

22 - 24 May 2024, Geneva Switzerland

UNECE

**Expert Meeting on Statistical Data Collection and Sources** 

# DATA COLLECTION OF THE ENVIRONMENTAL SURVEY IN CITIES: DATA VALIDATION

Domenico Adamo\*, Gianpiero Bianchi\*\*, Paolo Francescangeli\*\*, Lucia Mongelli\*\*\* \*Istat | DCAT – ATA \*\*Istat | DCRD – RDF \*\*\* Istat | DCRD – RDI

### Index

#### > Overview

#### Structure

#### Organization

- Data collection technique and methods
- Survey stages. Until 2022 edition
- Process innovations from ed. 2023. Validation automation
- Survey stages. After re-engineering
- > Automatic validation. Some types of rules for the detection of outliers. Some examples
- A generalized data editing for error detection
- Conclusions

Data Collection of the Environmental Survey in Cities: Data Validation



#### **Overview**

3



Data Collection of the Environmental Survey in Cities: Data Validation P. Francescangeli, D. Adamo, G. Bianchi, L.Mongelli

**Urban environment data** is a multiple data source statistical process, organized into **8 thematic modules**, that produces environmental indicators for **110 Italian cities** (the 109 provincial capitals +1, participating on a voluntary basis).

Most data are collected through a **CAWI survey** (Computer Assisted Web Interviewing) carried out annually by Istat since 2000. Data are collected directly from the Municipalities, via their statistical offices.

On several topics, survey data are integrated with **administrative data**, provided by other Agencies of the National statistical system (SISTAN).

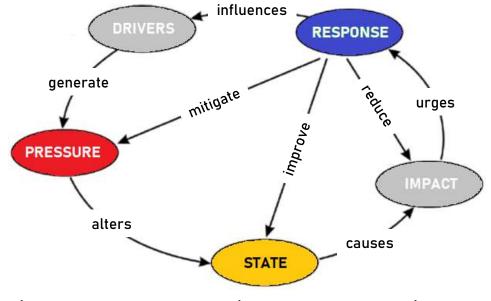


#### **Overview**

4



Data Collection of the Environmental Survey in Cities: Data Validation P. Francescangeli, D. Adamo, G. Bianchi, L.Mongelli



The process provides a comprehensive information framework for monitoring the **quality of urban environment** (state and pressure indicators, according to the DPSIR model ↑), and the environmental policies implemented by local administrations (response indicators).

## **Overview**

#### **LEGAL BASIS**

The survey on urban environment is part of the **National Statistical Programme** (PSN), managed by the SISTAN and updated every 3 years. Being included in the PSN as a **survey of public interest**, data collection is performed by law, and the response is mandatory for the reporting units.

The PSN provides also the legal basis for the **use of administrative data for statistical purposes**. Specific agreements are made between Istat and the data holders to define the features and timing of data provision, within the SISTAN regulatory framework and under the rules for the protection of personal data.

Use of administrative data allows reducing costs and the burden on respondents.

#### **REPORTING AND ANALYSIS UNITS**

The survey's reporting units are the **Municipal statistical offices**. Every year, before opening the data collection, Municipal offices appoint a coordinator and one or more reference persons for each thematic questionnaire.

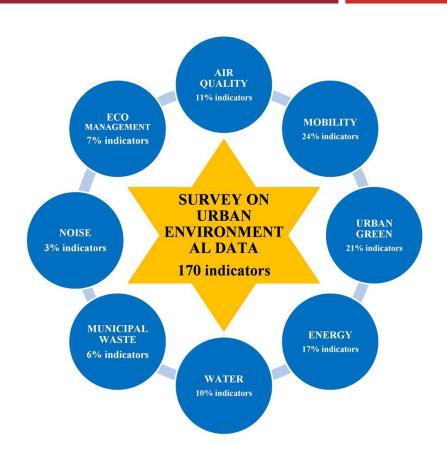
Depending on the topics, reference persons collect the data directly from the Municipal administrations, or request it from other local Agencies (e.g. public transport companies). All indicators are disseminated by municipality, aggregated estimates by geographical area are also provided.

Data Collection of the Environmental Survey in Cities: Data Validation



#### **Structure**

6



Data Collection of the Environmental Survey in Cities: Data Validation P. Francescangeli, D. Adamo, G. Bianchi, L.Mongelli

**Survey data** are collected through **thematic questionnaires** (except for Water, since 2018 part of the *Urban water census*, covering the whole national territory).

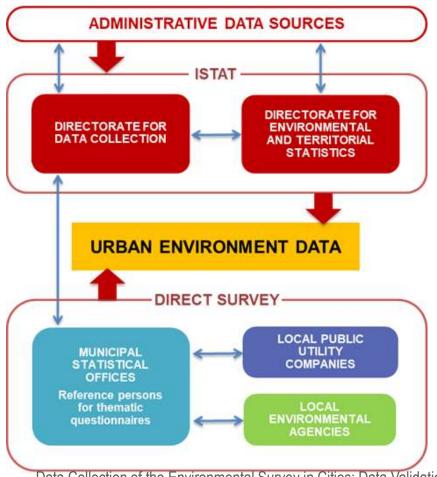
Four modules integrate survey data with **administrative data** on particular topics (provided separately by data holders).

Administrative data are acquired, stored and made available to statistical production units by the <u>Data collection directorate</u>. Being collected for other purposes, administrative data need to be processed and validated in order to produce relevant statistical information. This task is performed by the <u>Environmental and territorial statistics directorate</u>, which also assist Data collection in the relationship with data holders.

The entire process manages about **500 elementary variables**, to produce **170 indicators** (2024 release), 13% of which based on administrative data. Some indicators are part of the lstat set of statistical measures for monitoring the SDGs in Italy.



# Organization



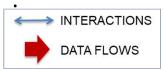
Data Collection of the Environmental Survey in Cities: Data Validation

7 P. Francescangeli, D. Adamo, G. Bianchi, L.Mongelli

The **Directorate for data collection** is responsible for the entire process of data collection, the organization and monitoring of the survey and the acquisition of data from administrative sources.

**Urban environmental data** is part of the **Directorate for environmental and territorial statistics**, which is responsible for survey design, information contents, data validation and dissemination.

The wide range of topics covered makes the process of gathering information at the local level quite complex. Thus, **assistance to respondents** and a strict **monitoring of survey operations** are key factors for data quality.





## Data collection: technique and methods

In accordance with the Code of Digital Administration (d.lgs 82/2005) for the collection "Environmental data in cities", from year 2008, a major innovation and simplification in the data collection process has been introduced.



Introduction of the CAWI (Computer Assisted Web Interviewing) technique for data acquisition in electronic format, through the Gino++ portal (Gathering Information Online).

Data Collection of the Environmental Survey in Cities: Data Validation



## Data collection: The GINO++ system

**GINO++ is a generalized software**, robust and flexible, allowing you to:

- 1. create electronic questionnaires for data collection with controlled data acquisition;
- 2. to set up the customized site for the investigation;
- 3. manage the survey (population of respondents and all actors involved, monitoring of questionnaires and registrations, contacts for reminders and reminders, reports).

Data Collection of the Environmental Survey in Cities: Data Validation P. Francescangeli, D. Adamo, G. Bianchi, L.Mongelli

q



#### Data collection: The GINO++ system

The data are collected by the Municipal Statistical Offices, which identify in the Administrations to which they belong the referents of the survey topics, who are given personal credentials to log in, enter, edit and save the data.

Through GINO++ the Municipal Statistics Offices and the contact persons of the various topics can send the data by filling in direct web questionnaires.

Automated checks are implemented to prevent the entry of inconsistent or invalid data or out of range, and the sending of questionnaires with missing responses.

Data Collection of the Environmental Survey in Cities: Data Validation P. Francescangeli, D. Adamo, G. Bianchi, L.Mongelli



# The GINO++ system

Access to the GINO++ acquisition system:

https://gino.istat.it/amburb/

In the *Home* are reported:

- ✓ the description of the investigation
- ✓ the obligation to reply
- instructions for access and completior the questionnaire (details in the 'Documents and instructions' section)
- ✓ the IT requirements
- ✓ the FAQ

Data Collection of the Environmental Survey in Cities: Data Validation P. Francescangeli, D. Adamo, G. Bianchi, L.Mongelli 11

imbra: Comp	oni	× 💀 GINO	× +								1	
⇒ C (	gino.istat.it/a	amburb/#						07	ß	☆		۲
		Documenti e istruzior						2				
	Jati ambi	entali nelle o	litta					Ist	at			
	comuni capolu statistica, hann	ogo di tutte le provi o l'obiettivo di fornire	e città", effettuata annualme nce italiane e delle città met e un quadro informativo a sup aministrazioni per assicurare la	tropolitane e al Co pporto del monitora	mune di Cesena. I da aggio dello stato dell'a	ti e l'informazione	Accesso al qu	uestiona	rio			
	La rilevazione,	in questa edizione 2	D23, si articola in cinque que	stionari d'indagine	- Aria, Mobilità, Rifiuti	urbani, Rumore e	mongelli					
	domanda e off	erde urbano - che raccolgono informazioni su: diffusione degli inquinanti atmosferici e misure per contenere l'inquinamento;					PASSWORD	PASSWORD				
	temporanee a	rilevante impatto acu	e la raccolta dei rifiuti urb ustico: verde urbano a gestion	ne pubblica, azioni	per lo sviluppo degli							
			ficazione relativi alle principali e per potenziare la qualità dei			ni dindagina sala	Password dimenticata	ACCE	DI			
	quest'anno non	viene richiesta la co	mpilazione dei questionari Eco ato nel modulo Acqua) ed	o management (ch	e include il Razioname	ento dell'acqua per	Modifica Dati Registrazio	ine		_		
			no 2022, l'indagine consente questionari Aria, Rumore e Ve									
□   <b>2</b> ← C □ Importa	raccolta relativa delle acque per Accesso al client W ti da Intern R	amente al 2021 per i uso civile" con caden deb Zimbra ×   © //gino.istat.it/question	questionari Aria, Rumore e Ve za biennale) la rilevazione avra	rde urbano. Per il m à luogo quest'anno. G Istat - Gino nitoraggio_risp.php	nodulo Acqua (inglobai	to nel "Censimento	AN 🗙	£ి≡ visa	¢.	-		D >
□   2 ← C ⊃ Importa ■ Ista Monitorage	raccolta relativa delle acque per Accesso al client V c https: ti da Intern R tt ti da Intern R tt Dati ambientali i	amente al 2021 per i uso civile" con caden leb Zimbra ×	questionari Aria, Rumore e Ve za biernale) La rilevazione avrà WhatsApp × ari/gestione2/monitoraggio/mor g Accesso al client W () (1) Wha	rde urbano. Per il m à luogo quest'anno. G Istat - Gino nitoraggio_risp.php	Acqua (inglobal	to nel "Censimento a Rilevazí 🖪 Login di Ad	obe Co 🗢 cartella condi urbani 🥆 Rilevazione j		@ 2023			- 🚺
C C C Importa Importa Monitorage Monitorage Regione	Accesso al client W Accesso al client W https: ti de Intern R It Dati ambientali i gio Dati Regist Provi	amente al 2021 per i uso civile" con carie de Zimbra × //gino.istat.it/question la Repubblica - Ne e new nelle citta* razione Report per sta ncia	questionari Aria, Rumore e Ve za biennale) la rilevazione avră whatsApp × ari/gestione2/monitoraggio/mor ș Accesso al client W  (1) What to indagine Report Registrati	Anegrafice Contatti	Acqua (inglobal + + sting - Zo • SGI   Scelt Gestione file Report vazione Nome/Cr.	to nel "Censimento a Rilevazi 🔀 Login di Ad V Tematica Riffuti t per Riapondenti	urbani 👻 Rilevazione j					Esci
C C C Importa Importa C C C C C C C C C C C C C C C C C C C	Accesso al client W Access	amente al 2021 per i uso civile" con caden eb Zimbra ×   • //gino.istatit/question Is Repubblics - Ne • new nelle citta" razione Report per sta ncia Comun NTI	questionari Aria, Rumore e Ve za biennale) la rilevazione avră whatsApp × ari/gestione2/monitoraggio/mor ș Accesso al client W  (1) What to indagine Report Registrati	Anegrafice Contatti	Acqua (inglobal + + sting - Zo • SGI   Scelt Gestione file Report vazione Nome/Cr.	to nel "Censimento a Rilevazi 🔀 Login di Ad V Tematica Riffuti t per Riapondenti		periodo [				Esci
☐ 2 ← C Importa Importa Internet Monitorage Monitorage Monitorage Monitorage Monitorage Monitorage	Accesso al client W Accesso al client W C https: it da Intern R It Ju Dati ambientali ii Glo Dati Regist Provi RAGGIO RISPONDE Rispondente Agrigento	Amente al 2021 per i uso civile" con caden //gino.istat.it/question la Repubblica - Ne e new nelle citta" razione Report per sta ncia Comun NTI b Regione * Sicilia	questionari Aria, Rumore e Ve za biennale) la rilevazione avra ari/gestione2/monitoraggio/mor g Accesso al client W  () What to indegine Report Registratt Provincia + Agrigento	Anagrafica Contatti Stat - rise Anagrafica Contatti Comune + Agrigento	Gestione file Report	to nel "Censimento a Rilevazi  Cuesto Constructional de la construcción de la construc	urbani V Rilevazione j Data utima modifica + Utente	período [ , + ⓒ3	2023	×		Esci
Codice + 19084001 19084001	Accesso al client VA Accesso al client VA i de Intera. R it de Intera. R It Dati ambientali i Glio Dati Registi Provi Rispondente Agrigente Alessandria	amente al 2021 per i uso civile" con cale de Zimbra × //gino.istat.it/question la Repubblica - Ne e new nelle citta* ressione Report per sta nella citta* Report per sta NTI Regione + Sicilia Piemonte	uppestionari Aria, Rumore e Ve za biennale) la rilevazione avra whatsApp × × ari/gestione2/monitoraggio/mor g Accesso al client W  (1) What to indegine Report Registrati Provincia * Agrigento Alessandria	Anagrafica Contatti Comune * Agrigento Alessandria	Gestione file Report Vazione Vazione * Inevaso	to nel "Censimento a Rilevazi 23 Login di Ad v Tematica Riffutti t per Rispondenti Stato rispondente Stato rispondente Attivo Attivo	urbani V Rilevazione j Data ultima modifica + Utente admin	período ( , + Q3 Q3	2023			Esci
Codice + 19084001 19084001 19084001 19084001	Accesso al dient W Accesso al dient W https: ti da Intern R It Dati ambientali i glio Dati Registi RAGGIO RISPONDE Rispondente Agrigento Alessandria Ancona	e new elie citta' resione Report per sta ncia Comun Regione * Sicilia Piemonte Marche	Accesso al client W	Ansgrafica Contatti Comune + Agrigento Anscandia Anona	Cestione file Report	a Rilevazi 💽 Login di Ad	urbani	período , * (Q) (Q) (Q)	2023			
<ul> <li>Monitorage</li> <li>Monitorage</li></ul>	Accesso al dient W Accesso al dient W ti da Intern R It Dati ambientali i Dati ambientali i Provi RAGGIO RISPONDE Rispondente Agrigento Alessandria Ancona Andria	e new e Report per sta neile citta* Regione * Skillia Piermonte Marche Puglia	to indagine Report Registrati  Provincia * Agrigento Agrigento Barietta-Andria-Trani	Ansgrafice Contatti Ansgrafice Contatti Comune * Agrigento Ansgrafica Ansgrafica	Gestione file Report	to nel "Censimento a Rilevazi  Cugin di Ad Tematica Rifluti t per Riapondenti dice rispondente Stato rispondente Attivo Attivo Attivo Attivo	Urbani	período , + කි කි කි කි ක ක ක ක ක ක ක ක ක ක ක ක ක				
	Accesso al dient W Accesso al dient W https: ti da Intern R It Dati ambientali i glio Dati Registi RAGGIO RISPONDE Rispondente Agrigento Alessandria Ancona	e new elie citta' resione Report per sta ncia Comun Regione * Sicilia Piemonte Marche	Accesso al client W	Ansgrafica Contatti Comune + Agrigento Anscandia Anona	Gestione file Report Vazione Stato rilevazione * Inevazo Inevazo Inevazo Inevazo Inevazo Inevazo	a Rilevazi 23 Login di Ad	urbani	periodo ( ) ද ද ලි ල ල ල ල ල ල ල ල ල ල ල ල ල ල ල ල				
Monitorage Monitorage	Accesso al client VA Accesso al client VA id a Intera. R id a Intera. R It Dati ambientali i Dati ambientali i Dati ambientali i RAGGIO RISPONDE Rispondente Agrigento Alessandria Anona Anota	e new  c	to indegine Report Registrati Provincia * Aprigento Alessandria Accesso Alessandria Ac	Anagrafica Contatti Comune * Aricente Aricente Ariagrafica Contatti Comune * Agrigento Alessandria Ancona Andria Aosta	Gestione file Report	to nel "Censimento a Rilevazi  Cugin di Ad Tematica Rifluti t per Riapondenti dice rispondente Stato rispondente Attivo Attivo Attivo Attivo	urbani Y Rilevazione j Data utima modifica + admin - admin - admin - admin	período ද ද ලි ල ල ල ල ල ල ල ල ල ල ල ල ල ල ල ල				
	Accesso al dient W Accesso al dient W https: ti da Intern R ti Dati ambientali i Dati ambientali i Previ RAGGIO RISPONDE Rispondente Alessandria Andria Andria Antraia	e new e zimbra ×	Agrigento Arsiandría Argigento Assendría Ancona Barietta-Andria-Trani Valle d'Aosta/Vallée d'Aoste Arsizzo	Anagrafica Contatti Comune + Agrigento Agrigento Agrigento Anesa Andra Ancona Andra Ancona Andra Ancona Andra Arezzo	Gestione file Report Vazione V Stato rilevazione + Inevaso Inevaso Inevaso Inevaso Inevaso Inevaso	a Rilevazi 💽 Login di Ad	urbani Y Rilevazione j Pate utima admin admin admin admin admin	periodo ( ) ද ද ලි ල ල ල ල ල ල ල ල ල ල ල ල ල ල ල ල				

# Survey stages. Until 2022 edition

	MUNICIP		STAGES	ISTAT			
		OFFICES	STAGES	Data collection	Environmental statistics		
			Survey design	Survey organization, Implementation of CAWI questionnaires	Information contents and metadata management		
states	Initial - before taking over by reference person	Registration		Controlled acquisition through Gino electronic			
Questionnaire sta	In process - after first opening by reference person			questionnaire with rules Monitoring of survey	Assistance to respondents		
	Sent - after completion by reference person	Data entry	Data collection	operations	and to Data collection staff		
	Checked - after preliminary check			Preliminary check and follow-up			
	Data processing				Data editing and validation		
Data dissemination				Data analysis and reporting			

Data Collection of the Environmental Survey in Cities: Data Validation



### Process innovations from ed. 2023. Validation automation

Concerning Istat's objectives to innovate the various statistical information production processes, concerning the progressive digitalization of data collection processes, it has become necessary to design the use of innovative solutions through the reengineering of the validation phase of the questionnaires, with the implementation of further automatic control rules different from those already provided by the validity and internal consistency checks of the Gino ++ system.

The new rules would manage, at least in part, the aspects so far entrusted to the review by expert colleagues: in particular the interception of measurement errors, discontinuities of time series and other outliers.

These rules would work on a dynamic basis, by comparing the data collected with those validated in the previous edition. To be fully operational, however, this process innovation requires a testing phase, which could be concluded over the next two editions of the survey (2023 and 2024).

Data Collection of the Environmental Survey in Cities: Data Validation



# 2023 Edition

During this experimental phase, to ensure the regular conduct of the survey while maintaining the quality standard of the data collected, a diversified frequency is expected for some thematic questionnaires, which do not produce indicators intended for institutional dissemination, as in the table below:

Thematic questionnaires	2023 Edition	2024 Edition
Air		
Eco-management	(suspended)	•••
Energy	(suspended)	····
Mobility	•••	
Municipal waste		(suspended)
Noise		(suspended)
Urban green		

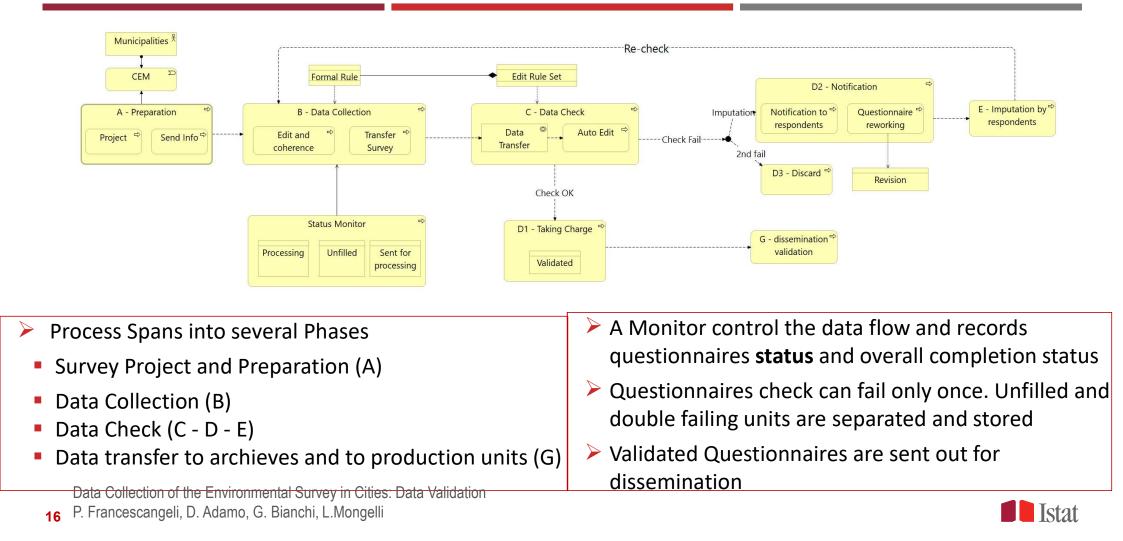
Data Collection of the Environmental Survey in Cities: Data Validation



## Survey stages. After re-engineering

		MUNICIPAL	STACES	ISTAT			
		OFFICES	STAGES	Data collection	Environmental statistics		
			Survey design	Survey organization, Implementation of CAWI questionnaires	Information contents and metadata management		
tes	Initial - before taking over by reference person	Registration	Data collection	Controlled acquisition through Gino electronic questionnaire with rules Automatic control of: • measurement errors, • historical series discontinuities • other abnormal values. Monitoring of survey operations			
aire states	In process - after first opening by reference person				Assistance to respondents and to Data collection staff		
Questionnaire	<b>Sent</b> - after completion by reference person	Data entry					
Que	<b>Checked</b> - after preliminary check			Automated return on respondents			
			Data processing		Data editing and validation		
			Data dissemination		Data analysis and reporting		
5	Data Collection of the Environmental Sur		Validation		Istat		

## **Process Outline**



# A generalized data editing for error detection (1/2)

#### Edits are usually represented by propositions:

Logical proposition expresses a logical conditions on values of a single field Examples:

 $(x_1 < 14)$ 

Mathematical proposition expresses a mathematical condition on values of at least two quantitative fields Example:

 $(x_1 - x_2 \ge 14)$ 

17

**Logical** edits are expressed only with logical propositions

Mathematical edits are expressed only with mathematical propositions

**Logical-mathematical** edits (or mixed edits) are expressed using both type of propositions

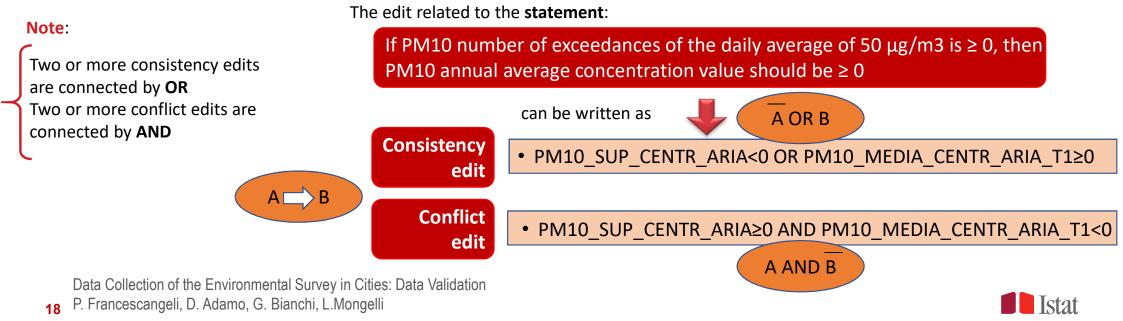
Data Collection of the Environmental Survey in Cities: Data Validation P. Francescangeli, D. Adamo, G. Bianchi, L.Mongelli



# A generalized data editing for error detection (2/2)

Each rule is translated into a generalized language, created on purpose. Such language is read by the editing system.

**Consistency** edits are represented by the **disjunction** of two or more propositions **Conflict** edits are represented by the **conjunction** of two or more propositions



# Conclusions

The main objective of the project will be:

- Design and implement the validation process from a generalized perspective, including reporting for error evaluation and analysis.
- Define the methodologies and algorithms necessary to carry out the automatic checks required by the validation process.
- Design and develop application components and database, including procedures for integrating the validation process with the capture and production environment.
- Analyze and validate the tools developed and the results of the new validation process through the definition of test cases and ongoing experimentation on the 2023 survey.

Data Collection of the Environmental Survey in Cities: Data Validation P. Francescangeli, D. Adamo, G. Bianchi, L.Mongelli



# Thank you for your attention

Domenico Adamo <u>adamo@istat.it</u> Gianpiero Bianchi <u>gianbia@istat.it</u> Paolo Francescangeli <u>pafrance@istat.it</u> Lucia Mongelli <u>mongelli@istat.it</u>





Istituto Nazionale di Statistica