Istat | Istituto Nazionale di Statistica

Online, 12 - 14 June 2023

UNECE Expert Meeting on Statistical Data Collection 2023

A SYSTEM-TO-SYSTEM DATA COMMUNICATION CHANNEL FOR MULTITECHNIQUE SURVEYS: THE CASE OF ITALIAN AGRICULTURAL CENSUS

Claudia Fabi – Maura Giacummo

Istat | Central Directorate for Data Collection & Central Directorate for Information Technology

Index

- The context: the data collection design for the Agricultural Census
- The data communication system: issues and solutions
- Results
- Conclusions

The Data Collection Design for the 7° Agricultural Census

The data collection design projected for the 7th General Census of Agriculture was based on an **integrated system**, entirely on **digital support**, which offered the possibility to adopt **three different survey techniques simultaneously**:

- **CATI** (Computer assisted telephone interviewing), in both «**inbound**» and «**outbound**» techniques;

- **CAWI** (Computer assisted web interviewing);

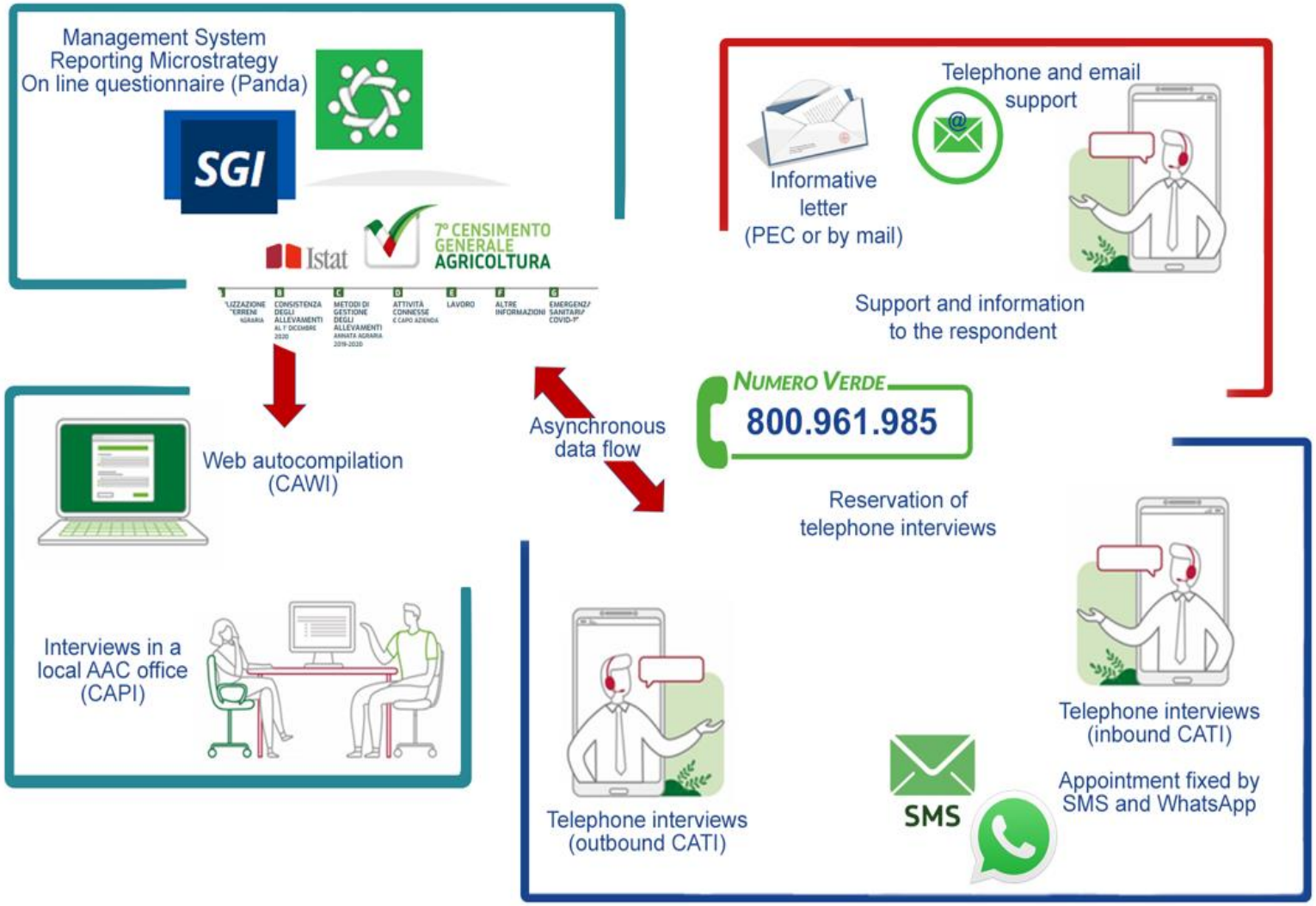
- **CAPI** (Computer assisted personal interviewing).

The Data Collection Design for the 7° Agricultural Census

The data collection process was based on SGI, an **online management system** developed by Istat, with many features dedicated to the networks involved, in order to **monitor, follow and evaluate** the work in progress with respondents.

CATI software for telephone interviews was based on a different System, developed and used by a Contact Center outsourcer, who support the data collection process providing nearly 400 telephone interviewers.

It was necessary to develop **an asynchronous data flow, to merge contacts and outcomes** from the two management Systems.



The Data Collection Design for the 7° Agricultural Census

The census list included approximately **1,700,000 units**, found through the use of Administrative Registers in the agricultural sector. The entire list was divided, before the start of the survey, into **two subgroups intended for a pre-assigned survey technique: CATI or CAPI**.

However, the pre-assignment was **not strictly binding**, but of a preferential nature. The criteria for pre-assignment were mostly based on the presence or absence of one or more telephone numbers. All respondents were able to choose to participate also through one of the **two techniques available on individual initiative.**

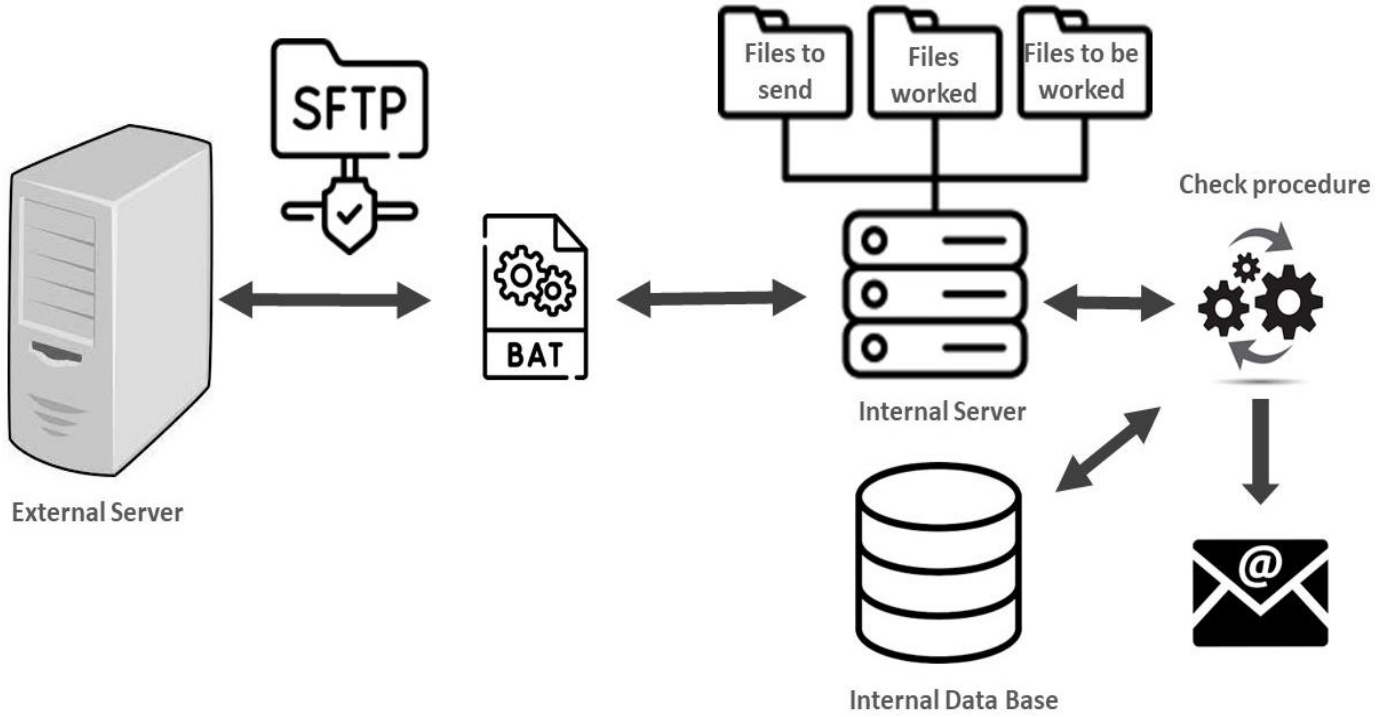
	Schedule of data collection activities	
	From 7th January 2021	*To 30th July 2021*
Techniques available on individual initiative	**CAWI**	
	Inbound CATI	
Pre-assigned Techniques	**CAPI**	
	Outbound CATI	

A System-to-System Data Communication Tool: the Project

As mentioned before, the survey was attested on **two distinct IT architectures**, the first dedicated to CAPI and CAWI techniques developed by Istat, the second to CATI techniques developed by an external outsourcer.

Consequently, it was necessary to design a **communication channel** between the two IT systems that would allow the results and the status of the questionnaires' compilation to be kept updated, almost on a daily basis.

The **efficiency and punctuality of these operations** were the key for **the success of the field survey**, ensuring that the simultaneous data collection techniques could effectively be interchangeable with each other.



A System-to-System Data Communication Tool: the Project

The design process of **an automated data exchange system**, through files in a predefined format, coming from SGI to the CATI system and vice versa, lasted about 6 months before the start of the survey.

The design specifically concerned the following issues:

a) **scheduling** of data exchange frequency: **twice a day**, in the morning and after lunch time;

b) exchange **file format**: ASCII format, .txt delimited by “pipe”;

c) **content** of the exchange files: **a minimal content, nearly 10 variables**, needed to merge contacts and outcomes from SGI to CATI and vice versa, rebuilding the contacts history of every farm/respondent in both Data Collection Management Systems;

d) **nomenclature** of exchange files: strictly binding, with an **alphanumeric nomenclature** that identifies the direction of the transmission (from CATI to SGI, or from SGI to CATI), the date and the time related to the outcomes included in every file;

e) **quality control strategies** for exchange files and identification of anomalous records;

f) **recovery plan**, in case of failure of the exchange procedures.

A System-to-System Data Communication Tool: Data Quality

One of the most important goal to be achieved, in a System-to-System Data communication, is to guarantee **Data Quality** and to find a right trade off between quality checks and quality data.

The system managed three type of check:

- **1st level**: to evaluate whether to accept or reject the entire file (types of controls: Correctness of the file name or structure);
- **2nd level**: to exclude non-compliant records and accept the compliant ones (some types of controls: questionnaire completed in other techniques, correctness of identification code, coherence of date with submission time window and name of the file);
- **3rd level**: to manage the creation of the file to send back to the outsourcer, only if 1st level check were ok.

A System-to-System Data Communication Tool: the Recovery Plan

There was also a **recovery plan** that allowed to guarantee a quickly restore of data in case of errors.

An **automatic system sent emails to a control team**, composed of Istat personnel and outsourcer personnel, in case of critical failure events (such as empty or incorrect deliveries).

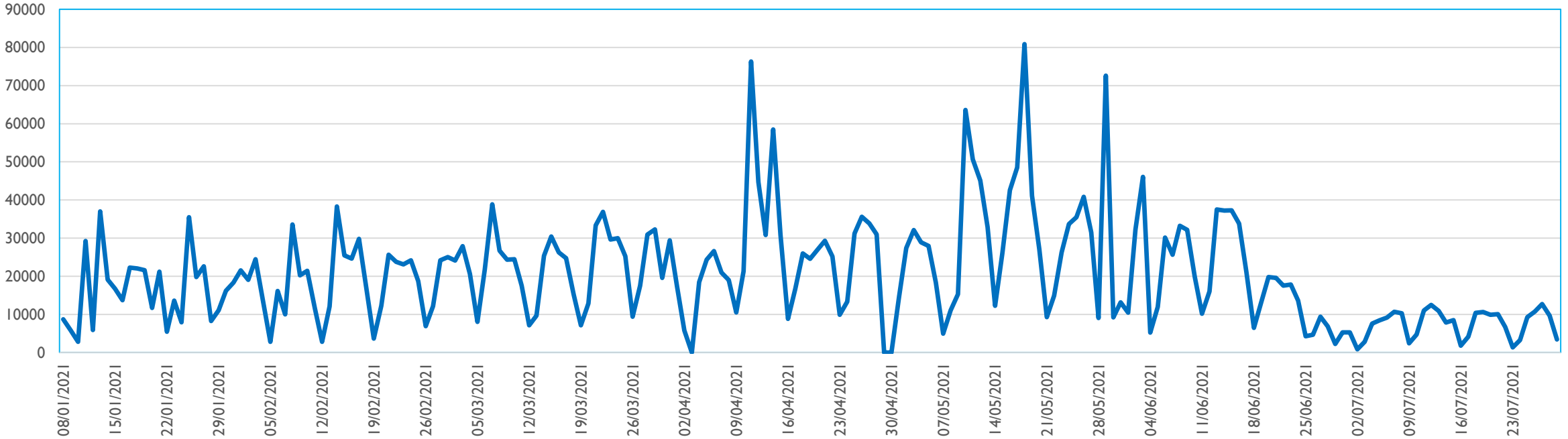
The system was designed to repeat the processing of any incorrect delivery, both entirety or partially, and to recover the **acquisition of more than one failed delivery simultaneously**.

Processed files and logs were stored in a **dedicated file system.**

Results: Daily records exchange

The amount of information exchanged daily can be useful to evaluate the size of the architecture and the data space needed to use a similar procedure to other surveys.

During 7 months, 412 files were exchanged, containing **over 6 million records**.
10,000 records is the average number of records processed **daily**, with peaks of **80,000** records during particularly intense fieldwork periods.



Results: Data Quality

Only **three** files were rejected at the first level of check, recovered by the outsourcer with new files very quickly.

The second level check found **45.351 errors (0.8%):**

- 73% coming from the change of technique, these were not true error but discrepancies due to the lack of a perfect synchronization between techniques
- **27% true errors** that lead to records rejection **(0.2% based on 6 million records exchanged)**.

All errors derive from the following data coherence issues:

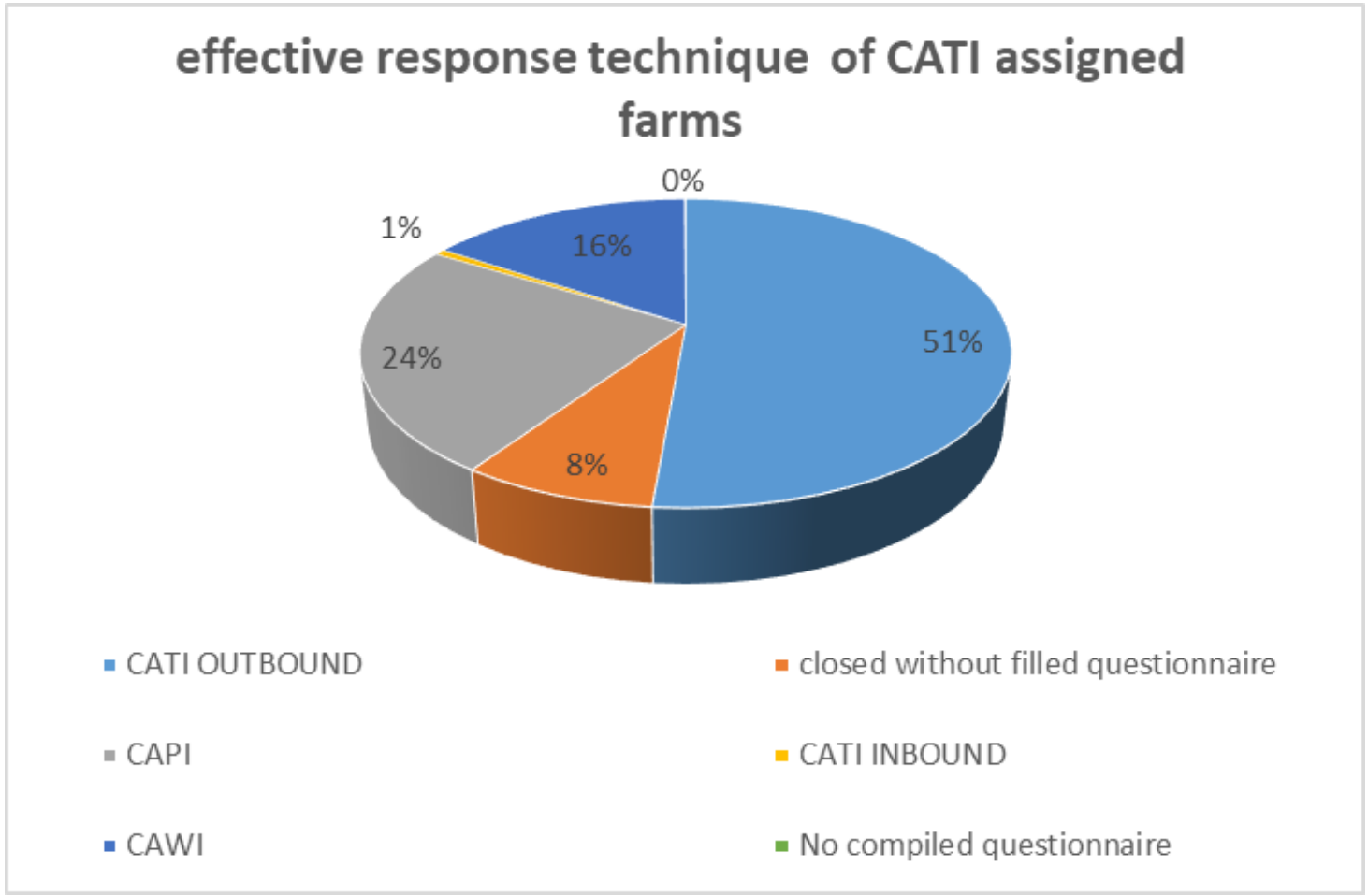
- internal date not compatible with the date of the file name;

- outcomes' codes not compatible with the technique (e.g., an outbound outcome code for an inbound attempt, etc.).

Results: CATI to CAPI transitions during the fieldwork

The farms pre-assigned to CATI technique were 550,000 (32% of the census list). Among them, only **282,536** interviews were completed **using the CATI** technique, slightly over 50%. The others preferred to fill the questionnaire with another technique: **around 16% to CAWI and 24% to CAPI.**

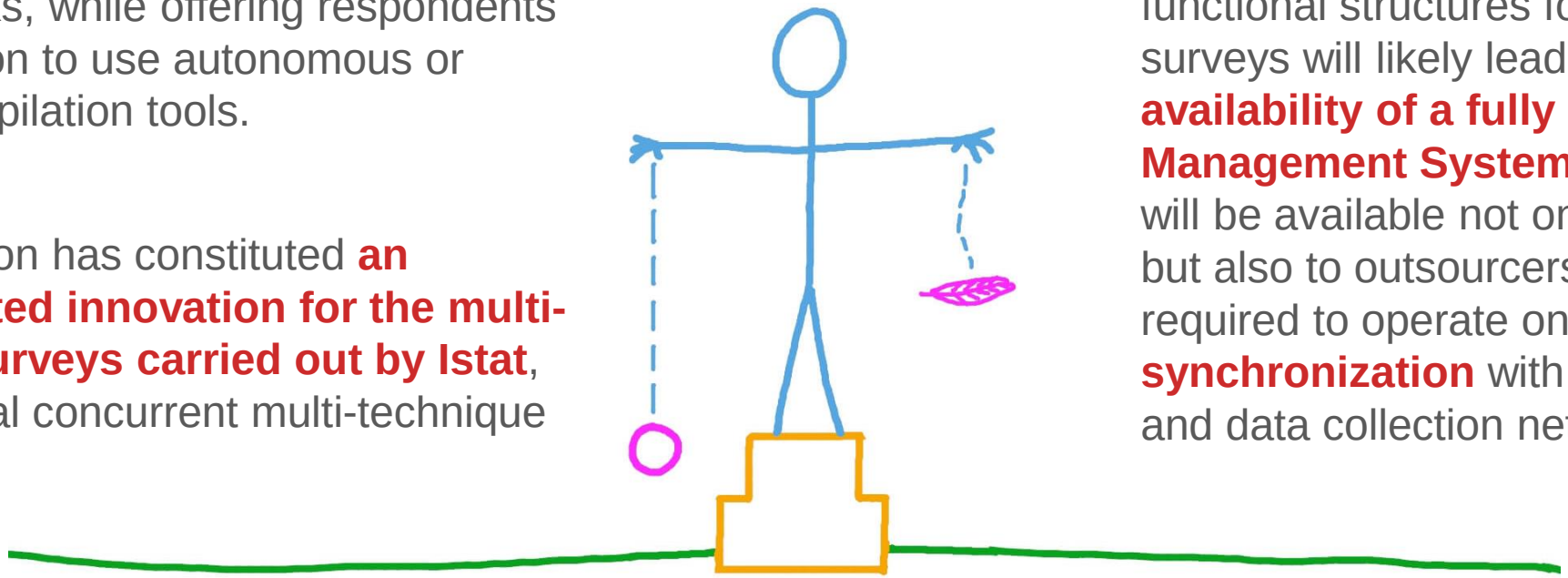
Thanks to the daily data exchange flow to CAPI network, **it was possible to recover 24% of the former CATI interviews, which otherwise would have been lost.**



Conclusions

Despite the limitation of representing **an approximation of real-time synchronization**, the asynchronous update has ensured a satisfactory smoothness in the data collection process both of CAPI and CATI networks, while offering respondents wide discretion to use autonomous or assisted compilation tools.

This integration has constituted **an unprecedented innovation for the multi-technique surveys carried out by Istat**, allowing a real concurrent multi-technique approach.



Over time, the continuous implementation of modules and functional structures for managing Istat surveys will likely lead to the **availability of a fully integrated Management System**. This System will be available not only to Istat users but also to outsourcers who will be required to operate on it in **perfect synchronization** with other techniques and data collection networks.

Thanks!

CLAUDIA FABI | claudia.fabi@istat.it

MAURA GIACUMMO | magiacum@istat.it

Istat | Istituto Nazionale di Statistica