

STATISTICAL SOURCES AND STATISTICAL SYSTEM IN THE INFORMATION SOCIETY

Carlo Filippucci

1. INTRODUCTION

The enormous amount of electronic information¹ collected in modern society in order to meet the needs of a variety of bureaucratic, administrative and managerial activities for companies, public authorities and institutions, is engendering two phenomena of particular relevance to statistical systems. The first concerns the increasing availability of archives in relation to many kinds of phenomena and subject matters. The second, largely a result of the first, concerns the growing interest in taking full advantage of these records by national statistical agencies and also by several institutional and private entities emerging as “new data producers”. All this significantly impacts on statistical systems, poses new problems and challenges to official statistics, but also requires thinking carefully about the qualities of the statistics drawn from the availability of these large pools of information.

The great availability of information to be used for statistical purposes can result in a real “revolution in data production”. Based on this premise, this paper tries to outline some of the main implications that this phenomenon may entail for official statistics and for the Italian statistical system; it also attempts to identify which issues may arise and how to ensure appropriate standards for data resulting from non statistical sources. Many issues arise but the paper is mainly devoted to identify those more general asking for a preliminary investigation. Finally, as among the new sources of information those coming from bureaucratic-administrative and managerial activities have an important place, the last part of this paper will focus on these specific statistical sources and on the issues to be addressed in order to ensure quality, comparability and coherence to the data coming from these sources, similarly to those coming from statistical surveys.

¹ In this note we define *statistical data/statistic* a factual piece of knowledge obtained through a statistical survey. We define instead *information/data* (without further specifications) the knowledge of facts, situations, etc. that can be acquired in a variety of ways. Of specific relevance in this paper are those deriving from the bureaucratic and/or managerial activity of institutions, public administrations and companies.

If we will succeed in taking the opportunity of this challenge, and to deal appropriately with the issues posed by the use of administrative and managerial sources for statistical purposes it could be possible to make a substantial progress and by outlining appropriate strategies and methods it would be possible to create the conditions for an increasingly rich, integrated and above all reliable statistical system.

2. CONTEMPORARY SOCIETY AND THE NEW CHALLENGES FOR STATISTICS

2.1. *The wealth of sources that can be used for data production*

In the last decade, civil society, institutions and the economy have changed dramatically and Official statistics is required to meet ever more demanding requirements as concerning new, more complex and diversified phenomena, growing complexity and interactions within civil society, the heterogeneity of social agents and the great turmoil in the processes at work.

In this context it is very important to have a statistical system and an organization of statistical activity capable of understanding the complexity of modern society, addressing the demands of many and diverse users, quickly adapting to the new needs and even anticipating them. A highly flexible system, therefore, characterized by the ability to dialogue systematically with the different parts of the civil society and its bodies, which is able to use all sources of information and adopt the idea of a continuous updating of objectives, surveys, sources and methodologies as a method. The creation of a system based on the principles mentioned above is still a challenge to be achieved that requires a specific political will, the collaboration of many factors, the contribution of various parties, the availability of adequate means and qualified human resources.

Among the elements of complexity with which a modern statistical system has to deal with, it should be included what could be called the “*abundance of data/information*”² coming from several sources. If we look at statistical production rather than at the organization of the statistical system and put ourselves in a scholar’s perspective, it is impossible not to notice some shortcomings in statistical production. The lack of data at the local level can be mentioned; it is possible to complain about an insufficient production of longitudinal statistics that would allow for a dynamic analysis of phenomena; it can be emphasized that the statistics needed in order to understand the evolution of production systems, of competitiveness and the effect of innovation are still insufficient. However, if we place ourselves in the shoes of a common citizen, a business agent or even a politician, one realizes that they deal everyday with a huge amount of information that is divulged daily by the media, international organizations, specific bodies, institutions, associations and businesses, or are produced by their current activities and on these they rely on to take many decisions, to these they must comply with, with these they have to deal with.

² The author also discussed about “*a flood of data: a new challenge for the quality of statistics*” in his talk at the X Statistics National conference, Rome, 15-16 December 2010.

Today quantitative information, largely of administrative and managerial source, is plentiful, accessible, often used to produce statistics and the user's issue is not only what to choose but also how not to be influenced or misled. It could be said that we are faced with an unimaginable and increasing availability of electronic information that a growing number of bodies and private parties collect, store, develop and spread through many different channels. Some examples could help to understand the size of the phenomenon. The Wall Mart chain stores company registers one million transactions per day equal to 2.5 petabytes³, and, albeit with different numbers, the other big chain stores are in a similar situation. Banks insurances and financial institutions collect databases on their customers and from these not only they drive risk analysis but also macro-economic indicators and forecasts of general interest. The Public Administration collects enormous databases on population, taxes, social security, health system, employees, on the building activity, and many others to which the governments refers to for many decisions. The opportunities offered by this wealth of information have certainly not been overlooked by the markets and indeed some large companies, i.e. Oracle, IBM, Microsoft, have invested heavily to acquire a *software* that allows for data and *database* management, thus preparing to provide a technology to build larger integrated information systems and to provide tools able to draw from them a large amount of data.

Together with this availability, which could be defined of managerial origin, there is the revolution that is relentlessly progressing with the evolution of the *web*. A transformation of the Internet that is characterized by all those *online* applications allowing for a high level of user-site interaction. Therefore the *web* becomes an instrument for bringing together the contributions of millions of people and for improving them, to create new way to interact (i.e.: social network). The web turns into an open platform that increases the free flow of information, enhances the tools for political and social debate and widens the diversity of views. *Web* users turn from simple information users to information distributors and producers. Moreover, with the new *cloud computing* technologies it will be easier to promote the mentioned processes allowing the access to big storage devices and the big data-bases.

All this suggests that we are facing what we might call a real "*electronic revolution of data*"⁴ that is extremely pervasive and is already producing, and will bring even more, an enormous availability of information that, among other things, is characterized by a kind of data so far rather rare: the micro-data. We might not be far off from the possibility to identify and explore new phenomena, households' and businesses' behaviors and strategies, to deal with new issues, thus considerably enhancing the understanding of phenomena stemming from the framework of macro-economic statistics. Such a wealth of information could reveal new perspectives, deepen and expand knowledge, particularly that pertaining to individual

³ A petabyte equals to one million GB.

⁴ Joe Hellerstein – computer scientist at the University of California, Berkley – used this definition, others computer scientists coined the new term "big data".

behaviours, encouraging the drafting of increasingly more targeted policies and *business* strategies. Is it not an exaggeration to say that the current widespread availability of information could open the way for the enrichment of a knowledge that is too influenced by the statistical data derived from conceptual models weakened by the evolution of modern society⁵.

The richness of information is only a prerequisite for an actual improvement of knowledge. Indeed, having many data does not imply that they are all equally relevant to demands for knowledge that we face today, nor this ensures a production of statistics able to respect the fundamental principles of impartiality, objectivity and reliability that the European Statistical Code places at the basis of the authoritativeness and effectiveness of a statistical system⁶. However, the production of such information is a fact and the tendency to use them is and will always be stronger because it is inevitably linked to the primary role that knowledge and therefore information, has taken in modern economy. It follows that exploiting the potentialities that we have mentioned is not trivial but considerable issues have to be faced. Among them a major problem produced by the mentioned processes could be the emergence of *competitors* for the official statistics or an incentive towards a fragmentation of the statistical function. Therefore, Official Statistic – and statisticians – must reflect on its role and organization in a modern statistical system. This reflection should concern technical and formal aspects, the phenomena on which to focus, but also extend to more efficient and less costly strategies for collecting information, to the identification of ways of collaboration, interaction and communication with the users and other entities that collect information and, last but not least, to the means by which to ensure the quality of the statistics and the accessibility to information for the entire civil society.

2.2. *Some overall issues lying ahead the use of administrative sources*

Does not seem useless to start recalling that not any quantitative information is a statistical data and certainly not a form of substantial knowledge. Statistics stem from a set of necessary conditions: a specific objective, a system of hypotheses, classification criteria and definitions, a strategy and controlled process of measurement and validation in order to measure data reliability through clear and objective indicators, understandable to all. Statistics are therefore a product that

⁵ It could be useful to remind that the various scientific revolutions have been often preceded by a revolution in measurement and observation methods. For example, just think about what the discovery of the microscope meant for biology.

⁶ Actually, the European statistical code applies to official statistics and takes into account some other principles such as: the “statistical secret” and the “*favourable cost-benefit relationship*”, principles that are not relevant for the subject of this paper. The definitions are: «impartiality»: statistics must be developed, produced and distributed in a neutral way, by ensuring an equal treatment for all users; «objectivity»: statistics must be developed, produced and distributed in a systematic, safe and unbiased way; this requires the respect of ethical and professional regulations and implies that the policies and procedures are transparent for the users and those participating in the; «reliability»: statistics must measure, as faithfully, accurately and consistently as possible, the reality that they intend to represent; this implies the use of scientific criteria in the choice of the sources, methods and procedures.

meets specific goals in terms of knowledge and whose reliability cannot be discussed on a strictly statistical technical level. As Fortunati (1978) taught "... at the basis of the differences in trends and opinions there are beliefs, not always explicit, on the type of knowledge related to the results of findings and statistical surveys, and the kind of information needs related to the new dimension of legislation and administration" and, I would add, of civil society. From the coexistence and interaction of these factors stems the heuristic value of empirical evidence, and Official statistics is legitimate if and as it explicitly defines the hypotheses underlying the data and all aspects of the measurement process and it does not limit itself to provide basic analyses of prepackaged data, but allows them to be used thus enabling users to elaborate them independently.

Other types of information derive instead from bureaucratic and/or managerial purposes and meet reliability criteria very different from those of the statistical survey, therefore their statistical utilization arise specific problems not simple to deal with. However they can be addressed, as we will explain later, but certainly not forgotten, neglected or treated superficially.

But there are other central overall problems that it is important to consider. "Non-statistical" sources, for the very reason of being produced by various parties for many different purposes, appear uncoordinated, occasional, made with different methodologies, suffering from frequent adjustments, generally characterized by a lack of attention to a well-specified definition of quality, without adequate documentation of the procedure adopted for collecting information and, more generally, of the process that led to the data provided. Moreover, looking at the data bases on internet, the order in which a search engine is listing sites does not take into account the reliability and quality of the database supplier and much less that of the data provided. However, these information databases allow you to easily acquire knowledge about many subjects and specific contexts that are not available in official statistics. Therefore these sources of information are appealing to an audience of users inattentive to the complex problems concerning gathering techniques and not adequately aware of the dangers of a kind of information affected by measurement errors if not by some manipulation. One could argue that everyone should be free to choose the information that deems useful, but will have to face the consequences of the choice of biased or unreliable information. The matter however is not so easy, especially if such sources, so to speak "alternative", concern the same phenomena covered by official statistics – or aspects related to them – i.e. those statistics that should guide government action, permit its evaluation and comparison, and therefore become foundations for shared information.

Therefore the risks are two, both different but equally dangerous for a democratic society that needs to make informed choices and that is increasingly based on factual knowledge. The first is an a-critical use of information that – at best – is useless to understand the phenomena considered, but could also be wrong or misleading because aimed at conditioning the opinions. The other, is a juxtaposition of conflicting data that could lead to a loss of credibility of statistics as a *modus intellegendi* for modern science and knowledge.

The weakening of statistics, as a technique as ever essential for the progress of knowledge in a context where no longer deterministic approaches and ideologies prevail, would be very useful to those who might think about comparing ideas escaping the evidence of shared facts to instead anchor itself onto ideological and biased prejudices. Moreover, it could impact on the basic right of individuals in the contemporary society to have access to information. Stiglitz, in a UNDP report, identified three primary reasons behind the need for the public nature of information: with regards to its utilization, because every citizen should have free access to information; regarding participation, because deciding which and how much information should be produced must take into consideration the views and needs of all stakeholders in an open process; with reference to the spread of benefits, because every citizen is entitled to benefit from it.

Such connotations regarding the public nature of statistics leave open the matter of whether it could or should be produced only by a single public subject. It seems very difficult to challenge the idea that the State should take charge of organizing and producing a system of public statistics as a tool to guarantee equality for citizens' rights. In fact if, in addition to the points highlighted by Stiglitz, the cost of collecting data is considered it seems evident that the absence of a public agency for statistics could generate dangerous informative asymmetries by putting the weakest agents (individuals, businesses, institutions, social groups) in a disadvantaged position due to the difficulty in obtaining statistics on the market.

2.3. Some key objectives for a modern statistical system

National statistical systems should form the core of statistical information, but they must also tackle the new characteristics that information is taking. In fact, we cannot plan to restrain the flow of information to which we have referred to, but we have to act positively. From this point of view, it would certainly be useful to outline a policy, a code of ethics and some rules for everyone producing statistics, but we also need strategies and actions that, on the one hand, can promote communication and interchange between the public system of statistics and other data producers and, on the other hand, can define concrete ways through which to ensure the reliability of the information.

In order to develop these aspects it is useful and necessary to discern between the identification of general actions from the in-depth examination of technical issues. In this latter, it is useful to consider that various issues that may arise from the use of non-statistical sources for statistical purposes can include very different sources: those strictly bureaucratic-administrative and those of managerial-administrative character⁷.

Regarding the first aspect we can identify three main areas calling for a change in the statistical system.

The dynamic with the demands for statistics and therefore with the users as well as to who co-operate to the statistical production is the first of such areas. A

⁷ The former refer to administrative sources produced by entities belonging to the Sistan; the latter refer to the sources generated by any other agent.

new way of relating to users is needed, a way that allows for a prompt detection of the needs occurring in all sectors of the civil society and for translating them into easily accessible statistics and able to ensure to return data to who contribute to collect data. Lets think, for example, to the demand of statistics from the local institution because of the new many tasks they have to accomplish in the federal organization of the State. This means creating the tools and spaces for a constant and effective dialogue with the *stakeholders*. As it is necessary to ensure a quick return of data to who has co-operate to data collection. We cannot ignore that from this point of view the Istat in particular must undertake new routes. An initiative that would meet a substantial slice of the demand is probably to make available official databases, and therefore to allow access to micro-data by finding appropriate ways to protect the *privacy*. Official databases have to be made available to anyone wishing to use them – and mainly who contribute to collect data –, thus forcing both producers and users of statistics to interact with each other regarding specific sources, and the producers to take on board criticism and suggestions for improving their products. This is already happening in the United States, Australia and New Zealand where set of “machine-readable data” are provided, and in the UK, where from the beginning of the current year a special site that anyone can access has become available. Returning timely statistics to the statistical system contributors should be a primary goal, but it is also a good way to strengthen the focus on the quality of the surveys and the transparency of the information produced. This can also generate considerable side-effects: the possibility for the whole of society to monitor the development of those phenomena of interest to them and to evaluate government action, and the possibility for governments to monitor the effectiveness of their bureaucracy. This can stimulate within society the emergence of independent data analysis centres, the creation of software for the management and analysis of databases and the user-friendly interfaces. These centers could also involve partnerships between public and private entities and the data producers with huge benefits for both producers and users of statistics in terms of outlining and exchanging know-hows, best practices and methodologies.

For Official statistics this is a revolution in the way of working that involves the need to focus on producing qualitatively reliable, transparent, timely and secure databases and this requires a shift from a control strategy of the surveys and the produced sources *a posteriori* to an ongoing control strategy of the production process⁸.

A second goal to be reached concerns updating the modalities through which public statistics divulgates, communicates and informs the civil society about its own products, their potential and their limitations. Accessibility of databases already meets this goal but it is not the only initiative to be performed, it is necessary as well to create a communication strategy based on straightforward and easily understandable messages that could clarify and document the procedures used,

⁸ This aspect would deserve an in-depth analysis; here we mention Morganstein, Marker (1997), Filippucci, Calia P. (2000) and Baigorri, Linden (2007) where the “Quality Assurance Framework for Eurostat” is presented.

in order to educate users to screen the characteristics of a data, but also its weaknesses. This communication strategy has to then find ways, other than the bulletin or the methodological notes on the Istat website, to know how to use the opportunities offered by the *web* to their full potential.

A third issue arises from the fact that in any case, official statistics and therefore the Istat cannot, like any other national institution, produce the enormous amount of statistics that are and will be required through its own surveys alone. It is already internationally well-established the idea that surveys are just one of the sources of statistics and that an ever increasing contribution could come from the use of administrative sources and their integration with surveys⁹, thus creating a statistical system capable of using “the contribution of different parties”. It cannot therefore be ignored the goal, which is very complex and longer term, of creating an interaction with all the other databases produced by parties other than the government and that could contribute towards meeting many specific needs for information. For this purpose, there are two fundamental steps to perform. The first one is re-structuring the national statistical system aimed at overcoming the “weakness” plaguing the Istat compared to other public administration bodies (see next section) thus realizing the objective of a concrete integration and interaction between the participating institutions and thus strengthening the significance and the prominent role of public statistics within a statistical system and as a reference point for relations with other producers of data¹⁰. The second is to establish an effective statistical authority that could assess all the parties converging into the system but also ensure that statistical activity, wherever undertaken, respect the fundamental principles of statistics.

It is important to remember also that the use of the many available databases, and in particular of the sources of administrative origin, today attracts the attention of many entities operating at the local level due to the increasing shift in government responsibilities and the not always adequate availability of statistics¹¹.

The Sistan should be the tool through which Official statistics can establish a relationship of collaboration, exchange of information, can learn from experiences developed when applying innovative methods of enquiry and, above all, can affect their production by outlining guidelines and suggesting widely tested methods and procedures.

⁹ A wide literature on this subject is available, a recent reference is the IAOS Conference on 2008, “*Smart data, innovative uses- reshaping official statistics?*”.

¹⁰ Many norms and regulations have been created to ensure a working interaction between the Istat and the PA, but concretely they have found a very limited application (Sestito, Trivellato, 2010, p. 17).

¹¹ The use of administrative sources by local bodies in Italy is illustrated by the reports presented at the IX Statistics National Conference, “Administrative sources: a primary resource for official statistics”. Rome.

3. THE CHALLENGE OF USING ADMINISTRATIVE AND MANAGERIAL SOURCES

This leads us to the second issue that we have mentioned, which involves the use of administrative and managerial sources, regarding which we will try to illustrate which principles and categories could be used to describe the quality of these sources and to guide users in their choice, and also to highlight the main issues raised by their use for statistical purposes¹².

There is no doubt that the use of administrative records in the field of the statistical production is not a new phenomenon. At least until after World War II, this data have been, together with the census, the main source for statistics (Brackstone, 1987) and, after a phase in which statisticians' interest has focused on undertaking large sample surveys, starting from the eighties it appeared throughout the world a renewed interest in the statistical use of administrative sources of data (Coombs, Singh, 1987). For over twenty years, in all major statistical agencies this "strategy of data collection" has not only been discussed and researched in-depth, but it has also been put into practice¹³. Recalling that some countries have basically abandoned census surveys by replacing them with an appropriate use of the administrative records could be enough, but let us also stress the many others possible statistical utilizations of archives. They can be used for sample frames, supply important auxiliary variables for surveys and give data to supplement surveys, provide original databases able to substitute surveys giving direct estimates. Last but not least, longitudinal data and highly disaggregated data at local and sectoral levels are achievable using administrative sources.

We can therefore state that their statistical use is not only old, but has also gained great importance in national statistical systems.

3.1. *Administrative sources and the statistical system between past and present*

The most recent reflection has received many contributions from outside of Italy not because Italian public statisticians are less "innovative" but probably because, while in most countries – at least those of Anglo-Saxon tradition – between the '20s and '50s it was deeply established the culture of inference and of probability sampling¹⁴, so the use of administrative records could seem a return to a neo-descriptive tradition, on the contrary in Italy this last approach has prevailed for a longer time.

At its origin the Italian public statistics was mainly aimed at representing the State organization and was characterized by a strongly descriptive setup, finding

¹² The decision to discuss together administrative sources produced by entities belonging to the Sistan and other managerial sources does not imply that we are underestimating the clear differences that still exist and that will be underlined any time it will be needed.

¹³ Conferences, seminars and published essays on the issue are many, here we will only mention the Eurostat Manual (1999); the *European Conferences on Quality of Statistics*, namely on 2008 and 2010; the international conferences organized by Statistics Canada; the IAOS Conferences, namely on 2008; the *Annual Conference by the Federal Committee on Statistical Methodology* in the USA, in particular on 2009; the Istat National Conferences.

¹⁴ For a brief history of sampling surveys see Hansen (1987).

its primary source in administrative archives¹⁵. The deep change that Italian official statistics made with the birth of the Istat in 1926 did not lead to a different approach in the methods of observation and data collection either. With the work of Corrado Gini, the modernization of statistics has been important, the openness toward observing new phenomena was considerable, but the orientation in terms of data collection was characterized by a specific focus on universal surveys and towards those administrative sources that seemed an appropriate surrogate for the former. It should be noted that this approach was aimed at extending official statistics beyond economic phenomena and at unifying the role of statistics in the country through the coordination and centralization of the statistical activity carried out by other administrations. The strategy was, and still is, quite ambitious but as a matter of fact it failed, at least in what concerns the ambition to create a unified statistical system, due to the corporative and independent oppositions of the various branches of the public administration (Sestito, Trivellato, 2010): as we can see not much has changed!

Many changes have taken place since then and we can state that Italian statistics, pushed by the profound changes that have characterized the country's economic and social development, is now for many aspects at same level of other National Agencies. However, it was with Rey as a president that a major change took place within Italian statistics (Favero and Trivellato, 2000). Rey's strategy, consistent with the careful and critic reflection in the well-known Moser Report (1983), aimed at developing the coordination and integration of data and the use of methodologies and statistical techniques, at achieving progress in the use of administrative data for statistical purposes; at developing a clear procedure for the planning the statistical activity carried out by the Istat and the other institutions of Public Administration.

New circumstances required a change of perspective and a diversification and expansion of statistical sources. The point was, and still is, the requirement to meet a growing demand for systematic and periodic statistics on a increasing number of phenomena, domains and specific sectors; to meet the new demand for individual data and longitudinal information; to rationalize, coordinate and make better use of available resources and deal with a shortage of resources; to limit "the statistical burden" borne by households and businesses. Administrative sources returned then, thanks also to the progress of computerization, to be a primary reference for public statistics.

The renewal has encountered obstacles that, however, as in the case of the Sis-tan, can lead us to talk about a still incomplete process if not, as someone is thinking, a failure.

This difficulty has had negative consequences for the use of administrative archives. In fact, their use for statistical purposes requires methods, techniques, and above all interaction between archives holders and statistics producers, and it

¹⁵ In 1861 Pietro Maestri founded the Italian Official Statistic that was developed under the direction of Luigi Bodio. A thorough and detailed history of public Italian statistics is in Sestito, Trivellato (2010).

takes place only if an institutional context, and a shared organizational and relational model exists and it is able to ensure a dialogue between Public Administration and statistics, as well as enabling an exchange of tools and methodologies. It is indeed important to standardize classifications, to enable the exchange of records, to know – even when it is not for the purpose of sharing – the specific procedures for gathering information.

The history of Italian Statistics has been characterized by factors that have encouraged or slowed down the creation of such a system, having an effect on the spread and the use of statistics. It was not only a matter of issues related to formal and technical choices but also of the influence that have had both official statistics' tendency toward centralization, particularly strong until the eighties, and the complex organization of the Italian Government. The interplay between the characteristics of the institutional organization and the centralization of statistical production – but not its integration at the central level – has certainly slowed down, or at least limited, the development of the statistical integration of various government agencies and a widespread use of administrative records. As consequence, the development of a thorough analysis on the problems that their use poses have been influenced. On the contrary, scholar's needs and their increasing sophistication, and the subsequent increased attention on data reliability and compliance with quality standards¹⁶ made it clear that the statistical use of administrative sources presents many challenges and therefore requires the development of strategies, categories, methodologies, and specific techniques.

The difficulties encountered by the Sistan are one of the main causes that are fostering an “informative disorder” and an inability to enrich our understanding regarding many phenomena happening both at a national scale and within the statistical production at local level. In relation to this, we can mention the difficulties in using the municipal registry that still today forces us to undertake very costly population census; those connected to the minimal use of employment-related data collected – i.e. by Inps and by the Ministry of Labour –; the difficulties that statistical activity encounters in the Italian regions who represent an emblematic case concerning one of the important parts in the Sistan¹⁷, which offers the possibility of documenting the various and serious problems confronting the implementation of a statistical system that has the characteristics that have been mentioned. Again arise the crucial challenge of the reorganization of Sistan. In this context, Official Statistics should work to formulate a general plan for the use of administrative sources, a plan able to create the institutional conditions needed in order to address issues associated with the exchange and the matching of administrative records, therefore able to put together into a system the various initia-

¹⁶ Awareness of the importance of data quality has established itself in time with the establishment of the idea that reliability and usefulness of a data depend on all the phases of the measurement process. For a detailed reflection on this process, refer to Filippucci (2000). Quality standards, procedures and methodologies are acknowledged by almost all statistics' national and agencies.

¹⁷ A thorough analysis of the statistical activity in the Italian regions and of the limitations it presents with regards to the act DL 322/1989 has emerged from the inquiry undertaken by Filippucci in 2007 for the Commissione per la Garanzia dell'Informazione Statistica, Cogis (2008).

tives to be used for statistical purposes. This task requires a strong political support, a new cultural awareness and technical skills. In other words, the issue is to create the appropriate contextual conditions to establish legal and procedural protocols for accessing and sharing sources. Finally, we need to ensure greater security for those who provide information through a law that protects the archives¹⁸. The two issues to be addressed are ensuring *privacy*, aimed at reaching an informed consent by those who hold information, and that of *confidentiality*, which is the strict specification of how to avoid misuse of the information contained in the files once it is made available for statistical purposes.

Computerization is an important condition in order to make such a system work, now information is generally stored on computerized devices but the effort still to be made concerns the development of a system able to effectively support communication and interaction between the its main parts. The reference to computerization gives us an opportunity for emphasizing that we also should not be misled by the idea of creating “universal containers” – *data warehouses* that only stand next to all the information from different databases – but we should rather focus on creating databases organized in view of systematizing the information relevant to the production of statistics and indicators of general interest, and those specific to the work undertaken by each agency or institution¹⁹.

3.2. *Statistical use and quality standards of administrative and managerial sources*

Statistical use of administrative sources is certainly positive and has great potential, but its implementation is not easy, and, above all, presents different connotations that require appropriate strategies and methodologies. In the 2008 IAOS conference, Nordbotten pointed out that technical issues related to the use of administrative records are by no means resolved, but much methodological work still needs to be done in various fields, in particular in order to ensure the quality of these sources. It is necessary to define conceptual models, methods and operating procedures resulting from the contribution and the combined efforts of different cultural approaches (legal, administrative, managerial and statistical) in order to ensure the production of data that are reliable, consistent with shared principles and standards, comparable across time and space and therefore able to comply with shared quality principles. The development of adequate principles and tools will be very useful even outside the Public Administration, i.e. for all the producers of data that do not come from a statistical survey. Finally, it will also be a reference point in guiding any user, and helping to judge the quality of the measurements he uses.

The quality of administrative sources is therefore a primary issue to deal with especially considering the appearance of many subjects that are referring to such databases for statistical purpose and the establishment of a society where infor-

¹⁸ Gates G. (2009) makes an in-depth reflection on this theme.

¹⁹ A suggestion in this sense also comes from a survey promoted by Commissione per la garanzia dell'informazione statistica “L'indagine sulla informazione statistica e gli indicatori per il governo della pubblica amministrazione a livello locale” coordinated in 2010 by S. Viviani.

mation circulates widely and is used by a variety of parties, requires an increasing attention to the quality of the information that are used. Data quality has been widely studied and discussed and many contributions have been produced, so broad consensus exists on its importance and statistical agencies pay now increasing attention on it²⁰. In contrast, the quality of sources of administrative-managerial origin has not been well considered, but it is only few years that scholars and national agencies focus on it²¹.

3.2.1. *A note on inference by administrative sources*

Before dealing with quality issues in administrative sources is suitable to distinguish between the use of an archive for directly drawing estimates on the population of interest, and any use intended to establish a database to be used for statistical purposes or for integrating, controlling a statistical source, or for building models aimed at the production of data²². In the follows this case is considered. The more general issue of quality will be deepened in the next section.

When direct estimates on the population of interest are drawn directly from an administrative source the problems of “population coverage”²³ with regards to the source and the handling of the potential under/over coverage prevail on the others. The matter obviously arises from the specific nature of the administrative-managerial sources that do not select the units according to a sampling procedure nor, generally, cover the whole population. From a statistical point of view, this could generate a problem of self-selection of the population units, which could have severe effects as it could lead to biased final estimations. It is clear that this issue must be addressed because from a statistical point of view the large amount of information coming from an administrative source rather than from a survey is not satisfactory. A large amounts of information is very important but not sufficient to produce an assessment of the data reliability. The variables of interest estimation can be made by recognizing in the first place what caused the *lack of presence* of population units. This phenomenon could be assimilated to the well-known problems of census under coverage and incomplete frames, nonresponse generating non-random selections and the case, increasingly frequent, of web surveys. Therefore classical approaches applied in these cases could be useful references to deal with our problem²⁴. However, measuring under coverage by specific sample surveys is not possible in general, then strategies used in those situations

²⁰ Significant general references to quality in official statistics are: the EFQM *excellence model*, published by the European Foundation for Quality Management in 2010 and the *European statistical code of practice* adopted by Eurostat on 2005.

²¹ To recall: the pioner paper by Vale, Perry, Pont (2001); the activity carried out by Eurostat (2003), the papers by some Scandinavian and Dutch researchers: Wallgren and Wallgren (2007), Baigorri, Linden (2007), Daas, et al. (2008, 2010), Laitila, Wallgren, Wallgren (2011).

²² For example, small area estimations.

²³ We could also talk about the representativeness of a source.

²⁴ Census under coverage has been widely studied, here we recall the two fundamental papers by Wolter (1986) and Cressie (1989), moreover the recent paper of Fortini, Gallo (2009) are cited because of its relation with an administrative source. About frame errors see Lessler, Kalsbeek (1992).

cannot be implemented. About methods to deal with frame errors it is to stress that the difference between target population and “administrative population” is in general more wide and not at random. The comparison with survey nonresponse is also not fully appropriate but the strategies adopted to study the nonresponse generating process and to deal with it are useful points of reference in the case of administrative sources as well. In fact, only starting from this analysis we can detect the risk of obtaining estimations that are biased or influenced by too much variability and what approach is to follow to solve the problems²⁵.

Roughly there are two options here: we have to limit ourselves to redefine the population on which inference is made; we can resort to some adjustment strategies. The former is a substantial decision not to address the problem, the latter is a substantial way to address it and has a specific statistical basis. Moreover, in the latter, there are two main alternatives: a strategy based on a weighting procedure²⁶, in this methodological context an interesting option is the *propensity score weighting*²⁷; we resort to a classical model based estimating strategy or a maximum-entropy estimating approach²⁸. Comparing these strategies is not the objective of the paper, moreover further methodological and experimental research remains to be done in order to comparatively evaluate the various strategies, but certainly the approaches suggested contain the basis to provide a solution to the problem of the lack of coverage in administrative sources.

3.2.2. *Features of quality in administrative-managerial sources*

Quality analysis needs definition of the general categories/dimensions²⁹ to precise its actual content, and to allow its measurement. These categories can have further specification so that a deeper analysis and appropriate indicators are allowed.

More in detail, the functions of quality dimensions for administrative and managerial sources are³⁰: i) to identify the criteria to be taken into account when evaluating a source or an information; ii) to provide a systematic framework for communication with the users; iii) to assign, if so it can be said, a *rating* for the sources. Then the goal to be pursued in suggesting qualitative categories for a source is not only to specify the main issues to be considered for an assessment of the source that we are about to use, but also to highlight if using that source in order to obtain effective results able to satisfy one’s needs is suitable. In other

²⁵ For the analysis of nonresponse generating process, see Little & Rubin (1987) and Groves et al (2002).

²⁶ These strategies were influenced by a work by Little & Rubin (1987) who suggested a method based on weighting.

²⁷ See Danielsson S. (2004).

²⁸ About the first (also predictive approach) see Särndal et al (1992), Bolfarine, Zacks (1992); some applications are in Filippucci, Drudi, (2000) and recently Thomsen, Li-Chun Zhang (2008). About the maximum-entropy approach see Filippucci, Bernardini (2000).

²⁹ Daaas et al. (2008) refer to “hyperdimensions” of quality.

³⁰ In terms of statistical use, the sources of information produced by Sistan’s agencies and by parties that do not belong to the Sistan present some similar issues. Differences will be stressed if arise.

words, the use of an administrative or managerial source for statistical purposes must be carefully evaluated through a sort of analysis of the benefits but also of the costs that this entails and the limitations it presents. In this perspective, the categories suggested below could form the basis for such assessment.

Various classifications of quality categories for administrative sources have been suggested; they can be synthesized distinguishing: i) quality of a source; ii) quality of data production process; iii) quality of output. However, in this paper, we suggest a preliminary distinction quality dimensions – discussed in this section – and errors stemming from the statistical utilization of an administrative source – discussed in the section 2.2.3 –. This classification seems useful in view of calculating errors. About quality dimensions we suggest to distinguish as to whether we refer to the quality of a source or that of the information produced.

The labels chosen for the categories are aimed at understanding the specific case of information/data contained within an administrative-managerial databases and therefore differ from those usually applied for statistical surveys. It is also to note that in this paper only the more general categories are considered even if, as it has been said, it could be possible to choose a deeper detail and to introduce an higher number of quality dimensions (see Laitila, Wallgren, Wallgren, 2011).

Regarding the *quality of a source*, six dimensions can be taken into account. *Legitimacy* indicates the conformity of a source to a “code of ethics” on how information is gathered. *Authority* indicates the prestige, independence and professionalism of the producer of a source. *Credibility-reliability*, through which to grasp the standards for ensuring compliance with general principles and practices for data collection and hence also the standards for collecting information according to specific established guidelines. *Transparency*, aimed at informing about the existence, quality and quantity of meta-data available from the source. *Usability* should indicate the possibility to easily obtain information in electronic format. *Stability*, aimed at assessing the permanence and continuity of the source, of the definitions and the information gathered (for example: changes in tax laws and/or more severity in fiscal controls change tax data). It could be said that this is about tracking back the source’s life cycle and therefore gives directions on the opportunity to invest in its use for statistical purposes.

With reference to the *quality of data* within the sources that we have analyzed so far, quality could be expressed through other six dimensions. *Relevance*, aimed at understanding the level of parallelism between statistical definitions and those used in non-statistical sources. *Precision*, concerning the closeness of the information gathered with the phenomenon in question and therefore the existence of assessment criteria for measurement errors. *Timeliness* regarding the time lag between when events occur and when the information is collected. *Punctuality* with the goal to clarify the time lag between the release date of data and the time when the event happened. *Compliance*, through which it should be possible to assess whether the various data collected are mutually consistent and integrated. *Comparability*, aimed at observing the existence of similarities and differences between the information produced in the course of time, space and for the relevant fields.

The work to be done in order to refine and make measurable the categories outlined above is certainly still relevant and it is clear that, even in this case, Official Statistic should play a key role by taking on the task of identifying the indicators to scale these classes. It should also take on the task of directing users regarding the meaning of quality of these administrative sources and the best practices to be used in order to integrate such sources and surveys, especially in the pursuit of producing data with high geographical and economic sector detail.

3.2.3. *The errors in the production process*

The general issue of assessing the quality of the production process of administrative data and the specification and evaluation of measurement error³¹ is the last topic considered.

It is obvious that the way in which the quality of an archive can be interpreted is different from the one used for a statistical source³². The producer of an administrative archive cannot have, due to the purpose and the method used to create archives, the same care, sensitivity and knowledge of the producer of statistics, much less, he can consider relevant the same categories as in the case of a statistician. In addition, the quality control of statistical data derived from an archive extends beyond the stage of information gathering, because it also concerns the necessary procedures for transforming it into a statistics and, in the event of the use of multiple sources, their integration.

Even in the case of an administrative source there is a more or less structured and complex measurement process, inevitably affected by errors and, it is perhaps worth emphasizing that admitting the existence of errors in a measurement process and then trying to assess them, also implies a clear conceptualization of the different types of error that can occur and consequently the definition of the different evaluation and correction approaches³³. If one were to accept the similarity between the two processes, it should be noted that there is not a general consensus when defining what is meant by measurement process control³⁴ and that among many archives users there is some superficiality in this respect. Quality assessment cannot be reduced simply to a few indicators, or to the description of the “*error profile*” related to the data elaboration procedure, nor it is to be found in the calculation of any component of the error or in the explanation of edit/imputation strategies, but it includes all these aspects and should also take into account the existence and effectiveness of procedures for process control and documentation of the strategies and methods adopted. We should also not forget

³¹ It would seem inappropriate to talk about non-sampling errors for administrative sources, but by this term we now indicate all those errors that take place in any phase of the measurement process, excluding sampling errors. For a description of the survey errors, see Grooves et al. (2004).

³² See Calzaroni (2008).

³³ For example, response error measurement and its correction have a very different foundation compared to the treatment of total or partial nonresponse. The error induced by a poor formulation of the questionnaire has little to do with the issues connected with listing errors, and so forth, see Lessler, Kalsbeck (1992).

³⁴ On this theme see the contributions by Filippucci (2000) and Baigorri, Linden (2009).

the usefulness of running appropriate “external” checks by using existing or *ad hoc* surveys.

The reference to external audits leads us to be cautious even of a simplistic approach that can sometimes generate misleading assessments. It is not uncommon to observe that in order to “validate” an administrative source such information is compared with data resulting from a statistical survey, perhaps considering only a few variables contained within the two sources. This is a purely empirical approach that applies to the situation in question and does not allow for generalization, above all it requires a systematic replication. It is quite obvious that it is not enough that two data resemble one another in a specific occasion, as this similarity may result from different errors that balance each other out. Therefore external verification too requires a strategy and specific methods that would be appropriate to share.

In summary, to talk about the quality of an administrative data it is important to clarify what are the aspects that define it, how to measure them and when you reach a level of acceptable quality. And, once there is agreement on all the aspects mentioned, a strategy able to ensure common behaviors, and therefore guidelines, indicators of best practices and methodologies are needed: it is clear that these problems cannot be addressed independently by each user.

The pursuit of quality is generally aimed at tackling the causes that have led to it and not only at providing a measurement of it. However, while those who design a statistical survey specify the quality control strategy (or at least they should), in the case of administrative sources, those who hold the archives are in general not interested in quality control and have a very bureaucratic concept of it. Hence, again, the need to develop a partnership between producers and users of administrative sources for statistical purposes, which is still not a real and widespread option.

Finally, it is worth identifying specific non-sampling errors that may occur in an administrative data or that may be generated through their use, in fact, despite the increasingly widespread use of non-statistical sources, the analysis and measurement of such errors has not been covered much and literature on this subject is poor³⁵.

In order to address this issue here we suggest a possible difference between *measurement errors* and *representation errors*.

Measurement errors generally stem from the difficulty to state concepts and definitions adequate for measurement, from the willingness/ability to provide the information by its holder, from data processing and editing and, finally, from the measurement methods and instruments³⁶. In principle, therefore, those measurement errors typical of statistical surveys can also generally be found in administrative sources and they need to be evaluated, documented and addressed also in the latter. However, there are also some other typical characteristics of measurement

³⁵ Some recent contributions are: Grünewald, Körner (2005); Wallgren, Wallgren (2007); Bakker, Linder, Van Roon (2008).

³⁶ On survey errors see the Federal Committee on Statistical Methodology (2001).

errors in the archives. These could be grouped into three main categories: i) processing errors arising from the treatment of data as a result of checks carried out in the phase of acquisition administrative information, the application of specific correction and treatment rules for their statistical use and the transformation of administrative variables into statistical ones; ii) Errors caused by the different “quality” that the many variables contained in one source have. Such difference depends on the higher attention the source owner gives to those variables relevant to the agency’s specific purpose, the error may occur as a result of using different strategies and methods to control variables. iii) Errors due to what we might call administrative delay: that is, the time lag between when the events are recorded or declared and when they have occurred.

Representation errors, according Grooves et al. (2004), can be attributed to: sampling, frame/coverage, lack of response, linking of population units and correction. In administrative archives, the first type of error does not arise, while the issue of coverage is associated with no response that we have mentioned above. So in this context they can be traced back to basically two types: linking errors and correction errors. *Linking errors*³⁷ concern the use of multiple sources and are attributable to missing or incorrect links. Faced with a missing link, as in the case of a missing data, we must ask ourselves what is the process generating the missing information because, as we have said, only by recognizing its nature we can specify an appropriate statistical method to address it. *Correction errors* derive from difficulties in obtaining a full link between archives and from the weighting of registers to be linked that contribute to generate errors if the methods adopted are not appropriate.

Identifying errors, documenting them, but also tackling them, by clarifying the consequences of the handling process is therefore an important aspect in the use of an administrative source. If in principle we can understand this need, we cannot deny the great practical difficulty in measuring errors both due to the scarce sensitivity toward the problem by those who manage the archives, but also for the low thoroughness of users who finds considerable obstacles on this road. Finding documentation of the errors when data are collected is difficult and therefore it is difficult, on a statistical, bureaucratic and economic level, to create a strategy to measure them *a posteriori*, but, once again, is it possible to ignore them?

We shall not forget that addressing errors is complex because it is made *a posteriori* and anyway it does not prevent other types of errors to happen in future occasions. So also when considering administrative sources we should look into the possibility of shifting from measuring the error to a *continuous control and improvement of the data production process*. Maybe this is a utopian prospect, however it cannot be excluded in this list of the issues connected to the use of non-statistical sources. The continuous improvement of the process requires appropriate methodologies and skills and this implies the existence of joint effort of all parties supported by a theoretical and applied kind of research. Thus, it emerges once

³⁷ See Arts, Bakker, Van Lith (2000) and Al, Bakker, eds, (2000).

again the importance of having a statistical system that works, otherwise any quality control and, let alone, the control of the production process, would never be possible.

We should be able to trust an administrative source in the same way we trust any statistical source, but in that case, we need to undertake all necessary checks in order to evaluate and document on its quality and therefore on that of the obtainable statistical data. Just as we acknowledge that a statistical data is always affected by an error component – measuring means accepting the idea of having a knowledge that has an element of uncertainty – but also that we can calculate, identify and explain the magnitude of that error and then the degree of trust that we can attribute to the measurement, in the same way we cannot accept administrative information without appropriate controls. An administrative source exempt us from the task of undertaking large, annoying and expensive surveys, but does not exempt us from some of these controls, nor by the need to resort to, if necessary, any appropriate survey to monitor and verify the available information.

4. CONCLUSIONS

The statistical system is facing major challenges: the need to represent an increasingly complex society and to satisfy a growing number of information needs with dwindling resources; a larger information offer coming from many entities with various purposes that, thanks to the *web*, seems particularly appealing; a mediocre statistical literacy that, among other things, also neglects the idea of data reliability; an advanced project of the national statistical system that has not reached its objectives.

These issues place at the centre of attention with renewed urgency the question of the quality of the information produced. The question we must ask is therefore how to ensure data quality and how to raise awareness of its importance. To answer these questions two areas for reflection have been suggested.

The first concerned the relationship between the public statistical system and the totality of information produced by any other source. The second, given the increasingly important use of non-statistical sources, focused on finding criteria and categories to evaluate the quality of administrative sources and on identifying the types of problems and errors that may arise when these sources are used for statistical purposes. Ample space was dedicated to the latter, which could seem to be the main issue. However the two issues are closely linked and it is difficult to imagine that the principles, strategies and suggested methods can be applied if do not create substantial innovation in the relationship between public statistics and civil society – as a set of subjects that can help to collect relevant information for statistical purposes – and the totality of all the users. Such project is not easy and must begin by reviewing the strategy with which Italy is now addressing the issue concerning the use of administrative sources.

Nowadays the Istat has remedied the relational and partnership problems with the public administrations with the purpose of intervening on the existing sources

but with little power of involvement on the definition of the archives content and management. In the short term this strategy has achieved significant results (e.g. just think about the production of ASIA), but has not solved all difficulties and especially does not provide an answer to the issues that we have raised, therefore, it is not and cannot be the long-term prospect for the statistical system. Unfortunately, in this perspective the main problem concerns a renewal of the management culture inside the Public Administration and it cannot be narrowed down to a few adjustments – though necessary ones – to the existing legislation, and to introducing the conditions to ensure that its directives do not go unheeded.

With regard to the issues related to the use of administrative sources and because of the relevance that they have assumed, we have emphasized the need for an increasing attention to the quality of these data. The difficulty of measuring errors and costs that this implies should lead to carefully consider the use of an administrative source. The sources to be used should be selected on the basis of the characteristics and opportunities they offer, but also on the efforts they imply, the cost they entail and the reliability that can be obtained for the data. The choice of using an administrative source is limited by several factors: by the contents and their consistency with the goals to achieve in terms of knowledge, by the reference units, by the schedule, by the quality of the information and, lastly, by the cooperation you get from the owner of the source. It is also limited by what may be called the “stability” of the contents that may affect the comparability of the data. Let us not forget that administrative sources are affected by frequent amendments due, for example, to normative change and to the characteristics of administrative directives that vary according to the policies and operations of entities that produce them.

We should also be aware that the use of sources born with an administrative and managerial purpose, thus based on information related to the well-established economic and social aspects, can become a “conditioning” on the knowledge of a phenomenon and the identification of the emergence of something new. An administrative source by its nature can only be one pillar of a statistical system. The risk is to flatten out the information that already exists. We must therefore ensure that the viewpoint represented by those information does not influence the capacity to improve and extend the knowledge to be pursued through a statistical system. In other terms, we must avoid that the level of analysis is dictated by the available administrative information and, in any case, we must not give up on providing an accurate assessment of the reliability of the statistics derived from that information.

Using non-statistical sources is not simply a loophole in modern statistical production but also an inevitable choice – there should be no misunderstanding about this –, however we must also be aware that this road is not free of costs and requires great attention and still much work.

REFERENCES

- AL P., BAKKER B.F.M. eds. (2000), Re-engineering Social Statistics by micro-integration of different sources. Themanummer *Netherlands Official Statistics*, vol. 15, pp. 16-22.
- ARTS, K., BAKKER B.F.M., VAN LITH E. (2000), Linking administrative registers and household surveys, In: P. Al & B.F.M. Bakker (red.), Re-engineering Social Statistics by micro-integration of different sources. Themanummer *Netherlands Official Statistics*, 15, pp. 16-22.
- BAIGORRI, A., LINDEN, H. (2007), A Quality Assurance Framework for Eurostat, *CCSA meeting*, September 10-11, Madrid.
- BAIGORRI A., LINDEN H. (2009), The implementation of quality assurance frameworks for international and supranational organizations compiling statistics, International Statistical Conference “*Statistics: Investment in the future*”, Prague, September 14-15, 492.
- BAKKER B.F.M, LINDER F. VAN ROOND. (2008), Could that be true? Methodological issues when deriving educational attainment from different administrative data sources and surveys, *LAOS Conference on Reshaping Official Statistics*, Shanghai, China, October 12-14.
- BOLFARINE H., ZACKS S. (1992), Prediction Theory for Finite Populations, Springer - Probability & Statistics, New York.
- BRACKSTONE, G. (1987), Statistical Issues of Administrative Data: Issues and Challenges, Proceedings of *Statistical Uses of Administrative Data - An International Symposium*, Statistics Canada, Ottawa, December.
- CALZARONI M. (2008), Le fonti amministrative nei processi e nei prodotti della statistica ufficiale, in *IX Conferenza nazionale di statistica*, Roma.
- COGIS (2008), Indagine sull'attività statistica delle Regioni”, Commissione per la Garanzia dell'informazione statistica, Roma, in press.
- COOMBS, J. W., SINGH M.P. eds. (1987), Statistical issues of administrative data, International Symposium, Statistics Canada, Ottawa.
- DANIELSSON S. (2004), The propensity score and estimation in non-random surveys - an overview”, Report n.18 project “*Modern statistical survey methods*” Department of Statistics University of Linköping.
- DAAS, P.J.H., ARENDS-TÓTH, J., SCHOUTEN, B., KUIJVENHOVEN, L. (2008), Quality Framework for the Evaluation of Administrative Data. *Q2008 European Conference on Quality in Official Statistics*, Roma.
- DAAS, P.J.H., OSSEN, S.J.L., VIS-VISSCHERS, R.J.W.M., ARENDS-TOTH, J. (2009), Checklist for the Quality evaluation of Administrative Data Sources. *Discussion paper 09042*, Statistics Netherlands.
- DAAS, P.J.H., OSSEN, S.J.L., TENNEKES, M. (2010), Determination of Administrative Data Quality: Recent results and new developments, *Q2010 European Conference on Quality in Official Statistics*, Helsinki.
- EUROSTAT (1999), Use of Administrative Sources for Business Statistics Purposes - Handbook on Good Practices, Luxembourg.
- EUROSTAT (2003). Quality Assessment of Administrative Data for Statistical Purposes; Working group “*Assessment of Quality in Statistics*”, Luxembourg, 2-3 October, Web publication.
- FAVERO G., TRIVELLATO U. (2000), Il lavoro attraverso gli ‘Annali: dalle preoccupazioni sociali alla misura della partecipazione e dei comportamenti nel mercato del lavoro, in P. GERETTO (2000). (a cura di), *Statistica ufficiale e storia d'Italia. Gli “Annali di Statistica” dal 1871 al 1997*, Annali di Statistica, Serie X, Vol. 21, Roma, Istat.
- FEDERAL COMMITTEE ON STATISTICAL METHODOLOGY (2001), Measuring and Reporting Sources

- of Error in Surveys, US Office of Management and Budget, Statistical Policy Working Paper 31, Washington DC.
- FILIPPUCCI C., DRUDI I. (1996), Model based estimates using longitudinal non-random surveys, in *Rivista di Statistica Applicata*, vol. 8, n. 4.
- FILIPPUCCI C. (2000), Qualità delle statistiche e controllo del processo di misura, in *Rivista Italiana di Economia Demografia e Statistica*, vol. LIV, n. 2.
- FILIPPUCCI C., BERNARDINI-PAPALIA R. (2000), Inference from non-random samples: a maximum entropy approach, Proceedings of "International Conference on establishment Surveys II", American Statistical Association, Buffalo, 17-21 June.
- FILIPPUCCI C., CALIA P. (2002), "Il controllo del processo della produzione statistica: una rassegna", in FILIPPUCCI C. (a cura di), *Strategie e modelli per il controllo della qualità dei dati*, Milano, F. Angeli (17-56), 2002.
- FORTINI M., GALLO G. (2009), Misure di sottocopertura anagrafica in base alla revisione post-censuaria del 2001, mimeo presentato a SIS-GCD, Febbraio.
- FORTUNATI P. (1978), Ancora a proposito di servizi statistici pubblici, in "Scritti in onore di G. De Meo", Istituito di statistica economica, Università di Roma, Roma.
- GROVES R.M., DILLMAN DON A., ELTINGE J.L., LITTLE R.J.A. (2002), *Survey nonresponse*, New York, Wiley.
- GROOVES R.M., FOWLER F.J., COUPER M.P., LEKPWSKY J.M., SINGER E., TOURANGEAU R. (2004), *Survey Methodology*, New York, Wiley Interscience.
- GATES G. (2009), Expanding statistical use of administrative data: a research proposal focused on privacy and confidentiality, in *Federal Committee on Statistical Methodology Research Conference*, Washington DC 2-4 November.
- GRÜNEWALD, W., KÖRNER T. (2005), Quality on its way to maturity: Results of the European conference on Quality and methodology in Official Statistics (Q2004), in *Journal of Official Statistics*, vol. 21, n. 4, pp. 747-759.
- HANSEN M. H. (1987), Some History and Reminiscences on Survey Sampling, in *Statistical Science*, vol. 2, n.2, pp. 180-190.
- LESSLER J.T., KALSBECK W.D. (1992), *Nonsampling Errors in Surveys*, New York, J. Wiley & Sons.
- LITTLE, R. E RUBIN, D. (1987), *Statistical analysis with missing data*. New York: John Wiley and Sons, pp. 50-60.
- MORGANSTEIN D., MARKER D.A. (1997), Continuous Quality improvement in Statistical Agencies, in Lyberg L. et al. (1997), in LYBERG L. et al. (eds), *Survey Measurement and Process Quality*, New York, J. Wiley & Sons (475-500).
- MOSER C. ET AL. (1983), *Aspetti delle statistiche ufficiali italiane. Esame e proposte*, [volume non numerato e senza indicazione della serie], Roma, Istat.
- SÄRNDAL C. E SWENSSON B., WRETMAN J. (1992). *Model Assisted Survey Sampling*, Springer, New York.
- SESTITO P., TRIVELLATO U. (2010), Indagini dirette e fonti amministrative: dall'alternativa all'ancora incompiuta integrazione, relazione presentata a "Giornata di studio in onore di G.M. Rey", Roma, 5 marzo, in corso di pubblicazione.
- THOMSEN I.B., LI-CHUN ZHANG (2008), A Predictive approach to representativity, *IAOS Conference on Reshaping Official Statistics*, Shanghai, China, October 12-14.
- VALE, S., PERRY, J. AND PONT, M. (2001), Developing a quality strategy for business registers: a UK perspective, *NTTS and ETK Conference*, Crete.
- WALLGREN, A., WALLGREN B. (2007), *Register-based Statistics: Administrative Data for Statistical Purposes*, Wiley Series in Survey Methodology, New York.
- WOLTER K. M. (1986), Some coverage error models for census data, *JASA*, 81, 394: 339-346.

SUMMARY

Statistical sources and statistical system in the information society

The aim of the paper is to analyze the impact on statistical systems and on official statistics derived from the increasing availability of archives in relation to many kinds of phenomena and subjects and the growing interest in taking advantage of these records by national statistical agencies and by several institutional and private entities emerging as “new data producers”. The effects of an enormous amount of electronic information extend to many aspects so that it seems possible to speak of a “revolution in data production”.

The main issues concerning the Italian statistical system and its key objects, and the topics arising to ensure appropriate standards for data resulting from non statistical sources are investigated and discussed in the paper. In particular, the new challenges arising from the statistical use of administrative and managerial sources are identified. Finally, in the last part of the paper, focus is on the particular meaning a characterization the categories concerning measurement errors, quality, comparability and coherence take on when these sources are used to produce statistics.