Data collection improvement for the Italian Road Accident Survey with fatalities and injuries 2022[1]

Francovich Lisa[2], Istat- Italian National Statistical Institute, francovi@istat.it
Santorsa Maria I.[3], Istat- Italian National Statistical Institute, santorsa@istat.it
Ielpo Roberto[4], Istat- Italian National Statistical Institute, ielpo@istat.it

Abstract

The working reorganization undergone in Istat in 2021 significantly changed the role and functions of the Istat territorial offices, with an important impact on the activities of the Central Directorate for Data Collection (DCRD). This led to review the data production processes to adapt them to the new organizational context, in particular of some processes that over the years had been decentralized on the territory, as was the case for Road Accidents Survey. In this work, we focus on the data collection new methods applied in 2022. The aim is describing them and highlight how they can guarantee and improve the efficiency in some phases of the process, right during a time of transition towards a new organizational model.

Keywords
Data collection, road accidents, process efficiency, quality of statistics, data correction, respondent

1. Introduction

Istat offers the cognitive framework on road accidents in Italy through two surveys: a monthly survey aimed at collecting detailed information on road accidents with fatalities and personal injuries and aimed at deepening the knowledge of the phenomenon. And a quarterly survey, carried out with the collaboration of the Municipal Police in about 200 municipalities throughout the country and made for collecting summary data on the number of accidents, deaths and injuries and producing preliminary estimates on road accidents in urban areas. Both are categorized as surveys of public interest, are included in the National Statistical Program, and provide for the obligation to respond for public entities.

In this paper the focus is on the monthly survey with the specific objective of describing the measures taken to standardize and automate the process of quality control of the collected data and the correction activities undertaken through the respondents' re-contact. To understand the actions carried out it is important to start from the survey's specificities and the description of context in which they took place.

2. The monthly Road Accidents with personal injuries and fatalities Survey and its organizational context

The survey is carried out with the collaboration of the police forces responsible for traffic control and traffic regulation on the roads, mainly Traffic Police, Carabinieri Stations and Municipal police.

---

[1] Extended abstract
[2] Paragraph 1, 4
[3] Paragraph 2,3,5
[4] Support in the production of the document and production of tables and graphs

Based on the definition of road accident[5] established by international standards and adopted in Italy, road accidents that fall in the survey field are those recorded by a Police Authority, occurred in streets or squares open to public traffic, with at least one vehicle involved, and resulting in injuries or fatalities (within 30 days). Therefore, are excluded from the survey those road accidents that do not result in fatalities or injuries or that do not occur in public traffic areas, or that do not involve vehicles. The survey unit is therefore the single road accident with fatalities or injuries to persons and the information collected refers to the time when the accident occurred.

For each road accident the Police Authority that recorded it must transmit to Istat a series of detailed information aimed at: locate the accident in time (date and time) and space (municipality, type and name of the road); describe the road characteristics (pavement, road surface, intersection/straight road, presence and type of road signs) and the weather and light conditions; reconstruct the accident dynamic by specifying the nature, the circumstances that supposedly caused it, the type and characteristics of the vehicles involved; specify the information on the driver of the vehicles involved (age, sex, nationality, type of driving license) and the consequences for persons (name of the injured or dead and hospital they were taken).

The survey is carried out with the cooperation of ACI (Automobile Club of Italy) and other local organizations in a complex and articulated context. In fact, since 1999, Istat has enhanced its collaboration at local level with provincial (NUTS3 level) or regional authorities (NUTS2 level) that actively participate in the survey phase, through special agreements (Memorandum of understanding and Bilateral Conventions). In addition, since 2007, a process of decentralization of the survey at a regional level has been enhanced, involving Istat Territorial Offices present in all regions (henceforth referred to as UT) in order to improve the level of coverage and quality of the collected information. This process concerned the Umbria, Campania, Basilicata, Marche, Molise and Abruzzo regions.

Here are the three organizational models that characterized the survey until 2021 (Figure 2):

- **Standard flow,** with direct data sending by the Municipal police to Istat, and is adopted in Valle d'Aosta/Vallée d'Aoste, Sicilia and Sardegna;

- **Data collection decentralization to UTs**, as well as monitoring, control and correction activities (Umbria, Campania, Basilicata, Marche, Molise and Abruzzo);

- **Decentralization to Province and Region authorities** of data collection and monitoring. It is adopted among regions adhering to the Memorandum of understanding (Tuscany, Piemonte, Lombardia, Emilia-Romagna, Puglia, Friuli-Venezia Giulia, Veneto, Liguria, Calabria and Lazio) or to the Bilateral Conventions in Bolzano/Bozen and Trento Autonomous Provinces.

Organizational specificities in the territory have also led to the adoption of a flexible data flow system. At present, there are different ways and timing of sending data to Istat: Traffic Police and the Carabinieri stations use a decentralized model on a national basis, irrespectively of the Region or Province agreements with Istat, while Municipal Police uses both the decentralized model and the direct data sending to Istat (Standard flow). Carabinieri stations and Municipal Police use a monthly transmission frequency, while the Traffic Police transmits data to Istat on a quarterly basis. (Figure 1).

---

[5] International definitions (European Commission, Eurostat, OCSE, ECE, etc.) of road accident state that a road accident is "that event in which at least one vehicle is involved and that happened on the road network and that causes fatalities or injuries to people" (Vienna Conference, 1968).

**Figure 1 - Data flow system from Police authorities to Istat (standard model and decentralized model)**
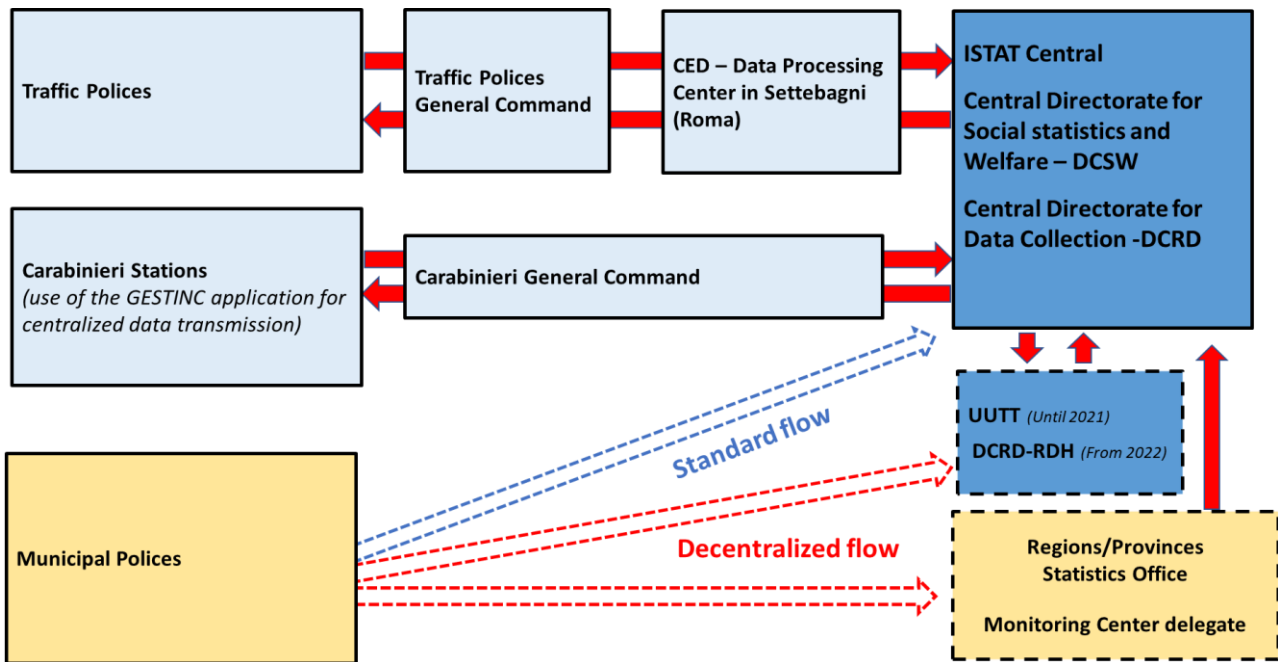
The channels used for data transmission consist mainly of two data acquisition system:

- **GINO**++ (Online Survey Management) dedicated to Municipal polices. Through this system, police can transmit the data of road accidents by registering them in a online Data Entry or uploading the files generated by their management software. The system, in use since 2019, ensures the correct output of the information collected thanks to internal quality and consistency checks on data that avoid the delivery of partial or incorrect information.

- **INDATA** portal, addressed to local police corps that use their own management software, where *filedata* can be uploaded, using a record layout prepared by Istat, that however need to be reviewed and corrected, not being guaranteed a controlled data entry. Municipal Police that have not yet adapted to the new standards of GINO++, the Carabinieri and the Traffic Police use this system.

The organizational structure described so far has been further modified as a result of a major reorganization that has involved Istat and its UTs in September 2021. Therefore, starting with the 2022 survey, data collection activities in regions where the survey was decentralized to the UTs (Abruzzo, Basilicata, Campania, Marche, Molise and Umbria) and in the standard flow regions (Sicilia, Sardegna and Valle d'Aosta/Vallée d'Aoste) have passed to the Central Directorate for Data Collection (DCRD) and specifically into the 'RD-Road Accidents' working group that is part of the Data Collection Service for Demographic, Social and Welfare Statistics (RDH). The new organizational framework at present has two macro-areas, the one in regions adhering to Memorandum of understanding or special agreements (11 regions in total) and the area in nine regions centrally managed by DCRD-RDH[6] (Figure 2).

---

[6] The transfer of the survey to RDH took place gradually, as the data collection activities for a given year of data *t* are carried out from March in year *t* to May in year *t+1*, to allow UTs to complete the activities related to the 2021 survey. It began in January 2022 in Abruzzo, Basilicata, Molise, Sardegna, Sicilia and Valle d'Aosta/Vallée d'Aoste, where

Figure 2 – Road Accident Survey: organizational models, before and after Istat 2021 re-organization
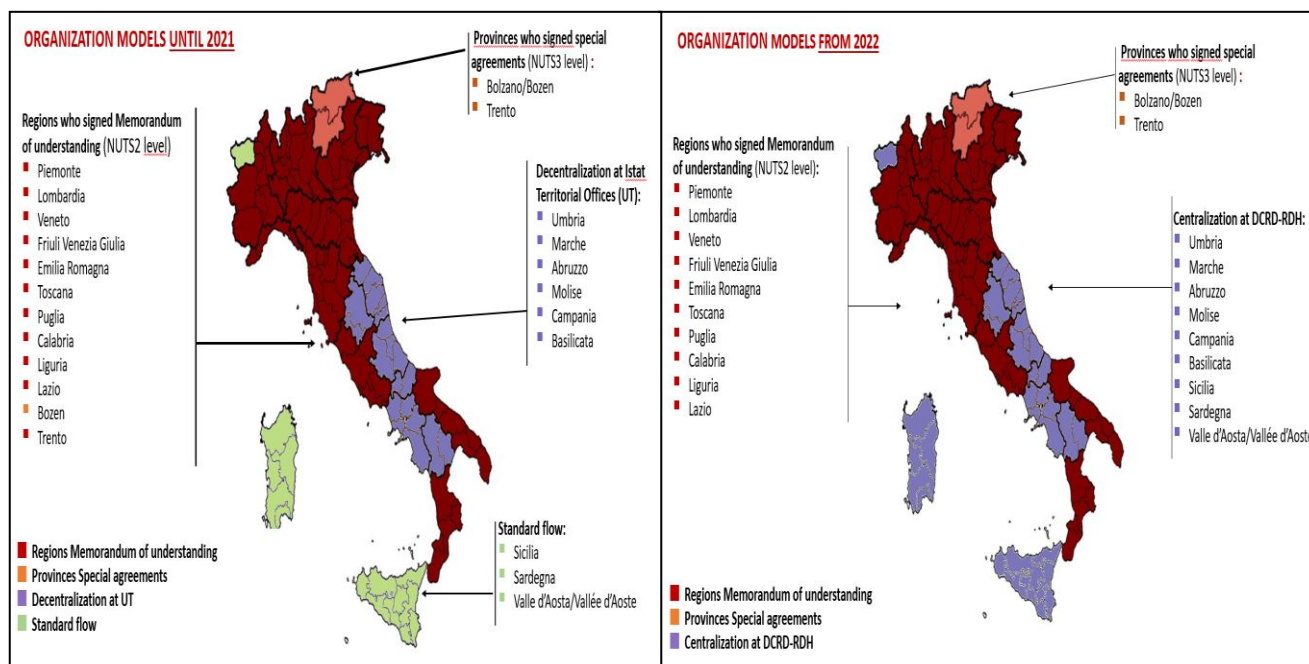
3. The Road Accident Survey in the new organizational framework: main objectives and actions

The latest organizational changes needed a redefinition of production processes in order to adapt them to the new context. The transition of the survey management from many subjects on the territory (UTs) to a single entity (DCRD-RDH) also led to review the organizational system of the survey that in the regions with UT decentralization presented different models with different impact on the processing of the collected data, on the timing and type of datasets returned to thematic service (DCSW-SWC)[7].

Thanks to a reconstruction carried out in collaboration with the thematic service DCSW-SWC, it emerged, in fact, that in some UTs the decentralization process concerned only the activities of monitoring and recovery of total non-responses as regards Municipal Police; in others UTs, the decentralization affected all the police corps (Municipal Police, Carabinieri, Traffic Police) but only some stages of the process; in others, however, it covered both aspects, with a positive impact on the quality of the data collected and on the timeliness and coverage of the information produced.

Changes in the survey organization, if on the one hand it has been a necessity and a major challenge given the high level of quality and efficiency achieved in some regions, on the other hand, it was an important opportunity to standardize and harmonize data collection on a territory represented made of 9 different regions, with about 2,300 municipalities and an annual average of over 36,500 accidents and a share of fatal accidents equal to 25% of the value recorded at the national level (Table 1).

RDH was responsible for completing the survey in 2021, to conclude in August 2022 with the delivery to DCRD-RDH of the Campania regions, Marche and Umbria, for the 2022 survey.

[7] Central Directorate for Data Production - Integrated Service for health, care and welfare system.

Table 1- Numbers of road accidents collected in regions assigned to DCRD-RDH in 2022. Years 2010-2021.
Absolute values

Regions	Number of towns	Year											
		2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
		Total road accidents											
Valle d'Aosta/Vallée d'Aoste	74	370	299	295	315	295	283	285	256	267	313	194	247
Umbria	92	2.913	2.856	2.363	2.402	2.258	2.285	2.382	2.361	2.385	2.306	1.699	2.001
Marche	228	6.728	6.535	5.482	5.549	5.422	5.333	5.185	5.484	5.216	5.399	3.695	4.663
Abruzzo	305	4.099	4.058	3.671	3.603	3.429	3.217	3.037	2.946	3.145	3.160	2.205	2.729
Molise	136	657	639	581	507	511	461	479	510	478	555	378	421
Campania	550	11.129	10.225	9.698	9.103	9.182	9.111	9.780	9.922	9.721	10.058	7.088	9.014
Basilicata	131	1.147	1.054	949	888	936	936	945	848	979	903	677	918
Sicilia	390	14.255	13.283	11.790	11.823	11.366	10.864	11.067	11.056	11.019	10.702	8.053	9.943
Sardegna	377	4.206	3.785	3.472	3.664	3.492	3.537	3.508	3.425	3.461	3.633	2.479	3.200
Total	**2283**	**45.504**	**42.734**	**38.301**	**37.854**	**36.891**	**36.027**	**36.668**	**36.808**	**36.671**	**37.029**	**26.468**	**33.136**
Italy	**7904**	**212.997**	**205.638**	**188.228**	**181.660**	**177.031**	**174.539**	**175.791**	**174.933**	**172.553**	**172.183**	**118.298**	**151.875**
		Road accidents with fatalities											
Valle d'Aosta/Vallée d'Aoste		11	9	10	7	13	6	3	7	9	4	-	1
Umbria		74	59	48	57	45	59	33	44	43	50	43	52
Marche		106	120	95	79	98	92	97	90	86	93	67	81
Abruzzo		78	78	86	67	72	77	75	66	73	75	56	73
Molise		27	18	17	22	25	21	15	27	12	21	24	15
Campania		235	232	229	213	208	215	208	235	193	205	170	203
Basilicata		45	31	42	20	39	40	40	29	36	26	18	33
Sicilia		260	247	211	229	192	211	179	197	195	194	155	205
Sardegna		97	91	90	111	91	103	99	84	99	69	89	86
Total		**933**	**885**	**828**	**805**	**783**	**824**	**749**	**779**	**746**	**737**	**622**	749
Italy		**3.871**	**3.616**	**3.515**	**3.161**	**3.175**	**3.236**	**3.105**	**3.178**	**3.086**	**2.982**	**2.275**	**2.737**

With the aim of improving the process, the main task of DCRD-RDH was to ensure a harmonized and standardized system of the data collection process and the activities, overcoming the differences of each region without however renouncing the good practices adopted in the territory.

Initially, it was planned to use the complete organizational work model, which concern the entire data collection process and all the local police corps, and was implemented before 2022 only in Basilicata, Campania and Umbria. Subsequently, on request of the DCSW-SWC service, the activity of data collection was restricted only to Municipal Police, thus excluding accidents recorded by Traffic Police and Carabinieri stations. The reasons are essentially linked to the data delivery timing, given the need for the DCSW-SWC service to anticipate the return by RDH of the so-called "annual consolidated data" that is the complete data, checked and corrected through the contact of respondents[8]. The reasons are also linked to the need to reduce the statistical burden on respondents and to the need to ensure compliance with the times at different stages of the process.

Therefore, the commitment of RDH from 2022 concerned the process of data collection from Municipal Police with the dual objective of ensuring the total coverage of the survey (also using a dedicated call center service for inbound and outbound activity with respondents) and the quality of the data collected.

Specifically, the following activities have been undertaken:
- Collection of information from Municipal police;
- Training, assistance and support during data collection and as regards the use of GINO++;

---

[8] The survey decentralization to UTs included the sending the complete annual data by May 31st of the year following the reference data year. By switching data collection management to the DCRD-RDH service, the DCSW-SWC team has requested to anticipate of the data transmission to the first half of April of the year *t+1* and to restrict the quality control and the total coverage control to the accidents reported by the Municipal Police.

- Monitoring and control of the survey total coverage[9];
- Contacts with the local police forces aimed at recovering the total non-response;
- Quality control of the data collected and contact with the Municipal police to correct the errors found in the data transmitted;
- Final control of the over/under coverage of the phenomenon on the basis of historical data series and with other sources, with consequent contact of local police forces in case of significant differences [10].

The quality control of the collected data consists in verifying that all the information required in the questionnaire for each accident is complete and consistent with each other. In some cases, Municipal polices are contacted to request clarification and to proceed with the correction and/or integration of the missing and/or incorrect information.

As stated before, this activity concern Municipal Police data files transmitted through INDATA portal, while the data transmitted by GINO++ are excluded from quality control. Although very few Municipal police use INDATA, the proportion of accidents transmitted is still relevant. The analysis of the data collected by Municipal police in regions followed by RDH in 2022 highlights, in fact, that the transmission of data through the INDATA portal accounts for just 1% of Municipal police but with a significant share of accidents transmitted, equal to 23,5% of the total, that is over 4,900 incidents transmitted in INDATA by only 21 local police, all operating in municipalities with at least 20,000 inhabitants (Table 2).

Table 2 - Number of Municipal polices and percent of road accidents by data transmission system. Year 2022

	Data transmission system						
Region	Municipal Police				% records transmitted		
	GINO++	INDATA	INDATA-GINO++ (a)	Total	INDATA	GINO++	Total
Valle d'Aosta/Vallée d'Aoste	74	-	-	74	-	100,0	-
Umbria	91	-	1	92	1,7	98,3	100
Marche	214	8	3	225	45,9	54,1	100
Abruzzo	304	1	-	305	24,6	75,4	100
Molise	136	-	-	136	-	100,0	-
Campania	544	4	2	550	12,5	87,5	100
Basilicata	128	3	-	131	79,6	20,4	100
Sicilia	386	4	1	391	25,4	74,6	100
Sardegna	376	1	-	377	26,9	73,1	100
TOTAL	2.252	21	7	2.281	23,5	76,5	100

(a) Municipal polices migrated in GINO++ during the year.

It is easy to understand that the process of data check and correction is challenging, especially with a complex and long questionnaire like in this survey, with many variables that are subject to dissemination. Specific computerized SAS® procedures have been thus developed, that are iterative and made to prepare an accurate error map in order to simplify the correction activity (through the re-contact of Municipal police) and with the aim of:

a) identify duplicate records and records out of the survey field[11] ;
b) extract records that are incorrect or do not meet minimum quality requirements;
c) identify and describe the errors in each record.

---

[9] In the absence of accidents with fatalities or injuries, the municipal police must transmit to Istat a communication of "negative outcome". The control of the total coverage, aimed at reducing the error of total non-response, is carried out during the data collection phase by monitoring at a municipal level the number of road accidents and urging for monthly missing data.

[10] This check is carried out at the end of the data collection phase by comparing total number of road accidents absolute values and percentages per municipality, per local police corps and per month, with data of the corresponding historical series and with the quarterly survey (number of accidents, deaths and injuries).

[11] Accident with no fatalities nor injury and/or that have not vehicle involved or have not occurred in a public traffic area are out of the survey field.

Moreover, to allow the Municipal police to proceed independently with the record integration and correction, a specific area in GINO++ is under construction, separate from the production area[12]. Its use would avoid contact between Istat and the Municipal police for error correction and would allow acquiring the correct data in real time and in complete safety. The joint use of SAS® control procedures and GINO++ correction area would speed up and simplify the whole process through the following steps:

- the SAS® procedures check the data files received on the INDATA portal and extract the records with errors;
- incorrect data for each accident will be uploaded by Istat in the new GINO++ area dedicated to correction;
- an email will be sent to local police corps indicating that there is information to be corrected;
- local police corps will connect to the GINO++ specific area and make the correction by intervening only on the variables that the system reports as incorrect.

Before the development of the SAS® mapping procedures, variables to be corrected by re-contact of the Municipal polices where selected, as well as it was necessary to select errors to take into account. The SAS® mapping procedures involved all variables of the questionnaire, while the correction activity, in agreement with DCSW-SWC, was addressed to a minimum set of variables, and were those related to date, time, place, location, nature, presumed circumstances of the accident, vehicles and injured persons. This selection was made in order to find a fair compromise in the cost-benefit evaluation, that is, between the need to produce a dataset as complete and 'clean' as possible and the need to reduce the statistical burden on respondents, knowing that the local police corps re-contact would generate an excessive response load with a negative impact on the organization and speed of the survey.

Since the GINO++ specific correction area is under development, the correction work on the 2022 data during data collection was carried out through the re-contact of the Municipal police. In order carry out this activity (which has taken place at different moments) in the best way and to facilitate communication with the Municipal polices in order to limit errors as much as possible, the re-contact work was divided among colleagues by assigning to each one of them the same municipalities in the different correction rounds and making available to them: the descriptive error map, containing the data identifying the accidents to correct with the indication of the wrong variables and the description of the errors; the data base with the accidents to correct and the questionnaire Access mask, where the questionnaire is displayed and the corrections can be made.

Being aware that re-contact with the respondent is challenging and that it requires professionalism, attention and mastery of the contents of the questionnaire and of the tools used for data correction activity, particular attention was also paid to the training of colleagues in charge of the correction, who were 'prepared' for the job through a formation seminar dedicated to deepening the questionnaire knowledge and the use of tools, with the help of simulations and practical exercises.

4. The computerized procedure for collected data quality monitoring

In the survey waves before 2022, the information completeness and correctness control, implemented in the data collection phase, was set up in some regions with different tools (SAS, Access, Excel) according to the informative and operational needs that in concrete emerged locally during data collection. During 2022 data collection, given the allocation of nine regions to DCRD-

---

[12] A separate area in GINO++ is necessary because GINO++'s record layout is different from the record layout of file transmit through INDATA, not only in the structure (4 csv files are needed in GINO++ to describe one accident) but also in the collected information. In addition, the questionnaire in GINO++ allows managing many vehicles involved in the road accident, while INDATA text file is structured to contain information up to a maximum of three vehicles.

RDH, it was necessary to adopt a procedure that would allow the management of the data quality control process in a more systematic and comprehensive way. The guiding key concepts have been to standardization, simplification, and automation, where possible, of all the activities in the process. Therefore, an error data mapping was structured, as exhaustive and systematic as possible ("internal" error profile Filippucci C., 2002). Below we present the error mapping and the logical path through the criteria definition and methods of re-contact of Municipal polices.

The error classification resulting directly from the structure of the questionnaire (Manzari A. 2022 and Istat 2004) and from *a priori* knowledge on variables is described below:

A. **Missing errors**: a variable has a missing value. This can happen in two cases, when the variable must be present in all records, and when it is under condition, that is the error exists depending on a filter question.

B. **Domain errors**: when the variable returns an ineligible value, that is, out of the range of its possible values, given the answer modalities in the questionnaire.

C. **Not-due answers (NDA)**: given a filter variable, the NDA happens if a question outside the filter has been answered.

D. **Incompatibility between variables**: or consistency errors, also called 'conditions of incompatibility between variables'.

Table 3 displays 2022 percent distribution of accidents that have been *mapped*, that is the data acquired through the INDATA portal, by local police corps[13].

Table 3 - Road accident percentage sent to DCRD -RDH through INDATA, by region and by local police corps. Year 2022

Region	Local Police Corps				Total
	Traffic police	Carabinieri	Municipal police	Total	
Abruzzo	27,9	53,1	18,9	100	9,3
Basilicata	17,1	51,6	31,3	100	4,4
Campania	28,9	56	15	100	24,1
Marche	25,1	41,4	33,5	100	18,5
Molise	28,8	71,1	0	100	1,6
Sardegna	19,8	58,1	22,1	100	10,7
Sicilia	25,6	36,7	37,7	100	25,1
Umbria	36,3	61,5	2,2	100	5,2
Valle d'Aosta/Vallée d'Aoste	31,8	68,2	0	100	1,1
Total	**26,2**	**48,9**	**24,9**	**100**	**100**

The accidents' record layout contains about 180 variables, several of them logically interdependent[14]. Writing for each variable the control rules (errors definitions) in SAS® described above, we ended up with a consistent number of possible errors (in total 367), 181 missing, 140 domain errors, 7 not-due answers and 39 incompatibilities. The application of these rules to the datasets sent by local police corps leads in theory to define two sets of records, those with at least 1 error and those without any error.

The analysis of the incorrect records by type of error in the nine regions managed by DCRD-RDH is displayed in Table 4, distinguishing between all police corps (Municipal Police, Traffic Police

---

[13] Data on the number of incidents in 2022 are deliberately expressed in percentage values, as they have not yet been disseminated.

[14] The high number of variables is depending on the fact that information about vehicles, drivers and passengers are repeated for all vehicles involved in the accident, up to a maximum of three.

and Carabinieri) and the Municipal Police alone. The table also reports the percentage of duplicated records and of those 'off-field observation' (OFO)[15].

Table 4- Percentage of error, by type of local police corps. Regions assigned to DCRD-RDH. in Year 2022

Type of error	All Local Police corps	Municipal Police
Number of accident (*absolute value*)	19.327	4.811
1 Missing at least	100%	100%
1 Domain at least	11,90%	26,80%
1 NDA at least	6,50%	5,70%
1 incompatibility at least	26,40%	10,90%
Duplicated records (*absolute value)*	16	2
Off-field observation-OFO	0,20%	0,08%
Re-contacts	49,50%	42,50%

The quality control of the data was done on 19,300 accidents' records, of which approximately 4,800 coming from Municipal Police and on all the variables. The analysis of errors shows that: all the records contain at least one missing error; 11,9% of records have a domain error, with a higher percentage for the Municipal Police (26,8%); the presence of at least one NDA error is recorded in 6,5% of all cases, while the presence of at least one error of incompatibility is lower for Municipal Police (10,9%) in comparison with the general percentage (26,4%). On the territories, there is a regional peculiarity in Sardegna and Sicilia as regards domain errors. The other errors show no specific particularities. We now restrict the analysis to a subset of variables (that also destinated to dissemination and are of public interest), focusing on the fundamental information and its dynamic and consequences. It should be noted that the following data concern all local police corps.

The missing error mapping shows that the missing answers in the variables describing the timing of the accident are very few, only 9, due to the missing 'hour' and 'minute' variables. Even the localization on the territory (Province and Municipality variables are never missing) and the nature of the accident have very few missing errors, with not even 1% of answers missing. Also missing errors on 'type of vehicle' variable are very low for vehicle A and B, respectively 0,24% and 0,87%. Missing errors occur mainly in variables related to the presumed circumstances of the accident: 13,4% of cases for vehicle A, in 24,0% for vehicle B/pedestrian or obstacle; missing values in at least one of the variables related to vehicle A driver (age, gender and accident consequences) is present in a negligible percentage and always below 2%; moreover, the simultaneous absence of this information affects 1,6% of cases. In the case of vehicle B, omissions have a greater impact but do not exceed 3,0%. These results, in the opinion of the authors, are an indication that the controls on the variables province, common, time, nature of the accident and the presence of at least one vehicle involved in the accident are basic in most of the software used by all police corps.

Most NDA errors, which account for 6,5% of the total number of cases, are related to accidents involving a moving vehicle with a parked one, while information about the parked vehicle is not required.

Domain errors (12,0%) mainly concern the circumstances of vehicle A (1,2%) or vehicle B (0,5%), and, in third position, but far away by incidence, the nature of the accident (0,3%). There is no domain error in the timing and in the localization of the accident. At a regional level, the highest incidence of domain errors is recorded in Sardegna (4,2%) and Sicilia (4,8%).

The most frequent errors are incompatibilities that amount to 26,4% of the records.

---

[15] Cfr. note 11.

The greatest number of inconsistency errors are observed on the variables relating to the presumed circumstances of the accident which relate in particular to the incorrect indication of the circumstances of vehicle B, pedestrian or obstacle depending on the nature of the accident; their percentage ranges from 9,2% to 1,7%. Errors in the compilation of variables related to the presumed circumstances of the accident are due to the fact that for these variables the local police corps management systems often do not provide any type of control nor a guided compilation that can help in the form filling-in phase and, moreover, to the fact the presumed circumstances questionnaire section is less intuitive.

The procedures ran on the variables related to fatalities and injuries made it possible to identify non-eligible records (OFO), which are 35 with no indication of dead or injured persons, particularly concentrated in Abruzzo and Marche; the procedure also identify 23 records with mismatch between the total number of fatalities and injuries (reported in a specific summary section of the questionnaire) and the same number deduced from variables related to the consequences of the accident (dead and injured persons) for drivers, passengers of the vehicle and pedestrians involved in the accident. The analysis of the Municipal police data has also highlighted a specificity in the town of Messina, where there are twenty records with at least one injured pedestrian but without the indication of the vehicle; these cases result to be OFO, after the re-contact with the Municipal police.

As stated before, all the records received by Istat through INDATA undergo the error mapping procedure, but the re-contact of the respondent for the correction of the errors has concerned only the Municipal polices, and concretely it became necessary in 42,4% of those cases for a total of 2,072 records to be corrected. If we had extended the correction activity to all local police corps, the number of records to be corrected would have risen to over 9,000. Below is displayed the frequency error distribution for the accidents that needed corrections, by variable, region and type of error (Table 7).

In conclusion, the use of mapping procedures has greatly simplified the process of quality control during data collection, allowing to quickly identify wrong records and to reproduce a map of the errors for each record. The errors analysis has allowed to direct the attention towards the Municipal police that more critics for incidence and typology of errors and to identify the systematic errors, allowing a focused re-contact on respondents. They also allowed the production of an operational report (complete and easy to use) in use by colleagues involved in the re-contact and error correction phase.

Table 7 - Frequency of errors found in the variables to be corrected by region and type of error - Municipal Police

Information about the accident	Variables and Type of error	percentage values							Total Errors Detected	
		Abruzzo	Basilicata	Campania	Marche	Sardegna	Sicilia	Umbria	absolute values	values %
WHEN (data, time)	HOURS *(missing/out of domain)*	-	-	-	-	-	-	-	**0**	-
	DAY *(missing/out of domain)*	-	-	-	-	-	-	-	**0**	-
	MONTH *(missing/out of domain)*	-	-	-	-	-	-	-	**0**	-
WHERE (location of the accident)	ROAD TYPE *(missing)*	-	-	-	0,3	-	0,1	6,7	**4**	**0,1**
	ROAD TYPE *(out of domain)*	-	-	-	-	-	-	-	**0**	**0,0**
	ROAD NUMBER* *(missing)*	-	12,7	-	15,4	-	0,9	3,3	**86**	**2,5**
	STREET NAME *(missing)*	-	-	-	-	-	-	-	**0**	**0,0**
	PROGRESSIVE KILOMETERS* *(missing)*	-	5,1	0,2	9,6	2,8	0,7	3,3	**62**	**1,8**
TYPE of accident	NATURE OF THE ACCIDENT *(missing)*	-	-	0,2	0,6	-	3,2	6,7	**63**	**1,9**
VEHICLES involved	VEHICLE**A *(missing)*	-	-	-	-	0,4	1,7	-	**33**	**1,0**
	VEHICLE** B *(missing)*	-	-	0,2	0,3	2,8	3,8	3,3	**80**	**2,4**
	CIRCUMSTANCES*** VEHICLE A *(missing/domain/incompatibility)**	48,6	44,1	46,2	33,7	29,5	45,0	23,3	**1447**	**42,8**
	CIRCUMSTANCES*** VEHICLE B *(missing)*	37,8	20,3	35,0	16,0	4,3	30,3	10,0	**930**	**27,5**
	CIRCUMSTANCES*** PEDESTRIAN/OBSTRUCTION *(missing)*	13,5	16,9	17,8	20,1	19,6	8,2	20,0	**423**	**12,5**
	CIRCUMSTANCES*** VEHICLE B - intersection *(incompatibility)*	-	-	-	2,0	38,8	0,5	-	**126**	**3,7**
CAUSES of the accident	CIRCUMSTANCES*** VEHICLE B - non-intersection *(incompatibility)*	-	0,8	-	1,7	1,8	5,0	10,0	**107**	**3,2**
	CIRCUMSTANCES*** PEDESTRIAN *(incompatibility)*	-	-	-	-	-	0,1	3,3	**2**	**0,1**
	CIRCUMSTANCES*** VEHICLE IMPACT. STOP/TRAIN/OBSTACLE *(incompatibility)*	-	-	-	0,3	-	0,1	-	**2**	**0,1**
	CIRCUMSTANCES *** LISTING/FALL *(incompatibility)*	-	-	-	-	-	0,4	-	**8**	**0,2**
CONSEGUENCES of the accident to people	INJURED/OUTCOME *(incompatibility between summary and outcome)*	-	-	0,2	-	-	-	10,0	**4**	**0,1**
	DEATHS/OUTCOMES *(incompatibility between summary and outcomes)*	-	-	-	-	-	-	-	**0**	**0,0**
Total errors *(absolute values)*		**362**	**118**	**409**	**344**	**281**	**1833**	**30**	**3377**	**100,0**
Total Record incorrect (absolute values)		**206**	**80**	**275**	**248**	**225**	**1021**	**17**	**2072**	**43,2**

* This variable is considered missing for accidents that occurred on motorways, national, regional, or provincial roads.

** For each accident it is possible to enter up to a maximum of 3 vehicles involved (A, B and C) sorted by degree of responsibility in the dynamics of the accident.

*** The presumptive circumstances of accidents are intended to understand the accident dynamic. They refer only to two moving vehicles (A and B). In the case of accidents involving a single moving vehicle, they refer to vehicle A and the pedestrian in case of pedestrian collision; to the parked vehicle/train/obstacle in case of collision; to the obstacle not collided in case of sudden slip/braking or fall.

5. Conclusions

The SAS® procedures represent the process innovation implemented in 2022 that enabled the workload and the new organization to be addressed. In fact, the use of these mapping procedures has resulted in a significant reduction in time and resources involved in the data quality control process and in the correction activities with positive implications for the results. It is not possible to determine precisely the advantages obtained in terms of recovered time and resources by not having elements of comparison with the past (substantial number of regions to be treated, decentralization on many different territories and offices). But certainly, we can say that the use of the mapping procedures allowed the achievement of the objectives in the activities planned on the survey 2022

respecting deadlines data delivery timetable to DCSW-SWC, despite the small number of resources allocated on this activity.

Moreover, the reports produced by municipality, and in particular the map of errors, allowed to identify the critical spots on the territory and to manage in a more flexible and targeted way the re-contacts with respondents; they also allowed to find and adopt specific methods of re-contact for bigger municipalities and for those with many errors. The map of municipalities with the highest number of errors and the most frequent errors also allows to plan targeted information and training actions, focusing on the most critical aspects of the questionnaire. The error map could then be a useful tool to be extended to all local police corps re-contact activities, in order to improve the data quality, in particular if used in conjunction with the new correction area in GINO++.

The contact with respondents in the data correction phase is confirmed to be very useful helping to strengthen the collaboration relationship with the Municipal polices and often becoming a moment of training on the job, with the consequences of significantly reducing some types of errors, in particular in municipalities where the accidents' data entry information is managed by the same person.

References
Filippucci, C. (edited by), *Strategie e modelli per il controllo della qualità dei dati*, Franco Angeli, 2002.

Riccini Margarucci, E. (edited by), *Concord V.1.0 Controllo e correzione dei dati. Manuale utente e aspetti metodologici*, Istat, 2004.

Istat, *Linee guida per la qualità dei processi statistici che utilizzano dati amministrativi*. Version 1.1, August 2016. https://www.istat.it/it/files/2010/09/Linee-Guida-fonte-amministrativa-v1.1.pdf

Istat, *Manuale Concord V. 1.0 Controllo e correzione dei dati: manuale utente e aspetti metodologici*, Roma, Istat, 2004.

Manzari, A., *Aspetti generali sulle procedure di controllo e correzione dei dati*, Presentation made in ISTAT on 13-06-2002.

Brancato, G., Boggia, A. Ascari, G., *Linee Guida per la Qualità delle Statistiche del Sistema Statistico Nazionale. Ver. 1.0*, Istat, March 2018. https://www.istat.it/it/files//2018/08/Linee-Guida-2.5-agosto-2018.pdf.

Casale, D. (edited by), *CLAG: verso un software generalizzato per l'acquisizione controllata dei dati via Web e l'organizzazione autonoma e flessibile della rete di rilevazione*. Istat, 2010.