Statistics at the Ministry of Culture and Communication: an interview

Philippe Chantepie* and Jeannine Cardona**

* Philippe Chantepie is the head of the Department of Studies, Forward-Looking Analysis, and Statistics (Département des Études, de la Prospective et des Statistiques: DEPS) at the French Ministry of Culture and Communication.
** Jeannine Cardona is in charge of statistical operations at DEPS.


Courrier des statistiques: How is DEPS organized, and how have its operations and work program changed since the 1970s?

Philippe Chantepie and Jeannine Cardona: At its founding in 1963, our unit was called Studies and Research Office (Service Études et Recherche: SER). It turned into the Department of Studies and Forward-Looking Analysis (Département des Études et de la Prospective: DEP) in 1986 and, later, the Department of Studies, Forward-Looking Analysis, and Statistics (Département des Études, Prospective et des Statistiques: DEPS). In 2004, it became one of the four departments of the new Delegated Office for Development and International Affairs (Délégation au Développement et aux Affaires Internationales: DDAI) at the Ministry of Culture and Communication, after having functioned for over twenty years as a unit reporting to the Ministry’s General Administration (Administration Générale).

Since its inception, the Department’s mission has been to build up economic and social knowledge of cultural life in France. DEPS does this by (1) compiling statistics on culture, (2) producing analyses on the economics of cultural activities, and (3) preparing studies on the changes in cultural participation generated by the evolution in lifestyles. Thanks to these activities, DEPS provides quantitative and qualitative insights for defining the nation’s cultural policies, charting their course, and assisting in decision-making.

More specifically, DEPS carries out surveys and studies on French cultural life from two angles. The first is a sectoral approach, i.e., broadly reflecting artistic fields: surveys on museum attendance, art schools, movie audiences, and actors; studies on leveraging cultural assets, the economics of music, and so on. The second comprises approaches of a more cross-sectional nature, such as public funding of culture, tracking of cultural employment, households’ cultural expenditures, and cultural participation of the French.

The mix of these two categories has changed constantly, depending on the shifts in the Ministry’s remit and in cultural-policy priorities—and also, of course, because society itself has evolved.

In the past two decades, we have set up new cross-sectional projects aimed at gaining fuller knowledge of cultural industries, artistic and cultural employment and occupations, art education, and other areas.

In 2005, the Department’s generalist and cross-sectional mission was reasserted and linked to new priorities such as the analysis of changes in
the economics of culture (with special emphasis on cultural industries) and the analysis of the uses of culture, cultural expectations, and cultural behaviors. The 2005/2006 DEPS studies and statistics program is accordingly divided into five chapters: “economics and cultural industries,” “employment and education,” “analysis of policies, public action, and regulation,” “participation, audiences, and consumption,” and “modes of transmission and legitimation; transmission of knowledge.” The program is implemented by some twenty staff members working in three units: studies, statistics, and publications.

The DEPS work program is drawn up every year on a proposal from its staff members (who, by definition, are well placed to have relevant opinions on the issue) and in consultation with the Ministry’s other directorates. A Studies Council, set up in 1986, defines work guidelines, monitors program progress, and makes sure that results are disseminated. The Council, chaired by the Minister or his/her chief of staff (directeur de cabinet), comprises the heads of the Ministry’s directorates, a regional director of cultural affairs, and qualified individuals, including a representative of INSEE’s Business Statistics Directorate under the terms of reference of a project group comprising the Directorate’s “Register and Demography of Enterprises and Local Units” Unit, which manages the national SIRENE business register.

DEPS is asked by INSEE to participate in large-scale official-statistics projects such as the inter-departmental mission on the economics of membership organizations or, more recently, the revision of the French classification of economic activities (Nomenclature d’Activités Française: NAF) and the European Union classification of economic activities (NACE). For the cultural sector, the revision will notably offer a better coverage of artistic activities, which have now been divided into three sub-classes.

Agreements with INSEE allow DEPS to use national surveys such as the Family-Budget Survey. DEPS works regularly with the Institute’s Directorate for Demographic and Social Statistics on the changes in the variable sections of the Continuous Survey on Living Conditions (Enquête Permanente sur les Conditions de Vie: EPCV), and on the definition-in-progress of the content of the future Statistics on Income and Living Conditions Survey (Statistiques sur les Revenus et les Conditions de Vie: SRCV [the French transposition of SILC]). This cooperation has given DEPS opportunities to work with statistical offices in other ministries as well: the Ministry of Youth Affairs and Sports for the 2004 survey on cultural and sports participation; the Ministries of Employment and Solidarity, of Youth Affairs and Sports, of Education, and others for the 2002 survey on participation in associational life.

Lastly, we can mention our cooperation with the Directorate for Assessment, Forward-Looking Analysis, and Performance (Direction de l’Évaluation, de la Prospective et de la Performance: DEPP), which, under the terms of an agreement, allowed DEPS to use the panel of children entering first grade in 1997 for a longitudinal observation of cultural participation of children and adolescents.

Since 1995, when DEPS launched an initiative to promote EU cooperation in culture statistics, our Department has developed ties with Eurostat—first as part of a Leadership Group (LEG), then as a member of a European experts group on culture statistics. The work undertaken has led to the definition of a methodology using NACE and ISCO for measuring cultural employment in the EU and a methodology for measuring cultural participation. Eurostat is expected to publish the first available results in 2007.

CDS: What are the links between DEPS and official statistics?

PC & JC: The link with INSEE is materialized on a long-term basis by the Institute’s participation in the DEPS Studies Council, as mentioned above, but also by the DEPS’s membership—as a ministerial statistical office and producer of statistics—in the “Demography and Living Conditions” task force of the National Council for Statistical Information (Conseil National de l’Information Statistique: CNIS). Our Department presents its annual work program to the task force as well as descriptions of its planned surveys so that the task force can rule on their usefulness (avis d’opportunité).

Also, in its relations with INSEE, DEPS relies very routinely on the Institute’s Division for Relations with Ministerial Statistical Offices (Division des Relations avec les Services Statistiques des Ministères: DRSSM) and, more generally, with the Statistical Coordination Department. The latter provides DEPS with effective support and facilitates contacts with other INSEE units on specific projects, such as requests for use of survey data files.

One of the major current projects at the Ministry of Culture is the creation of a “cultural register” aimed at establishing a sampling frame of all units engaged in the production of cultural goods and services. DEPS receives methodological support from INSEE’s Business Statistics Directorate under the terms of reference of a project group comprising the Directorate’s “Register and Demography of Enterprises and Local Units” Unit, which manages the national SIRENE business register.

CDS: Reading the 1977 article by Augustin Girard and Odile Timbart, the striking fact is that, about thirty years ago, culture statistics had to be built from scratch, starting with the very definition of the cultural field. Do DEPS researchers and statisticians still need to ask such questions today, or is there now a consensus on what “culture” is?

4. DEPS won a Eurostat tender to prepare the first statistical yearbook on culture in the European Union, which Eurostat will publish in November 2007.
PC & JC: We’ve come a long way from the construction period. Thanks to a series of statistical projects, DEPS has achieved a vital legitimacy and role in the advancement of knowledge on the cultural environment. In our now well-established publication series—Notes et études statistiques, Notes et études sur l’emploi culturel—our recurrent statistical studies on funding by local government, the cultural participation of the French, employment in the performing arts, the activities of music schools, and other subjects have contributed a corpus of essential knowledge to the Ministry of Culture and Communication and, beyond, to cultural actors in the public and private sectors alike.

As noted earlier, a large section of our area of study concerns short-term phenomena such as public-policy initiatives—i.e., financial intervention in favor of culture—or particularly sensitive fields such as cultural employment (which, in the performing arts, owes its singular profile to a specific unemployment-benefits system). Concerning our work on modes of consumption, household spending on culture and the media, and the cultural participation of the French, these are objects of social analysis tied to structural factors such as demographics, educational attainment, and income levels. Because they are structural, these factors change slowly, which is sometimes disappointing for public decision-makers who learn about our work!

In any event, all of these projects concern an “established” cultural field, which is that of public cultural action—in the broad sense—undertaken by a ministry that, when you think of it, is very young. Yet the cultural field necessarily remains an open, problematic issue, which raises many questions.

To begin with, we cannot neglect a more anthropological (and hence more encompassing) approach to culture: many social phenomena are actually cultural phenomena—or are, at the very least, informed by cultural factors.

Let us give you a few examples at random from recent years: audiovisual and digital usages, particularly the exchange of content; the importance of interpersonal communications; the lifestyle of the young, who are entering the workplace at an ever later age; intergenerational relationships; and the expansion of amateur cultural and sports participation.

These usages are not the easiest to measure statistically. This is a challenge that should be taken up through a sharing of insights among public players: the Ministry’s Directorate for Media Development, the National Film Center (Centre National de la Cinématographie: CNC), the National Audiovisual Institute (Institut National de l’Audiovisuel: INA), the Ministry of the Economy, Finance, and Industry, and, of course, INSEE and the ministerial statistical offices concerned. That is a potential horizon for our Department’s statistical function: DEPS will then need to rely on tools of a more cross-sectional nature (possibly developed by third parties), as it does already, for instance, in the longitudinal analysis of children’s panels conducted in cooperation with DEPP.

---

Another possible initiative is the study of technical change, which surely has powerful consequences on the modes of access to culture. This will lead DEPS to overhaul its investigations of cultural participation. They should be largely extended to media and digital usages, particularly Internet access; they should also take account of the time scale of consumption (the fact that consumption times vary according to usage), and the dynamics of complementarity, substitution, and renewal of cultural participation.

The cultural field is changing not only on the demand side, i.e., among the audience for cultural offerings. It is also changing on the supply side itself, chiefly in regard to the commercial offering, which has become central to cultural access. A second statistical horizon should therefore open up here: the exploration of the economic field of public- and private-sector cultural activities. Here as well, we need to develop suitable tools. That is why one of the major undertakings of DEPS consists, as we have seen, in compiling a cultural register, structured along the same lines as INSEE’s SIRENE business register, but using identifiers specific to the cultural sectors concerned. This is obviously a “heavyweight” project, which relies on the full range of information systems at the Ministry of Culture and Communication and will take time to complete. Ultimately, the tool’s content will shape the development of relevant statistical surveys on the economic world of culture.

We have already taken a first step by preparing analyses of cultural industries—using data from the Annual Enterprise Surveys (Enquêtes Annuelles d’Entreprise: EAE)—published in an initial Aperçu statistique des industries culturelles (Statistical Overview of Cultural Industries) in early 2006. These studies, as well, will provide an indispensable base for preparing deeper examinations of individual sectors and themes. Beyond this, a better coverage of cultural industries by DEPS ties in with a reappraisal of the cultural field. Several countries—in North America, Asia, and even Europe—have committed their culture, or at least their “cultural industries,” to an approach based on international competitiveness. This comes at a time when cultural diversity has been internationally recognized at UNESCO. The measurement of a cultural field rebuilt around the notion of creativity could therefore become a novel and highly complex statistical object. The U.N., OECD, UNESCO, and the European Commission are weighing this possibility. Without passing judgment for the moment on whether such changes are justified, DEPS will, in the coming years, be involved in the review: as one of the world founders of culture statistics, our Department has a part to play in their evolution.

Therefore, and to answer your initial question, yes, the consensus on the cultural field is already being reappraised, at the very time when entire sectors of culture—such as the performing arts—are not yet sufficiently well known in statistical terms.

Nevertheless, our work now rests on a sufficiently solid foundation to keep our statistical system moving in the right direction, preparing tools and studies with greater explanatory power. This will give us a better grasp of new objects or even a new field.

Bibliography: recent DEPS publications


Les loisirs culturels des 6-14 ans, Sylvie Octobre, Paris: La Documentation Française, 429 pp.

Regards croisés sur les pratiques culturelles, Olivier Donnat (ed.), Paris: La Documentation Française, 348 pp.

Développement culturel (a periodical)
151 - Les publics des Archives nationales et leurs attentes (January 2006).
150 - La fabrique sexuée des goûts culturels. Construire son identité de fille ou de garçon à travers les activités culturelles (December 2005).
147 - La féminisation des pratiques culturelles (June 2005).
146 - L’économie de la billetterie du spectacle vivant (March 2005).
143 - Transmettre une passion culturelle (February 2004).


Notes de l’Observatoire de l’Emploi
45 - L’emploi dans les professions culturelles en 2004 d’après l’enquête Emploi de l’INSEE (December 2005).

Notes statistiques du DEPS
19 - L’observation culturelle en région: 2 - le financement public (June 2006).
20 - L’observation culturelle en région: 3 - les publics (June 2006).
16 - Aperçu statistique des industries culturelles (January 2006).
The information produced by official statistical agencies derives from two sources: surveys and administrative records. Sample surveys allow statisticians to tailor questions to specific needs. However, they are costly—both financially and in time (particularly for respondents)—and ill-suited to geographic distributions. Administrative sources are assumed to be less expensive and, being often nearly exhaustive, they can be broken down in greater detail. But their variables are determined by regulatory categories. Our article seeks to broaden this standard view of the characteristics of the two sources. In a brief historical section, we look at a few cases where the two have coexisted: statistics on population, poverty, unemployment, income and wages, and crime. An important criterion is the allocation of the formatting work, i.e., the construction and coding of categories and variables: these tasks are carried out either “upstream” from the statistical system or within it. A question often asked is: do registers reflect society or the activity of government departments? To escape this apparent paradox, we propose an analysis in terms of the uses of statistics. We reconstruct the circularity and interaction between the interpretations of society, the actions performed on society, and modes of description. In this framework, administrative sources tend to reflect existing actions, while surveys are better suited to expressing new aspects of society.

The information disseminated by official statistical agencies is produced from two types of sources: direct surveys (such as censuses and sample surveys) and administrative records (such as management databases and registers). The advantages and drawbacks of each category are well known—and reciprocal. Surveys allow statisticians to tailor questions to specific needs. However, they are expensive and (particularly for respondents) time-consuming. In addition, sample surveys are hard to distribute geographically, given the requirements of sample representativeness. By contrast, administrative sources are assumed to be less costly, since they are previously compiled for other purposes; being often nearly exhaustive, they can be “territorialized”; but the regulatory definitions and categories on which they depend differ from those required by statistics users, and they can complicate comparisons between countries or between periods.¹

The issue of the breakdown of sources in official statistics is generally reduced to these few concepts. The present context is dominated by two factors: (1) budget restrictions and (2) the increased reluctance of individuals and enterprises to respond to surveys (i.e., the “response burden”). In these conditions, the use of administrative sources is strongly recommended, particularly in certain European countries (Northern Europe and Netherlands). In France, the 1951 Act on legal obligation, coordination, and confidentiality in the field of statistics has been recently amended to accommodate this shift. Previously, government departments had the option of transmitting individual data to official statistical agencies [...]. Now, they are required to do so if the agencies ask for the data.

¹ I am grateful to Michel Blanc, Emmanuel Didier, Michel Glaude, Gérard Lang, Jean-Pierre Le Gléau, Jean-Pierre Puig, and Pascal Rivière for their suggestions on this text.
the use of administrative sources for reasons of cost and exhaustiveness; in particular, they allow a flexible adjustment of geographic divisions.

We shall return to this issue by examining the strength of the standard arguments rehearsed above; we shall then broaden the discussion beyond these arguments—which, admittedly, do have some relevance—and incorporate it into a wider reflection on the cognitive content and social role of statistical institutions’ products. Before spelling out in greater detail the content of the two types of sources, we shall provisionally describe as a survey a collection operation performed by an entity specializing in statistics, for the purpose of producing an aggregated set of information of general import. By contrast, an administrative source is obtained from an institution whose mission is not to produce such information, but whose management activities involve the keeping of individual files or records in accordance with generally accepted rules. Their aggregation is only a by-product; their key component is individual information, particularly for the persons or enterprises concerned. The criteria and methods for assessing the qualities of these two types of sources, from the standpoint of their initial uses, are different. Surveys can be assessed, in particular, by applying the well-established theory of sampling errors; for administrative sources, no general error theory yet exists (Platek and Sarndal 2001; Boydens 1999).

Even setting aside the fact that the distinction between the two sources is less obvious than it may seem at first sight, their comparison shows that the ways in which these statistics shape the social world depend heavily on the channels through which they are recorded and coded: the “surveys versus records” distinction is only one of several aspects. The prerequisite for statistical production is the construction, in cognitive and political terms, of a conventional equivalence space. For administrative sources, this space is strongly determined by national factors. The national institutional and regulatory contexts determine these sources even more directly and explicitly than they determine surveys. This has consequences on the growing demand for international comparability of statistical data. Regulatory changes disrupt the continuity of data series. Another recurrent request concerns the evaluation of public policies, which implies distinguishing between ways and means. Administrative sources are focused on means and public policies more than on the state of society. This creates a difficulty, of which crime statistics offer a well-known example.

This issue may seem abstruse if presented solely in its “utilitarian” aspect through the list of classical arguments. One might conclude that it is of interest only to official statisticians. It is not much addressed by academic statisticians or by researchers, economists or sociologists—I.e., statistical-data users—except, occasionally, to point out gaps or potential bias in the data due to their administrative origin. However, by using the issue as a starting point, we can take a fresh and unconventional look at official statistics, most notably by analyzing simultaneously their contribution to knowledge and the social channels through which they are shaped and used. The recent growth in demand for statistics is linked, for example, to three types of changes in their use: European integration, French decentralization, and the reforms in the management of French government services, which have introduced quantitative performance evaluations. These developments offer opportunities to address the issues—which are as old as official statistics—in new terms.

We begin with a historical review of some important cases where the two types of sources coexisted or competed: population statistics (civil registration and censuses); nineteenth-century analyses of poverty; the measure of unemployment; income and wage distribution; and crime. We then examine the standard criteria, and the distinction between the two categories of sources will be criticized or at least put into perspective. The two sources are shown to be different but complementary. An important notion is that of the distribution of the data-formatting work (construction and implementation of categories and variables; coding). This work can be undertaken either before the data enter the statistical system (preformatting) or inside the system. Lastly, we shall analyze the uses of the statistics, reconstructing the circularity and mutual relationships between conceptions of society, actions on society, and modes of description. Paradoxically, this approach seems to call into question the externality (both epistemological and social) of description relative to action: this externality is a core assumption behind the necessary relative independence of official statistics from centers of economic and political authority. Thus the question of the duality of types of sources is not only technical. It is also a means to highlight a vital tension inherent in official statistics: on the one hand, the ethics of official statistics are rightly predicated upon their independence; on the other hand, their production and usage channels are closely embedded.
in public policy-making. From this standpoint, administrative sources are more apt to reflect existing policies, while surveys alone can document new aspects of society. The point here is not to supply the “right turnkey solutions” to the tension in a normative manner, but to spell it out as clearly as possible.

Between administration and observation: some historical precedents

The measures of population have formed the historical core of official statistics since the eighteenth century. In France, population flows (births, marriages, deaths) are measured from civil registration records introduced in 1792, which were based on the parish registers established by royal edict in the sixteenth century. The total population “stock” has been measured by means of censuses since 1801. This distinction between civil registration and censuses seems to foreshadow the duality between registers and direct surveys. Today’s census combines an (administrative) enumeration and a socio-demographic survey. In the nineteenth century, one could already view it as a survey conducted by specialists, soon called statisticians. From the 1830s onward, the five-year census was the main (if not the only) specific activity of the new statistical bureaux that multiplied across Europe, at the behest of Adolphe Quetelet (Desrosières 2004). All the other statistics published by the bureaux at the time were compilations of tables supplied by the main government departments, as by-products of their specific activities. Through the population censuses, statisticians gradually carved out a domain of their own, laying the groundwork for a new and innovative professional status. They even had to routinely reassert their difference from supervisory departments—most notably the tax authorities—in order to avoid being confused with them. This implied the gradual formulation of a claim to autonomy for the new institution: the “statistical bureau.” The main building-block for autonomy was the construction of a corpus of “general knowledge” not tied to a specific management-related use; that is indeed the case of the population census. The same demand arose later for surveys. The assertion of autonomy—both institutional and cognitive—characterizes official statistics, a space simultaneously distinct from a conventional government department and from the academic world engaged in producing scientific knowledge. This dual distinction is its novel feature.

The census is a cumbersome, expensive administrative operation requiring the participation of many players, well beyond statisticians alone. Until the 1930s, it was almost the only direct survey conducted by the official statistical system. At that point, the “relegitimation” of the sampling method, discredited throughout the nineteenth century, would utterly transform the activity of national statistical institutes (NSIs) by creating a very wide range of opportunities for them to perform direct investigations (Desrosières 1998). Three historically important and well documented examples—poverty, unemployment, and crime—effectively show that the alternative between “registers versus surveys” is not only a practical matter of availability and cost of sources. It also involves measures whose nature and significance are very different.

Administrative sources are more apt to reflect existing policies, while surveys alone can document new aspects of society.

Before the poor could be counted, they had to be coded, i.e., assigned to an equivalence class. This task can be performed either by a specific institution, for practical purposes other than the pursuit of knowledge, or by survey statisticians, using a more theoretical definition consistent with information needs. A logical problem emerges when the institution’s management method is used as an “explanatory variable” for the poverty evaluations that are based on the same data. We find this in the work of Yule (1895), who used the breakdown between relief provided in workhouses (indoor relief) and in homes (outdoor relief) to “explain”—by means of a regression equation—a total poverty that was nothing other than “total relief.” This type of paradox is classic in the studies combining registers and surveys.

We can raise similar questions about the emergence of the measure of unemployment, between 1900 and 1950 (Salais et al. 1986). In France, the spread of municipal agencies providing relief to the unemployed (the forerunners of the Agence Nationale Pour l’Emploi—ANPE—the “national employment agency”) gave unemployment a social existence and a tangible dimension that made it measurable. Then, starting in 1950, “labor-force surveys” allowed unemployment to be measured differently, in accordance with the three ILO criteria: out of work, immediately available for work, seeking work. Both measures lived
Statistics on income and wages go back a long way. The studies on income by Pareto and Bowley relied mainly on tax sources, since it was accepted that direct surveys on the topic were almost impossible. They therefore used administrative data, gathered for purposes that raised suspicions of under-reporting or fraud. This did not prevent statisticians from producing very refined studies of them (Piketty 2001). In 1956, INSEE conducted an initial “income survey” that actually covered a sample of households’ tax returns and so did not properly constitute a direct survey (Bégue 1987). This offers an example of re-use of transferred data, checked in particular by means of probabilistic sampling procedures. Another example is provided by wage statistics, collected since the nineteenth century by employees’ unions and employers’ unions for purposes of submitting claims, wage bargaining or wage regulation. Meanwhile, since 1947, French statisticians have exploited employers’ annual statements of payroll data (Déclarations Annuelles de Salaires: DASs) (Volkoff 1987). In other words, there have been at least three channels for measuring wages and salaries. The three may, of course, be subjected to a critical comparison and assessment of their sources. That is what national accountants do. But a sociology of the production and uses of numbers can also take this diversity as its corpus, observing how the result of these statistical syntheses goes on to circulate, as if cast off from its origins, embarking on new adventures, and telling another story featuring other characters.

The measurement of crime is another historical example of the difficulty of linking administrative sources with surveys. In France, the administrative sources go all the way back to the “General Justice Account” (Compte Général de la Justice) of the 1820s, which published statistics on court activity (Robert et al. 1994). More recently, statistics on police, gendarmerie, and court activities have also been used. But these sources are often criticized, as some observers precisely argue that they reflect the work of the producer entities more than criminal activity and victims’ damages. An increase in crime statistics is thus perceived as reflecting stepped-up efforts by agencies in charge of protecting individuals. Observers accordingly speak of the “black hole” in these statistics, i.e., unreported crimes and law-breaking. To remedy this, statisticians conduct “victimization surveys” of individuals, and even anonymous surveys of persons concerning their potential law-breaking activities. This debate, which has been going on since the nineteenth century, is comparable to those mentioned earlier on poverty and unemployment. Some specialists (such as U.S. ethno-methodologists: Cicourel and Kitsuse 1963) go as far as claiming that crime statistics only reflect the activity of the repressive system.

A subtler view is that these data reflect the intersection of the activities of the system and those of law-breakers—just as, in a product market, trading volume and prices lie at the intersection of the supply and demand curves. Similarly, the poverty statistics used by Yule in 1895, taken from the management files of the Poor Law Unions, may be regarded as an intersection of the Unions’ policies and the rise in poverty in England in the 1880s-1900s; likewise, the number of free meals handed out by the French charity Les Restos du Cœur (Restaurants of the Heart) lies at the intersection between the charity’s efforts and the resurgence of deep poverty in France in the 1980s-2000s. This approach to administrative sources avoids the pitfall of the relativism that threatens some data readings by the most radical ethno-methodologists. But it also enables us to go beyond a still common exercise—namely, the comparison of sources in a strictly realistic, quantitative, and one-dimensional manner. This implicitly assumes the existence, prior to its various recordings, of some “right number” independent of the conditions in which it is produced. The stake in these debates is to find a way of navigating between a somewhat naive realism (vigorously demanded by the public) and a relativism induced by certain uses (critical or not) of the notion of social construction (Hacking 1999).

The examination of the concrete problems faced by statisticians in comparing direct surveys and administrative sources offers a field for empirical and sociological studies of the opposition—often seen as purely theoretical and philosophical—between realism and constructivism.

These selected historical examples have enabled us to spell out some of the issues raised by the linkage between the two types of sources, in terms of interpretation and usage, beyond the list of technical arguments typically put forward. We must now review and discuss these arguments, most notably in the light of the questions raised by the examples described. In particular, we need to elaborate the concept of administrative source, which we have mentioned above only in connection with practical examples.
Differences, continuities, and combinations between the two types of sources

What common ground is there between (1) disparate compilations of tables produced in different government departments and published in statistical yearbooks until the 1950s, and (2) the sophisticated processing and analysis now performed by statistical offices on individual data forwarded by government departments? In both cases, we are dealing with “administrative sources,” but the main difference is the transmission of individual data to specialized statistical offices. This turning point in the history of official statistics was predicted by René Carmille, on the strength of his experience in the French defense administration, before he even created the National Statistical Office (Service National de Statistique: SNS) in 1941. In a book published in 1936, La mécanographie dans l’administration, he advocated a harmonization of records managed by different government departments through the use of common identifiers, for individuals and enterprises, to allow cross-checking of these various information sources. This task could be achieved, he argued, thanks to the punched-card tabulating machines manufactured by Hollerith, the direct ancestor of IBM. Carmille thus invented information processing before computers. He even began to implement the project at the SNS between 1941 and 1944, in particular by designing what later became the NIR (13-digit Social Security number), but the punched-card technology of the time hampered significant progress. Carmille—who died at Dachau in 1945—had also helped to introduce the then-new sampling method at SNS, under the influence of Jean Stoetzel. After the founding of INSEE in 1946, the two approaches—punched-card registers and surveys—seemed to be in competition. But, between 1946 and 1960, sampling was regarded by the young statisticians trained at the new École d’Application de l’INSEE (ancestor of the École Nationale de la Statistique et de l’Administration Économique: ENSAE) as more modern, mathematically elegant, and suitable for exploring a wide range of new issues. At the time, they were temporarily preferred to “Carmille’s bulky card files,” seen as cumbersome, bureaucratic, and—technically and intellectually—not very innovative.5

The relative perceptions of the two types of sources evolved from the 1960s on, as a result of three changes in official statistics: (1) the spread of information technology, (2) the key role of national accounting in the 1960s and 1970s, and (3) the corollary of the latter, i.e., the promotion of statistical coordination. At that point, the concept of “administrative source” was integrated into the official statistical system. When using such sources, statisticians no longer merely reproduced tabulations prepared by others “as is”; they now processed and transformed in their own way the material supplied by these institutions. In English, the general term data editing is used to denote the complex of data checks, verifications, file cleaning, adjustments, imputations, and related operations. Through these technical procedures, the files of individual data from government departments were transported from one world of meanings to another. They lost their initial meanings to acquire others, on which statisticians placed their own stamp. In particular, statisticians looked at the whole as such, and no longer at individuals, each in his/her right—as is the case, for example, for a file of social-benefit recipients.

The transformation described above is not only symbolic and semantic, but also physical, through the process of data editing (for which there is no true equivalent in French). Last but not least, it is costly, in terms of money, worktime, and gray matter. This puts into perspective the still widespread notion that administrative sources are “economical” because they save the cost of the initial collection—as if “you just have to bend down to pick them up.” The re-use of administrative sources is in fact a heavy investment. The notion that they are virtually free is largely false. By contrast, their extensive—often nearly exhaustive—coverage theoretically allows territorial divisions that can be modulated flexibly thanks to the dissemination of personal computers. This possibility, however, is subject to compliance with strict confidentiality rules, which forbid geographic disaggregation to excessively fine levels of detail.

Different ways of combining surveys and registers

But the problem of the “relative economy” of the two collection methods can be rephrased differently when we address the issue of the technical and social division of labor in formatting work, i.e., the implementation of equivalence conventions. This shaping of information can be done at different stages of a long process and involves successive transformations, either before the statistical collection, during the collection, or after in the data editing and summarizing phases. For instance, the statistical offices of some ministries conduct “surveys” of healthcare or educational institutions by gathering information on their activities already recorded and formatted by the institutions themselves. More generally, the “structural surveys” of businesses (in France, the Annual Enterprise Survey: Enquête Annuelle d’Entreprise or EAE) collect information on accounts and employees already formatted by the enterprises’ own management.6 Another example is the “income survey,” in which a sample of

5. The argument that files posed a threat to the protection of data confidentiality and privacy does not seem to have been invoked at the time except incidentally, even though the previous period had witnessed dramatic examples of file abuse by totalitarian regimes in several countries. (For an overview of such uses, see Seltzer and Anderson [2001].) These issues would not be raised and widely discussed until the 1970s, when information management systems came into use (French law on Information Technology and Civil Liberties, establishment of the National Commission on Information Technology and Civil Liberties [CNIL] in 1978). 6. Indeed, that is why the notion that these surveys could collect “better” information than the surveys based on tax sources was gradually abandoned.
responses to the labor-force survey is matched against the tax returns of the corresponding households. This is partly a by-product of an administrative source—not a survey in the ordinary sense. In all these cases, the term “survey” has a rather different meaning from the situation where statisticians collect information that is not yet formatted. A continuum therefore exists because the formatting work is distributed down a long chain, in which, for example, national accountants step in later to complete a further stage in the transmission-transformation of the information.

The continuity between the two forms of data collection and formatting can also be examined in terms of the sociology of organizations. The French statistical system comprises a central institution, INSEE, and ministerial statistical offices (Services Statistiques Ministériels: SSMs), which employ nearly 40% of official statisticians. The chronic problem for official statisticians is to be able to intervene at the earliest possible stage in the organization of information systems in the ministries; that way, they can ensure that their specific needs—such as definitions of variables, classifications, and so on—will be taken into account. This depends on the type of cooperation and mutually useful exchanges that they can establish with other offices in their ministries. A good example is provided by the evolution of the Customs Statistical Office (Service Statistique des Douanes), whose activities had been at first heavily threatened by the abolition of internal customs barriers in the European Union. But a new system, called the “surveys versus registers” continuum, this time on the “surveys” side, by examining the work of interviewers. They take on a share of the formatting task mentioned above by fitting the sometimes incommensurable diversity of the situations they encounter into the Procrustean bed of their questionnaires. The cost of this alignment—in psychological terms, or simply in time spent—is often significant (Nivière 2003). The interaction between interviewer and respondent varies according to the type of questions. Of the range of possibilities, we can mention three: (1) opinion questions, (2) “factual” questions that are not supposed to require special searches by the respondents, and (3) questions requiring references to documents, administrative or not (pay slips, social-security documents, doctors’ prescriptions, bills, sales slips, etc.). In the first two cases, the original informant is the respondent. By contrast, the third case has something in common with enterprise surveys, which ask for accounts and balance sheets. Here, the interviewer’s job is to help the respondent to pick through the written documents and translate them into the questionnaire language.

The “informant” concept usefully reminds us of the historical development of the sample surveys on agricultural production forecasts in the United States in the 1920s and 1930s (Didier 2002). Initially, the “informants” were farmers who were assumed to be well acquainted with their neighbors and colleagues; they would report the information gathered from their contacts. In other words, informants were both interviewers and respondents. Their “representativeness” was on the same order as that of the informant for an anthropologist studying a population of Amazon Indians. The informant speaks for the tribe members and transmits to the anthropologist a body of information translated—in every sense of the word—into the anthropologist’s language. In the case of U.S. farmers, random sampling procedures were later gradually introduced in the 1930s. Informants became respondents, speaking only for themselves. Their reports were now generalized by virtue of their random selection, and no longer because they were trusted by their neighbors and the Department of Agriculture emissaries. One form of representativeness, that of the random sample, replaced another, that of the well-chosen informant—who is closer to the elected official acting as spokesperson for his or her voters.8 This substitution notably implies (1) a standardization of questionnaires, which makes them resemble administrative forms, and (2) a redistribution of the formatting work.

The entire history of the sampling method can be reread as a manifestation of the tension between the survey approach and the reuse of earlier information, derived in particular from exhaustive administrative sources. From the outset, statisticians had to arbitrate between two justifications of the method consisting in the observation of “a part for the whole”—known respectively as the random-selection method and the purposive-selection

7. A somewhat comparable transformation has occurred with the shift in statistical activities concerning the preparation of the balance of payments by the Bank of France since the creation of the euro and the European Central Bank (ECB).

8. It would be useful to study the extent to which persons interviewed in a political opinion poll see themselves as spokespersons for other citizens.
method—and then combine them in increasingly sophisticated ways:

- Reference to the law of large numbers (selection in a Bernoulli urn), regulated by the determination of a confidence interval expressed in probabilistic terms: random selection, formalized by Arthur Bowley in 1906. Yet the simplest random selection requires a sampling frame consisting of a register or census.

- Use of external information to construct the sample, then to improve the estimation accuracy: purposive selection (Gini 1928), followed (in their distinct ways) by stratified sampling (Neyman 1934), the quota method, and, more recently, modeling based on administrative sources, which enables the information provided by a survey of a national sample to be used at refined geographical levels (Jean-Claude Deville).

These various means of combining survey data and administrative sources have been considerably developed in the past thirty years; they have revolutionized the sampling method by blurring the distinction between the two types of sources. The recent redesign of the French Population Census is an example, since, for municipalities of 10,000+ inhabitants, a survey has been conducted on a sample of dwellings consisting of the “Localized Buildings Register” (Répertoire d’Immeubles Localisés: RIL), a database managed and updated by INSEE with the help of municipalities.

The social division of formatting work

But these increasingly elegant combinations of the two tools presuppose a harmonization of formatting rules, classifications, and definitions of variables used on either side. While theoretically enjoying a greater degree of freedom with surveys, statisticians often prefer to collect data already formatted, for example to comply with legal requirements. That is one of the reasons for their reluctance to endorse “subjective” surveys such as opinion surveys, since, in this case, the task of “categorization” will fall on their shoulders. However, that does not always occur, particularly for surveys that have long been standardized and routinized. One example is confidence surveys of business owners, in which the closed-end three-choice opinion questions and the procedures for interpreting them are practically institutionalized and no longer pose categorization problems.

The usual distinction between surveys and administrative sources may therefore seem to be grafted onto this complicated universe in accordance with institutional segmentations. But it is precisely this social division of formatting work that has sociological significance. Its careful study implies a temporary sidestepping of the issue of “reliability,” on which professional statisticians typically focus for obvious reasons. The successive formatings conducted by different players are informed by deeply dissimilar—if not incommensurable—motives, rationales, technical expertise, and “tricks of the trade.” For instance, as discussed earlier, many management files were initially intended to preserve traces, recorded in conformity with standard rules, of individual agents’ rights, obligations, receivables, debts, and performances: that is typically the case with corporate financial statements and balance sheets. However, when these files are retrieved by statisticians, something is circulating through the pipes, but the “data” flowing into and out of them do not look “the same,” setting aside any possibility of deliberate or involuntary deception. The shift in attention from the individual case to the vision of a whole sums up the transformation taking place. Yet something has traveled from one end of the pipe to the other: the generality principle, the legal rule, the classification, the chart of accounts—all of the frameworks that organized the set of individual cases upstream will enable statisticians, downstream, to prepare their tabulations, even if they rearrange the initial classification principles differently.

This alchemy is even more complex, since the “data” converging from several pipes are then combined and “made coherent.” For example, France’s “Unified System of Enterprise Statistics” (Système Unifié de Statistiques d’Entreprises: SUSE) links the Annual Enterprise Survey (EAE) and the tax returns for firms in the “industrial and commercial” sector (Bénéfices Industriels et Commerciaux: BIC). Such combinations and syntheses of diverse data, derived from surveys and registers, are increasingly common. The most exemplary illustration is provided by national accounting. While this modus operandi is now widely accepted, it is striking to note that, until the 1960s, some statisticians specializing in sample surveys were very reluctant to countenance the notion that their evaluations, whose margins of error were duly bounded by probability theory, should be mingled with and drowned in administrative data, in the national accountants’ tables.

The entire history of the sampling method can be reread as a manifestation of the tension between the survey approach and the re-use of earlier information, derived in particular from exhaustive administrative sources.

9. The revision of the classification of socio-occupational categories between 1978 and 1982 was an opportunity to compare the two situations, one involving the preformatting in quasi-legal terms for categories defined by their status, the other involving the necessary formatting of categories by statisticians themselves, for social groups less well defined in institutional terms (Desrosières and Thévenot 1988).

10. Consider the immigrant from Eastern Europe landing on Ellis Island one morning in 1900. Often, (s)he had to change his/her name. Did the person remain “the same”? That depends on the point of view—existential or sociological. The same may apply to the circulation-transmutation of data in the “pipes” analyzed here. This admittedly far-fetched analogy aims to suggest the question: what does it mean to “keep one’s identity”? There is no single correct answer, and that is what makes the question interesting.
There were two metrologies at work: on the one hand, the computation of confidence intervals derived from a well-built sample; on the other hand, the tradeoffs required for the calibration of the sources-and-uses balances in the national accounts. Their telescoping was indeed the result of what we have examined here: the comparative evaluation of several approaches to linking sources, i.e., surveys and registers. The social solution to this paradox is that the synthetic measure looks like a new object, which has cut the umbilical cord with its parents (registers and surveys), sailed away from its birthplace, and begun another life. This alchemy—which transforms the vile lead of a myriad of individual information items into the (nearly) pure gold of general knowledge—(INSEE 1996, p. 9)—is made possible, in particular, by the technical social authority of official statistical institutions. Historically, the construction of national accounts has been a powerful driver for the comparison and reciprocal critique of disparate sources, chiefly surveys and administrative sources. But our assumption here is that these syntheses, which are justified by the goal of producing an overall, consistent image of macroeconomic flows, have come at a price: the deletion, so to speak, of the sources’ specificities in their initial usage contexts. This does not disqualify the “re-use,” which is justified by its later utility. However, it is an incentive to study the transmutation more closely, through the transfer of the numbers from one usage context to another—a task performed by official statistical agencies. The agencies’ “value added” resides in their strict enforcement of the transfer procedures.

### Circularit of knowledge and action

Administrative sources are derived from management files whose purpose is to process individual cases. But what is important for their later use for statistical purposes is that the processing operations are carried out in accordance with standardized general rules, which are recorded in a variety of general laws and regulations. The initial process was the uniform application of the principles of justice or of management economics. In the downstream stage, it becomes the social and technical foundation of equivalence classes, without which no statistical work could be done. An affinity therefore exists between (1) the forms of generality implemented by the management files and (2) the conventions on which statistical work relies. These affinities are often perceived as a drawback of administrative sources, as they restrict statisticians’ creative imagination; yet they are also the condition of possibility of their work.

---

11. This question will become vital with the implementation of the LOLF, which requires public policies to be endowed with indicators of objectives, resources applied, and results obtained. It can be solved only through a collective examination and complex negotiations involving many parties. When the bill was unanimously voted into law in 2001, its sponsors may not have fully realized that the definition and concrete measurement of indicators was not only a technical matter, to be outsourced to specialists, but a complex operation bringing the system’s entire philosophy into play. In this field, some research on the sociology of quantification would be welcome.

---

An interpretative hypothesis: co-construction (thought-action-description)

How can we move out of this apparent contradiction? One possible way is to define a “triplet” of three elements that are simultaneously co-constructed and interconnected, with the proviso that no single
one determines the other two. The three elements are: (1) the mental perceptions of a society; (2) the forms of action in that society; and (3) the modes of description, particularly statistical description. This construct presupposes a circularity between action and description, which gives a fairly accurate rendering of the logical problems discussed above. Some changes in the modes of production and social uses of official statistics since the 1980s involve, in particular, a greater reliance on administrative sources. These changes are an opportunity to test the co-construction hypothesis proposed here. To explore it, we begin with the vocabulary that developed along with the gradual dissemination of a partly new triplet (Boltanski and Chiapello 1999). One term, admittedly well-established, occurs with ever greater frequency in the social uses of official statistical products: indicator. In the context of these uses, the term “indicator” is tied to three semantic universes, each associated with the three elements of the triplet. Here are some examples of the terms often used in each universe:

- **Mental perceptions of a society**: market, self-regulation, network, competition, governance, quality, competence, uncertainty, temporality, cycle, short term, mobility, flexibility, etc.

- **Modes of action**: incentive, project, innovation, expectation, evaluation, judgment, objective, result, performance, diagnostics, activation of social benefits, best practice, etc.

- **Descriptive tools**: indicator (quantitative, economic, cyclical, social), warning light, dashboard, tracking, observatory, symptom, scale, classification, ranking, benchmarking, microeconometric model, panel, “pure” effect of a variable with all other things being equal, microsimulation, employment rate, etc.

The market/incentive/quantitative-indicator triplet characterizes the space to which official statistics has been partly reassigned since the 1980s. We could conduct an exercise using the vocabulary of the earlier triplets to contrast the present space word-for-word with other spaces that have certainly not disappeared. The words used today result from blends of these different spaces. For example, the name of Claude Gruson could symbolize the “star” triplet of the 1950s and 1960s: Keynesianism; planning; national accounting and measurement of inequality. However, while the first two elements no longer possess quite the same aura, the last two are very much with us, although their meaning and use have changed (for national accounting, see Vanoli 2005). One example of direct interaction between statistical indicators and forms of public action is the now significant use of such tools in the successive treaties that have marked the construction of the European Union (EU). Since the 1980s, Member States’ financial contributions to EU budgets have been linked to their GDP, under what is known as the “fourth resource” procedure. Later, the Maastricht Treaty set guidelines for public deficit and debt in EU countries. To join the common currency—the euro—Member States were obliged to keep their deficit and debt below the levels prescribed by the Treaty. These procedures make national-accounting aggregates play a different role from their earlier one, which did not have such institutional implications.

However, we can study other forms of interactions between “measures” in the public-policy sense and “measures” in the metrological/statistical sense in a more distant past. For instance, the notion of mutual co-construction of thought, action, and description was already explored by Salais et al. (1986) in connection with French unemployment in the 1930s. The authors observed a close correlation between the unemployment rate recorded in each French département in the 1936 census and the département’s urbanization and industrialization rates. In the départements comprising large industrial firms (particularly metalworking and mining companies) and large cities, jobless-assistance offices were active and gave the unemployed a visible, recognized status. This made the reporting of unemployment status in the census questionnaire all the more plausible. The situation underscored the complex interaction between the survey (1936 census) and the activity of the institution (unemployment office), yet this did not foster the relativistic notion that the offices “would create” unemployment. We have here an example of analysis in terms of an intersection (already mentioned)—in this case, between job losses due to the 1930s depression and the action taken to aid its victims. The product of the intersection is the “number of unemployed.” More recently, the “European Strategy for Employment” (ESE), promoted by the EU Summits of Amsterdam and Luxembourg (1997) and Lisbon (2000), has entailed a shift in the indicators deemed relevant to tracking employment. In particular, the classic unemployment rate (unemployed/labor force) has given way to the employment rate (members of labor force in employment/total population in an age group). The ESE, the related indicators, and the discussions about them have been well documented in recent research commissioned by the Directorate for Research, Studies, and Statistics (Direction de l’Animation, de la Recherche, des Études et des Statistiques: DARES) at the French Ministry of Employment, Labor, and Continuing Education: Salais, Raveaud, and Grégoire 2002; Barbier and Sylla 2002; Salais 2004. These studies provide a concrete, detailed
tracking of the gradual implementation of the new triplet outlined earlier, which thus ceases to be a mere working hypothesis. There is a total interaction between (1) job-seeking incentives and assistance programs for the unemployed and (2) the employment structures reflected in administrative sources and labor-force surveys. This observation does not seek to invalidate the indicators based on the two kinds of sources, but rather to incite analysts to apply stringent criteria when interpreting them, particularly for international comparisons.

It would be important to conduct similar research, combining different disciplines, on European policy-making areas and on the official statistics co-constructed about them. Among these policies, we need to distinguish between (1) those that are spelled out in the Treaties as being EU prerogatives (for example, currency and competition) and (2) those that, under the subsidiarity principle, remain under the control of Member States (social policies, employment, education, etc.). For the latter group, however, the European Council and Commission can decide to launch coordinated incentive actions, of which the ESE is precisely an example. In such conditions, the statistical indicators used to evaluate and rank national performances become decisive. That is implicit in the term benchmarking, so often used by EU institutions. Just as with the French LOLF, the discussion on the choice of indicators and of the procedures for measuring them becomes critically important: it is during these negotiations that the thought-action-description triplet is effectively co-constructed. We cannot totally separate a political level where indicators would be “chosen” from a technical level at which their measurement would be “outsourced” as is still often the case.

The cognitive and professional autonomization of the official statistician has long since been demanded and achieved. But it sometimes fosters a sort of implicit naive realism resulting from this division of labor. Of the six “official quality criteria for European statistics,” “relevance” is nominally the responsibility of the commissioning bodies (or “customers”), while the other five (accuracy, timeliness, accessibility, comparability, and coherence) are supposed to be enforced by statisticians. This allocation of tasks of administrative sources and their uses for statistical purposes. It is not enough to regret that they were designed for other purposes; statisticians also need to understand those other purposes well. To do so requires going into the rationales of the institutions involved. In France, this is already a familiar concern for statisticians assigned to the Ministerial Statistical Offices (Services Statistiques Ministériels: SSMs), if they wish to engage in the prior formatting of registers and files later re-used by official statistics. A greater use of administrative sources would be usefully enlightened by an attentive examination of the work performed in a highly active field of political science: policy science.

### A side-trip to the United Kingdom

It is instructive to compare the French context with that of another country, the United Kingdom, where the issues raised here have been discussed since the nineteenth century, but differently. British official statistics go back a long way—with a strong tradition of social surveys—and they are used extensively in public debates, no doubt more than in France. But they are also widely debated and criticized, most notably in scientific circles and the press (Dorling and Simpson 1999). For the past thirty years, the official statistical system has been discussed in many parliamentary and government reports, and has been repeatedly reorganized. In the 1980s, Margaret Thatcher’s conservative government decided that statistical offices should work exclusively to meet government requirements. This position was later eased by the governments of John Major and Tony Blair, but the question (whom do “official statisticians” work for?) continues to be debated, particularly by academics.

The striking phenomenon in France is that academics’ requests and criticisms directed at official statistics (particularly within the National Council for Statistical Information...
Surveys versus administrative records: reflections on the duality of statistical sources

[CNIS)] largely concern access to data 
files, which, they claim, has been made 
too restrictive by confidentiality rules, 
in particular the ones laid down by the 
National Commission on Information 
Technology and Civil Liberties (CNIL). 
In the U.K., by contrast, the recurrent 
criticisms is far broader in scope.16 
can one trust a statistical system that 
works mainly “for the government” 
and not “for society”? This opposition 
does not, of course, overlap with the 
distinction between registers and 
surveys, as some surveys may be 
conducted “purely for government 
needs.” But British academics stress 
that “official statistics” is inherently 
conservative” in the sense of policy- 
confirming, i.e., it reflects the goals 
that government sees and the way in 
which it acts; because it is already 
acting on them.17 This notion 
obviously applies to administrative 
sources, which, by definition, fit into 
that category. Official statistics, the 
academics claim, is less capable of 
identifying new phenomena, which fall 
more within the scope of academic 
research and surveys. Admittedly, 
the U.K. has a strong, long-standing 
tradition of academic surveys and 
expertise. This is rather different from 
the situation in France, where such 
surveys and expertise are more likely to 
fall within the scope of government— 
a reflection of the French tradition of 
cadres (corps) of technical experts. 
Indeed, this explains the behavior of 
French academics, who, for the 
past twenty-odd years, have devoted 
themselves to major empirical studies 
and complain of their lack of easy 
access to data produced by official 
agencies.

But, despite the difference between 
the two national contexts, the 
British academics’ emphasis on the 
opposition between information “for 
the government” and “for society” is 
interesting. It echoes the comments 
of some heads of French ministerial 
statistical offices (SSMs) (Blanc and 
Desrosières 2003). Initially set up to 
facilitate the use of administrative 
sources produced in the ministries, 
the SSMs realized, at a certain point in 
their growth, that to find out about 
the state of society and unresolved 
problems within their sphere of 
competence, those sources do not 
suffice. Surveys are needed. 
This position is clearly voiced, for 
example, by the statistical offices of 
the ministries of social affairs and 
health, of labor, and of education. 
Thus the British academics’ intuition 
has been expressed, albeit differently, 
within the French SSMs themselves. It 
reminds us that surveys and sources 
based on administrative records do 
not meet the same knowledge 
requirements. Surveys explore the 
state of society, while administrative 
sources document public policy. 
As we saw in the second section, 
registers, in a sense, complement 
surveys and can improve their quality. 
But they differ substantially in terms 
of what they show and what they 
help to express.

Conclusion: we need to 
examine the respective 
contributions of surveys 
and registers

France has recently amended 
its 1951 Statistics Act to require the 
transmission of administrative 
data to official statistical agencies 
that request them. The change will 
facilitate the re-use of administrative 
data. The option brings French 
practices closer to those of some 
Northern European countries, where 
the method is more common. 
However, France is not abandoning 
its distinctive tradition of surveys on 
new and under-documented topics, 
particularly in the social field. Rather, 
the amendment is an opportunity to 
take a closer look at the specificities of 
the two types of sources. There 
are two classic arguments in favor of 
administrative sources: (1) they are 
cheaper; (2) their scope of coverage, 
allowing regional and local processing. 
On the downside, they intrinsically 
reflect existing policies, and are ill- 
suited to addressing new issues or 
even to evaluating the effects of 
public policies. France’s recent choice 
therefore caters to the satisfaction of 
new needs (demand for localized 
data) in some cases, and steps away 
from them in others (need to explore society, 
and not only government policies; 
risk of breaks in statistical series if regulations change).

The classic argument that surveys 
offer greater freedom to formulate 
questions must be qualified. Cultural 
and organizational constraints weigh on surveys too, albeit less 
directly than on registers. For 
example, questionnaires heavily rely on the 
cognitive and linguistic categories 
of respondents, or, in the case of 
business surveys, on forms of prior 
information organization, such as the systems of accounting standards 
and practices. There is also a 
frequent, strong demand for surveys of longitudinal panels of individuals or 
families. But it difficult to provide such 
tracking over time through repeated surveys; the term attrition 
denotes the inevitable shrinkage of samples for a variety of reasons. On this 
score, panels based on administrative 
sources undergo fewer losses and 
offer greater security.

However, France should not relinquish 
its system of sample surveys, which 
has proven its worth particularly in the

16. These debates are well reflected in Ray 
Thomas (1999). Some of the ideas are found at 
http://www.socresonline.org.uk/socresonline/ 
1/3/3.html. 
17. This notion is consistent with the thought- 
action-description co-construction hypothesis 
formulated above.
social field. Recent examples include the “Homeless” survey (in response to an explicit CNIS request), and the surveys on “Handicaps, disabilities, and dependencies” (HID), “Life histories, construction of identities,” “Literacy,” and “Neighborhood life.” These surveys largely rely on registers, either as a sampling frame or to build models that promote an improvement in the accuracy of localized estimates. But the surveys also allow the introduction of new questions and topics not explored by registers. One of the main reasons is that the preparation and processing of the surveys involve a wide variety of players: not only statistical offices, but also user organizations (which are often the co-funders), non-profit groups, and social scientists. This form of involvement by users, particularly academics, in the design and processing of social surveys is fairly recent. It partly closes the earlier—and typically French—gap between university research and statistical offices. It helps to design a landscape that remains different from the British one. It may lead to a relationship between academics and statisticians that would not be confined to the requests (admittedly, often justified) for access to individual data, but would integrate the various players into the entire design and production chain. The issue of the duality of sources—surveys and registers—would cease to be a purely technical matter left to official statisticians. It could now be discussed in a thoughtful and contradictory manner by all participants in the process.

The questions raised here do not aim to indicate a preference for one source over another. Rather, they are an invitation to think more deeply about the social and cognitive status of the data produced by the official statistical system, and about the diversity and complementarity of its sources. The use of administrative records is growing, and the value added by statisticians’ data editing raises the quality of the source to a level comparable to that of surveys.¹⁸ In closing, however, it is useful to diversify the assessment of source quality according to the six standard criteria. We have discussed the relevance criterion here, distinguishing between the concepts of information on society and information on public policy. Accuracy cannot be measured with the same tools: on the one hand, there is survey theory (collecting, sampling, etc.); on the other, checking and cleaning data files. Timeliness (or “freshness”) is now an essential criterion. From this standpoint, registers, being managed on a continuous basis, are valuable if not indispensable, but institutional changes often pose interpretation problems for the time series. Accessibility raises the issue of the release of meta-data, which are of an entirely different nature in the two cases. Lastly, comparability and coherence cannot be handled identically for both sources, since the formatting of register-based data depends less on statisticians than the formatting of survey-based data. The comparison between the two types of sources thus requires taking account of diverse quality criteria. It cannot be summarized in simple terms, and calls for an overall reflection on the channels of production and use of official statistics.

¹⁸. However, the technical documentation and teaching materials for these data-editing methods are scarce by comparison with the abundant literature on sample surveys.
REFERENCES


Boydens, I., 1999: Informatique, normes et temps, Évaluer et améliorer la qualité de l'information: les enseignements d’une approche herméneutique appliquée à la base de données "LATG" de l'ONSS, Brussels: Bruylant.


INSEE 1996: Cinquante ans d'INSEE... ou la conquête du chiffre, Paris: INSEE.


Shortage and oversupply of doctors in France: contribution of demographic projections

Traditionally, human-resources planning has sought to forecast the gap between future supply and demand so that resources for closing the gap can be implemented in a timely manner. The healthcare organization and financing system designed in nineteenth-century Germany and later adopted by France does not allow an approach of that kind. The experience of recent decades shows that the gap between personnel supply and demand—i.e., shortage or oversupply—in such a system is highly elusive and even uncertain. Nevertheless, demographic projections of healthcare professions remain very useful, as they allow a preview of system participants’ collective perceptions. These can lead to decisions whose consequences might not be the most desirable two or three decades later.

The experience of the developed countries shows that the demand for medical care is, so to speak, infinitely extendable—owing, among other reasons, to technical progress and everyone’s desire for greater well-being. Only two factors can, if not halt, at least dampen demand in the health sector: the availability of local caregivers and the price that patients themselves must pay.

Nowadays, however, most developed countries have adopted healthcare funding systems in which the patient’s deductible is very low or even zero. Therefore, the race to medical care now depends only on supply—chiefly the supply of doctors.

Theoretically, when supply exceeds demand, the system is overstaffed. When supply fails to meet demand, the system is understaffed. Reality is more complex, as the French experience of recent decades shows.

### Lessons of the past

In the 1980s, there was a broad consensus in France regarding the excess number of doctors. Between 1980 and 1993, the number of “omnipractitioners” (i.e., general physicians) in the private sector alone rose 36%, that of specialists 64% (table 1). During those thirteen years, the French population grew only 7%. Alarms sounded everywhere. Yet doctors’ activity, instead of falling, enjoyed a remarkable expansion. In the same period, the annual number of visits (home and office) rose 6% per general physician and 21% per specialist. Unsurprisingly, therefore, the total volume of visits rose 45% in general medicine and almost doubled in specialties between 1980 and 1993. The number of care providers—i.e., doctors—admittedly grew, but their activity expanded even faster.

Let us now examine the other alternative: a sharp, fast rise in healthcare demand in a stable-supply situation. In this scenario, medical labor paid on a fee-for-service basis quickens its production pace “naturally” (so to speak), preventing any shortage crisis whose visible expression would be an inordinate lengthening of waiting lists.

### Table 1: Annual number of doctors and medical procedures in private sector (1980-1993)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of general physicians</td>
<td>44,123</td>
<td>52,467</td>
<td>58,159</td>
<td>58,947</td>
<td>59,519</td>
<td>59,967</td>
<td>35.9</td>
</tr>
<tr>
<td>Procedures per doctor</td>
<td>4,331</td>
<td>4,454</td>
<td>4,374</td>
<td>4,460</td>
<td>4,610</td>
<td>4,612</td>
<td>6.5</td>
</tr>
<tr>
<td>General-physician procedures (million)</td>
<td>191</td>
<td>234</td>
<td>254</td>
<td>263</td>
<td>270</td>
<td>277</td>
<td>44.7</td>
</tr>
<tr>
<td>Number of specialists</td>
<td>31,748</td>
<td>37,873</td>
<td>49,655</td>
<td>50,662</td>
<td>51,513</td>
<td>51,963</td>
<td>63.7</td>
</tr>
<tr>
<td>Procedures per doctor</td>
<td>3,131</td>
<td>3,546</td>
<td>3,617</td>
<td>3,796</td>
<td>3,807</td>
<td>3,793</td>
<td>21.1</td>
</tr>
<tr>
<td>Specialist procedures (million)</td>
<td>99</td>
<td>134</td>
<td>180</td>
<td>192</td>
<td>196</td>
<td>197</td>
<td>98.3</td>
</tr>
</tbody>
</table>

Source: Système National Inter-Régimes (SNIR) data published by Caisse Nationale d’Assurance-Maladie des Travailleurs Salaries (CNAMTS), metropolitan France.

*Bui Dang Ha Doan is the Director of the Center for Medical Sociology and Demography (Centre de Sociologie et de Démographie Médicales: CSDM), member of the International Institute of Statistics (IIS), and former Secretary-General of the Federation for International Cooperation of Healthcare Services and Systems Research Centers (FICOSSER).

2. In France, the terms “omnipractitioner” (médecin omnipraticien) and “general practitioner” (médecin généraliste) are typically regarded as synonymous. We follow this usage in our article. However, the statistical publications of the French sickness-insurance organizations use a slightly different definition: “the term ‘omnipractitioner’ encompasses both general practitioners and physicians not qualified as specialists who restrict their practice to a specific field such as acupuncture, phlebology, etc.”
The 1960s shortage offers a striking example of the process. The first national agreement between private medical practitioners and social-security agencies was signed in May 1960. The compact allowed a substantial lowering of the money barrier to healthcare for a vast majority of French people. As a result, demand exploded within a few years. Medical careers were suddenly perceived as an eldorado and attracted the young irresistibly. The number of medical-school freshmen rose from 9,880 in October 1963 to 24,448 in October 1967, an increase of almost 150% in four years. This surge vividly illustrates the race to medical care. Yet nowhere in France, at the time, did one hear reports of an extraordinary lengthening of waiting lists or a proliferation of dramatic cases—as occurred in the United Kingdom in the 1990s. But it is a fact that French doctors were far busier in 1968 than at the end of the Fourth Republic, ten years earlier.

These two reminders from the recent past show how hard it is to provide irrefutable proof of a tangible shortage or excess of healthcare providers.

**Reality and collective perception**

However, if actual shortage or oversupply resembles *L’Arlésienne*, the belief in shortage (or oversupply, depending on the period) is very real. During the second half of the 1970s, not a single voice was raised in France to complain about a lack of doctors, the consensus on the glut being shared by all system players. Starting in 2000, a reverse consensus gained ground, with everyone in France agreeing on a severe present or future shortage. The two periods of national consensus are somewhat paradoxical: in 1979, there were 200 practicing physicians registered with the French Medical Council (Ordre National des Médecins) per 100,000 inhabitants in metropolitan France; 25 years later, in 2004, there were nearly 350. In a quarter-century, anxiety of the void replaced the fear of excess numbers, even though the physician/population rate had risen by three-quarters: a remarkable collective achievement! Does this paradox have an explanation?

**The dictatorship of youth**

When, over a decade, the number of new graduates steadily rises every year, the trend unfailingly triggers reflexive fear—fear of the emergence of unbridled competition in ambulatory care, fear of the lack of jobs in hospital departments. As the two concepts of excess numbers and devaluation are commonly linked in people’s minds, a robust and lasting influx of young doctors in the market induces an acute perception of glut. In the opposite case, a prolonged, sizable decline in the number of new entrants per year leads people to believe in a shortage. Even private practitioners—who, by nature, are rather reluctant to see their ranks expand—do not escape the collective belief.

In other words, what conjures up the specter of flooding or drought is the size of the wave of young graduates. In both cases, *neither the number of doctors, nor their density, nor the comparison between healthcare supply and demand* has a perceptible impact, except in certain well-defined geographic areas and for short periods, pending the adjustment between supply and demand. This result, to which our team’s observations have led, is obviously at odds with the views currently expressed in various textbooks. Later, the new theory will find its place there—but only later. For now, an interesting implication has emerged.

**The value of demographic projections**

One of the merits of demographic-projection techniques is to provide a picture—subject to certain assumptions—not only of the future size of the country’s population, but also of its distribution by age and sex. Applied to the medical profession (or any other occupational group), these techniques give an idea of the total number of doctors and their age and sex distribution in ten, twenty, thirty or more years. In a country where the influx of new doctors is controlled by a quota
system, projections enable us to build scenarios that suggest the future consequences of the quota set today. In other words, we can foresee the size of the future influx of young doctors that will result from a given quota adopted now.

As that size is precisely the main determinant of the collective perception of a shortage or oversupply of healthcare providers, we can see the value of using demographic projections.

**The first quarter-century**

In 2003, we prepared a demographic projection of the medical profession on the basis of assumptions concerning university intake, i.e., the number of students authorized by inter-ministerial decision to enroll in PCEM2 (second year of medical studies). This quota, once very low, has been gradually raised since 1999 (chart).

In the first five years of the century, the quota was lifted 61%, the collective perception being dominated by the specter of shortage. In the 2003 projection, at a time when the quotas for 2004 and beyond were not yet known, we relied on two hypotheses. The first was that the quota would gradually rise to 6,000 by October 2007 then level off. The second was that the ceiling would be higher, on the order of 8,000 (table 2). The starting point for the projection was December 31, 2000. At that date, there were 200,800 practicing physicians in metropolitan France (mainland + Corsica) and French “overseas collectivities” (Collectivités d’Outre-Mer).

Under the 6,000 quota hypothesis, the number of doctors would keep rising to 209,300 in 2010 then decline to 183,200 by end-2025. With the quota set at 8,000, the number of doctors would not change in 2010 but the subsequent reduction would be far smaller: there would be 205,700 physicians by 2025, i.e., slightly more than in 2000.

But, far more than the total number of doctors, attention should focus on the number of young practitioners under forty years old. There were 46,600 of them in 2000 but, ten years later, in 2010, the projection puts them at only 17,000. The plunge is obviously due to the low university intake in the 1990s. It explains the reversal of opinion in 2000-2001, when the fear of shortage replaced the obsession with oversupply, as outlined above.

In 2025, the number of young practitioners aged under forty would reach 25,400 if the quota were set at 6,000, but would stand at 33,700 for a quota of 8,000. In other words, in the fifteen years between 2010 and 2025, the number of young doctors would rise by half on the first assumption; on the second one, the number would double. As for the number of older practitioners (aged 55 +), it would remain virtually unchanged over the fifteen-year period irrespective of the quota.

### Table 2: Change in population size of medical profession in France (metropolitan France and overseas collectivities)

<table>
<thead>
<tr>
<th>At December 31</th>
<th>2000</th>
<th>2010</th>
<th>2025</th>
<th>% change (2010-2015)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quota raised to 6,000</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of practicing doctors, of whom:</td>
<td>200,793</td>
<td>209,285</td>
<td>183,154</td>
<td>- 12.5%</td>
</tr>
<tr>
<td>&lt; age 40: number</td>
<td>46,664</td>
<td>16,968</td>
<td>25,431</td>
<td>+ 49.9%</td>
</tr>
<tr>
<td>%</td>
<td>23%</td>
<td>8%</td>
<td>14%</td>
<td></td>
</tr>
<tr>
<td>age 55+: number</td>
<td>28,631</td>
<td>85,904</td>
<td>85,992</td>
<td>+ 0.1%</td>
</tr>
<tr>
<td>%</td>
<td>14%</td>
<td>41%</td>
<td>47%</td>
<td></td>
</tr>
<tr>
<td><strong>Quota raised to 8,000</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of practicing doctors, of whom:</td>
<td>200,793</td>
<td>209,285</td>
<td>205,681</td>
<td>- 1.7%</td>
</tr>
<tr>
<td>&lt; age 40: number</td>
<td>46,664</td>
<td>16,968</td>
<td>33,735</td>
<td>+ 98.8%</td>
</tr>
<tr>
<td>%</td>
<td>23%</td>
<td>8%</td>
<td>16%</td>
<td></td>
</tr>
<tr>
<td>age 55+: number</td>
<td>28,631</td>
<td>85,904</td>
<td>86,634</td>
<td>+ 0.9%</td>
</tr>
<tr>
<td>%</td>
<td>14%</td>
<td>41%</td>
<td>42%</td>
<td></td>
</tr>
</tbody>
</table>

Source: 2003 projection by Center for Medical Sociology and Demography (Centre de Sociologie et de Démographie Médicales: CSDM).

6. In France, since 1971, the number of students allowed to enroll in medical school is set every year by an inter-ministerial decision. Previously, any holder of the baccalauréat (see note 3) could enroll in medical school without preconditions. The system that consists in setting annual university-intake quotas is called numerus clausus (in Latin, “closed number”).
In October 2005, the number of students authorized by inter-ministerial decision to enroll in PCEM2 (second year of medical studies) was 6,200 and the quota for 2007 reached 7,100. Our hypothetical 6,000 has thus been largely exceeded. Today, the 8,000 scenario seems far more probable. One of its consequences should be a surge in the number and proportion of young doctors in the next two decades.

As in the past, the process will invariably trigger a new collective perception of oversupply by 2025—a likely prelude to a new drastic cutback in university intake.

References


A tool for territorial segmentation applied to the health sector

Brigitte Baccaïni*

In 2004, the Regional-Action Service Cluster (Pôle de Service de l’Action Régionale: PSAR) for “Territorial Analysis” supplied INSEE Regional Offices (Directions Régionales: DRs) with a tool enabling them to divide their region into “health territories” under the terms of their partnership with their Regional Hospitalization Agency (Agence Régionale de l’Hospitalisation: ARH).

INSEE set up the six PSARs in 2001-2002 as part of its policy of “consistent regional offering” (Offre Cohérente aux Régions: OCRE). Each cluster handles a specific segment of the Institute’s offering of products and services. Their mission is to help our Regional Offices to provide a consistent and efficient response to requests from players in the regional public debate.

The “Territorial Analysis” Cluster is located at the INSEE Regional Office for the Provence-Alpes-Côte d’Azur region. One of its areas of study is the functioning of territorial units. For this purpose, it offers all Regional Offices the results of its investments in the form of tools such as databases, analytical methods, computer programs, map modules, and standard studies. Since 2005, it has incorporated the “Study and Dissemination” (Cartographie d’Étude et de Diffusion: CED) mission, formerly handled by the CED cluster.

The Cluster’s work program is defined each year at a meeting of the Regional Action Steering Committee (Comité d’Orientation pour l’Action Régionale: COPAR). Two or three investments are programmed and delivered each year to the Regional-Office network. Since its inception, the Cluster has worked on a variety of topics, including transfers of local units, spatial change in urban areas, student migration, the structuring of rural space by “living area” (bassin de vie), and health-sector zoning. All these studies include a strong methodological component.

The segmentation tool applied to the health sector, delivered to the Regional Offices in 2004, is being used in the preparation of the third-generation Regional Health Organization Plans (Schémas Régionaux d’Organisation Sanitaire: SROSs), described in a Health Ministry circular of March 5, 2004.

The “health territories” were intended to replace “health sectors” (secteurs sanitaires), with the aim of improving the organization of healthcare supply. The SROS was also asked to participate in the identification of local-healthcare “access territories.” The Regional Hospitalization Agencies (ARHs) were therefore responsible for revising these territorial boundaries, and many turned to INSEE for help in segmentation.

The tool offered to the INSEE Regional Offices contacted for help by their ARHs comprises four parts: (1) a study of the territory’s coverage by local healthcare facilities (isochrone map); a proposal for segmentation into “health areas” (bassins de santé) based on the analysis of observed flows (hospital stays and visits to doctors, compiled from PMSI³ and URCAM⁴ data); (3) a description of the resulting “health areas,” determined by an analysis of the available socio-demographic data; (4) a comparison of the segmentation with other divisions such as cantons, pays, and “living areas.”

The product was used immediately on delivery by more than half of INSEE’s Regional Offices in partnership with their ARHs. One example is the Languedoc-Roussillon Regional Office.

The latter was contacted in September 2004 by the local ARH with an invitation to collaborate in defining health territories, as part of the preparation of the third-generation Regional Health Organization Plan (SROS). The INSEE Regional Office accepted the proposal and, given the complexity and specificity of the request, decided to participate in the national investment supervised by the “Territorial Analysis” Cluster.

* Brigitte Baccaïni heads the Regional-Action Service Cluster (Pôle de Service de l’Action Régionale: PSAR) for “Territorial Analysis” in the Studies and Dissemination Department (Service Études et Diffusion: SED) of INSEE’s Regional Office (Direction Régionale: DR) in Marseille. She was previously deputy SED head at the INSEE Regional Office in Lyon and a researcher at the National Institute for Demographic Studies (Institut National d’Études Démographiques: INED), where she specialized in mobility issues: residential migration and commuting between home and work.

2. An isochrone map gives a graphic display of access times to facilities. Each facility is surrounded by curves linking points representing locations with identical access times to the facility (Editor’s note).
3. Program for Information-System Medicalization (Programme de Médicalisation du Système d’Information).
4. Regional Union of Sickness-Insurance Funds (Union Régionale des Caisses d’Assurance Maladie).
In June 2005, the segmentation was validated. The regional territory is now divided into:

- 19 “proximity territories” (*territoires de proximité*: map opposite), the first level of hospitalization, which combines (1) technical capability and safety of routine non-specialized hospital care and (2) proximity to “living areas.” Among them, three territories are closely tied to a facility located outside the region.

- 8 “backup territories” (*territoires de recours*), a second tier of hospital facilities to provide more specialized care, as well as an infrastructure and organization that guarantees 24/7 handling of hospital emergencies.

The divisions obtained are consistent with existing territorial units such as “old-age areas” (*bassins gérontologiques*), patient areas (*zones de patientèle*), and “living areas” (*bassins de vie*). The boundaries of the health territories were drawn using PMSI data. Uncertain divisions were settled by an ARH-led working group applying clearly defined criteria.

Example of regional publication using these tools:
“Les bassins de fréquentation hospitalière en Pays de la Loire”
http://www.insee.fr/fr/insee_regions/pays-de-la-loire/publi/publiweb/pagebassins1.htm

Sources: SROS 3 - ARH - INSEE LANGUEDOC-ROUSSILLON
The Office for Health Professions
(Bureau des Professions de Santé: BPS):
an interview¹

Sabine Bessière*

¹ At the time of this interview, Sabine Bessière was deputy head of the Office for Health Professions (Bureau des Professions de Santé: BPS) at the Directorate for Research, Studies, Assessment, and Statistics (Direction de la Recherche, des Études, de l’Évaluation et des Statistiques: DREES), Ministry of Health and Solidarities.

Courrier des statistiques: How does the Office for Health Professions fit into DREES, and what are its missions and partners?

Sabine Bessière: The Office for Health Professions (Bureau des Professions de Santé: BPS) forms part of the Sub-Directorate for the Observation of Health and Sickness Insurance (Sous-Direction de l’Observation de la Santé et de l’Assurance Maladie) at DREES. This Sub-Directorate performs statistical operations and prepares studies on health policies and the health system in France. It compiles research work on population categories and their needs, healthcare facilities and professionals, and the monetary sums involved—analyzed from the standpoint of both funders and recipients.

Within the Sub-Directorate, BPS is tasked with describing the demographics, characteristics, and geographic distribution of (future) health professionals, their activity, working conditions, and income. To this end, BPS works with many partners: the Ministry’s other directorates (Social Security Directorate, Directorate-General for Health, Directorate for Hospital Care and Healthcare Organization), health professionals’ associations, the national sickness-insurance fund for payroll workers (Caisse Nationale d’Assurance Maladie des Travailleurs Salariés: CNAMTS), and the National Federation of Regional Health Observatories (FNORS). BPS maintains close contacts with the National Observatory on the Demography of Health Professions (Observatoire National de la Démographie des Professions de Santé: ONDPS), contributing studies every year to its report. BPS also has ties to other agencies such as the Institute for Research and Documentation on Health Economics (Institut de Recherche et de Documentation en Économie de la Santé: IRDES), researchers, and sociologists to whom it outsources qualitative studies.

BPS regularly prepares medical-demography projections using a micro-simulation model developed in-house. We presented and discussed the 2004 projections in the ONDPS report. As in previous projections, our aim was to chart the medium-term change in the number and distribution of doctors, assuming that past behavior patterns and public decisions remained unchanged over the entire period. In the latest edition, we also simulated scenarios for change that diverged from the baseline “constant-behavior” scenario.

Entry into medical professions is regulated. The number of students allowed to enroll in the second year of medical school is subject to a quota set by the authorities. Likewise, the distribution of students by region and specialty is controlled by an entrance exam at the doctoral level (3rd cycle). Moreover, medical studies are very long: about ten to twelve years, on average, are needed to train a general practitioner or specialist. It therefore takes at least a decade for official decisions to influence the size of the population of health professionals. That’s why it’s essential to project the changes in the number and distribution of doctors.

In parallel, BPS carries out studies on doctors’ behavior—for example, regarding their retirement decisions, their choice of specialty in the national ranking exams (in 2004, a new exam at the start of the doctoral level of medical studies replaced an...
older internship-based system), or their geographic distribution. These studies enable us to track behavioral changes, but also to update and enhance the hypotheses underlying our projection model.

We publish an annual assessment of medical demographics, and very detailed statistics on doctors by specialty, mode of practice, geographic location, sex, age, and other criteria. These data enable us to respond to many requests for information.

BPS is continuously engaged in harmonizing sources. There are many participants in the health sector, and a multiplicity of information sources. The credibility and value of data for shedding light on the public debate depend on their reliability.

We extract medical-demography data from the ADELI register, managed by DREES, in which all doctors (as well as all health professions regulated by the Public Health Code) are required to record their degrees. These data undergo an annual statistical edit, as recommended in the September 2002 report by the National Council for Statistical Information (Conseil National de l’Information Statistique: CNIS) on the harmonization of medical-demography data. The total number of doctors published by the French Medical Association (Ordre National des Médecins) and DREES are totally consistent.

Meanwhile, ONDPS has conducted several methodological studies on source comparisons in the past two years: comparison of population size by specialty or mode of practice, and comparison of number of hospital physicians. These studies have usually allowed a reconciliation of any discrepancies between sources due to differences in concepts or coverage.

**CDS: What are the major projects ahead at BPS, and what issues will they address?**

**SB:** We shall keep contributing actively, through our studies, to the work of ONDPS. In particular, we shall continue to enhance and update the demographic-projection model to take into account (1) the recently observed shift in medical students’ specialty choices since the introduction of national ranking exams, and (2) the expected changes in practice start-up behavior, given the massive retirements that will soon take place in the profession.

BPS is also building demographic-projection model for other health professions: physical therapists, nurses, and, most recently, pharmacists. These projections will be repeated periodically to take account, for example, of public-policy changes. Tomorrow, we plan to undertake in-depth studies of current and future demographics of yet other health professions.

Looking ahead to the longer term, BPS is contributing to the project for the future shared register of health professionals. In a few years, this information system will compile sources from different institutions and harmonize them by offering a single entry point for health professionals, which will simplify their official paperwork. The result should be a better knowledge of their activity and demographics.

Besides demography, BPS studies many other aspects of health professions, such as training, practices, and income. With regard to healthcare practices, we are preparing a survey on the use of specialists, which will tell us more about the nature and motives of their use and the ways in which it varies according to patients’ healthcare histories. Another project concerns the reconstruction of doctors’ total income. This should enable us to analyze all of their professional earnings: at present, only their private-practice income is known.
INSEE’s Permanent Demographic Sample (EDP)¹

Christine Couet*

For nearly forty years, INSEE has been collecting socio-demographic information on a sample of individuals representative of the population of metropolitan France. At present, the Permanent Demographic Sample (Échantillon Demographic Permanent: EDP) is enriched each year with data from annual census surveys and civil-registration records. Panel members were selected to ensure that the sample would remain representative and permanent in time. About 900,000 individual life histories are now being tracked, offering significant opportunities for social and demographic studies. A project is also under way to enlarge and enhance the panel so as to make it an even more useful and practical resource.

In the late 1960s, INSEE set out to keep a record of various information, largely of a demographic nature, that it had been accumulating over the years. It thus decided to create a long-term representative sample—in other words, a panel—of the French population. The 1968 census provided the opportunity to start the project. Thus was born the “Permanent Demographic Sample” (Échantillon Demographic Permanent: EDP), whose evolution-oriented design has made it a major base of individual data.

At the international level, the EDP was not an isolated initiative: its launch coincided with the implementation of its British equivalent, the Longitudinal Study (LS), based on the 1971, 1981, 1991 and 2001 U.K. censuses. In those years, INSEE established many contacts with teams at the City University of London and the Population and Health Department of the Office of National Statistics (ONS), which jointly manage LS production and processing.

The EDP has been steadily enhanced since the 1960s

Today, the French sample has been enriched with more than thirty years’ information on nearly 900,000 individuals who reside or have resided in metropolitan France (mainland + Corsica). The introduction of data from the 1999 population census marks the end of a stage in the sample's history: the period in which the EDP relied on exhaustive censuses—those of 1968, 1975, 1982, 1990, and lastly 1999—supplemented by vital statistics. Since 2004, the annual census surveys seek to enumerate one in seven individuals on average. Henceforth, additions to the EDP will therefore be made on a partial basis but annually.

Like the British sample, the French EDP tracks individuals from one census to the next and is enhanced, when possible, with data from civil-registration records. Its value lies in offering perspective on a set of information items that, while available elsewhere (census and vital statistics), were not previously interconnected. By combining several statistical sources, the EDP also enables specialists to carry out methodological studies, which are difficult if not impossible to conduct with a single source: we can, for example, observe individuals’

* Christine Couet works in the Demographic Surveys and Studies Division of INSEE’s Demography Department.

memory errors, or the distortions due to information-collection conditions. Another valuable aspect of the EDP is its wealth of information on geographic, occupational, and demographic changes concerning its members.

**Individual data are anonymous and available to researchers**

The EDP study database contains no variable allowing direct identification—such as name, date of birth, or registration number in the National Identification Register of Private Individuals (Répertoire National d’Identification des Personnes Physiques: RNIPP).

Data access has been facilitated by the activities of an “exploitation group” tasked with promoting EDP-based studies. Special guidelines allow academics and researchers from outside the Institute to examine the data on INSEE premises: four visitors were hosted in 2005.

**How is the sample constructed?**

One of the main goals of EDP implementation has been to develop a sample that can be continuously enhanced with newly available information. Hence the need to apply the simplest possible selection method while maintaining its cross-sectional representativeness (i.e., at all times). The selection criterion chosen was the date of birth. The database contains all the civil-registration records and all the information consigned in successive census forms for persons born on one of the four reference days: October 1, 2, 3, and 4 of each year. Consequently, at any given point in time, the database is closely equivalent to a 1:100 sample (4:365) of the population residing in metropolitan France.

Sample replacement is automatically ensured by births or by the arrival in metropolitan France of new immigrants documented either in a census or in an event mentioned in a civil-registration record.

Conversely, the sample ceases to track an individual when (s)he dies or emigrates from metropolitan France. However, the person’s trace remains in the sample via a detailed list of the demographic events that have occurred during his or her stay in metropolitan France.

**Type of information collected**

Two sources feed the EDP database: (1) the population census, which characterizes individuals at fixed dates; (2) civil-registration records, which gather information, also of an individual nature, when major family-life events occur.

First, for each sample member, the EDP database collects the information recorded in the individual schedule filled out by the person concerned in the last five exhaustive population censuses of 1968, 1975, 1982, 1990, and 1999. The information gathered covers demographic and social variables: sex, marital status, date and place of birth, nationality, département and municipality of current and previous residence, educational attainment, social category, employment and economic activity, household and family composition, etc. These data are supplemented by information from the census dwelling schedule.

Second, the introduction of events from civil-registration records further enhances the family histories of sample members. Major demographic events initially recorded in civil-registration documents and then incorporated in EDP updates include:

- birth of an individual within EDP coverage (i.e., by definition, birth occurring on one of the four reference days)
- recognition by a parent (if applicable)
- marriage of EDP individual’s parents, an event that can legitimate the individual
- individual’s own marriage(s)
- births of individual’s children
- recognition of individual’s children
- individual’s death.

Typically, these documents describe the circumstances (place and date) of the event and provide the person’s characteristics: sex; date and place of birth; municipality, département, and country of residence; marital status; social category; nationality.

**The current EDP in figures**

In January 2003, the EDP comprised nearly 900,000 individual records, including more than 600,000 persons present in the 1999 population census. Among them, 260,605 were present in the last five censuses (table 1). This figure is high enough to allow studies, including on specific population groups—previously impossible with a survey of only a few thousand people.

Another 128,493 records concern individuals already deceased, whom INSEE nevertheless wanted to keep in the database, and 27,752 records cover persons never enumerated in a census, and hence incorporated into the EDP via a civil-registration record that mentioned them. Of the total sample, nearly 752,000 members (85%) were born in France, versus 135,000 born abroad. The population is described by 1,622 variables, of which a significant number track the same status through diversified sources.

**Problems encountered in EDP data collection**

To track individual life courses properly in the EDP, we need to collect census data in their “raw” form, i.e., before editing. But, to assess the quality of the source, we need to be able to compare the person’s EDP profile with the census coding. Hence the importance of also retrieving the
variables adjusted for non-response, flagged as such in the sample. This dual requirement has not always been fulfilled. For instance, the individual schedules collected for the EDP in the first three censuses were captured then entirely recoded on the basis of the raw data. By contrast, the other census documents—most notably the dwelling schedule—were not captured, which explains the lack of raw family and dwelling data before 1990.

But the nature and scope of information collected on an individual’s status at the time of a census are also determined by whether or not the census database and the EDP are matched. The following instances should be noted:

– The lack of matching after recapture in 1968, bearing in mind than many variables in this census were captured only for a 1:4 sample. As a result, edited variables for the 1968 census are lacking at all levels, both individual and family.

– In 1975, by contrast, the census data were processed exhaustively, and INSEE was able to match the EDP records with the census. This allowed the introduction of variables adjusted for non-response, concerning not only individuals but also their environment (family, household, building, dwelling, and geographic location).

– In 1982, the scope of family-environment data collection was narrowed. Family and household data were extracted only in the “heavy” (or “main”) processing of the census. But the latter was performed on a one-quarter sample of the population. As a result, INSEE was able to retrieve the information only for EDP members included in the one-quarter sample.

Unfortunately, the 1990 census yielded fewer standard raw variables for the EDP than earlier censuses. On the positive side, several newly created variables were introduced in 1990, in particular concerning employment: employment indicator, time spent seeking work to date, job position and detailed description of occupation, hours worked, employment status, and so on.

The 1999 population census saw the introduction of new study criteria for topical issues, such as the year of arrival in metropolitan France for persons born outside metropolitan France, educational attainment, and information on lifestyle, dwelling arrangements, and means of travel to place of work (table 2).

### Gradual extension of census data collected for EDP

Beginning with the 1990 census, the EDP was incorporated into the census processing chain in order to achieve maximum automation of EDP enhancement.

INSEE decided to include in the “heavy processing” of census data all the dwellings containing at least one EDP member. This enabled INSEE to collect raw and edited data on all persons in the EDP sample and on their environment—unlike what was done in 1982.

### Collection channels and dissemination media for EDP data

From 1968 to 1989, EDP data were collected entirely by hand. During that period, when an individual was added to the sample, the INSEE Regional Office (Direction Régionale: DR) covering the person’s place of birth would open a “paper file.” The

### Table 1: Presence of EDP members in population censuses (PCs)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>99PC</td>
<td>617,195</td>
<td>497,699</td>
<td>399,915</td>
<td>327,705</td>
</tr>
<tr>
<td>%</td>
<td>100</td>
<td>65</td>
<td>53</td>
<td>42</td>
</tr>
<tr>
<td>90PC</td>
<td>588,428</td>
<td>470,808</td>
<td>387,580</td>
<td>309,839</td>
</tr>
<tr>
<td>%</td>
<td>100</td>
<td>80</td>
<td>66</td>
<td>53</td>
</tr>
<tr>
<td>82PC</td>
<td>551,211</td>
<td>450,965</td>
<td>362,682</td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>100</td>
<td>82</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>75PC</td>
<td>526,014</td>
<td></td>
<td>418,025</td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>100</td>
<td></td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>68PC</td>
<td>485,009</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How to read this table: Of the 617,195 sample members present in the 1999 census, 497,699 (or 81%) were already present in the 1990 census, 399,915 (65%) were present in the 1990 and 1982 censuses, 327,705 (53%) in the 1990, 1982, and 1975 censuses, and 260,605 (42%) responded to all five censuses.

### Table 2: Type and scope of information collected in successive population censuses (PCs) to feed EDP

<table>
<thead>
<tr>
<th></th>
<th>Raw data</th>
<th>Edited data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual level</td>
<td>Family + dwelling level</td>
</tr>
<tr>
<td>68PC</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>75 PC</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>82 PC</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>90 PC</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>99 PC</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

How to read this table: Raw data from 68PC exist for individuals but not for family and dwelling.

3. In 2004, the census began collecting data on year of arrival in France (and no longer on year of arrival in metropolitan France) of persons born abroad (and no longer on persons born outside metropolitan France).
automation of EDP update channels began in the run-up to the 1990 census.

The redesign project aimed to modernize the media for recording individual files, with magnetic media gradually replacing paper. The creation of a computerized database eased the maintenance burden and facilitated channel reorganization. Census schedules could continue to be managed by the DR responsible for collecting them, avoiding their transfer to the DRs covering the individuals’ birthplaces.

The project also enabled the automatic EDP update to be placed at the end of the census-schedule processing chains, after the schedules had gone through the vital-statistics and RNIPP4 processing channels. This arrangement ensured a continuous update of the EDP database, in a manner consistent with the overall processing of vital statistics.

These advances in EDP management allowed INSEE to make available to users a set of study files updated annually, i.e., without an excessive lag relative to any events that might have occurred in the lives of sample members. To meet the very different needs of EDP staff and researchers, INSEE decided to set up two separate databases: a management database and a study database to be used with SAS software. The first version of the study database was released in 1995. It contained information for the period 1968-1990 extracted from the 1968, 1975, 1982, and 1990 censuses and from civil-registration records from 1968 to 1989.

**Incorporation of 1999 census data**

After the March 1999 population census, the main priority was to enhance the EDP with information on the 633,243 newly enumerated individuals born on one of the four dates covered by the sample. Given the need to identify the enumerated person, the addition of the new data required a matching of the two sources on the basis of the NIR. Unfortunately, the census data are inadequate to allow that identification.

INSEE therefore developed a system for solving the problem, but, for practical reasons, it has proved to be relatively complex. We set up an image base compiled by means of optical reading, which includes a manuscript record of the individual’s vital statistics: last name; first names; sex; date of birth; municipality, département or country of birth. We captured the data by videocoding and selected them in the EDP coverage. Using an image number, we then merged them with the corpus of raw, edited, and coded census data. The person could then be identified via the RNIPP, using first and last names and date and place of birth as search engines. After retrieving the person’s NIR, we were able to incorporate the new data into the existing base.

The automatic identification program has sometimes produced rejects. The total number of contentious items is nearly 75,000, a reject rate of about 12%. After manual processing, many of these items were eventually incorporated and 4,700 discarded. The final failure rate is therefore very low, at less than 1%.

When these operations were completed, INSEE made an initial study database available at its National Computing Center in Orléans in early 2003. It contains census data from 1968 to 1999, births from 1967 to 1999, and marriages and deaths from 1968 to 1997.

In 2003, we compiled a second study database. Besides the information...
already present in the January 2003 sample, it comprises (1) information retrieved on missing deaths, (2) additional vital statistics for 1998 and 1999 (on recognitions and deaths), and (3) several area indicators—such as Tabard socioeconomic typology of neighborhoods and municipalities, urban areas, and “sensitive urban zones”—in order to incorporate specific local data into the study field. The current version of the “study” database is supplemented by the full set of vital statistics up to 2003.

Meanwhile, INSEE has produced a new EDP user manual to make the database easier to exploit for statistical and analytical purposes.

**Day of birth: attraction effect of first day of the month**

The number of sample members varies slightly according to the day of birth. Whatever the census, we consistently find a few more persons born on October 1 than on October 2, 3 or 4 (chart 1). This may indicate the attraction exerted by the first day of the month on individuals who do not know their date of birth. Significantly, the group born on October 1 contains the highest percentage of persons born abroad, particularly in countries where the quality of the civil-registration system is known not to match that of France. The differences in the number of individuals according to day of birth are indicative of small imperfections in the sample’s construction.

**Slow improvement in EDP coverage ratio**

In March 1999, the sample represented 1.05% of the population of metropolitan France, with 617,195 EDP members present in the 1999 census out of 58,513,700 enumerated in metropolitan France.

The EDP coverage ratio increases slightly with each new census and has been trending up in successive cohorts (chart 2). This suggests progress in the identification of EDP members over time and—among younger entrants in
the sample—fewer errors in indicating their date of birth and a lesser lack of knowledge of date of birth (a problem mainly encountered among persons born abroad).

This coverage ratio by birth cohort corresponds, as we have seen, to the number of births occurring on the four reference days as a proportion of total births in the 365 or 366 days of the year. Ignoring birth seasonality (for example, the fact that March births outnumber October births), the EDP should, on average, comprise 1.1% of the population (4/365), a slightly higher coverage than the actual one. This seems to accredit the notion of an omission rate in the sample.

In the past twenty years, the fluctuations in the EDP coverage ratio by birth cohort are fairly consistent with the ones in the ratio of births in metropolitan France occurring on the four reference days to the annual total. These variations are largely due to the presence or absence of a Sunday among the four reference days of a given year: the presence of a Sunday significantly reduces the daily number of births, since the French healthcare system is organized to avoid births on that day. The fluctuations are amplified among recent cohorts because of the increasingly disruptive effect of Sundays on birth levels.

**Longitudinal tracking of individuals across successive censuses**

One of the valuable features of the EDP is that it enables us to explain the behavior of individuals belonging to proximate birth cohorts—for example with regard to migrations, occupational mobility, and fertility—using information collected from censuses and to make intergenerational comparisons.

The sample tracks the history of many cohorts from the 1870s to the present. But they are unevenly represented in the EDP in quantitative terms, and are not equally covered by the five censuses.

Because the sample began with a “snapshot” taken in 1968, the sizes of the oldest cohorts were defined only by the number of survivors at that date (chart 3).

However, from the 1900 birth cohort onward, the percentage of the deceased has fallen to the point where the number of EDP members by cohort is close to that of the number of births registered in the vital statistics for the four reference days, adjusted for migration flows.

Furthermore, since 1968, each intercensal period has seen a massive arrival of new cohorts and a more modest inflow of immigrants of all ages.

Chart 3 displays the number of persons in each birth cohort who are present and absent in successive censuses.

The chart is built around a “census presence” indicator. Taking the 32 possible combinations as our starting point (2^32 cases according to whether the individual is present or not in each of the five censuses), we have aggregated these combinations into 12 more general and relatively homogeneous situations.

The indicator is thus a 5-digit number, each consecutive digit indicating one of the 5 consecutive censuses: read from left to right, the indicator displays the information ranging from the 1968 census to the 1999 census. The digit 1 denotes that the individual is present in that particular census, 0 that (s)he is not. The letter x indicates that the person is “present or absent” in the given census (in other words, the conjunction of the two possible events, “present” and “absent”).

For instance, the indicator begins with the digit 0 for all persons born after 1968—who, by definition, could not be present in that year’s census. Likewise, the indicator begins with
two 0s for persons born after 1975, and so on.

Chart 3 therefore enables us to visualize the links between longitudinal tracking (membership of a given birth cohort) and collection of cross-sectional information (the cohort’s breakdown by presence in successive censuses). This information is valuable for intergenerational comparisons.

Studies made possible by EDP

The EDP’s main purpose is to allow the reconstruction of individual event histories, analyze status changes between two censuses (typically in regard to geographic or socio-occupational mobility), explain the influence of specific circumstances (precariousness, immigration) on life courses, and identify the determinants of a given phenomenon or variable (voter turnout, residential mobility, etc.). The sample can also be used to study fertility (Robert-Bobée 2006).

Another of the EDP’s strengths is that it can help to assess the quality of statistical sources. For example, the identification of enumerated EDP members may lead to an evaluation of double-counting in population censuses.

Similarly, the longitudinal tracking of individuals in the sample gives an estimate of omissions in censuses, mainly among population groups that are hard to capture: the very elderly and young adults. It also enables us to judge the consistency of responses and hence their quality.

The EDP is sometimes used as an initial database for selecting other samples. Its advantage lies in making available to researchers a corpus of census information that may explain a wide variety of behaviors. For example, the matching of election rolls and the EDP is used as a basis for the periodic selection of persons who will contribute to the study of voter turnout.

Lastly, thanks to its composition, the EDP allows the large-scale study (for 900,000 individuals) of multiple family and social histories—at reasonable cost. This is because the initial production of the sample and its later updates, which use data already collected, have been achieved at a far lower cost than if a survey had been necessary. Information collection does not require continuous questioning of sample members, since it needs no additional interview. And data quality is higher, for the collection method does not rely on memory, which is very often unreliable.

Specific constraints to bear in mind when using EDP

Despite the regular addition to the EDP of events from civil-registration records, the compilation of EDP members’ event histories is not a strictly continuous process. Rather, it takes place in stages, sometimes with leaps of several years.

For example, a change in the place of residence or in occupation is not necessarily due to the occurrence of a demographic event. While a census enables us to register the change once and for all, it does not allow us to date it, nor does it avert the risk of overlooking other intercensal changes. Also, the EDP does not take account of events occurring outside metropolitan France.

Longitudinal tracking is not facilitated unless the description of the variables remains stable. But the past thirty years have witnessed, for example, a redesign of the classifications of economic activities in 1975 and occupations in 1982. However, the social category is also available in its most recent definition (1982) for all five censuses concerned, as the 1968 and 1975 schedules were not captured and coded until the 1980s, by which time the new classification was available.

The frequent revision of the geographic boundaries of French municipalities (communes) is a source of inconvenience with respect to the incorporation of census information on current and previous residence. To overcome this problem, the EDP codes municipalities according to the definitions current at the time of each census but also using an updated division that reflects creations and mergers of municipalities.

A promising future

The wealth of information contained in the EDP and the fact that it is a panel of individuals have promoted its use as a source for numerous studies, although the sample still seems to be under-exploited.

The appendix on the following pages lists the published studies based on the EDP. It shows the substantial share of topics concerning occupational and geographic mobility of individuals. By contrast, relatively few studies have addressed “strictly” demographic subjects.

True, the sample has contributed to demographic studies on, for example, differences in mortality according to particular criteria. But the EDP has probably suffered from competition from other sources—mainly the Family Survey (Enquête Famille) and the “Mortality Samples” (Échantillons de Mortalité)—that are better suited to highly focused demographic studies or more easily accessible (to examine EDP data, non-INSEE researchers must file an application with the National Commission on Information Technology and Civil Liberties [Commission Nationale Informatique et Libertés: CNIL];7 access to other databases is less restricted).

Yet these sources do not offer the same wealth of sociodemographic indicators at different dates as the EDP. One of the goals of the “EDP exploitation group” is, therefore, to promote access by researchers to its data. This will enhance the studies undertaken on such topics, as is already the case with mortality studies.

7. Today, this requirement has been waived for researchers who have signed a standard agreement giving them access to the data on INSEE premises.
An EDP redesign is in the planning stage. Its aim is to enlarge and enrich the sample, whose size should be quadrupled from 1% to 4% of the population. For this purpose, the sample is expected to select individuals born on 16 days of each year versus 4 days at present. This will partly make up for the shortfall due to the reduction in the number of persons enumerated in the annual census surveys (on average, one person in seven every year) that replaced the exhaustive general censuses in 2004.

EDP coverage will be widened to French overseas départements, hitherto excluded from the sample. It may also be enlarged to new sources. Initially, these will consist of the election rolls and the INSERM database on causes of death. Looking further ahead, INSEE will examine the technical and legal feasibility of matching the EDP against employers’ annual statements of payroll data (Déclarations Annuelles de Données Sociales: DADSs) and the civil-service payroll database.

The EDP redesign will also include file restructuring to make the database better suited to research work. The sample will comprise interlinked modules, each grouping together the data by topic.

The EDP redesign is scheduled for 2008, concurrently with the implementation of the new application developed for the “database for registers of private individuals” (Base des Répertoires des Persons Physiques: BRPP). 8

---

8. Institut National de la Santé et de la Recherche Médicale (National Institute of Health and Medical Research).
Appendix: chronological list of published studies that have used the EDP

1988

1989

1990

1991

1992

1993

1995

1996

1997

1998
1999

2000

2002

2003

2004

2005

2006
Ten years of statistical surveys in France – 1995-2004

Claude Poulain, Pierrette de Fouchier, and Marie-Mila Grissonnanche-Tartière*

How many statistical surveys are there in France? What are they about? Who conducts them? What are their frequencies, the size of their samples and questionnaires, the types of population they study, their collection methods? How have they changed in a recent ten-year period? To answer these questions, we compiled and analyzed a register describing statistical surveys in France; we confined our study to nationwide surveys included in the program of official statistical surveys published each year in the Journal Officiel (the government paper of record). This article draws on our work to present an initial overview of official statistical studies, although the relative scarcity of information has often narrowed the scope of our analysis and—most significantly—restricted its level of detail.

In 2004, 680 surveys were listed in the program of official statistical surveys published in the Journal Officiel, the government paper of record. Most of them (62%) are conducted only once a year, the rest at different frequencies (half-yearly, quarterly, monthly, and one even weekly). In all, the surveys involved about 2,500 data-collection operations in the field.3

INSEE produces 6% of official statistical surveys, but 70% of data collected

Surveys are like enterprises: some are very small, comprising a few questions put to a handful of statistical units; others are very large—because of the target-population size, the questionnaire size or, sometimes, the two criteria combined. The breakdown of surveys by size is thus at least as important as their absolute number. To measure survey size, we therefore used two indicators: (1) size of population covered by the survey, measured by the number of statistical individuals; (2) questionnaire size, measured by the number of questions asked. The product of the first multiplied by the second gives an initial indication of the “quantity of information” collected by statisticians. The absolute values are of little interest: we therefore broke down this “quantity of information” into percentages among the different types of entities producing statistical surveys (table 1).

The results show that, in terms of the number of surveys and collection operations, INSEE produces only a small proportion (6%), whereas in terms of the quantity of information collected, it clearly predominates (70%). Conversely, the other official producers and entities approved for producing official surveys generate a large number of them (43% of surveys, and as many as 53% of data-collection operations) to yield a very small volume of information (1%).

<table>
<thead>
<tr>
<th>Producer entities</th>
<th>Surveys</th>
<th>Collection operations</th>
<th>“Quantity of information”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage</td>
<td>Number</td>
</tr>
<tr>
<td>INSEE</td>
<td>39</td>
<td>6</td>
<td>140</td>
</tr>
<tr>
<td>Ministerial statistical offices (MSOs)</td>
<td>346</td>
<td>51</td>
<td>1,038</td>
</tr>
<tr>
<td>Other official or approved producers</td>
<td>295</td>
<td>43</td>
<td>1,336</td>
</tr>
<tr>
<td>Total</td>
<td>680</td>
<td>100</td>
<td>2,514</td>
</tr>
</tbody>
</table>

Coverage: statistical surveys included in 2004 program of statistical surveys by French government agencies published in Journal Officiel.

* At the time of writing, Claude Poulain was tasked by INSEE’s Statistical Coordination Department with a mission to set up a register of sources for official statistics. Pierrette de Fouchier and Marie-Mila Grissonnanche-Tartière were assisting him at the INSEE Regional Office in Auvergne.

3. A greater precision, sometimes lacking, should be taken into account: for example the fact that all the surveys announced as monthly are in reality conducted only eleven months of the year, but so-called “quarterly” surveys are performed sometimes three times a year, sometimes four. Also, if a new, high-frequency survey (for example, monthly), is launched in mid-year and not in January, the number of its occurrences will necessarily be less than eleven.
4. Such as the producers of surveys of individual industries (enquêtes de branches).
93% of surveys concern enterprises, but 84% of data collected concern surveys of private individuals.

While enterprise surveys are by far the largest category (93%), the bulk of data collected (84%) comes from personal surveys—i.e., of private individuals or households—(table 2).

Three-quarters of official statistical surveys are surveys of individual industries.

Of the 680 surveys carried out in 2004, 508—i.e., three-quarters—were surveys of individual industries (enquêtes de branches industrielles). Their purpose is to provide two types of information: (1) structural data, gathered by annual surveys on production levels broken down by product; (2) short-term data collected by sub-annual surveys tracking changes in production. For the sake of precision, the information must concern well-identified products, hence the large number of specific questionnaires tailored to each industry or to groups of fairly similar industries. Besides providing information for trade organizations and business firms themselves, these surveys are ultimately intended for national-accounting purposes.

The structure of their questionnaires is highly consistent: they always aim to measure production in real (i.e., volume) and nominal (i.e., value) terms. The differences between questionnaires concern the list of products specific to the industry covered. They must be accurately identified and observed in the required statistical units when the goal is to determine production volume.

Most of these surveys are conducted by industry federations or trade organizations among their members with the approval of CNIS, SESSI, and SCEES. This coordination leads both statistical offices to conduct some surveys of their own in industries not covered by trade organizations, or in industries not surveyed by their representative bodies.

Industry surveys have been compulsory since their introduction, a characteristic that requires their inclusion in the programs of official surveys published in the Journal Officiel. Under current arrangements, this requires their review by various CNIS units, their examination by a specialized CNIS “task force” (to obtain a “usefulness opinion”) and by its Quality Label Committee (to receive “public interest” and “statistical quality” certification). The large number of industry surveys has led CNIS to adopt a streamlined procedure. The Council defined a quality charter in 2001, to which industry organizations have been invited to comply. Registration with the Journal Officiel is now processed on a collective basis: surveys conducted with identical frequency and by the same type of producer are reported together, with a listing of the classification codes of the industries covered.

But this simplified paperwork is exceptional and temporary. At the earliest opportunity—for example, when a survey is redesigned or the questionnaire is expanded beyond...
its initial objective—industry surveys must follow the standard procedure for official-survey programming. For instance, the *Journal Officiel* for 2004 lists eight individually declared industry surveys, with the survey title, producer entity, frequency, and coverage.

Information on industry surveys certified under the simplified procedure is less easily available than information on surveys certified under the standard procedure. For the purposes of our study, we have therefore decided not to compile fuller documentation on the first category. The task would have taken a long time to complete, for a modest information yield relative to our goal. Consequently, the results below concern only the surveys identified in the official-surveys program. In addition to the eight industry surveys approved under the standard procedure (table 3), our study will therefore focus on surveys other than those of specific industries.

The official statistical system (INSEE + ministerial statistical offices) conducts 90% of official surveys other than surveys of individual industries

Under the scope of coverage defined earlier, we enumerated 180 surveys in 2004, of which 131 were compulsory and 49 were not (table 4).

<table>
<thead>
<tr>
<th>Non-compulsory surveys</th>
<th>Compulsory surveys</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ent</td>
<td>indiv</td>
<td>ent+emp</td>
</tr>
<tr>
<td>INSEE: DSDS(^{10})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INSEE: DSE and DESE(^{11})</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Agriculture-SCEES</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Labor-DARES(^{12})</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Health-DREES(^{13})</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Education-SDES(^{14})</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Infrastructure-SESP(^{15})</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Industry-SESSI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other MSOs</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cooperative survey</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>16</td>
</tr>
</tbody>
</table>

ent: enterprise emp: employees

When a survey is compulsory, a surveyed unit that fails to respond after a final reminder is liable to an administrative fine. This rule is effectively enforced for enterprises, whose non-response can significantly impact survey results.

“Cooperative surveys” (table 4) denote surveys conducted jointly by INSEE and a ministerial statistical office (MSO; in French: *service statistique ministériel* or SSM) in five out of six cases, or by several MSOs in a single case.

The breakdown by producer entity reveals the weight of the official statistical system, comprising INSEE and the MSOs: the system conducts 163 out of 180 surveys, or 90% of official surveys excluding surveys of individual industries.

**Surveys by size of population surveyed**

The size of populations covered (in this instance, samples) by personal surveys is generally larger than that of populations covered by enterprise surveys: the first cover a population whose modal value ranges between 5,000 and 10,000 statistical units (table 5, p. 42), versus under 500 for the second category. As regards enterprise surveys, the sample sizes would certainly have been smaller had we separated healthcare and educational institutions from market-sector enterprises. DREES and SDES produce 26 surveys of “enterprises” (table 4 above) that are, in fact, healthcare and educational institutions; about 10 of these surveys cover samples of more than 5,000 units.

On average, questionnaires cover more variables in household surveys than in enterprise surveys

The most meaningful indicator of “quantity of information” collected would probably be the number of quantified variables included in the file generated from the survey. This information is not easy to

11. Direction des Statistiques d’Entreprises (Business Statistics Directorate) and Direction des Études et Synthèses Économiques (Economic Studies and National Accounts Directorate). Short-term business surveys are produced by DESE.
gather, owing to the difficulty of distinguishing between “primary” variables, “computed” variables, variables retrieved from other sources, and so on. We therefore estimated the indicator from the initial survey questions and the “code boxes” (bacs de chiffrement) that appear in the questionnaires. The enumeration is not always easy, so the figures that follow should be viewed as an approximation. Absent other gauges, they at least provide an initial estimate.

Personal surveys tend to use large questionnaires (47% of them collect more than 200 variables; table 6), whereas enterprise surveys use much shorter questionnaires (57% of them have questionnaires that collect 100 variables or less). Our earlier remark on the influence of the presence of DREES and SDES surveys among surveys of local units (establishments) and enterprises remains valid (the two MSOs survey healthcare and educational institutions respectively). Here as well, the result is that the average questionnaires are bigger than those used in surveys of market-sector enterprises.

**Mail-only surveys for enterprises, interviewer surveys for households**

A rather logical conclusion from the findings above is that personal surveys are mainly conducted by interviewers (alone or backed up by mail and telephone), whereas most enterprises surveys use mail only (table 7).

---

16. Particularly because of the fact that many questions are asked only on the basis of responses to previous questions (these are known as “filtered” questions). In this case, we have made assumptions about the frequency of occurrence of sub-questionnaires.

17. Sometimes, there is a major discrepancy between the number of variables collected and the number of variables in the file generated by the survey. New variables are created by combining initial variables, others are recoded, and a third group is introduced from other survey files or administrative sources.

---

### Table 5: Surveys of persons and enterprises: breakdown by number of statistical units surveyed

<table>
<thead>
<tr>
<th>Number of units</th>
<th>Persons Number</th>
<th>Percentage</th>
<th>Enterprises Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 200</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>201-500</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>501-1,500</td>
<td>2</td>
<td>5</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>1,501-5,000</td>
<td>7</td>
<td>16</td>
<td>37</td>
<td>29</td>
</tr>
<tr>
<td>5,001-10,000</td>
<td>15</td>
<td>35</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>10,001-50,000</td>
<td>13</td>
<td>30</td>
<td>26</td>
<td>20</td>
</tr>
<tr>
<td>&gt; 50,000</td>
<td>6</td>
<td>14</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>43</td>
<td>100</td>
<td>128</td>
<td>100</td>
</tr>
</tbody>
</table>

Coverage: official surveys excluding surveys of individual industries, whose statistical units are enterprises or persons, conducted in France in 2004 and for which information on the number of units was available.

### Table 6: Surveys of persons and enterprises: breakdown by questionnaire size (number of variables collected)

<table>
<thead>
<tr>
<th>Number of variables collected</th>
<th>Persons Number</th>
<th>Percentage</th>
<th>Enterprises Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 50</td>
<td>4</td>
<td>9</td>
<td>39</td>
<td>30</td>
</tr>
<tr>
<td>51-100</td>
<td>5</td>
<td>12</td>
<td>35</td>
<td>27</td>
</tr>
<tr>
<td>101-200</td>
<td>14</td>
<td>33</td>
<td>24</td>
<td>19</td>
</tr>
<tr>
<td>201-500</td>
<td>15</td>
<td>35</td>
<td>27</td>
<td>21</td>
</tr>
<tr>
<td>&gt; 500</td>
<td>5</td>
<td>12</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>43</td>
<td>100</td>
<td>128</td>
<td>100</td>
</tr>
</tbody>
</table>

Coverage: official surveys excluding surveys of individual industries, whose statistical units are enterprises or persons, conducted in France in 2004 and for which information on the number of variables collected was available.

### Table 7: Surveys of persons and enterprises: breakdown by collection method

<table>
<thead>
<tr>
<th>Collection method</th>
<th>Persons Number</th>
<th>Percentage</th>
<th>Enterprises Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mail</td>
<td>8</td>
<td>19</td>
<td>92</td>
<td>72</td>
</tr>
<tr>
<td>Interviewers</td>
<td>23</td>
<td>53</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Interviewers + telephone</td>
<td>5</td>
<td>12</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Mail + interviewers</td>
<td>7</td>
<td>16</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>43</td>
<td>100</td>
<td>128</td>
<td>100</td>
</tr>
</tbody>
</table>

Coverage: official surveys excluding surveys of individual industries, whose statistical units are enterprises or persons, conducted in France in 2004 and for which information on the collection method was available.

### Table 8: Surveys of persons and enterprises: breakdown by frequency

<table>
<thead>
<tr>
<th></th>
<th>One-time</th>
<th>&gt; 1 yr</th>
<th>Annual</th>
<th>HY</th>
<th>Q</th>
<th>M</th>
<th>W</th>
<th>N.D.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persons</td>
<td>9</td>
<td>7</td>
<td>17</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td>43</td>
</tr>
<tr>
<td>Enterprises</td>
<td>5</td>
<td>19</td>
<td>56</td>
<td>5</td>
<td>15</td>
<td>22</td>
<td>1</td>
<td>5</td>
<td>128</td>
</tr>
</tbody>
</table>

Coverage: official surveys excluding surveys of individual industries, whose statistical units are enterprises or persons, conducted in France in 2004. HY: half-yearly survey; Q quarterly; M: monthly; W: weekly.

**Survey frequency**

High-frequency surveys are designed to meet the needs of short-term analysis. They are rare among household surveys (only four), but common among enterprise surveys (nearly one-third) (table 8).
Topics covered by surveys

The range of topics covered by survey questionnaires would deserve a study in its own right. The compilation of questionnaires and their recording in digital form offer a valuable resource for such an investigation. At a basic level, it allows the questionnaires to be explored with search engines. When a new survey is being prepared, designers can thus examine the existing corpus far more thoroughly than by relying on survey titles alone or on lists of topics covered—when such lists are available. At a higher level, using linguistic-analysis software, one can “map” the subjects dealt with in the questionnaires and the intensity of their coverage. Pending such analyses, we shall confine our examination here to a summary breakdown cross-tabulating the main survey topic and the surveying entity.

The combination of the two criteria—producer entity and main topic—gives an image of the specialization rate in the official statistical system (table 9). SCEES, the statistical office of the Agriculture Ministry, conducts 92% of surveys on agriculture. DARES and DSDS together produce 81% of surveys on employment and income. DSDS produces 58% of surveys on housing and living conditions, SESP and INSEE’s DSE + DESE together conduct 57% of non-farm enterprise surveys. SDES, the statistical office of the Education Ministry, performs 64% of surveys on education. DREES carries out 60% of surveys on health, and so on.

The figures therefore confirm the system’s specialization, but—with the exception of agriculture—we consistently find several producer entities working in a single broad area.

Changes between 1995 and 2004

How has the program of official statistical surveys changed in the past decade?

It is not easy to give a precise answer to this question. First, as we have seen, surveys of individual industries made up a sizable share of total surveys in 2004 (more than 500 out of 680). The compilation of data on individual-industry surveys in 2004 was not a straightforward task and required specific queries to SESSI and SCEES. To obtain the same information on the situation in 1995 would have called for even greater investigative work (not to mention that, in order to document changes during the ten-year period, we would need data not only for 1995 and 2004, but also for all the years in between). The experts in the area, at SESSI and SCEES, believe that the number of individual-industry surveys is broadly identical, and that most of the change involved the gradual transfer from approved organizations to ministerial statistical offices (MSOs)—which, incidentally, has led to greater consistency in questionnaires and processing.

The following analysis of change therefore remains confined to surveys other than those of individual industries.

From 1995 to 2004, the number of surveys rose by nearly 50% from 121 to 180. This increase is all the more significant—and surprising—as the period studied witnessed far more stringent budget constraints than earlier periods.

18. N.D.: not determined. That is not for lack of trying: we were unable to determine the frequency of eleven official surveys!
19. Such as Tropes.

Table 9: Breakdown of surveys by main topic covered and producer entity (column percentages)

<table>
<thead>
<tr>
<th></th>
<th>Agriculture</th>
<th>Industrial production</th>
<th>Employment, income</th>
<th>Education</th>
<th>Housing, living conditions, demography</th>
<th>Health</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture-SCEES</td>
<td>92%</td>
<td>2%</td>
<td>5%</td>
<td>8%</td>
<td></td>
<td></td>
<td>23%</td>
</tr>
<tr>
<td>INSEE: DSE and DESE</td>
<td>37%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13%</td>
</tr>
<tr>
<td>Infrastructure-SESP</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>Cooperative surveys</td>
<td>3%</td>
<td>5%</td>
<td>9%</td>
<td></td>
<td></td>
<td></td>
<td>4%</td>
</tr>
<tr>
<td>Industry-SESSI</td>
<td>6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2%</td>
</tr>
<tr>
<td>Labor-DARES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>63%</td>
<td>8%</td>
</tr>
<tr>
<td>Education-SDES</td>
<td>10%</td>
<td>5%</td>
<td>64%</td>
<td></td>
<td></td>
<td></td>
<td>13%</td>
</tr>
<tr>
<td>INSEE-DSDS</td>
<td>3%</td>
<td>18%</td>
<td></td>
<td>58%</td>
<td></td>
<td></td>
<td>9%</td>
</tr>
<tr>
<td>Health-DREES</td>
<td>5%</td>
<td></td>
<td>8%</td>
<td>60%</td>
<td></td>
<td></td>
<td>6%</td>
</tr>
<tr>
<td>Other MSOs</td>
<td>5%</td>
<td>2%</td>
<td>4%</td>
<td></td>
<td></td>
<td></td>
<td>3%</td>
</tr>
<tr>
<td>Other producers</td>
<td>10%</td>
<td>16%</td>
<td>16%</td>
<td>40%</td>
<td></td>
<td></td>
<td>9%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td></td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Commentary: the rows have been arranged to display a diagonal showing the relative specialization of producers, from SCEES, which covers 92% of agricultural topics, to DREES, which deals with 60% of health-related topics.
However, the breakdown by producer entity reveals major differences between (1) "established" producer entities (such as INSEE, SCEES, and SESSI), whose number of surveys posted a below-average increase (number unchanged for SESSI, multiplier coefficients of 1.2 for SCEES and 1.2–1.3 for INSEE DSE + DESE), and (2) more recent entities, the other MSOs, DREES, DARES, and other producer entities, whose number of surveys grew much faster (multiplier coefficients of 2.5 for the other MSOs, 2.8 for DREES and DARES, and as much as 3.4 for the other producer entities—for which the initial number was admittedly low) (table 10). These figures therefore suggest a “catch-up” effect: if the new entities were set up, the aim was to fill a knowledge gap, particularly by conducting surveys.

But the size of the increase in the number of surveys in such a short period (ten years) is also explained by a purely enumeration-related phenomenon: of the sixty-odd additional surveys, most are certainly "new" in that they did not exist previously, but others are old surveys whose "novelty" is merely due to their having been recognized and recorded in the official-surveys program. It has long been accepted that only compulsory surveys needed to be included in the program. As a result, many non-compulsory surveys did not appear in the program published in the Journal Officiel. 20

When CNIS set up a Quality Label Committee to examine all surveys applying for inclusion in the official program, it was not possible to add these surveys overnight. The Journal Officiel list has thus incorporated old surveys that predate the establishment of the Quality Label Committee and have therefore never been reviewed by it; similarly, it does not comprise existing and more recent surveys that have not been processed via the simplified ad hoc procedure.

What percentage of official statistical surveys is not (yet) included in the official program?

Given the difficulty of obtaining information on official surveys, it is understandably even harder to gather information on the others! However, some information does exist, notably in reports on the status of statistics in particular fields. For example, since 1995, three reports have been produced, one of whose intermediate goals was to “provide an overview of available sources.” The three documents are (1) a report by a CNIS working group on “statistical information on population health status,” 21 (2) a report by the Center for Studies on Employment (Centre d’Études de l’Emploi: CEE) on behalf of DREES and DARES entitled “Assessment of sources in the field of health and labor,” 22 and (3) a report by another CNIS working group on “statistical information on medicines.” 23 For all three groups, a large share of the work consisted in preparing an inventory of existing sources—a task regarded as a prerequisite for analysis and issuing recommendations.

In all three reports, the search for sources naturally applied quality criteria to justify their inclusion in the “asset base” of statistical surveys: for example, the survey methodology and questionnaire had to be accessible and the results made public.

We can summarize the information on sources listed in the three reports as follows (table 11).

### Table 10: Number of statistical surveys by producer entity: change between 1995 and 2004

<table>
<thead>
<tr>
<th>Producer Entity</th>
<th>1995</th>
<th>2004</th>
<th>Multiplier Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other producer entities</td>
<td>5</td>
<td>17</td>
<td>3.4</td>
</tr>
<tr>
<td>Labor-DARES</td>
<td>5</td>
<td>14</td>
<td>2.8</td>
</tr>
<tr>
<td>Health-DREES</td>
<td>4</td>
<td>11</td>
<td>2.8</td>
</tr>
<tr>
<td>Other MSOs</td>
<td>2</td>
<td>5</td>
<td>2.5</td>
</tr>
<tr>
<td>Infrastructure-SESP</td>
<td>16</td>
<td>23</td>
<td>1.4</td>
</tr>
<tr>
<td>Education-SDES</td>
<td>13</td>
<td>18</td>
<td>1.4</td>
</tr>
<tr>
<td>INSEE: DSE and DESE</td>
<td>19</td>
<td>24</td>
<td>1.3</td>
</tr>
<tr>
<td>Agriculture-SCEES</td>
<td>33</td>
<td>41</td>
<td>1.2</td>
</tr>
<tr>
<td>INSEE: DSDS</td>
<td>14</td>
<td>17</td>
<td>1.2</td>
</tr>
<tr>
<td>Industry-SESSI</td>
<td>4</td>
<td>4</td>
<td>1.0</td>
</tr>
<tr>
<td>Cooperative surveys</td>
<td>6</td>
<td>6</td>
<td>1.0</td>
</tr>
<tr>
<td>Total</td>
<td>121</td>
<td>180</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Commentary: entities are ranked by decreasing multiplier coefficient.

### Table 11: Breakdown of unofficial sources by type (statistical or administrative), and by report in which they are cited

<table>
<thead>
<tr>
<th>Type of Source</th>
<th>Report on “statistical information on population health status”</th>
<th>Report on “health and working career”</th>
<th>Report on “statistical information on medicines”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical processing of management files</td>
<td>15</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Statistical surveys</td>
<td>28</td>
<td>62</td>
<td>10</td>
</tr>
<tr>
<td>of which: published in Journal Officiel</td>
<td>8</td>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>43</td>
<td>68</td>
<td>15</td>
</tr>
</tbody>
</table>

20. For example, INSEE has conducted short-term business surveys since the 1960s. Initially non-compulsory, they relied on voluntary cooperation by respondent enterprises, and did not enter the Journal Officiel program until 1995. Some became compulsory in 2004.


The first type of survey consists in the statistical processing of files prepared for administrative management purposes. This type of source has been expanding steadily in the past twenty years or so, but is not yet covered by a programming procedure and is not yet publicized in the same way as statistical surveys.24

Statistical surveys remain the most abundant sources. We observe that a very large majority of surveys described in these reports do not appear in the official-surveys program. But who, then, are their producers?

Thirteen surveys are prepared by CNRS25 research units, thirteen others by INSERM,26 and a further thirteen by other national institutes such as INED,27 INRETS,28 and INVS.29 Ten are produced by ministerial statistical offices (DREES and DARES), and two by European Union organizations (Eurofound30 and the European Commission’s Directorate-General for Employment and Social Affairs). The three reports also included about ten private sources31 in their censuses.

The involvement of producer entities in survey-programming procedures thus differs according to their proximity to INSEE: for units with the closest ties—the ministerial statistical offices (MSOs)—this participation is almost routine. For institutes with somewhat looser ties to the Institute (INED, CEREQ32, CEE, CERC,33 etc.), participation depends on the survey’s subject and importance; many other agencies that are more specialized or more remote from INSEE seldom become involved.

The situation changed during the period studied. In 1995, only five surveys in the official program were conducted by “outside” official producers: the National Civil-Engineering Industry Federation (Fédération Nationale des Travaux Publics) (two surveys), the French Building-Industry Federation (Fédération Française du Bâtiment), the Bank of France, and the Energy Observatory (Observatoire de l’Énergie).

In 2004, we find seventeen surveys carried out by the same producers (the Bank of France now with two surveys), plus five others: CEREQ (three surveys), INED (two), OFDFT34 (two), CREDES,35 and AFSSA.36 These entities were sufficiently associated with CNIS activities to submit their surveys to the standard procedures, either because they engage in frequent exchanges—particularly of researchers and statisticians—with INSEE (examples include INED, CEREQ, and OFDFT), or because they cooperate on specific projects such as the ten-year health survey, which led CREDES and INSEE to join forces.

### Changes in selected survey characteristics, 1995-2004

Between 1995 and 2004, the number of personal surveys has risen faster than of enterprise surveys (table 12). The breakdown between “compulsory” and “non-compulsory” surveys shows a mild shift toward the second category, from 21% to 27%.

---

24. For example, CNIS, while often informed of their existence, does not have a say in these surveys, which are not reviewed by the Quality Label Committee either.25. Centre National de la Recherche Scientifique (National Center for Scientific Research).

26. Institut National de la Santé et de la Recherche Médicale (National Institute of Health and Medical Research).

27. Institut National d’Études Démographiques (National Institute of Demographic Studies).


29. Institut National de Veille Sanitaire (National Health Watch Institute).

30. European Foundation for the Improvement of Living and Working Conditions.

31. They consist of surveys conducted by pharmaceutical firms or pharmacy trade organizations among pharmacies or prescribing physicians.

32. Centre d’Études et de Recherche sur les Qualifications (Center for Studies and Research on Skills).


34. Observatoire Français des Drogues et des Toxicomanies (French Observatory for Drugs and Drug Addiction).

35. CREDES: Centre de Recherche, d’Études et de Documentation en Économie de la Santé (Center for Research, Studies, and Documentation on Health Economics), now IRDES: Institut de Recherche et de Documentation en Économie de la Santé (Institute for Research and Documentation on Health Economics).

36. AFSSA: Agence Française de Sécurité Sanitaire des Aliments (French Food Safety Agency).

---

<table>
<thead>
<tr>
<th>Type of statistical unit</th>
<th>Number, 2004</th>
<th>Number, 1995</th>
<th>Multiplier coefficient measuring 2004/1995 change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non compulsory</td>
<td>Compulsory</td>
<td>Total</td>
</tr>
<tr>
<td>Persons</td>
<td>16</td>
<td>27</td>
<td>43</td>
</tr>
<tr>
<td>Enterprises</td>
<td>27</td>
<td>101</td>
<td>128</td>
</tr>
<tr>
<td>Enterprises + employees</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Other stat. units</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>131</td>
<td>180</td>
</tr>
</tbody>
</table>

Coverage: official surveys excluding surveys of individual industries, whose statistical units are enterprises or persons, conducted in France in 2004 and for which information on listed characteristics was available.
The breakdown by topic was distorted by the emergence or build-up of new MSOs, particularly in the health and education spheres (table 13).

When a new survey was conducted, it was submitted to a CNIS task force beforehand. This gives us information not only on its objectives, scope of application, collection method, and so on, but also on its frequency. When the frequency was not specified, we have assumed here that the survey would not be repeated. From 1995 to 2004, “one-time” surveys—developed for a single data-collection operation—posted a spectacular increase: their number was multiplied by 3.3 (table 14).

In the period studied, the largest increase in personal surveys concerned medium-sized samples (in particular, 1,500-5,000 statistical units: table 15). By contrast, sample sizes for enterprise surveys remained practically stable (the steep apparent increase is actually due to DREES and SDES surveys of healthcare and educational institutions).

What happened between 1995 and 2004?

After examining the 2004 surveys, the 1995 surveys, and the changes in selected characteristics of the surveys at both dates, we need to explore what happened between the two dates.

A survey may be viewed as a “machine” built to produce results from a chosen population in the form of a data file. This “machine” is the result of a design process, of an investment (questionnaire, methodology, computer programs, documentation, etc.) that will serve once (if the survey is a one-time operation) or, on the contrary, will be written off over several “editions” if the survey is repeated at a given frequency. But a “machine” can wear out or become obsolete, and its producer may replace it with another one that is more modern or better suited to present needs. Surveys therefore experience a “life cycle” whose analysis may be quite valuable in economic terms but also from a functional perspective in order to understand changes in the statistical system.

While 121 surveys were “active” in 1995 and 180 in 2004, we find 330 that were “active” at one time or another during the period.

<table>
<thead>
<tr>
<th>Table 13: Number of surveys by main topic: 1995-2004 change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main survey topic</strong></td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>Health</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Employment, income</td>
</tr>
<tr>
<td>Housing, living</td>
</tr>
<tr>
<td>conditions,</td>
</tr>
<tr>
<td>demography</td>
</tr>
<tr>
<td>Industrial production</td>
</tr>
<tr>
<td>Agriculture</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Coverage: official surveys excluding surveys of individual industries, whose statistical units are enterprises or persons, conducted in France in 2004 and for which information on main topic was available.

<table>
<thead>
<tr>
<th>Table 14: Number of surveys by frequency: 1995-2004 change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survey frequency</strong></td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>One-time or unknown</td>
</tr>
<tr>
<td>Five-year</td>
</tr>
<tr>
<td>From over 1 year to</td>
</tr>
<tr>
<td>less than 5 years</td>
</tr>
<tr>
<td>Annual</td>
</tr>
<tr>
<td>Quarterly</td>
</tr>
<tr>
<td>Half-yearly</td>
</tr>
<tr>
<td>Monthly</td>
</tr>
<tr>
<td>Weekly</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Coverage: official surveys excluding surveys of individual industries, whose statistical units are enterprises or persons, conducted in France in 2004 and for which information on frequency was available.

<table>
<thead>
<tr>
<th>Table 15: Number of surveys by sample size and type of statistical unit: 1995-2004 change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of statistical units in surveys</strong></td>
</tr>
<tr>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>&lt; 200</td>
</tr>
<tr>
<td>201-1,500</td>
</tr>
<tr>
<td>1,501-5,000</td>
</tr>
<tr>
<td>5,001-10,000</td>
</tr>
<tr>
<td>10,001-50,000</td>
</tr>
<tr>
<td>&gt; 50,000</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Coverage: official surveys excluding surveys of individual industries, whose statistical units are enterprises or persons, conducted in France in 2004 and for which information on number of statistical units was available.
Ten years of statistical surveys in France – 1995-2004

96 surveys survived the decade: they were present in 1995, and still in existence in 2004 (table 16). This is a sizable figure, implying that a survey’s amortization period now exceeds a decade—in contrast to the situation prevailing in the 1970s-1980s, when advances in information technology induced more frequent reprogramming. However, we should not rule out the possibility that the number is due to a purely “statistical” vision of surveys: if one of them has been redesigned only in its IT aspects, with the statistical design remaining strictly identical, the change will likely go unnoticed. It is the same survey, IT strictly identical, the change will likely have given rise to a new version (or new “generation”) of the survey.

Some of these redesigns provided an opportunity to split a single earlier survey into several new ones (hence the increase from 32 to 34).

5 surveys were active in 1995 and were later terminated: these surveys were discontinued without being replaced.

32 surveys were active in 1995 and were replaced by 34 others, present in 2004.

The reprogramming of the “machine” was accompanied by a redesign of the questionnaire, the method, the sample, etc., and these changes have given rise to a new version (or new “generation”) of the survey. Some of these redesigns provided an opportunity to split a single earlier survey into several new ones (hence the increase from 32 to 34).

58 surveys are new surveys launched during the period: they are also repetitive surveys that, even if not “active” in 2004, will be so in 2005, 2006 or beyond if their frequency exceeds one year.

105 surveys are one-time surveys, i.e., non-recurring surveys or surveys repeated only once. In theory, they are unlikely to be conducted again.

Progress needed in knowledge of sources

The few data presented here illustrate the limits of our knowledge of our own surveys.

True, each survey designer has the information pertaining to his or her work, but, it seems, neither the organization for which the designer works, nor—for certain—the official statistical system has any consolidated information on existing, planned, or discontinued surveys. By contrast, there are many partial, specialized, and often ephemeral lists that are hard to locate and even harder to exploit. Ultimately, the current situation with regard to surveys is identical to that of business statistics before the introduction of SIRENE (the register of French enterprises and local units).

Ideally, the small set of characteristics described here should be obtainable not from a specific survey requiring several months’ work, but every year, as a direct by-product of the preparation of work programs on which CNIS is required to give its opinion.

We could even go further and spell out several points.

The first manifest information deficiency concerns costs. The cost of a survey can be calculated, as can its production cost. But only some surveys yield usable information about costs. This blind spot is all the more problematic given the intensification of budget constraints.

The second deficiency is the lack of information on the surveys’ possible integration into an international setting, particularly with regard to European Union countries. National status (“surveys in the public interest,” compulsory survey, etc.) is admittedly a useful information item but so is EU status: some surveys are labeled as “Community [i.e., EU] surveys” in the Journal Officiel; others, according to CNIS Quality Label documents, are “stipulated by EU regulations”; a third category results from a “gentleman’s agreement,” and so on—but we are still a long way from having a comprehensive and usable corpus of information, even amid the growing influence of EU integration on national statistical programs.

Although all the surveys examined here are national in scope, we lack information on the geographic levels for which detailed results could be produced. Yet it would be interesting to know if the survey will yield results at the national level alone, or also for French regions, départements, or even smaller territorial units.

Information on the response burden is very inadequate as well. The indications given earlier on sample and questionnaire size are far too succinct. It would be desirable to distinguish the units surveyed from the statistical units to be observed. Questionnaire size should be broken down into the number of variables collected for each statistical unit in the survey. These statistical units may be of various types (a “vehicle-week,” a night’s accommodation, a purchase,

Table 16: Classification of surveys by type of change during observation period (1995-2004)

<table>
<thead>
<tr>
<th>How surveys changed during period</th>
<th>Observation period</th>
<th>Number of surveys concerned</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1995</td>
<td>...</td>
</tr>
<tr>
<td>No change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terminated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terminated and replaced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduced then discontinued</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

37. This information is useful both for the surveyed unit and for the litigation department, which deals with non-responses to compulsory surveys of enterprises.
and not only consist of enterprises or persons. When analyzed in connection with the response burden, non-response rates provide valuable information considering how high they sometimes are.

Not enough is known about survey life cycles. As mentioned earlier, the information should enable us to distinguish between its “birth” (date of the survey’s first launch) and its “recognition” (date of inclusion in official-surveys program and—where applicable—date declared compulsory). It is useful, too, to know how surveys were merged or, on the contrary, split up, and how and when they were redesigned, for purely technical or broader purposes.

Lastly, even if this information is more technical, it would not be unreasonable to know more about the production processes used: which surveys use online data collection, which surveys outsource collection to research firms, which use computer-assisted collection through face-to-face or telephone interviews, and so on.

And all of this is relatively basic information. We could also gather more elaborate data, such as date of publication of initial results, availability of files and their documentation, studies conducted on the survey data, and integration into databases or panels.

The official statistical system did not wait to have such an organized and comprehensive corpus of knowledge in order to expand and coordinate its activities. Back in the 1930s, France already had a system of business statistics that—it is said—was not contemptible. Yet the introduction and build-up of SIRENE have brought advances without which the current system of business statistics could not exist. Likewise, within limits, a register of statistical surveys and, more generally, of statistical sources would certainly have a role to play.
Methodological appendixes

Survey coverage

We extracted the file used for our study from a database entitled “register of sources” by selecting surveys that were:

- “active” at any given time between 1995 and 2004,
- included in official program as published in the Journal Officiel,
- national in scope (with or without overseas territorial units), excluding regional surveys.

For 2004, the coverage can be measured against the total set of official statistical sources available in France using the following diagram:

![Survey Coverage Diagram](image)

Statistical surveys handled by official producer organization as defined in Article 2 of CNIS decree (500-1,000?)

a) The study covers the 680 national surveys included in the 2004 program of statistical surveys by government agencies published in the Journal Officiel.
b) Excluding the surveys of specific industries (enquêtes de branches), which are not individualized in the survey program published in the Journal Officiel, our study applies to the other surveys, which are individualized in the program. Their total number is 180, of which 108 are described on the CNIS website (www.cnis.fr) for 2004.

to cover the entire set of statistical sources handled by an official producer entity as defined by Article 2 of the CNIS decree, we need to add:
c) The regional surveys included in the Journal Officiel program (9 in 2004)
d) Statistical sources constructed from files compiled for administrative purposes (we identified 84 but there are certainly more)
e) Sources produced by matching administrative sources against statistical sources (we identified 40 but there are certainly more)
f) Statistical surveys performed by public-sector entities or research institutes (we identified 125 but there are certainly more).

Sources used for the study

Our primary source consists of the ministerial decisions (arrêtés) published in the Journal Officiel. This is an administrative source that requires special processing to allow statistical analysis: declarations aggregating several surveys on the same line have been segmented; purely methodological surveys (typically, questionnaire tests) have been removed; surveys for administrative purposes (such as supplementary censuses and vital-statistics filings) have been removed as well. From the resulting survey list, we fed the database by using several documentation sources.

First, we used the Journal Officiel, which consistently gives the survey title, name of producer entity, survey coverage, and—fairly often—frequency and sample size.

Our second source is the fact sheets on surveys in progress, published on the CNIS website. These documents, prepared by the CNIS secretariat and checked by the producer entities, are reliable. But only two-thirds of the field is covered and, even for the covered part, the documents contain only the current-year surveys. Moreover, this source never describes the survey questionnaires.
For questionnaires and descriptions not found in the two sources above, we have used the producers’ publications, available in paper form or—increasingly—on their websites, as a growing number of producers describe their sources and put their questionnaires online.

Fourth, as most producers are government bodies, we have been able to access some of their intranets, which often provide fuller documentation on their surveys.

_**Lexicon**_

**Statistical survey:** observation system constructed to supply statistically usable data for the purpose of producing knowledge. Survey preparation largely consists in defining a population and writing a questionnaire.

**Collection:** administration of questionnaire to chosen population. A given survey may involve a single collection (the survey will be described as “one-time”) or, on the contrary, collections repeated at more or less regular intervals (the survey will be described as “periodic” or “repetitive”).

**Producer:** entity acting as principal for the survey and its collection (and therefore responsible for it in legal and intellectual terms). The producer can outsource specific phases of the survey or collection (such as computer programming, fieldwork or tabulation of responses) to different contractors.

**Designer:** person or group of persons in producer entity responsible for designing the survey, tracking its production, and analyzing its results.

**Official producers:** “The producer entities [of the official statistical system] are the Institut National de la Statistique et des Études Économiques [INSEE] and ministerial statistical offices; government bodies; official organizations or organizations majority-controlled by government entities; private-sector organizations in charge of a public service to collect or process economic and social data; and private-sector organizations in charge of collecting or exploiting economic and social data.” (Article 1 of Decree no. 2005-333 of April 7, 2005, on National Council for Statistical Information [CNIS] and Statistical Confidentiality Committee)

**Box: should one say personal survey or household survey?**

The expression “household surveys” is more often used than “personal survey” in the French official statistical system, i.e., at INSEE, of course, but also in most MSOs and in bodies such as CNIS. The first designation refers not to the surveys themselves but to the way in which most samples of these surveys are picked: the process consists in a random selection of dwellings using criteria that combine population-census variables. The statistical unit, therefore, is the dwelling. However, as INSEE maintains a strict equivalence between household and dwelling, the statistical unit is ultimately the household—although the household present in the dwelling at the time of the census is not necessarily the same as the one present at the time of the survey!

Yet, at INSEE as well as in the MSOs, there are surveys whose samples are composed of persons and not households, for example if the statistical unit is an employee: this is true of INSEE surveys based on employers’ annual statements of payroll data (Déclarations Annuelles de Données Sociales: DADDS) and DARES surveys based on the files of the unemployment-benefits agencies (UNEDIC).

In formal terms, therefore, the expression “personal survey” seems preferable to that of “household survey,” the first being more general than the second. Households are indeed always composed of persons, whereas not all persons necessarily belong to a household—either because the person is surveyed as an employee and not as a household member (case described above) or because the person lives in an institution (boarding school, barracks, prison) and therefore not in a household.

However, the expression “personal survey” has its drawbacks too: first, to be strictly correct, we should speak of “natural person” or “private individual” (personne physique) rather than “person,” for a “person” can also designate a “legal person,” i.e., an enterprise. Moreover, there are surveys rightly qualified as “enterprise surveys” because the statistical units surveyed are enterprises, but in which employees are surveyed at the same time as the enterprises themselves, for example in order to determine pay structure. A “personal survey” may thus actually be an “enterprise survey in which persons are questioned.” As we can see, the expression “personal survey” may generate confusion between “statistical unit surveyed” and “respondent.”

French legislation cannot help us settle the issue either: true, the 1951 Statistics Act draws a distinction based on the quality of information collected, since Article 6 distinguishes between (1) individual information “regarding personal and family life and, generally speaking, facts and behavior of a private nature” and (2) “economic and financial” individual information. The first category is broadly covered by surveys of persons or households, the second by surveys of enterprises. Except that here, the distinction is based not on the statistical unit surveyed but on the questionnaire content!

Conclusion: while the expression “household survey” is preferred for reasons of usage, and the expression “personal survey” is preferable for formal reasons, neither is totally satisfactory. The authors of this article prefer the second and have thus decided to use here.

*[Box text prepared by the editors, with the assistance of INSEE’s Statistical Coordination Department]*
Geographic origin of individuals in French population censuses

Olivier Monso* and Thibaut de Saint Pol**

In the nineteenth century, France introduced political and legal criteria aimed at establishing who was or was not a member of the national community. To this end, questions on geographic origins were added to the nominal lists and individual schedules in French censuses. By the end of the century, the emergence of Nation States had redrawn the political map of Europe. Amid rising immigration, France’s concern was not only to count the population but also to draw a statistical distinction between French and foreign residents, and between French citizens by birth and naturalized citizens. In the second half of the twentieth century, statisticians increasingly sought data about origins; this information now needed to cover the life histories of new entrants to the national community. After reviewing the historical background, this article compares the treatment of origins in France’s general population censuses with examples from other countries. The authors thus shed light on the distinctiveness of the French approach—in particular, its classification of persons by national citizenship and by means of French nationality acquisition.

The general population census counts the set of persons forming the population of a State at a given moment, by means of an operation in which all the State’s residents must take part. It is not, strictly speaking, an enumeration of individuals but a tool providing information on different characteristics of the population. The nominal lists and, from 1876 on, the individual schedule were compiled for the purpose of describing the demographics of each municipality (commune): How many men and women live in it? How many young and elderly people? How many economically active and inactive persons? Questions were later extended to other topics such as the number of payroll employees and business owners, and the number of children attending school in their municipality of residence. Analyzing the population by origin was not an initial priority. Rather, it was gradually incorporated into the questionnaires owing to factors as diverse as variations in migration flows, changes in French nationality law, and French foreign policy.

The concept of nationality emerged at the time of the first censuses

Under the Ancien Régime, officials compiled comprehensive hearth rolls (relevés des feux) for the kingdom, but the notion of a direct population census for all of France was not introduced until the French Revolution, with the “police act” of July 22, 1791. But “if, by census population, we mean a detailed, per-capita listing of a country’s inhabitants, we can say that no true general census was conducted in France before the nineteenth century.”

In 1833, France established a Statistical Office (Bureau des Statistiques) and decided to conduct a census every five years, in years ending in 1 or 6. This period saw the construction of the French Nation-State. The voter-enfranchisement issue crystallized the authorities’ will to determine who belonged to the community and who did not, and the issue of individuals’ origins emerged in the census.

Conceptualizing the “foreigner” in terms of his or her “nationality” is a fairly recent approach. As Patrick Weil notes (2002), “when the Revolution began, there was no explicit definition of a French person. Under the Ancien Régime, the jurisprudence defining the boundary between the French person and the foreigner developed in incidental fashion, through legal disputes

* Olivier Monso is an INSEE staff member with the rank of administrateur. He works in the Employment Division of the Institute’s Demographic and Social Statistics Directorate.

** Thibaut de Saint Pol is an INSEE staff member with the rank of administrateur. He works at the Quantitative Sociology Laboratory of the Institute’s Center for Research in Economics and Statistics (Centre de Recherche en Économie et Statistique: CREST), in liaison with the Sociological Observatory of Change (Observatoire Sociologique du Changement), a joint unit of the National Foundation for Political Science (Fondation Nationale de Sciences Politiques: FNPS) and the National Center for Scientific Research (Centre National de la Recherche Scientifique: CNRS).

2. Résultats statistiques du recensement de la population en 1901. The volume’s introduction provides a detailed history of French censuses and their methods.
arising from estate problems." At the
time, the foreigner was, above all,
the person one did not know. Local
recognition of the individual was
the decisive criterion, establishing
membership in a community. The
Revolution broke with this tradition
by abolishing the symbols of power
vis-à-vis the foreigner, namely, droit
d'aulaire (the sovereign's right to
inherit the property of foreigners
residing in France at their death)
and the authority to grant French
nationality.

The first law on nationality—
although the concept itself was
not actually used—was contained
in the Napoleonic Civil Code. The
term itself did not appear until the
early nineteenth century. In France,
it was first used in administrative
documents around 1820 and in a
literary text, by Madame de Staël, in
1830. The appearance of a notion
of "nationality" therefore coincided
with the implementation of the first
general censuses of the French population.

In 1851, the very year when access
to nationality was modified by the
enactment of the double jus soli
double droit du sol),3 a question on
nationality was added for the first
time to the census, along with a question
on religious affiliation. Naturalized
French persons were distinguished
from other French persons, and
foreigners were classified by
nationality. This innovation was
motivated by the State's need to
determine who belonged to the
nation. In the Ancien Regime, persons
born in a given place were deemed
to "belong" to that place. One example
of the consequences of this practice
is that children not born on French
soil could not inherit from their French
parents (a form of droit d'aulaire). In
the nineteenth century, a new need to
differentiate between nationals and
foreigners arose. Place of birth or
place of residence would not suffice
as a criterion. Official statisticians
accordingly introduced a question on
nationality in the census.

In response to protests, the questions
on religion and nationality were
withdrawn in the following census
(1856) (Dupâquier 1965). They were
restored in 1861 in the household
schedule (itself introduced in 1856)
and in the tables published at the
time, although they did not appear
in the nominal lists.4 They were kept
in 1866 and in 1872. At the latter
date, only nationality reappeared in
the nominal lists. The question on
religion was finally eliminated in 1876
with the introduction of the individual
schedule.

In the second half of the nineteenth
century, nationality and place of birth were
therefore gradually adopted as
criteria for distinguishing between
French residents. In the same period,
international statistics—in the wake of
the 1853 Brussels Congress,
attended by Adolphe Quetelet—
turned to the linguistic criterion to
draw the same distinction, found, for
example, in the Prussian censuses
(Labbé 2006).6

### Nationality categories

#### proposed: a pragmatic,
#### political approach

France used seven categories to
classify foreigners in 1851. They
mainly concerned neighboring
countries, from which most foreigners
came: Belgians, Spanish, British,
Germans, Italians, and Swiss; the
seventh category—Poles—was the
exception to the rule. The choice of
nationality criterion was ambiguous:

The question's introduction was
aimed at producing a census of
foreigners living in France, without a
genuine interpretation of the results.
The census-takers were mainly interested in foreigners' places of
settlement by origin: people from
Spain lived in south-west France,
Belgians in the north, and so on. The
census showed that foreigners stayed
in regions near the border with their
country of origin. Hence the desire
(1) to identify the reasons prompting
foreigners to settle in France and
(2) to determine the country's most
attractive regions.

These categories were also sensitive
to political stakes, particularly in
the sphere of international relations. For
instance, the inclusion of Poles in
the list of nationalities was not a self-
evident choice, given the Russian
Empire's tutelage over the kingdom
of Poland. But official French support
of the Poles in their 1830 insurrection
(in the name of cultural and religious
values and of a common history
dating back to the Napoleonic period)
led to the statistical recognition of
Polish nationality (Goussef 1997). This
shows a determination to assert a
political choice through the nationality
categories used in the census: military
defeat turned the Poles into political
refugees welcomed by France in the
name of humanist values inherited
from the Revolution.

The Second Empire witnessed the
political authorities' growing control
over official statistics, most notably
the general population censuses.
This is illustrated by the 1861
census, which added Russians,
Romanians, Greeks, and Turks to
the existing nationality categories in
order to reflect France's geopolitical
vision of Europe. The introduction of
"Russians" in the list coincides with the end of the Crimean War
and the advent of Tsar Alexander II. It also forms part of a broader French-Russian rapprochement. The end of tensions in Crimea and the Balkans led to the emergence of nationalities protected from Ottoman and Russian ambitions: Romanians (also designated as Moldo-Valachians) were introduced into the 1861 census, whereas the independence of Romania, previously under Ottoman suzerainty, would not be recognized until 1878. This controlling influence of political factors was compounded by statisticians’ resolve to take fuller account of the diversity of foreigners present in France. The same reasons explain the addition of Danes, Norwegians, Swedes, and Americans to the list in the 1861 census.

The late nineteenth century marked a turning point. The international scene had changed, and statistics needed to keep up with the latest developments. France’s defeat by Prussia in 1871 led the authorities to revise nationality categories as well as the overall approach to the census of foreigners that had prevailed until then. The political class and statisticians alike focused on a new urgent concern: the need to clearly designate the enemy. In consequence, the “Germans” category was split up into “Germans,” “Austro-Hungarians,” and “persons of Alsace-Lorraine who have not opted for French nationality.” Just like the identification of the danger, the search for potential allies took on ever greater importance. In the eyes of political, military, and diplomatic leaders, the Tsar became an essential partner in the struggle against the German empire, after a long period in which the Russian empire had been neglected or, at the very least, ill regarded. For diplomatic reasons, the “Poles” category was eliminated, leaving vast statistical room for “Russian subjects” (Goussef 1997). Through the nationality categories, the very spirit of the census was changing. The removal of categories of hitherto protected population groups essentially meant a break with the humanistic and universalist values of the 1789 Revolution, which had informed the first censuses of foreigners (1851, 1861, and 1866). Revolutionary values and political vision cohabited so long as French ascendancy in Europe persisted. Henceforth, “universalism was blamed for undermining patriotism” (Brubaker 1993), and the revolutionary message gave way to polarization over the destiny of the French State.

Thus France, in the late nineteenth century, shifted away from a vision open to the outside world, marked by a desire to protect oppressed populations. It moved toward a “less generous” approach, amid less favorable political circumstances. As regards the census, a transition occurred between a straightforward approach based on an enumeration of foreigners and a far-reaching interpretation of results, at a time when France was registering a large increase in the number of its foreign residents.

**The emergence of the immigrant figure raised the question of nationality acquisition**

From the late nineteenth century onward, the perception of the foreigner as a citizen of another State with an “objective” existence (and not a fictive one, as in the case
of Poland discussed above, which survived in the censuses between 1863 and 1876 even though the country was now only a province of the Russian empire) became enshrined in the censuses. Emphasis would now shift from the “foreigner” to the “immigrant,” i.e., the foreigner who had come to France in order to settle there.

France in the 1850s sought to be a land of asylum for all those who were fleeing conflicts in Europe: emigration of Poles after the November 1830 revolt against the Russians, Spaniards fleeing the Carlist Wars that began in 1833. But the reality of immigration upset this vision, most notably with the 1881 census, which revealed a foreign population of over one million. Beyond this quantitative increase, another fact stood out: the motives that led foreigners to settle in France had changed. Immigration for strictly political reasons had become a minority phenomenon, overtaken by immigration of workers, often driven (particularly Italians) by economic problems in their home countries. The temporary nature of these foreigners’ stays in France was strongly challenged, posing in no uncertain terms the issue of the integration of new arrivals (including whether or not they should be allowed to obtain French nationality).

The integration issue concerned the workplace, where the French often saw foreigners as rivals for access to employment—an access made difficult in periods of deep economic crisis such as France experienced in the late nineteenth century. In Parliament, members from the regions most concerned by wage-labor immigration (the north and the Mediterranean south) complained of this competition, which was aggravated, in their view, by foreigners’ frequent willingness to work for lower wages. The other focus of discontent was that foreigners were not subject to certain “civic” duties—in particular, military service, compulsory since 1872. This argument worked in favor of the naturalization of children born in France to foreign parents, who could hitherto keep their foreign nationality and so avoid conscription.

The response came in the form of the 1889 Act, which facilitated naturalizations of immigrants’ children by strengthening *jus soli.* A person was now French if he or she was (1) born of a French parent (*jus sanguinis,* effective since the 1804 Civil Code), (2) born in France of a foreign parent also born in France (“double *jus soli,*” in force since 1851), or (3) born in France and residing there when reaching legal majority. The Act allowed an increase in the number of naturalizations.

The growing prominence of foreigners, both in numerical terms and because of the social problems due to immigration, led statisticians to pay greater attention to the alien population. The first-ever publication devoted exclusively to foreigners appeared in 1891. The most significant change in the census questionnaire itself was the introduction, in 1881, of a question on nationality with a choice of three answers: “born of French parents,” “naturalized French person,” and “foreigner of what nation?” The novelty was the addition to the individual schedule—even before the 1889 Act—of the “naturalized French person” choice as an answer to the nationality question: this separated persons having acquired French nationality during their lifetime from those who possessed it at birth. The distinction had already been drawn in the nominal lists of 1851 and repeated in the household schedules of 1861, 1866, and 1872. Now, it became a lasting feature of individual schedules in the censuses of the late nineteenth and twentieth centuries. This is puzzling if we consider the rather commonly accepted image of French nationality based on *jus sanguinis* (although, as noted earlier, France actually practiced a blend of a form of *jus sanguinis* and *jus soli*). The distinction seems even more baffling if we refer to the ideal expressed by Ernest Renan of a nation founded on a “daily plebiscite.” Why, then, the need to separate the French by nationality of origin?

The “naturalized French person” category rested on the series of laws defining access to nationality passed at the end of the nineteenth century and in the twentieth century. It defined the position of the person in transition between the status of foreigner and that of a full-fledged citizen endowed with all the rights attached to French nationality. The distinction was primarily legal, as the 1889 Act imposed a ten-year waiting period for recently naturalized French persons before they could hold a seat in Parliament. Other civic rights restrictions were added in the twentieth century, in particular a five-year ban on voting after naturalization, introduced in 1938 (Weil 2004). In other words, statisticians did not follow the law but anticipated it by the end of the nineteenth century by introducing a distinction between French citizens according to their origins via a question on the means of nationality acquisition.

In the early twentieth century, the nationality question in the census underwent few changes: the “foreigner? from what nation?” choice became “foreigner? from what country?” in 1910. That same year, the “born of French parents” option became “born French,” then “French by birth” in 1921. As Alexis Spire and Dominique Merllié point out (1999), the latter amendments were, no doubt, due to the fact that, before 1927, a French woman would lose her nationality when she married a foreigner (by virtue of the 1803 Civil Code): “The Civil
Code stipulated that French women marrying foreigners lost their French nationality; the instructions on the individual schedule asked them to declare themselves both French at birth and of foreign nationality; the three nationality options were therefore not all mutually exclusive. Statistically speaking, the problem thus arose from the fact that two women in the same situation—namely, having become foreigners by marriage—could give different answers to the nationality question in the census schedule. This ambiguity was partly lifted by the 1927 Act, which allowed women marrying foreigners to keep their French nationality. Consequently, only those women who had “expressly” renounced their French nationality were obliged to make the double declaration mentioned above.

Beyond these slight adjustments, the main outcome was the persistence of the nationality question in the census, i.e., the fact that “beyond the form of the question, the principle of a distinction between French citizens based on the means of nationality acquisition became a permanent feature of the census” (Spire and Merlié, op. cit.). In practice, therefore, France’s opening to foreigners, promoted by its legislation, was offset by a greater legal separation (1) between French citizens and foreigners (for example, certain jobs, particularly in the civil service, were reserved for French citizens) but also (2) between the French themselves according to whether they were naturalized or French by birth. Both of these separations were enshrined in the population census.

The Vichy regime marked a break in the attitude toward nationality law, as in the use of the census. The Occupation was one of the main moments of what Patrick Weil (2002) calls the “ethnic crises of French nationality.” From 1940 to 1944, the policy enacted by the regime ran counter to the priorities of the Third Republic. Whereas the latter had encouraged naturalizations (cf. the 1889 and 1927 Acts), Marshal Pétain’s government severely restricted them and challenged the naturalizations granted since 1927.

During the Vichy years, the term “census” was often associated with anti-Jewish policy, as in September 1940, when the Germans ordered a census of Jews in the “northern zone” (i.e., the part of France under direct German occupation). Two “Jewish statutes” (October 3, 1940, and June 2, 1941) defined a person as Jewish on the basis of the number of his or her Jewish grandparents (the statutes speak of “Jewish race”).

10. This distinction between French citizens by means of nationality acquisition grew sharper until World War II. In addition to the ten-year ineligibility period (introduced in 1889 and extended to all elected offices in 1927) and the five-year voting ban after naturalization, the authorities introduced restrictions on employment categories open to new citizens. In 1935, lawyers’ and doctors’ lobbies convinced the government to ban recently naturalized French citizens from entering their professions for a “probation” period (ten years for lawyers, five years for doctors) after naturalization (Noiriel, 1988).

11. It should be pointed out that the notion of “Jewish race” was imposed not by the Germans (who, in the September 27 ordinance, referred only to the Jewish religion), but by the Vichy government itself.
The second statute added a religious criterion. These criteria paved the way for the census of Jews in the Vichy-administered “southern zone” and the German-occupied northern zone by late 1940. Another consequence was the introduction of a “Jews” category in the “census of occupational activities” conducted in the southern zone in 1941 (Levy 2000). This census is not comparable to the pre-war censuses, particularly as it was confined to the population aged 13-65 and as the introduction of question 11, among others (“do you belong to the Jewish race?”), was motivated by the Vichy regime’s anti-Jewish policy.

Coverage of geographic origins was enhanced by that of individual life histories

The post-war years saw a restoration of the overall status quo ante in the field of nationality law, albeit with slight amendments to the very liberal provisions of 1927. The survival of most of the inter-war legislation may explain why the questions on individuals’ origins continued to be addressed in similar terms in the census. The 1946 individual schedule did, however, feature some minor changes: for example, the choice “French by naturalization” was now followed by a list, in parentheses, of specific means of acquiring nationality: “including marriage, declaration, choice.” In other words, statisticians continued to define naturalization in a “broad” sense, comprising both (1) procedures that the 1945 nationality code qualified as naturalization in the strict sense (nationality granted not de jure but by administrative decision), and (2) “automatic” acquisition of nationality, resulting from the application of a right (jus soli for acquisition of nationality at majority, or right arising from marriage to a French spouse, etc.). By contrast, starting in 1954, individual census schedules clearly separated naturalization from other means of nationality acquisition. The choice “French by naturalization” was followed by the words in parentheses “or by marriage, declaration, [or] option,” the initial “or” replacing the “including” of the 1946 schedule.

The distinctions between French citizens by means of nationality acquisition therefore persisted in law and in the census. One of its most striking applications was in colonial administration. The example of Algeria, discussed by Alexis Spire and Dominique Merlié (1999), is particularly enlightening as it shows the value of studying the treatment of geographic origin not only in the individual schedule but also in census-takers’ practices. Back in the 1921 census, “coding instructions” had already recommended that all persons with Muslim last names and first names should be classified among “natives, whether French subjects or persons under French protection, even if they had declared themselves French” (Singer-Kerel 1986). In the postwar years, the 1946 constitution granted the status of “French” to persons from the “French Union” (i.e., French overseas possessions). This provision, followed by the “organic statute for Algeria” (Act of September 20, 1947), led to a situation where “persons with a civil status as French under ordinary law” (Français de statut civil de droit commun) lived alongside “persons with a civil status as French under personal law” (Français de statut civil de droit personnel), in other words, “Muslim French persons.” The latter were subject to local law (most notably, to the jurisdiction of the “kadi’s court” in localities where such courts existed) but were deprived of most political rights granted to the French. This “degraded” citizenship was merely the perpetuation of the “native code” (code de l’indigénat) instituted in 1887 and abolished in 1946, except in Algeria, where it lasted almost until the country’s independence in 1962. However, the distinction between “persons with a civil status as French under ordinary law” and “Muslim French persons” did not appear as such in the census: again, the statistical administration offices relied on approximation, using a list of “the most common Muslim first names observed in civil registers in Algeria” (Lacroix and Thave 1997). This enabled official statisticians, in the tables published in 1954 and 1962, to divide “French persons by birth” into two categories, “Muslims of Algerian origin” and “Other than Muslims of Algerian origin.”

In the 1960s, decolonization made the distinction between French “citizens” and French “subjects” obsolete. The shift is perceptible via the question on place of birth, worded almost identically from the late nineteenth century to 1946. Since 1901, persons born in the colonial empire had been asked to indicate their “colony” of origin (the 1896 census specified “colony or possession” but the term “possession” was later removed). The term “colony” survived to 1946. The 1954 and 1968 schedules used the expression “overseas country” (pays d’outre-mer), whereas the 1962 schedule made no reference to the territories of the former colonial empire. From 1975 onward, only French overseas territories (territoires d’outre-mer: TOMs) were distinguished from French départements (including the overseas départements d’outre-mer: DOMs) and other countries. Questions on persons’ origins were thus closely linked to questions on territorial boundaries.

The legal distinction between French by birth and naturalized French was also challenged, and totally eliminated in the early 1980s: all French citizens were now equal by law regardless of the means by which
they had acquired their nationality. In particular, restrictions on employment and the five-year ban on voting after naturalization were waived in 1978. The ten-year waiting period before running for elected office was abrogated in 1983 (Weil 2004). True, the census continued to reflect a differentiation of French persons by origin—and to cover it in fuller detail. But given the changes described above, we may assume that the distinction was kept not to perpetuate a legal inequality but to undertake deeper research on individual life histories. The distinction was strengthened by adding specific questions for naturalized French citizens. A question on previous nationality was introduced in 1962. It remained until 2004, when it was rephrased to cover “nationality at birth.” In 1954, 1962, and 1968, persons living outside metropolitan France (abroad or in an “overseas country”) at the time of the preceding census were asked to indicate the year of their arrival in France (whether they were French or not). This question was eliminated in the following census, then restored in 1999 in a different form: all persons born abroad or in the DOM-TOMs had to give the year of their arrival in metropolitan France. In 2004, the qualifier “metropolitan” was deleted and the question was put only to persons born abroad. The contributions provided by these questions (nationality at birth, date of arrival in France) are largely informed by inputs from demographers and sociologists seeking to learn more about the composition of the French population. A fuller knowledge of the geographic origins of individuals allows a more detailed study, for example, of the integration of immigrants.

The French tradition compared with practices in the English-speaking countries

The American vision of origins is diametrically opposed to the French concept of the nation as an inseparable whole. Unlike their French equivalents, U.S. censuses clearly emphasize the differences in origin that appeared at the nation’s foundation, if only in mythical terms.

The issue of origins should be viewed in the context of the struggle by ethnic

15. At the same time, however, immigration policy was tightened, with wage-labor immigration suspended in 1974. Stricter control of immigration was accompanied by a policy of integration of foreigners already in France (granting of rights such as full freedom of association in 1981) and, for naturalized foreigners, access to the same rights as French citizens by birth.

16. Note that the individual schedules of the overseas départements (DOMs), created in 1946, have exhibited only minor differences from the schedules used in metropolitan France since 1980 and are identical in the redesigned census implemented in 2004. “Overseas collectivities” (Collectivités d’Outre-mer [COMs], formerly the overseas territories [TOMs]) are in a different situation and would deserve separate analysis, particularly owing to the persistence of a question on the language spoken.
minorities and the assertion of their identity. Minorities unite in pressure groups to demand better statistical coverage—as a major step toward political recognition. The English word race is thus used directly in American censuses, and the 2000 census questionnaire [individual schedule] offered a large number of choices: White, Black, American Indian, Alaska Native (respondents were asked to report their tribes), Asian Indian, and so on. The construction principle for these categories reflected the dual meaning of the term race, which denotes both (1) a group defined by inherited physical characteristics and (2) an “ethnic group” characterized, instead, by the sharing of a common culture. Hispanic-American lobbies have obtained the right to be listed as a separate entity in the national statistics, even though they do not constitute a “race” but, in effect, a cultural and linguistic community.

British censuses, for their part, reflect strong hesitations on the value and necessity of including respondents’ geographic origins in the national statistics. Unlike in America, British politicians, academics, and community representatives were long skeptical on the issue. In the 1960s, several tests were performed using indirect questions. An initial attempt was made in 1966 for persons from the Commonwealth, on the basis of residents’ places of birth. In 1971, a question on the origin of parents was introduced. Finally, in 1991, the government and the statistical authorities gave up indirect questioning in favor of “self-declaration” from a list of ethnic origins: White, Black African, Black Caribbean, other Black, Pakistani, Indian, Bangladeshi, Chinese, other ethnic groups. In 2001, the division by ethnic category was even more detailed: broad categories distinguished between Whites (among them, “British” and “Irish”), mixed (“White and Black Caribbean,” “White and Black African,” etc.), “Asians or Asian British,” “Blacks or Black British,” and lastly “Chinese or other ethnic group.” Each of these categories is divided into more detailed subcategories.

The Canadian experience shows the difficulty of asking census respondents for their geographic origins. A decade earlier, respondents were asked to provide the father’s ethnic or cultural group at the time of his arrival on the continent. In 1981, this became a question on “you and your ancestors” with a choice of fifteen possible answers. The government wanted to know the ethnic and geographic origins of the population, but part of that population responded in terms of national sentiment: the “Canadian” option was not on the list, yet a significant number of respondents chose it under the heading “other.” In the 2006 census, after a note indicating that “the census has collected information on the ancestral origins of the population for over 100 years to capture the composition of Canada’s diverse population,” the questionnaire asked “what were the ethnic or cultural origins of this person’s ancestors?” In parentheses, it suggested over twenty examples of origins, starting with “Canadian” and followed by “English” and “French.” Further down, the list also included “Inuit (Eskimo)” and “Jewish.” Respondents were free to define their own origins. The question on geographic provenance of ancestors became a question of opinion, reflecting a sense of attachment to an origin. The question was supplemented by, among other things, information on language, but also by a choice resembling U.S. practice in its blend of “race” and nationality (White, Black, Arab, Filipino, etc.), and by a question on the parent’s country of birth. The latter, last asked in 1971, was reintroduced in 2001.

How can we explain the differences with the questions on geographic origin of individuals asked in French censuses? The first possible answer is specific to the history of France (and to the period of the Vichy regime in particular), and centers on the fear of misuse of ethnic data. The same motive has generated a reluctance to envisage a population register listing changes of residence, as in Belgium and the Nordic countries (Héran 2004). However, our glimpse of the history of population censuses shows that prudence with regard to data on geographic origin was visible in France by the late nineteenth century. This followed a period of hesitation in which statisticians sometimes faced protests, for example on religion and nationality questions in 1851.
The French choice must also be viewed in terms of the intrinsic nature of the census. While the inclusion of certain information items on individuals’ geographic origin in French twentieth-century censuses may have been deemed inappropriate, this does not mean that such issues were not of interest to government officials. Several INSEE surveys, for instance, include questions on the language, religion, and geographic origin of respondents’ parents. Perhaps the best illustration is the “Study of Family History” survey: performed concurrently with the 1999 census, it comprised questions on geographic origin (including parents’ country of birth), on the language spoken, and on family and social trajectories (Cassan et al. 2000). Questions that could not be incorporated into the census itself have been asked “on the side” as part of a sample survey. Thus the attitude of French statisticians should not be seen as reflecting an impossibility for them to use statistics on geographic origin, nor even as a form of self-censorship, but as a genuine effort to match these statistics with census goals.18

17. INSEE’s 2003 “Life-History Survey” includes, for example, questions on religion. However, respondents are not asked to name their religion explicitly.
18. This study was carried out with valuable advice from Alain Desrosières and Michel Armatte and the assistance of Jean-Yves Icole. The authors also wish to thank Jacques Cazenave, Morgane Labbé, and Laurent Wilms for their suggestions and encouragement.
References


Website: http://census.ined.fr/