

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS

**UNECE Expert Meeting on Statistical Data Collection – Towards a New Normal?**

26 to 28 October 2022, Rome, Italy

---

## **The use of administrative data in Istat**

Grazia Di Bella (Istat)

[dibella@istat.it](mailto:dibella@istat.it)

### *Abstract*

The strengthening of the use of administrative data for the production of Official statistics is a strategic axis in Istat. The results of the innovations introduced are highlighted by some figures concerning the methods of producing statistics with respect to the use of administrative data in the last 5 years. The progressive replacement of survey data with administrative data as input of production processes, the combination of survey and administrative data and the integration of administrative data from multiple sources are the aspects explored in this short paper.

### *Introduction*

In Italy, the use of administrative data as input of the production of Official statistics is very intense with the aim of optimize production processes in terms of information content, costs and quality.

The first experiences dating back to the 90s with economic statistics and the Business register (Istat, 2021) gradually consolidated and then involved the other production sectors, until 2016 when the modernization process was started with the launch of the construction of an Integrated system of statistical registers based largely on administrative sources (Istat, 2016).

Meanwhile, Istat has invested a lot in this area: various centralized services for production processes are available at the Central Directorate of Data Collection with the aim of facilitating and supporting the use of administrative data and guaranteeing an adequate level of quality for the acquisition and treatment processes and for the data themselves. In particular, the Arcam system has the task of centrally acquiring data in a secure manner, the SIM (System of Integrated Microdata) takes care of the generalized pre-treatment of the data with the aim of maximizing the protection of personal data and the Quality Report Card of Administrative data (QRCA) manages the administrative data documentation reporting on data and processes quality guaranteeing the functionality of the information flow through the interoperability of the various IT- Tools involved (Di Bella, 2022).

In this paper I will present some figures relating to the current situation which can provide some elements for reflection. Two information sources were used. The first one is the National Statistical Program (NSP) approved each year by law and containing the list of statistical processes of public interest at the national level providing the country and international organizations with Official statistics. The Italian National Statistical System (Sistan) is the network of offices that contribute to the definition of the NSP with their statistical activities. It includes the National Institute of Statistics (Istat) and other statistical offices. For each activity, the NSP reports the main features of the processes, data sources and informative goals (Scanu, Sorvillo, 2022). The second source of data

is the QRCA, already mentioned. The QRCA portal, available in the Istat intranet, produces 22 reports and 49 quality indicators automatically updated mostly in real time for each supply of administrative data acquired by Istat and therefore contains a large amount of information (Di Bella, 2022). In this paper some first elaborations are derived.

### *Survey and administrative data*

To get an overall picture of the data used to produce Istat statistics, the Istat statistical processes listed in NSP (about 280 projects each year) have been classified according to the type of data used as input: processes that use Survey data and not Administrative data; processes that use Survey and Administrative Data, processes using Administrative Data and not Survey data and, finally, processes whose input consists only of Statistical data defined as outputs of other statistical processes<sup>1</sup>.

Table 1 shows the number of NSP Istat processes by type of data collection from 2017 to 2022. In 2022 the modal category of processes is the exclusive use of Administrative data (34.9% of processes), followed by the category of Survey data (24.6%) and the category for which Survey and Administrative data are combined together (22.9%). It is after 2018, when most of the processes still used exclusively Survey data (26.7%), that the trend is reversed.

The use of administrative data has grown over time, reducing the use of survey data collection: from 31.8 % of processes based on the exclusive use of survey data to 24.6%. The increase has been progressive to stabilize over the last three years. The use is associated both in combination with survey data, from 14.4 to 22.9 of the statistics production process in 2022, and as an exclusive source, from 19.1 to 34.9. The source of the statistical data too has been partially and progressively replaced by the use of administrative data.

**Table 1 – Statistics production processes in Istat by type of data collection – Years 2017-2021**

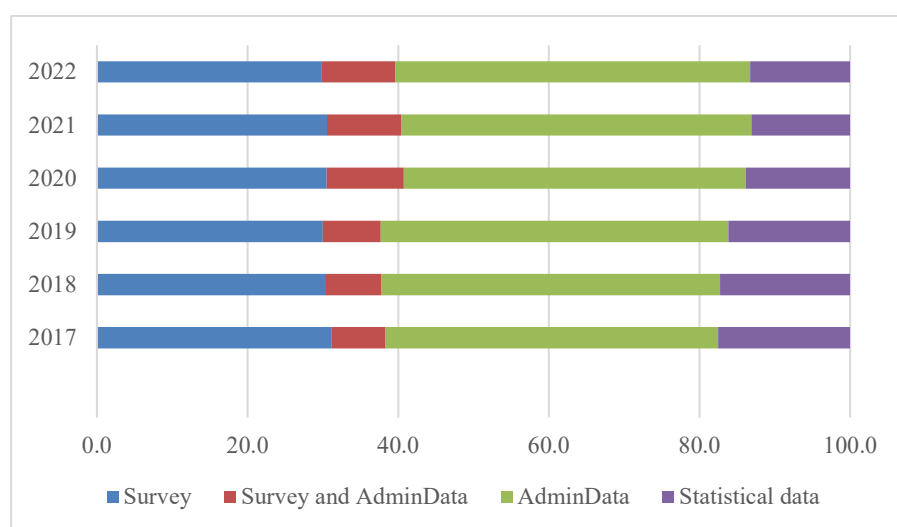
YEAR	Survey data	Survey and Administrative Data	Administrative Data	Statistical data	Total
2017	31.8	14.4	19.1	34.7	100.0
2018	26.7	21.1	23.0	29.3	100.0
2019	26.6	22.3	28.5	22.6	100.0
2020	24.5	23.5	34.3	17.7	100.0
2021	25.2	23.4	34.2	17.3	100.0
2022	24.6	22.9	34.9	17.6	100.0

It is interesting to observe the dynamics of the statistical production processes for the institutions other than Istat participating in the Sistan (Figure 1).

<sup>1</sup> NSP includes more than 800 projects each year, of which about 60% owned by Istat. The Feasibility studies were not considered in the elaboration and the projects analyzed have been about 740 each year, of which approximately 280 owned by Istat.

The scenario is different. Usually the Sistan offices that produce statistics carry out elaborations on the administrative data they own. Thus, the use of administrative data as input to processes is the modal category already in 2017 (44.2%). In the following years a growth trend is observed, also in the integrated use of administrative data and survey data, which corresponds to a decreasing trend in the use of survey data only.

**Figure 1 – Statistics production processes in Sistan, without Istat, by type of data collection – Years 2017-2021**



Source: Elaboration on NSP data

### *Multiple use of administrative data*

Istat acquires data from about 190 administrative sources, for a correspondent of more than 460 datasets<sup>2</sup> acquired. Table 3 shows the time series from 2017 to 2021. It is possible to observe a progressive increase during these 5 years. A large amount of data whose management requires a high standardization of processes.

**Table 3 – Administrative Source holders, Subsets, Datasets acquired per year – Years 2017-2021**

YEAR	Administrative Source holders	Administrative sources	Subsets	Administrative Datasets
2017	45		173	435
2018	52		188	469
2019	51		186	452
2020	53		191	479
2021	57		196	463

Source: Elaboration on QRCA data

<sup>2</sup> The Administrative sources Subset are defined by the extraction rules from the Administrative source determined by the set of units and variables of interest for statistical purposes. The higher number of Administrative datasets depends on how many data supplies are requested during the year (annual, half-yearly, quarterly, etc.) so each supply from the same source differs for the time reference of the data or for the possible characteristic of being provisional data followed by the supply of definitive data.

An element of optimization of the use of administrative sources derives from the fact that, in some cases, more Istat processes use data from the same source, each for the specific uses aimed to produce the corresponding target statistics. In Table 3, the administrative sources acquired are distributed with respect to the number of processes using them. If most of the acquired sources are used by a single process (48.3%), the remaining number of sources is used by several statistical processes, 14.1% by 2 processes, 9.8% by 3 processes, up to 3.2% of sources used by more than 16 processes. These are the tax and the social security data sources which have a high informative value.

**Table 3 –Administrative Sources by number of uses– Year 2021**

Number of uses	% of administrative sources acquired by Istat
1	48,3
2	14,1
3	9,8
4	8,6
5-10	10,1
11-15	6,0
+16	3,2
Total	100,0

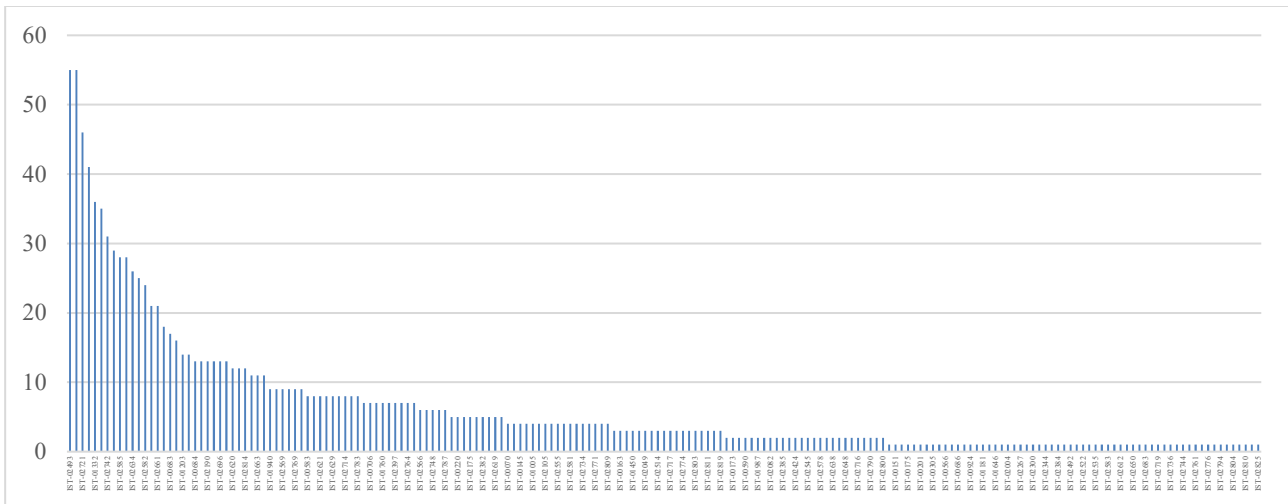
Source: Elaboration on QRCA data

The main feature of administrative data that allows its exploitation for statistical purposes is its granularity. Adopting the appropriate data protection measures, within the framework of European and national legislation, the opportunity to integrate microdata makes it possible to reach very useful information contents for the investigation of economic, social, demographic and environmental phenomena<sup>3</sup>.

Using the QRCA information it is possible to calculate the number of administrative sources used by each process (the number of sources whose use is declared in the NSP). In Figure 2, the statistical processes are listed on the X axis, while the number of administrative sources used are shown on the Y axis (processes are sorted by decreasing number of sources used).

<sup>3</sup> In Istat, data integration takes place through the use of pseudonymous codes assigned by the SIM system to the statistical target units, individuals and companies (Runci et al., 2016).

**Figure 2 – Number of administrative sources used by each statistics production processes in Istat– Year 2021**



Source: Elaboration on QRCA data

In 2021, the maximum number of sources used by a process is 55 (the Population and Housing Census). The following processes are mainly those aimed at building statistical registers (population register, labour register, business register, etc.) with a high numbers of sources used.

Some summary measures of the distribution shown in Figure 3: the median value of the distribution is 3, the 1st quartile is 1 and the 3rd quartile is 7, the average number of sources used by a statistical process is 3.5 (in the context of processes using administrative data).

**Conclusions**

These data show how the production system has actually innovated a lot in recent years, bringing significant advantages in reducing costs and statistical burden.

Regarding the reduction of the statistical burden, the use of administrative data from the public administration (tax data, social security data, education data, etc.) to produce statistics also meets the once-only principle that aims to ensure that citizens, institutions, and companies have to provide information to the authorities and administrations once, always in full compliance with the legislation on the personal data protection that regulates the exchange of data and the production of Official statistics.

## *References*

Di Bella G. (2022) Systematic reporting on the quality of administrative data as input of the statistical production process: the Italian experience, European Conference on Quality in Official Statistics (Q2022), 8 -10 June 2022, Vilnius, Lithuania. <https://q2022.stat.gov.lt/scientific-information/papers-presentations/session-24>.

Istat (2021) Registro statistico delle imprese attive, Anno 2019 Nota metodologica (in Italian) <https://www.istat.it/it/files//2021/11/Nota-metodologica-Registro-2019.pdf>.

ISTAT 2016. Istat's Modernisation Programme.

[https://www.istat.it/en/files/2011/04/IstatsModernisationProgramme\\_EN.pdf](https://www.istat.it/en/files/2011/04/IstatsModernisationProgramme_EN.pdf)

Runci M. C., Di Bella G., Galiè L. (2016) Il sistema di integrazione dei dati amministrativi in Istat, Istat Working Papers, N. 18/2016. <https://www.istat.it/it/archivio/193056>, Istat, Roma (in Italian).

Scannapieco M., Garofalo G., Radini R. (2017) The Italian Integrated System of Statistical Registers: On the Design of an Ontology-based Data Integration Architecture, presented at New Techniques and Technologies for Statistics (NTTS) Conference, 14-16 March 2017, Brussels, Belgium.

Scanu M., Sorvillo M. P. (2022) A new approach to monitoring the quality features of projects in the Italian statistical programme, European Conference on Quality in Official Statistics (Q2022), 8 -10 June 2022, Vilnius, Lithuania. <https://q2022.stat.gov.lt/scientific-information/papers-presentations/session-8>.