ROBUSTNESS TECHNIQUES FOR THE EVALUATION OF MUNICIPAL DATA: AN EXAMPLE ON THE "A MISURA DI COMUNE" INFORMATION SYSTEM[1]

Agata Maria Madia Carucci, Valeria Marzocca, Roberto Antonello Palumbo

Abstract. The "A misura di Comune" information system is the result of the "Misure di benessere e programmazione a livello comunale" project started in 2016 and it was created with the explicit objective of supporting public decision-makers in intercepting the needs of a territory. Through a complex set of indicators useful for the planning, programming and management tasks of Local Authorities, it is also a support tool for the preparation of the Single Programming Document (SDP).

In this work, after having presented the main methods of collecting statistical information, the robustness techniques of the proposed indicators will be analysed.

From 2022, the indicators are widespread in a time series and therefore alignment with the new geography and robustness analysis of the indicator are necessary annually not only in cross-sectional terms but also longitudinally. In particular, the robustness tests of the indicators depend on the type of indicator produced and three examples will be presented in this work.

1. Introduction

Territorial development and growth policies must necessarily start with analytical and continuously updated knowledge of the social, economic, demographic and environmental conditions of the area on which action must be taken. At the basis of good governance there are data and the ability to make decisions based on exact knowledge of the area. The availability of territorial statistical data is an indispensable requirement for defining, implementing and evaluating policies. To obtain measurable policies, it is necessary to have quality, timely and comparable data across time and space (Istat, 2018).

Among the objectives of official statistics there is certainly that of expanding the offer of territorial statistics on the main environmental and socio-economic phenomena of the country, making timely information available that reads the transformations of Italian society (Crescenzi F., Lipizzi F., 2020).

[1]The paper was carried out by the joint work of the authors. More in detail, the single paragraphs are attributed as follows: paragraphs 3 and 3.1 to Agata Maria Madia Carucci; paragraphs 1, 2, 4 to Valeria Marzocca; paragraphs 3.2 and 3.3 to Roberto Antonello Palumbo.

"A misura di Comune" represents an information tool designed by Istat which collects numerous municipal indicators on the main socio-economic and environmental issues. The tool, easy to consult, is also of central interest in the preparation of the Single Planning Document[2].

In this work, after having illustrated the structure and potential of "A misura di Comune" and the related use of indicators by municipal authorities, the techniques used to control and validate the constructed indicators will be illustrated through examples.

The growing demand for data at municipal level is also justified by the presence, in Italy, of a strong territorial lack of homogeneity which is linked to different access to services. Small and very small municipalities (less than 1,000 inhabitants) represent 25% of the municipalities in Italy and over one million inhabitants live in them. In the latest annual report, Istat dedicates an entire chapter to the territories, with the view that "The territory, with its economic, demographic, social and cultural specificities, represents a moment of synthesis of the complex transformations taking place at a national level and global, highlighting specific potential and constraints compared to what is illustrated at a national level. The theme of accessibility, understood as the possibility of access for citizens and businesses to various services, is closely linked to that of the peripherality of territories and to policy strategies oriented towards territorial planning" (Istat, 2024).

2. The information system "A misura di Comune"

The thematic structuring of "A misura di Comune", reported in Table 1, presents a strong coherence with the classification of indicators for measuring well-being and sustainable development and it tries to respond to the main information needs described by the Single Programming Documents. The information system, with the latest update in March 2024, collects 72 indicators for 13 thematic areas.

"A misura di Comune" is a multi-source system, in which sources of an experimental nature are valorized alongside other more consolidated ones. Among the experimental sources, an important place goes to the databases created within the ARCH.I.M.E.DE Project, which deals with the construction and updating of databases for territorial analysis within the Integrated Microdata System of the Istat and whose data are disseminated for municipalities with more than 5,000 inhabitants. A significant contribution also concerns the use of Open Data made available by other Sistan entities, such as the Ministry of the Interior, the Ministry of Economy and Finance, the Italian Institute for Environmental Protection and Research (ISPRA), the Italian Automobile Club (ACI) and the National institute for the evaluation of the education and training system (INVALSI). Other sources are

[2] https://www.istat.it/statistica-sperimentale/aggiornamento-degli-indicatori-del-sistema-informativo-a-misura-di-comune/

included in the current statistical production of Istat, such as for example demographic statistics, statistical registers of active businesses and non-profit institutions, the "Dati ambientali nelle città" survey (Table 2).

Table 1 – *"A misura di comune" Indicators by thematic area, release 2022 and 2024*.

THEMATIC AREA	N. Indicators 2022	N. Indicators 2024	% Indicators by thematic area 2024
Popolation	10	10	13,9
Household	0	6	8,3
Education	3	4	5,6
Labour	0	4	5,6
Economic well-being	1	4	5,6
Politics and institutions	5	5	6,9
Culture	2	3	4,2
Gender Gap	0	10	13,9
Social services	0	2	2,8
Territory and environment	5	9	12,5
Settled economy	5	6	8,3
Research and innovation	1	1	1,4
Infrastructure and mobility	7	8	11,0
TOTAL	39	72	100

Source: Authors elaboration based on ISTAT data.

Table 2 – *"A misura di comune" Indicators by source*.

Source	N. Indicators	% Indicators by source
ISTAT	**59**	**81,9**
ARCH.I.M.E.DE (Integrated administrative sources)	*6*	*8,3*
ASIA Business Register	*11*	*15,3*
PERMANENT CENSUS OF POPULATION AND HOUSING	*7*	*9,7*
Current statistics	*35*	*48,6*
OPEN DATA	**13**	**18,1**
TOTAL	72	100

Source: Authors elaboration based on ISTAT data.

In the latest dissemination, the indicators were presented in a time series starting from 2014 and 80% of them are available at the individual municipality level.

3. Indicator Selection: construction and control methods

The choice of indicators, in addition to being derived from the information needs of the institutions, is linked to the availability of the data and its quality (Istat, 2012). In general, the indicators were chosen if they meet the general criteria of:

- Timeliness
- Coverage
- Completeness and
- Consistency.

According to the type of indicator and the source that feeds it, macro and micro controls have been envisaged on the input data and on the output produced.

In general, for each input dataset, it is verified that the underlying definitions have not changed to ensure the comparability of the information with each other and subsequently a univariate and multivariate analysis is carried out of the time series data.

With univariate analysis, the single time series is verified in its temporal dynamics. The analysis starts from the basic data, on which the indicator is built, and is supported by the graphic representation of the values for each indicator, starting from the regional domain and gradually descending towards territorially finer domains.

From the graphic representation and through synthetic measures, calculated on each individual basic data, qualitative considerations can be drawn on the series and "causal factors" are "recognised" within the temporal dynamics of the data (in the period considered for example, many trends affected by the pandemic event), aimed at capturing the characteristics of the historical series.

With multivariate analysis, the connections and relationships between the observed phenomenon and others considered correlated (population trend, value added of enterprises trend, etc.) are subsequently verified.

Once the consistency of the data in the historical series has been verified, the coverage is analysed at a territorial level, reconstructing the territorial perimeter of the series at the last reference year and favouring the data for which there is exhaustive municipal coverage.

The analysis of the basic data is followed by the construction of the indicator and again the verification of the indicator produced through simple and complex statistical methods to identify irregularities and inhomogeneities in the indicators produced.

The analysis, conducted on three structurally very different indicators, will be shown below:

- Taxable income per taxpayer,
- Entrepreneurship rate
- Libraries registered in the National Library Registry per 100,000 inhabitants.

3.1 Taxable income per taxpayer

The average income per taxpayer is an indicator from an open data source, the Ministry of Economy and Finance, which annually disseminates statistical data on tax returns by taxpayer characteristics, type of income and municipality.

At a territorial level the data is exhaustive, all municipalities are surveyed annually. Furthermore, the data are disseminated every year, in the month of April, on the MEF's open data platform; they are comparable over time and space as it was not subject to any change in definition or data collection methodology over the period considered.

After analysing the input data with descriptive statistical analysis techniques, the trend of the historical series was compared with the trend of the main "well-being" indicators published by official statistics to identify anomalies or particularities in the trend of the data and to highlight the territorial characteristics of information.

Below are some simple examples of statistical analysis on the input data and the indicator that, expressed in euros, is calculated as the ratio between Taxable income and Taxable income taxpayers.

On average, an Italian taxpayer has an income of 21,000 euros, stable over the years, confirming the well-known Italian North-South dualism. In particular, the income per taxpayer in the North-West of Italy is 24,600 euros while in the South is around 17,500 euros (Table 3).

Table 3 – *Taxable income per taxpayer by territory. Years 2014-2021.*

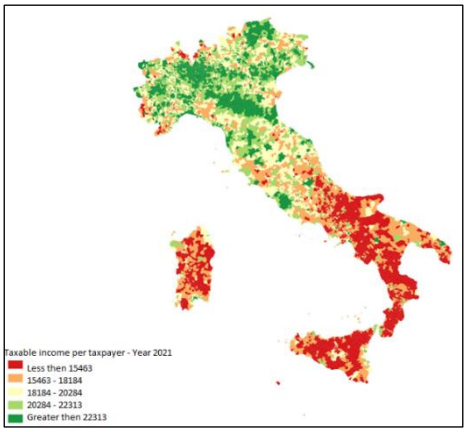
TERRITORY	2014	2015	2016	2017	2018	2019	2020	2021
Nord-West	22,335	22,741	22,910	23,074	23,608	23,789	23,518	24,664
Nord-East	20,734	21,119	21,413	21,565	22,043	22,285	22,099	23,120
Center	20,566	20,782	20,996	21,078	21,519	21,713	21,566	22,428
South	15,872	16,105	16,244	16,326	16,658	16,831	16,864	17,566
Islands	16,131	16,232	16,383	16,513	16,845	17,004	17,016	17,672
ITALY	19,720	20,017	20,214	20,355	20,800	20,993	20,852	21,777

Source: Authors elaboration based on MEF data.

To identify any outlier values of the income indicators at municipal level, the analysis was paid particular attention to the marked territorial configuration of the income phenomenon in Italy. Figure 1 represents the municipal values of the Income per taxpayer indicator.

The distribution of the municipalities in the five income classes is immediately consistent with the well-known regional differences in per capita income, also confirmed by the MEF data. The analysis of the data of the municipalities that are in the two tails of the distribution of per capita income also allows us to highlight the different cases within the general pattern of pronounced lack of homogeneity of the North-South area.

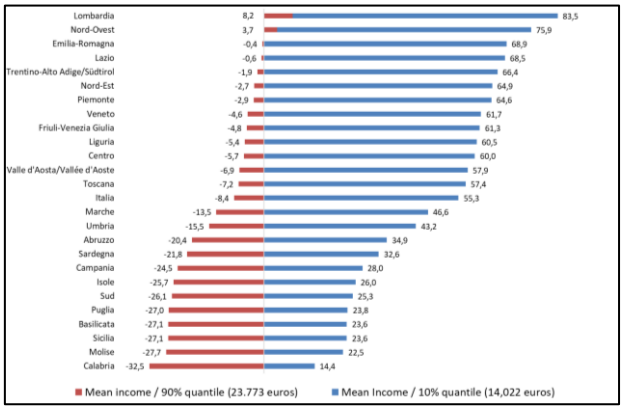
Figure 1 – *Taxable income per taxpayer- Year 2021.*



Source: Authors elaboration based on MEF data.

Considering, for the year 2021, the total of Italian municipalities for which information is available, the ninth decile threshold is equivalent to 23,773 euros, which is 9.2% higher than the national average value, but even lower than the average value for Lombardy. Conversely, this threshold is more than 20% higher than the income per taxpayer of all the regions of Southern Italy. Examining the first decile threshold (equal to 14,022 euros), it is 35.7 percentage points lower than the national average value, but the average income in Lombardy is 83.5% higher. On the contrary, the average income in the South is 25% higher than this threshold, in Calabria the average income is just 15% higher than the first decile (Figure 2).

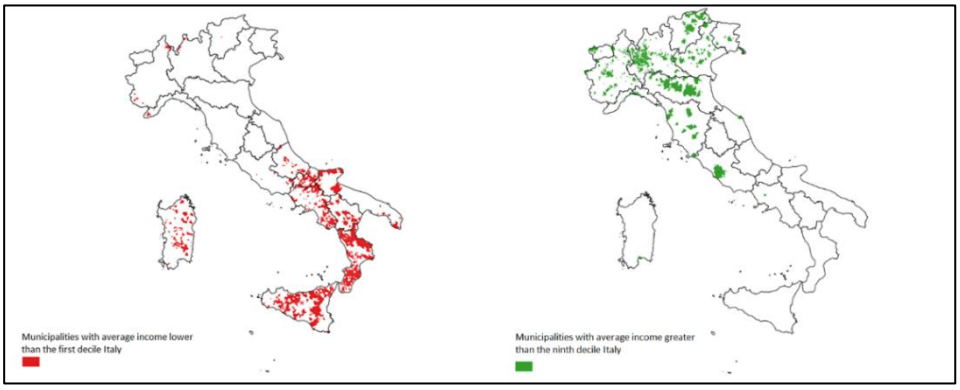
Figure 2 – *Taxable income per taxpayer: percentage distance between Mean income of the territory and first decile or ninth decile of Italy. Year 2021.*



Source: Authors elaboration based on MEF data.

At the level of the individual municipality, the municipalities in the first decile are almost exclusively in the regions of the South, apart from a few cross-border municipalities, while the municipalities whose indicator falls in the ninth decile of the distribution are in the regions of the North and Lazio, particularly in the Roman belt (Figure 3).

Figure 3 – *Taxable Income Per Taxpayer- Municipality in the first and ninth decile. Year 2021.*



Source: Authors elaboration based on ISTAT data.

In conclusion, the results of this analysis support the belief in the "spreadability" of the indicator, which is stable over the years, complete in terms of coverage, timely due to the availability of the source and coherent at a territorial level.

3.2 Entrepreneurship rate

The entrepreneurship rate is a classic indicator used for the analysis of the production system at a territorial level and the data appear to be particularly "solid" as they derive from the Business Register ASIA (Eurostat, 2010). The indicator therefore does not present coverage and timeliness problems, and the data is available for the entire historical series. The trend of the indicator was compared with the trends of the main indicators on the structure of enterprises released by official statistics and did not highlight any "irregularities" in the information collected. In addition, the output analysis cannot ignore the territorial analysis, as although the basic data are very robust by construction, the indicators at municipal level could highlight outliers value. On average, there are just over 70 enterprises per thousand inhabitants and this value is substantially stable over the years (Table 4). In the North of Italy, there are 81.2 enterprises per thousand inhabitants while in the South there are 64.1 enterprises.

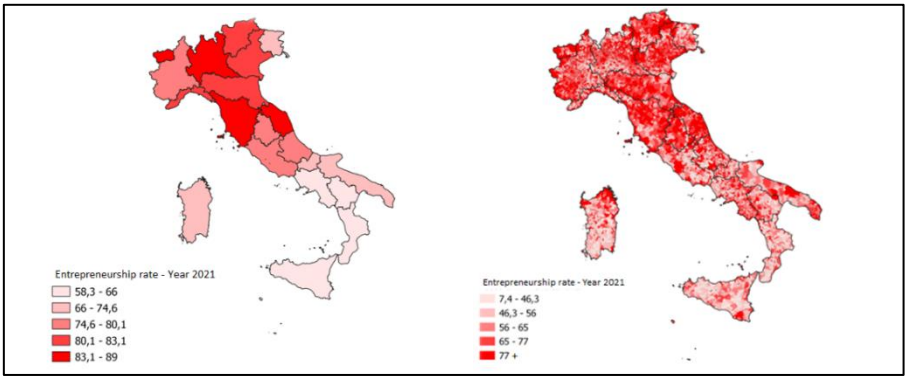
Table 4 – *Entrepreneurship Rate- enterprises per 1,000 inhabitants. Years 2014-2021.*

YEARS	2014	2015	2016	2017	2018	2019	2020	2021
Mean of the indicator	72.3	72.0	73.0	73.3	73.6	72.2	73.5	75.7
50% quantile	58.8	58.4	58.9	58.8	58.9	57.3	58.8	60.6
90% quantile	85.2	84.9	85.5	85.4	85.5	83.3	84.9	87.0
Max value	267.2	273.4	276.9	269.4	272	271.7	260.2	255.6

Source: Authors elaboration based on ISTAT data.

From the regional map, we can see the classic North-South dualism with two exceptions: Friuli-Venezia Giulia and Piedmont. Among the regions of Southern Italy, Abruzzo records the highest indicator value. Dualism is confirmed by the municipal representation even if some regional specificity is highlighted (Figure 4).

Figure 4 – *Entrepreneurship rate, Regional and municipal map, Year 2021.*



Source: Authors elaboration based on ISTAT data.

There is a low variability of the indicator between years (coefficient of variation less than 5% in 61% of the municipalities) and this variability is even lower as the municipal size increases, the 88.6% of the municipalities with a coefficient of variation greater than 30% has fewer than 1000 inhabitants (Table 5).

Table 5 – *Entrepreneurship rate: CV of the municipalities by demographic size.*

COEFFICIENT OF VARIATION CLASSES	N. OF MUNICIPALITIES (class of inhabitants)				
	Totals	< 1,000	> 1,000	> 5,000	> 10,000
Less then 5%	4,822	313	4,509	2,216	1,143
5%-10%	2,221	938	1,283	148	59
10%-20%	717	624	93	4	2
20%-30%	100	94	6	2	0
Greater then 30%	44	39	5	1	0

Source: Authors elaboration based on ISTAT data.

3.3 Libraries registered in the National Library Registry per 100,000 inhabitants.

The last indicator analyzed is the number of libraries per 100,000 inhabitants. It is a particularly different indicator compared to the previous ones.

From the point of view of the availability and timeliness of the data, they are provided periodically for all Italian municipalities, upon specific request, directly from the National Library Registry. As regards the quality analysis, it mainly involves detecting the presence or absence of the library (zero or one library, 27.3% and 58.5% respectively). On average, there are 20 libraries per 100,000 inhabitants and the number is stable over the years (Table 6).

Table 6 – *Libraries Per 100,000 Inhabitants.*

YEARS	2016	2017	2018	2019	2020	2021
Freq Mcp with value=0	2,023	2,139	2,320	2,535	2,452	2,158
Mean	24.1	23.2	22.9	20.5	20.8	21.9
50% quantile	20.9	19.8	18.3	16.1	17.0	19.7
90% quantile	108.0	105.5	103.8	101.4	104.2	111.4
MAX	2,816.9	1,470.6	1,626.0	3,076.9	2,739.7	2,439.0

Source: Authors elaboration based on ISTAT data.

From the cartograms we do not find a territorial dichotomy but rather a strong relationship between the number of libraries and the municipal size. In very small municipalities the presence of even just one library leads to a significant increase in the indicator. The maximum value, for example, is recorded in the municipality of Moncenisio (TO) where there is a library and 42 inhabitants.

The areas with the highest value of the indicator are in fact generally those in which there is a greater presence of small and very small municipalities (Table 7).

Table 7 – *Number of libraries per 100,000 inhabitants (I) by demographic size of municipalities (Mcp)- Year 2021.*

	< 1,000 inhabitants		1,000-5,000 inhabitants		5,000-10,000 inhabitants		10,000 + inhabitants		Total	
	# Mcp	*I*	# Mcp	*I*	# Mcp	*I*	# Mcp	*I*	# Mcp	*I*
North-West	401	174.5	1,027	41.6	362	16.8	278	17.9	2,068	24.4
North-East	101	162.6	565	43.7	298	18.1	269	24.1	1,233	26.4
Center	58	155.9	292	41.7	137	21.4	211	22.2	698	24.0
South	158	164.7	456	46.4	174	20.3	276	16.4	1,064	21.0
Islands	143	174.0	321	43.4	92	17.9	127	14.3	683	20.7
ITALY	861	169.7	2,661	43.0	1,063	18.4	1,161	19.4	5,746	23.6

Source: Authors elaboration based on ISTAT data.

The analysis of the presence or absence of the library can also be conducted through simple graphic representations, the so-called sparklines which allow cases of inconsistency to be intuitively highlighted (Table 8).

Table 8 – *Sparklines of the municipal libraries*.

2016	2017	2018	2019	2020	2021	Freq	% Freq	
0	**0**	**0**	**0**	**0**	**0**	**1,648**	**20.85**	
0	**0**	**0**	**0**	**0**	**1**	**168**	**2.13**	
0	**0**	**0**	**0**	**1**	**1**	31	0.39	
0	**0**	**0**	**1**	**0**	**0**	1	0.01	
0	**0**	**0**	**1**	**1**	**0**	6	0.08	
0	**0**	**0**	**1**	**1**	**1**	**130**	**1.64**	
0	**0**	**1**	**1**	**1**	**1**	27	0.34	
0	**1**	**0**	**1**	**1**	**1**	1	0.01	
0	**1**	**1**	**1**	**1**	**1**	10	0.13	
1	**0**	**0**	**0**	**0**	**0**	3	0.04	
1	**0**	**0**	**0**	**0**	**1**	5	0.06	
1	**0**	**0**	**1**	**1**	**1**	104	*1.32*	
1	**0**	**1**	**1**	**1**	**1**	15	0.19	
1	**1**	**0**	**0**	**0**	**0**	4	0.05	
1	**1**	**0**	**1**	**1**	**0**	1	0.01	
1	**1**	**0**	**1**	**1**	**1**	218	*2.76*	
1	**1**	**1**	**0**	**0**	**0**	469	5.93	
1	**1**	**1**	**0**	**0**	**1**	151	1.91	
1	**1**	**1**	**0**	**1**	**1**	55	*0.7*	
1	**1**	**1**	**1**	**0**	**0**	2	0.03	
1	**1**	**1**	**1**	**1**	**0**	23	0.29	
1	**1**	**1**	**1**	**1**	**1**	**4,831**	**61.13**	

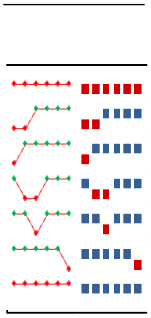
Source: Authors elaboration based on ISTAT data.

In the entire period, 20% of Italian municipalities do not have a library, just as 61% of municipalities always have at least one library and this behavior of the indicator can be said to be reasonable. Equally reasonable is that the event is present from a certain point onwards: in 2% of cases, the library is taken over starting from 2021, as well as in 1.6% of cases from 2019. Especially due to pandemic events, even the closure of a library from a certain year onwards cannot be considered an anomalous event.

The anomalous values, on the other hand, on which it is appropriate to reflect, are those for which the event is present in some years, then is no longer present and then appears present again, a condition graphically represented with a "V" in the graph. For the years under review, approximately 5% of the municipalities exhibit behavior of this type. In this case, it is necessary to understand whether, due to an exogenous event, the library was temporarily closed or whether this apparent inconsistency hides an error in the detection of the basic data.

To understand whether this is an acceptable phenomenon, given that it affects a relatively low percentage of municipalities, the same sparkline analysis was conducted at a regional level. From the regional analysis, there are no inconsistencies in all regions except in Sardinia, where 40% of municipalities lacked the signal in 2018 (Table 9). This certainly "dirty" data is affected by incorrect recording of the information. Data of this type, which cannot be justified by the presence of exogenous events, is probably due to an inconsistency in data input.

Table 9 – *Sparklines of Sardinia libraries.*

2016	2017	2018	2019	2020	2021	Freq	%Freq Italy	%Freq Area	
0	**0**	**0**	**0**	**0**	**0**	6	0,08	1,6	
0	**0**	**1**	**1**	**1**	**1**	1	0,01	0,3	
0	**1**	**1**	**1**	**1**	**1**	1	0,01	0,3	
1	**0**	**0**	**1**	**1**	**1**	1	0,01	0,3	
1	**1**	**0**	**1**	**1**	**1**	**153**	**1,94**	**40,6**	
1	**1**	**1**	**1**	**1**	**0**	3	0,04	0,8	
1	**1**	**1**	**1**	**1**	**1**	212	2,68	56,2	

Source: Authors elaboration based on ISTAT data.

This inconsistency did not make it possible to disseminate the indicator for the entire reference period. While waiting for a timely check on the data, in the March release, it was decided to release the indicator for all municipalities but only for the period 2019-2021. The same analysis technique, simple and easy to interpret, was used for other qualitative indicators: for example, the number of anti-violence centres in a municipality. Also in this case, the method was particularly effective in identifying anomalous situations of signal absence in particular years.

4. Conclusion

For a decade, official statistics have been committed, both at a national and international level, to producing data that is as exhaustive as possible from a territorial point of view and timely. Register data and open data have certainly contributed to increasing the richness of information. Nonetheless, the dissemination of data for specific territorial domains presupposes important analytical work as it can bring out anomalous values that would not be detected at a macro-area level. A fundamental prerequisite for the dissemination of municipal or sub-municipal indicators is the robustness analysis of the indicators. Depending on the type of indicator chosen, it is advisable to provide an analysis, whether transversal, longitudinal or which aims to analyse the presence or absence of a signal.

In the next years, the even more massive use of the integration between statistical and administrative data will lead on the one hand to the dissemination of more data even at sub-municipal level and on the other to the possibility of testing the available data more effectively. This will be combined with the use of new sources deriving from the web and usable through scraping techniques and the current use of big data.

The public decision-maker now needs data that allows him to comprehensively describe his territorial reality and needs indicators for monitoring public actions. This is not only a need for larger municipalities but also for small and very small municipalities, distant from essential services which therefore require targeted cohesion policies. The new frontier of official statistics is undoubtedly the integrated use of statistical, administrative and big data which constitute a very important information basin especially for municipal and sub-municipal analyses. Official statistics is investing heavily in this direction, through numerous research projects, aimed at identifying the big data sources useful for official statistics, at transforming big data into statistical data by verifying their quality through robustness analysis techniques possibly also based on machine learning. Official statistics are following this path, and research must set up increasingly effective robustness techniques.

References

BIANCHINO A., CARUCCI A.M.M., VANNELLA G. 2018. Il nuovo sistema statistico "a misura di comune": un'applicazione sul territorio lucano, *Metodi e analisi statistiche,* Università degli studi di Bari Aldo Moro, pp. 83-108.

CRESCENZI F., LIPIZZI F. 2020. The integration of geographic and territorial data sources into the base register of territorial and geographical entities, *Statistical Journal of the IAOS*, Vol. 36, No.1, pp. 143-149.

EUROSTAT. 2010. Business Register. Recommendations Manual (2010 Edition). Methodologies and Working papers. Luxembourg.

ISTAT. 2012. *Quality Guidelines for Statistical Processes*, Versione 1.1, Dicembre 2012 – Istat –Roma Quality guidelines_English version_rev (istat.it)

ISTAT. 2018. *Linee Guida per la Qualità delle Statistiche del Sistema Statistico Nazionale*, Versione 1.0, Marzo 2018 – Istat –Roma

ISTAT. 2024. *Rapporto Annuale. La situazione del Paese*. Maggio 2024 –Roma

---

Agata Maria Madia CARUCCI, ISTAT, carucci@istat.it
Valeria MARZOCCA, ISTAT, valeria.marzocca@istat.it
Roberto Antonello PALUMBO, ISTAT, palumbo@istat.it