, “Compte rendu et relevé de décisions de la séance du 8 janvier 2004”, INED internal document.
- COMITÉ D'ARCHIVAGE DE L'INED, 2006, “Compte rendu et relevé de décisions de la séance du 14 décembre 2006”, INED internal document.
- COMITÉ D'ARCHIVAGE DE L'INED, 2009, “Compte rendu et relevé de décisions de la séance du 4 février 2009”, INED internal document.
- CORTI Louise, 2000, “Progress and problems of preserving and providing access to qualitative data for social research – The international picture of an emerging culture”, *Forum: Qualitative Social Research*, 1(3), <http://www.qualitative-research.net/index.php/fqs/article/view/1019>

- DATA WITHOUT BOUNDARIES, 2013, “Deliverable D7.1.Metadata Standards-usage and needs in NSIs and Data Archives”, Work Package 7 – Standards Development, prepared by: Merja Karjalainen, Mari Kleemola and Uwe Jensen ; with the contributions of Iris Alfredsson, Maurice Brandt, Michelle Coldrey, Claus-Göran Hjelm, Ørnulf Risnes and David Schiller, [www.dwbproject.org/export/sites/default/about/public\\_deliverables/dwb\\_d7-1\\_metadata-standards-usage\\_report.pdf](http://www.dwbproject.org/export/sites/default/about/public_deliverables/dwb_d7-1_metadata-standards-usage_report.pdf)
- DE MOOR Tine, VAN ZANDEN Jan Luiten, 2008, “Do ut des: Collaboratories as a new method for scholarly communication and cooperation for global history”, *Historical Methods*, 41(2), pp. 67-78.
- DOORN Peter, TJALSMA Heiko, 2007, “Introduction: Archiving research data”, *Archival Science*, 7(1), pp. 1-20.
- DUCHESNE Sophie, GARCIA Guillaume, 2014, “beQuali : une archive qualitative au service des sciences sociales”, in Cornu Marie, Formageau Jérôme, Müller Bertrand (eds.), *Archives de la recherche. Problèmes et enjeux de la construction du savoir scientifique*, Paris, L'Harmattan, pp. 35- 56.
- GAEREMYNCK Jean, 2009, “Le comité du secret statistique”, *Courrier des statistiques*, 128, pp. 15-18.
- GENERATIONS & GENDER PROGRAMME, 2014, “GGP at a Glance, no. 15”, Newsletter, [www.ggp-i.org/images/stories/GGP\\_AT\\_A\\_GLANCE\\_016-May-Jun-2014.pdf](http://www.ggp-i.org/images/stories/GGP_AT_A_GLANCE_016-May-Jun-2014.pdf)
- GREGORY Arofan, HEUS Pascal, 2007, “DDI and SDMX: Complementary, not competing, standards”, Open Data Foundation, [www.opendatafoundation.org/papers/DDI\\_and\\_SDMX.pdf](http://www.opendatafoundation.org/papers/DDI_and_SDMX.pdf)
- HANSEN Sue Ellen, IVERSON Jeremy, JENSEN Uwe, ORTEN Hilde, VOMPRAS Johanna, 2011, “Enabling longitudinal data comparison using DDI”, DDI Working Paper Series – Longitudinal Best Practice, 2, [www.ddialliance.org/system/files/EnablingLongitudinalDataComparisonUsingDDI.pdf](http://www.ddialliance.org/system/files/EnablingLongitudinalDataComparisonUsingDDI.pdf)
- HASTINGS Philip K., 1964, “Le Roper Public Opinion Research Center, centre international d'archives de données d'enquêtes par sondage”, *Revue internationale de sciences sociales*, 16(1), pp. 102-109.
- INED, 2002, “Orientations stratégiques de l'Institut national d'études démographiques pour la période 2002-2005. Présentées au conseil scientifique du 27 février 2002”, PARIS, INED.
- INED, 2006, “Orientations stratégiques de l'Institut national d'études démographiques 2006-2009. Approuvées par le conseil d'administration extraordinaire du 15 juin 2006”, Paris, INED.
- ICPSR (INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH), 2012, *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle* (5<sup>th</sup> ed.), Ann Arbor (Michigan), USA, [www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf](http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf)
- KING Gary, 1995, “Replication, Replication”, *PS: Political Science & Politics*, 28(3), pp. 444-452.
- KING Gary, 2006, “Publication, Publication”, *PS: Political Science & Politics*, 39(1), pp. 119-125.
- KRAMER Stefan, BANKS Randy, CHANG Vicky, SIEBER Ingo, VARDIGAN Mary, ZENK-MOLTGEN Wolfgang, 2011, “Presenting longitudinal studies to end users effectively using DDI metadata”, DDI Working Paper Series – Longitudinal Best Practice, 4, [www.ddialliance.org/system/files/PresentingLongitudinalStudiesUsingDDIMetadata.pdf](http://www.ddialliance.org/system/files/PresentingLongitudinalStudiesUsingDDIMetadata.pdf)

- LANG Gérard, 2008, "L'élaboration de la loi de 1951", *Courrier des statistiques*, 123, pp. 39-41.
- LE GLÉAU Jean-Pierre, ROYER Jean-François, 2011, "Le centre d'accès sécurisé aux données de la statistique publique française : un nouvel outil pour les chercheurs", *Courrier des statistiques*, 130, pp. 1-5.
- LEIGHTON Vernon, 2002, "Developing a new data archive in a time of maturing standards", *IASSIST Quarterly*, 26(1), pp. 5-9.
- MARKER Hans Jørgen, 2013, "Strengthening cooperation between European social science data archives: The evolving role of CESSDA", in Kleiner Brian, Renschler Isabelle, Wernli Boris, Farago Peter, Joye Dominique (eds.), *Understanding Research Infrastructures in the Social Sciences*, Zurich, Seismo Press, pp. 39-46.
- MOCHMANN Ekkehard, VARDIGAN Mary, 2011, "L'analyse secondaire de données et la documentation pour leur diffusion", in Chenu Alain, Lesnard Laurent (eds.), *La France dans les comparaisons internationales. Guide d'accès aux grandes enquêtes statistiques en sciences sociales*, Paris, Presses de Science Po, pp. 173-181.
- O'NEILL ADAMS Margaret, 2006, "The origins and early years of IASSIST", *IASSIST Quarterly*, 30(3), pp. 5-14.
- RASMUSSEN Karsten Boye, BLANK Grant, 2007, "The Data Documentation Initiative: A preservation standard for research", *Archival Science* 7(1), pp. 55-71.
- RHEIN Catherine, 2002, "Démographie et données statistiques", *Espace, populations, sociétés*, 20(1-2), pp. 125-132.
- RIANDEY Benoît, 2000, "La statistique 20 ans après la loi Informatique et libertés", in Héran François, Sebag Jean-Claude (eds.), *L'utilisation des sources administratives en démographie, sociologie et statistique sociale*, Paris, INED, Dossiers et recherches, 86, pp. 35-41.
- RIANDEY Benoît, 2003, "Centre Quetelet, deuxième anniversaire", *Courrier des statistiques*, 107, pp. 33-36.
- ROKKAN Stein, SCHEUCH Erwin K., 1963, "Conference on data archives in the social sciences", *Social Science Information*, 2(4), pp. 109-114.
- ROKKAN Stein (ed.), 1966, *Data Archives for the Social Sciences*, Paris, The Hague, Mouton & Co., 215 p.
- SZALAI Alexander (ed.), 1972, *The Use of Time. Daily Activities of Urban and Suburban Population in Twelve Countries*, Paris, The Hague, Mouton & Co., 868 p.
- SCHEUCH Erwin K., 2003, "Les services de données en sciences sociales : historique et perspectives", *Revue internationale des sciences sociales*, 177(3), pp. 433-449.
- SERVICE DES ENQUÊTES ET DES SONDAGES DE L'INED, 2004, Archives de l'équipe mise à disposition, INED.
- SILBERMAN Roxane, 1999, *Les sciences sociales et leurs données*, Ministère de l'Éducation nationale, de la recherche et de la technologie, 180 p., [www.ladocumentationfrancaise.fr/var/storage/rapports-publics/004000935/0000.pdf](http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/004000935/0000.pdf)
- SILBERMAN Roxane, 2011, "La protection des données individuelles en France et la recherche en sciences sociales", in Chenu Alain, Lesnard Laurent (eds.), *La France dans les comparaisons internationales. Guide d'accès aux grandes enquêtes statistiques en sciences sociales*, Paris, Presses de Science Po, pp. 183-204.
- SILBERMAN Roxane, 2013, "Transnational access to official micro-data: The Data without Boundaries European network", in Kleiner Brian, Renschler Isabelle, Wernli Boris, Farago Peter., Joye Dominique (eds.), *Understanding Research Infrastructures in the Social Sciences*, Zurich, Seismo Press, pp. 47-66.

- VARDIGAN Mary, HEUS Pascal, THOMAS Wendy, 2008, "Data Documentation Initiative: Toward a standard for the social sciences", *The International Journal of Digital Curation*, 3(1), pp. 107-113.
- VIKAT Andres, SPÉDER Zsolt, BEETS Gijs, BILLARI Francesco C., BÜHLER Christoph et al., 2007, "Generations and Gender Survey (GGS): Towards a better understanding of relationships and processes in the life course", *Demographic Research*, 17(14), pp. 389-440.
- WACKEROW Joachim, VARDIGAN Mary, 2013, "An established international metadata standard: The Data Documentation Initiative (DDI)", in Kleiner Brian, Renschler Isabelle, Wernli Boris, Farago Peter, Joye Dominique (eds.), *Understanding Research Infrastructures in the Social Sciences*, Zurich, Seismo Press, pp. 158-167.

### Arianna CAPORALI, Amandine MORISSET, Stéphane LEGLEYE • PROVIDING ACCESS TO QUANTITATIVE SURVEYS FOR SOCIAL RESEARCH: THE EXAMPLE OF INED

Sharing data and providing access to surveys are paramount to social research. Yet there is little knowledge of the archiving activities involved. This article focuses on quantitative survey data. It reviews the development of social science survey data archives from their creation in the 1950s and the emergence of international standards. France was a late starter in this domain. It then demonstrates the importance of archives for providing access to survey data via the example of the Surveys Department of the French Institute for Demographic Studies (INED), a co-founder of the Réseau Quetelet, which today centralizes access to most social science surveys in France. The work done to make survey data and metadata available for re-use is discussed, highlighting the importance of collaboration with data producers in a context of a strengthening trend towards open data.

### Arianna CAPORALI, Amandine MORISSET, Stéphane LEGLEYE • LA MISE À DISPOSITION DES ENQUÊTES QUANTITATIVES POUR LA RECHERCHE EN SCIENCES SOCIALES : L'EXEMPLE DE L'INED

L'accès aux enquêtes et le partage de données sont fondamentaux pour la recherche en sciences sociales. Pourtant, les activités d'archivage qui permettent un accès aux enquêtes mises à disposition sont peu connues. Cet article se concentre sur les données des enquêtes quantitatives. Il passe en revue le développement des archives de données d'enquêtes en sciences sociales depuis leur création dans les années 1950, ainsi que l'émergence de normes internationales. En France, les archives de données d'enquêtes ont été établies relativement tard. L'article montre ensuite l'importance des archives pour l'accès aux enquêtes en s'appuyant sur l'exemple du Service des enquêtes et des sondages de l'Institut national d'études démographiques (Ined), cofondateur du Réseau Quetelet, qui centralise aujourd'hui l'accès à la plupart des enquêtes quantitatives en sciences sociales en France. Les activités d'archivage et de mise à disposition des données et métadonnées d'enquêtes à des fins de recherche sont ensuite discutées, en montrant l'importance de la collaboration entre les producteurs de données dans un contexte d'ouverture croissante des données (*open data*).

### Arianna CAPORALI, Amandine MORISSET, Stéphane LEGLEYE • LA DISPONIBILIDAD DE LAS ENCUESTAS CUANTITATIVAS PARA LA INVESTIGACIÓN EN LAS CIENCIAS SOCIALES: EL EJEMPLO DEL INED

El acceso a las encuestas y el intercambio de datos son fundamentales para la investigación en ciencias sociales. Este artículo se concentra en las encuestas cuantitativas y examina el desarrollo de los archivos de datos de encuesta desde su creación en los años 1950, así como la emergencia de normas internacionales. En Francia, la creación de dichos archivos es relativamente tardía. El artículo muestra a continuación la importancia de los archivos para el acceso a las encuestas, apoyándose en el ejemplo del Servicio de encuestas y sondeos del Instituto francés de estudios demográficos (Ined), cofundador del Réseau Quetelet que centraliza hoy el acceso a la mayor parte de las encuestas cuantitativas en ciencias sociales realizadas en Francia. Las actividades de archivo y de puesta a disposición de los datos y metadatos de encuesta a fines de investigación son discutidas, mostrándose la importancia de la colaboración entre los productores de datos en un contexto de apertura creciente de estos últimos (*open data*).

---

**Keywords:** Social science data archives, data sharing, metadata, data documentation, Data Documentation Initiative, survey data.