An application of integrated statistical registers to produce new and systemic indicators on small territorial units

de Panizza Andrea[1], De Santis Stefano[1]

[1]ISTAT, Rome, Italy

Abstract

This paper proposes an application of integrated statistical registers to produce new and systemic indicators portraying small territorial units, which cannot be addressed by means of surveying. Indeed, statistical registers derived from administrative sources provide a full (census-like) coverage on the population under scrutiny, allowing to get (longitudinal) information on all the individuals and to define *à la carte* domains.

This comes at the price of major limitations in the information available in terms of underlying variables and their robustness, not to mention constraints on dissemination that may result from confidentiality rules. The development of registers is still limited in many European Countries. Like many other European NSOs, Statistics Italy (Istat) started developing a business register in the early Nineties. In the following decades, it managed to establish an ecosystem of registers addressing enterprises, a Base register on Individuals, a LEED for non-farm business employment, and some extensions to registers from additional administrative sources, with the objective to set up and maintain an Integrated Registers System (SIR). Until now, registers had mostly an infrastructural aim, to serve as basis for statistical production and analysis, and – to a limited extent – to produce register-based statistics and indicators drawing on one or another specific register. In the recent past, two releases of multi-register indicators provided some territorial information (up to the municipality level) on enterprises and the workforce. The two case studies proposed here show how Official statistics can provide insights even at very fine-grain territorial level and support policies. The first, focuses on the provision of a multi-dimensional perspective on Labour market areas (LMAs), joining information on individuals (employment, education, age, gender, household characteristics) and on productive units. We present an array of indicators, showing their mutual relationships and how they can define patterns of LMAs, which add to previous classifications made by Istat. Some examples of sub-communal level information are also provided for the three Italian FUAs of Rome, Milan, and Naples, with a specific focus on Rome.

These exercises serve as a basis to discuss the possibilities and limitations of this approach. Results ought to be considered experimental, and an anticipation of future work, which could also be performed in collaboration with other NSOs within the ESS.

Keywords: Statistical registers; Territorial indicators

1. Introduction

This paper presents some of the applications recently put in place and being experimented at Statistics Italy to portray socio-economic characteristics of small territorial units (Communes and sub-communal level), based on the integration of data from statistical registers being developed at Istat.

Traditional survey methods often fail to provide sufficient data for detailed regional or subgroup analysis due to cost and logistical constraints. To mitigate this limitation, the approach usually followed refers to Small Area Estimation (SAE) statistical techniques, used to produce reliable estimates for geographical areas or subpopulations where sample sizes are too small to provide accurate direct estimates. SAE addresses this issue by combining data from different sources, such as surveys and administrative records, and applying advanced statistical models.[1]

Statistical Registers are structured collections of data systematically compiled and maintained for statistical purposes. Unlike traditional survey data, typically collected through direct responses from individuals or entities, statistical registers collect information drawn from administrative records, originally produced for non-statistical purposes, such as tax records, social security data, or business registers.

Statistical registers offer several advantages, including:

- Comprehensive coverage: registers often cover the entire population or a very large portion of it, reducing sampling errors and improving the reliability of estimates;
- Cost-Effectiveness: Using existing administrative data minimizes the cost associated with data collection;
- Timeliness: Registers can be updated continuously or at regular intervals; this might result in the provision of more timely data compared to periodic surveys.

On the other hand, there are challenges associated with the use of statistical registers, such as ensuring data quality, dealing with data privacy concerns, and integrating data from multiple sources.

In practice, statistical registers play a crucial role in modern statistical systems. They provide foundational data for various statistical outputs, support the production of official statistics, and enhance the capacity for longitudinal studies. For instance, business registers track the number and types of businesses within a country, aiding in economic analysis and policy-making. Population registers, which include demographic information on residents, are essential for producing accurate population statistics and supporting census operations.

The integration of statistical registers with techniques like SAE further enhances the ability to generate detailed and reliable statistics for small areas, thereby supporting informed decision-making at all levels of government and industry.

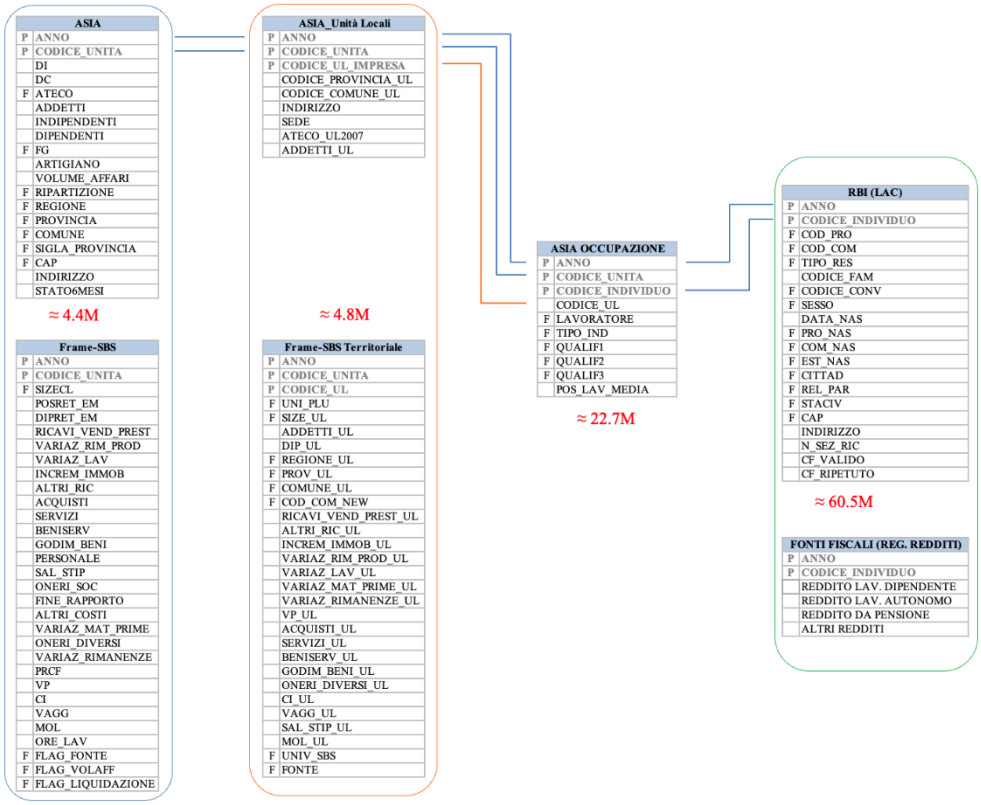
---

[1] The primary goal of SAE is to enhance the precision of estimates for small domains by "borrowing strength" from related areas or larger datasets. Typically, this is achieved by means of mixed-effects models, hierarchical Bayesian models, or empirical Bayes methods, which incorporate auxiliary information to improve estimation accuracy. SAE is widely used in various fields, including public health, economics, and social sciences, to inform policy-making and resource allocation.For example, in public health, SAE can provide more accurate prevalence rates of diseases at the county or district level, which is crucial for targeted intervention programs. In economics, SAE can help estimate unemployment rates or average income for small regions, aiding in the development of local economic policies.

2. Data Integration

The dimensions considered here include economic (business and local units) and population (socio-economic elements and employment/income status) characteristics of each area, obtained by means of a multilevel integration exercise, exemplified in Figure 1 below.

Figure 1: Integration scheme between statistical registers at micro level. Year 2016-2021.



The first step of the analysis concerns the micro level (enterprise, local unit, individual, etc.), followed by an integration on the domain variables, i.e. territorial units at the most detailed level. This consists of the about 7,900 municipalities, except for the exploratory work at sub-municipal level, for which we used postal codes (à regime, these will be substituted either by census sections polygons or by spatial coordinates). Units are subsequently aggregated to produce information at superior hierarchical levels: about 610 *Local systems* (functional regions), 107 *provinces* (NUTS3), 20 *regions* (NUTS2), 5 *macro-regions* (NUTS1), plus special aggregations (notably, *Functional Urban Areas* for the three largest agglomerations).

For each Base register, physical integration was carried out with the relevant extended/thematic registers (see Figure 1), followed by an aggregation at the minimum territorial level. The connection between businesses/local units and population generated a datacube, representing a relationship over the territory, i.e. at the postal code/municipality level. Relevant information is collected for the population (number of individuals and families, average age, etc.) and business units (number of local units/enterprises, sector, size, value added, etc.) in the same territorial context (see Table 1).

Table 1 - Integration of Population and Business Base registers (study domain-level)

Comune	Persone	Maschi	Stranieri	Famiglie	Scolarizzazione	Età	Unità locali	Imprese	Valore Aggiunto	Addetti
denominazione	*numero*	*numero*	*numero*	*numero*	*anni medi*	*anni medi*	*numero*	*numero*	*somma (€ .000)*	*numero*
Affile	1.512	732	58	632	9,28	46,4	77	76	6.064	220
Agosta	1.764	864	154	763	9,09	45,7	72	71	2.529	122
Albano Laziale	41.715	20.076	4.281	17.294	9,62	44,0	2.866	2.834	263.630	7.597
Anguillara Sabazia	19.432	9.574	1.932	7.739	9,72	43,2	1.130	1.121	61.522	2.187
Anticoli Corrado	897	442	76	420	9,29	46,2	43	43	1.331	60
Anzio	59.483	29.729	8.430	27.227	9,61	43,7	3.529	3.439	310.054	9.977
Arcinazzo Romano	1.350	664	29	647	9,35	49,4	45	45	4.325	148

Each relation allows for the description of the profiles of the territorial units, the calculation of key indicators, comparisons between units, temporal evolution, data driven territorial aggregations (i.e. regardless of administrative borders).

3. The work done to date

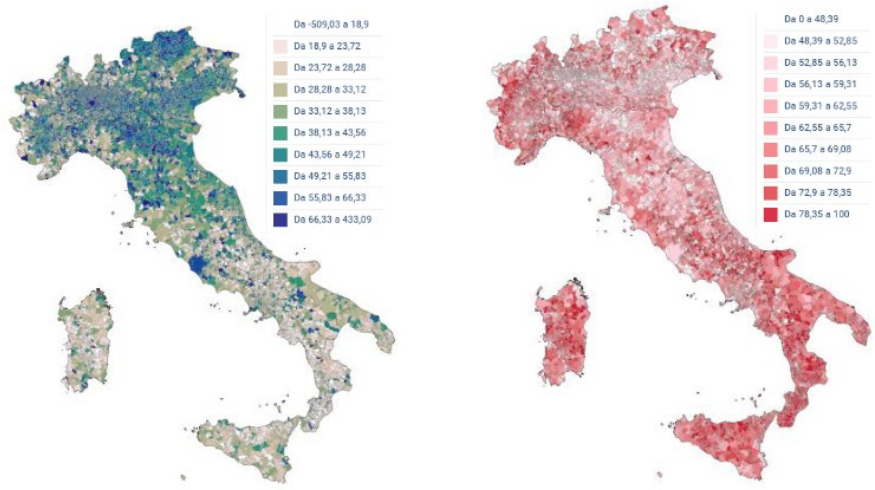
In recent years, territorial register-based information has been used extensively for analytical and dissemination purposes. On the other hand, the production of fine-grain territorial indicators is still an ongoing process, due to the still evolving nature of registers. For instance, until now the register of locations did not allow for a precise geolocation of all units throughout the country, thus impeding a full bottom-up aggregation strategy below the commune level, or the income register still lacks some modules to be 100% operational, so that we cannot yet produce official statistics at fine-grain territorial level that include overall income.

Nonetheless, about 90 indicators were already released, based on the integration of economic registers, and of the above with the extended population register, that offer a territorial perspective detailed to the municipality level, i.e. almost 8 thousand distinct units, exemplified in Figure 2 below.

Figure 2a. New territorial indicators: education of persons employed in non-agricultural enterprises (above, at NUTS 3 and NUTS4 levels; year 2021)

Source: https://esploradati.istat.it/databrowser/#/it/dw/categories/IT1,Z0900ENT,1.0/SIR_IND

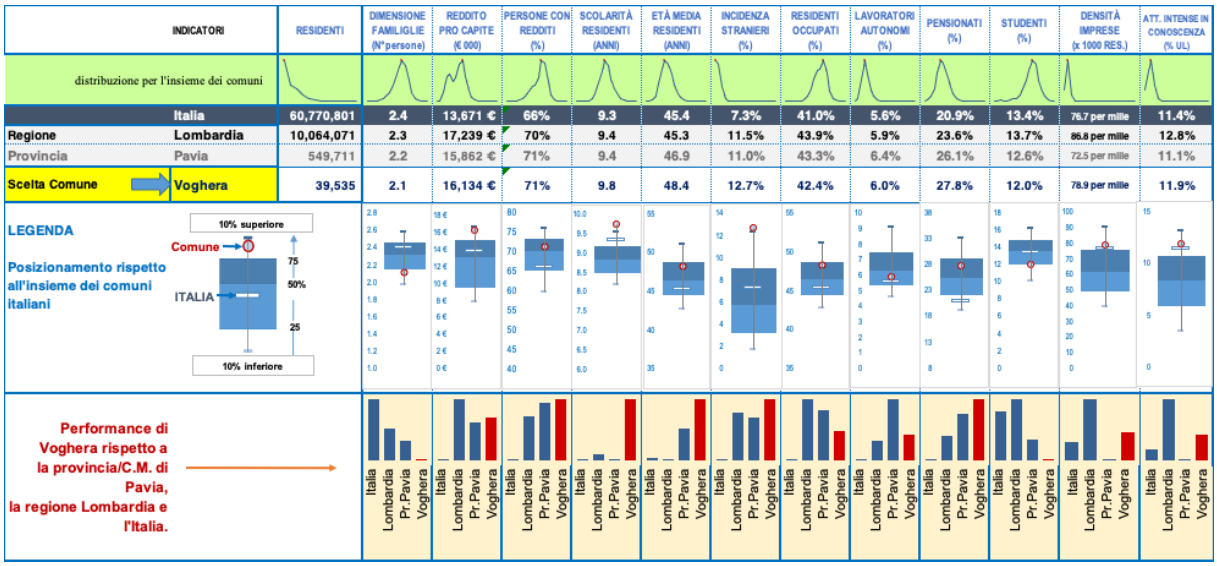
Figure 2b. New territorial indicators: value added per person employed (k€, left) and share of microenterprises in total (% values, right) (year 2020)



Source: https://esploradati.istat.it/databrowser/#/it/dw/categories/IT1,Z0900ENT,1.0/SIR_IND

These indicators can also be portrayed in dashboards, allowing for multivariate presentations and positioning of municipalities vs. their hierarchical superior entities or peers (Figure 3).

Figure 3 Positioning exercise with a battery of indicators at the municipality level

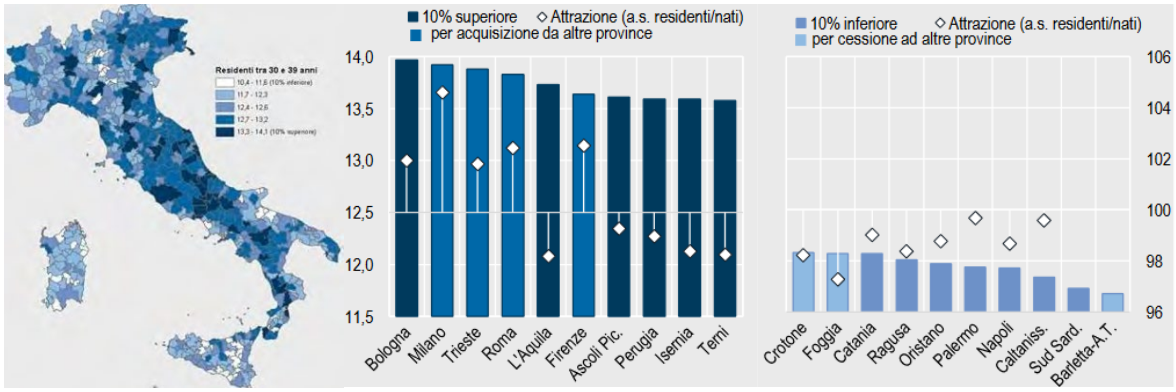


Source: Authors' elaboration on territorial indicators from integrated registers

Even before the release of these sets of indicators, some of them have been tested and used in territorial analysis. An example drawn from the first Istat *Report on the Territory* refers to the level of education (in legal school/academic years to achieve the highest qualification) of individuals aged 30 to 39, considering their birthplace and the place of residence, at the NUTS 3 level (Figure 4). This permits to show the impact on human capital stock of the attractive economic power of some territories and the ensuing drainage of human capital from other areas with weaker economic conditions. This example also allows recalling some limitations in the use of registers. In the first place, we could not meaningfully drill in beyond the NUTS3 level, as the birthplace reported in registers is that of hospitals (i.e. towns) rather than residence at birth. Secondarily, registers often lack information on the educational attainment of immigrants

(which are concentrated in northern regions). Finally, the notion of administrative residence presents some drawbacks, as people tend to change it only when they have some strong reason to do so (leading to an underestimation of population movements, especially if we were to consider younger cohorts).

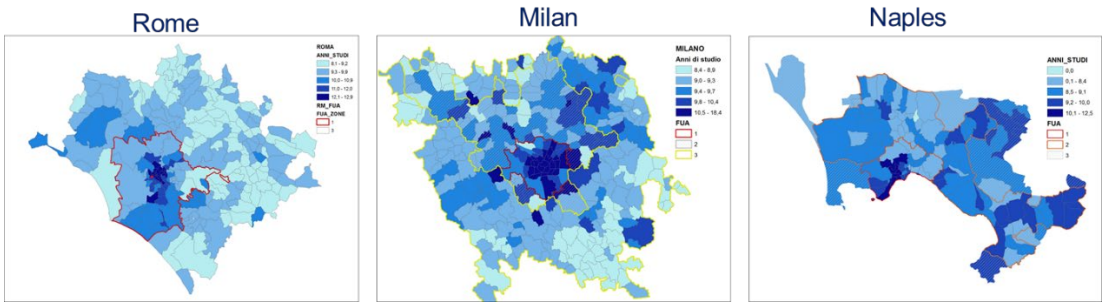
Figure 4 Education patterns, and the effects of human capital attraction and drainage. Top and bottom 10 NUTS 3 areas, year 2017.



Source: Istat, Rapporto sul Territorio 2020 (https://doi.org/10.1481/Istat.RapportoTerritorio.2020)

The educational attainment of individuals - defined as above – is used in Figure 5 below, to portray the three largest urban agglomerations, going below the municipality level by using postal codes as polygons (in red the administrative boundaries of the centre city, in yellow those of first and second circles of municipalities). The chart shows different patterns in the three FUAs, along common (and traditional) centre-periphery one, suggesting spatial distributions that go along with specific local characteristics (with a particular polycentrism in Milan FUA).

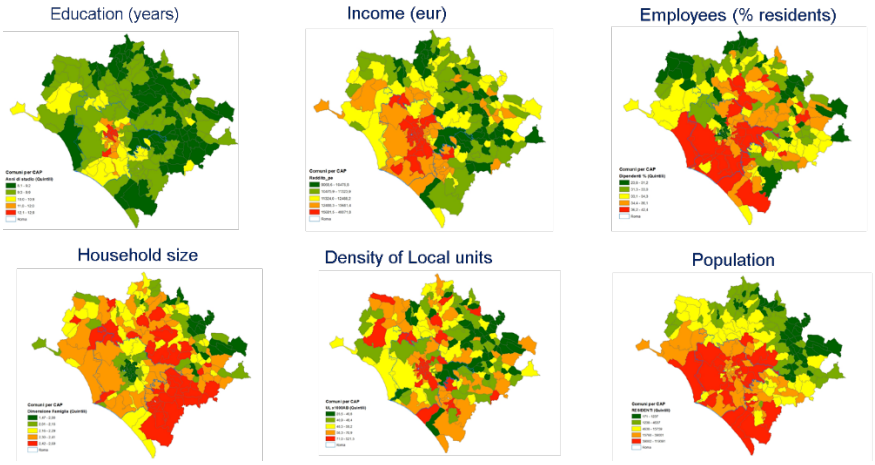
Figure 5 Education of residents in Rome, Milan and Naples *Functional Urban Areas*



Source: Authors' elaboration on territorial indicators from Istat extended Population Register (Base register + EDU)

With the same approach, in Figure 6 different socio-economic dimensions are juxtaposed drilling in Rome's FUA. These multiple portraits show similarities and differences within a specific urban setting: residents of the historical centre, as expected, are at the same time richer and live in smaller households (either because of age or because they are not in a traditional family setting), but for the rest, each of the variables considered shows its own spatial pattern.

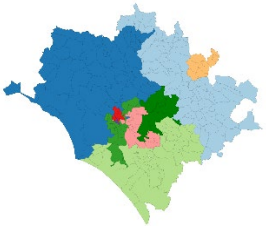
Figure 6 Spatial distribution of distinct socioeconomic features within Rome's FUA



Source: Authors' elaboration on territorial indicators derived from the integration of Istat statistical registers

In a subsequent step, the information on the same features presented in Figure 6 and on some additional socio-economic characteristics is summarised by means of a multivariate (spatial clustering) analysis (Figure 7). This latter, although presented for illustrative purposes only, permits to identify how joint distributions of income, education, age of residents, retired, self-employed, employees, foreign citizens, and economic activities tend to be spatially located within the urban area.

Figure 7 Multivariate (spatial clustering) distribution of socioeconomic characteristics within Rome's FUA



CL	RESIDENTI migliaia	DIMENSIONE Individui	Reddito_pc Euro	Alfabetizzazione Anni	ETA Anni	Stranieri %	Dipendenti %	Autonomi %	Pensionati %	UL_1000AB
1 (74)	2.562	2,19	10.995	9,07	47,22	9,13	32,83	5,14	23,46	45,22
2 (52)	16.248	2,28	12.075	9,59	44,43	13,84	33,32	5,83	18,02	56,36
3 (34)	27.011	2,42	12.253	9,53	43,04	10,83	35,11	5,12	16,72	54,80
4 (27)	40.115	2,01	20.728	11,49	47,63	11,23	34,56	7,36	21,95	84,56
5 (19)	37.611	2,20	14.466	10,14	45,78	11,95	36,15	5,58	19,66	63,52
6 (9)	18.551	1,92	32.898	12,58	48,34	13,37	28,82	11,29	20,86	288,10
7 (8)	292	1,71	11.716	9,10	52,70	9,48	28,68	5,34	32,80	36,41
8 (4)	94.225	2,44	11.037	9,23	41,82	15,24	36,04	4,54	14,80	50,89
9 (1)	28.167	1,47	27.675	12,20	42,96	46,52	23,75	8,19	11,87	191,00

Source: Authors' elaboration on territorial indicators derived from the integration of Istat statistical registers

A last example refers to changes over time: in this case, we considered business dynamics by economic activity (at the NACE Section level) from 2011 to 2019 in Rome and in a specific suburb ("Pigneto"), which during the 2010s underwent a process of gentrification (Table 2). The phenomenon is clearly visible in the data, in terms of differential growth of personal services, food and accommodation, and some professional services.

Table 2 Business dynamics in Rome and "Pigneto" suburb, by economic activity. Years 2011-2019 (2011=100)

Sezioni di Attività (ATECO 2007)		Roma	Pigneto
B	ESTRAZIONE DI MINERALI DA CAVE E MINIERE	82,1	
C	ATTIVITÀ MANIFATTURIERE	87,2	80,5
D	FORNITURA DI ENERGIA ELETTRICA, GAS, VAPORE E ARIA CONDIZIONATA	126,5	
E	FORNITURA DI ACQUA; RETI FOGNARIE, ATTIVITÀ DI GESTIONE DEI RIFIUTI E RISANAMENTO	103,5	300,0
F	COSTRUZIONI	91,9	96,3
G	COMMERCIO ALL'INGROSSO E AL DETTAGLIO; RIPARAZIONE DI AUTOVEICOLI E MOTOCICLI	96,9	103,5
H	TRASPORTO E MAGAZZINAGGIO	99,7	110,6
I	ATTIVITÀ DEI SERVIZI DI ALLOGGIO E DI RISTORAZIONE	129,5	144,8
J	SERVIZI DI INFORMAZIONE E COMUNICAZIONE	106,4	114,8
K	ATTIVITÀ FINANZIARIE E ASSICURATIVE	100,0	94,5
L	ATTIVITA' IMMOBILIARI	109,0	91,7
M	ATTIVITÀ PROFESSIONALI, SCIENTIFICHE E TECNICHE	111,9	136,0
N	NOLEGGIO, AGENZIE DI VIAGGIO, SERVIZI DI SUPPORTO ALLE IMPRESE	121,0	204,5
P	ISTRUZIONE	139,7	150,0
Q	SANITA' E ASSISTENZA SOCIALE	120,8	119,3
R	ATTIVITÀ ARTISTICHE, SPORTIVE, DI INTRATTENIMENTO E DIVERTIMENTO	111,9	151,4
S	ALTRE ATTIVITÀ DI SERVIZI	112,1	119,4
TOTALE		107,2	118,1

Source: ASIA Business Register

4. Remarks and conclusions

The examples presented in this paper provide some insights on the ongoing work on indicators and some analytical perspectives. It represents a relatively small part of what is already available and what could be available in the future, when the system of registers will be fully developed, at least in its main dimensions, and able to provide long-enough series to perform longitudinal analysis. We also have to underline that in the paper we had to use some incomplete information with respect to both income and employment (the coverage of all types of income and employment is being finalised now), and could not yet present a fully-fledged bottom-up methodology of spatial aggregation (the Base register of locations does not yet provide a 100% coverage). The development and interconnection of base and non base registers is an ongoing exercise and in the near future we hope to be able to enhance the production of new statistical information.

In addition, where quality will be sufficient, it could provide regular baselines for population and socioeconomic studies (e.g. snapshots of the entire population at a specific time, or of a population subset, or for specified geography).

References

Abdi H., Lynne J.W. (2010). “Principal Component Analysis.” John Wiley and Sons, Inc. WIREs Comp Stat 2: 433–59. http://staff.ustc.edu.cn/~zwp/teach/MVA/abdi-awPCA2010.pdf.

Aydin O., Janikas M.V., Assunção R., Lee T. (2018), SKATER-CON: Unsupervised Regionalization via Stochastic Tree Partitioning within a Consensus Framework Using Random Spanning Trees. Conference paper

Coppi R., Bolasco S. (eds.) (1989), Multiway Dala Analysis, North Holland, Amsterdam.

Escofier B., Pagès J. (1988), L'analyse factorielle multiple, Dunod, Paris.

Istat (2020) Rapporto sul Territorio. https://doi.org/10.1481/Istat.RapportoTerritorio.2020

Pagès, J. (2002). “Analyse Factorielle Multiple Appliquée Aux Variables Qualitatives et Aux Données Mixtes.” Revue Statistique Appliquee 4: 5–37.

Saporta G. (1990), Simultaneous Analysis of Qualitative and Quantitative Data, Società Italiana dl Statistica Atti della XXXV riunione scientifica, 18-21 aprile 1990 - Volume l, CEDAM Padova.