**Metadata driven monitoring of electronic data capture**

**Keywords:** Mobile data capture, data collection process integration, metadata driven monitoring system, census data collection.

# Introduction

As mobile technology is becoming more widely available, many statistical agencies have considered using mobile data capture for the 2020 round of population and housing census data collection. Adopting mobile data capture means changing from paper to electronic questionnaires on handheld electronic devices. This decision affects the entire census lifecycle and makes it possible to obtain benefits both in terms of data quality and savings of time and costs of census operations [1].

Mobile data capture allows users to take advantage of added features that can be programmed into mobile devices or linked to the data collection process. These features include, among others, integrated maps and Global Positioning System (GPS) and real time monitoring of fieldwork. As data from each device are sent to a central database, they can be used to monitor the progress of the enumeration activities, identify which areas have already been covered and produce real time reports on population structure. Having access to this information in real-time allows adjustments to be made to during the data collection to improve efficiency, coverage and data quality.

This paper shows how Istat has designed and implemented such a monitoring system, integrated with the census data collection process supported on Android devices by CSPro [2], the public domain software package developed by the U.S. Census Bureau.

The following paragraphs provide a high-level description of the data collection process and the implemented architecture in the framework of a cooperation project financed by the Italian Cooperation (AICS), to support Ethiopia in carrying out the population census, which is currently planned for the first quarter of 2019. The Ethiopian Census is proposed as a case study to describe the main features of the system implemented to support data collection and field work processes. The strong cooperation between Istat and the U.S. Census Bureau made such results possible.

# Data collection process

In order to implement a generalized system to support census fieldwork activities, it is necessary to split the data collection process into the following phases (Figure 1):

* **Primary data collection**: the enumerators collect data on tablets or smartphones using CSPro. Data are stored on the devices in a *plain text* format. Later, using CSPro sync functionality, the enumerators transfer the plain text files over the internet to a central database (CSWeb). This system allows data collection in offline mode in areas where the network is not available or is unreliable.
* **Data transformation:** CSWeb stores each *plain text* file, containing questionnaire microdata, in a table field. This means that, as the number of returned questionnaires increases, it becomes difficult to extract and elaborate real-time information from the database. In order to increase efficiency, it is necessary to store microdata in a more structured way (i.e. each variable in a separate column). The data transformation (from plain text to microdata in separate columns) is performed by the software component CSPro2Sql (described in the following section).
* **Field work monitoring:** data stored in the central database can be used to monitor the progress of the enumeration activities, identify which areas have already been covered and produce real time reports on population structure. In order to display reports, a web application (Dashboard) was developed. Moreover, the reports have been integrated with GIS maps. The link with the GIS data allows display of the status of the data collection on maps, so that the fieldwork process can be visually monitored.

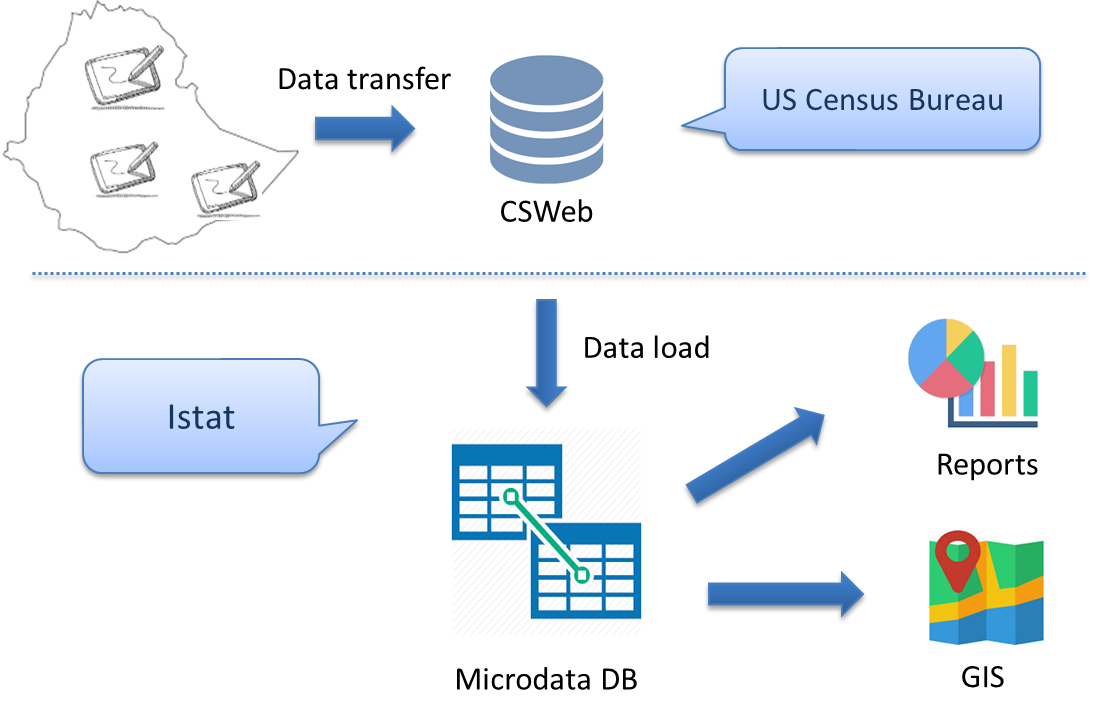


Figure 1 - Data collection process

# data collection architecture

In order to implement the data collection process described above, it is necessary to design a metadata driven architecture, which allows generation of the microdata database and the dashboard reports, parsing the questionnaire metadata (CSPro Dictionary). The software components needed, are described below:

* **CSPro data dictionary**: the data dictionary contains all the questionnaire metadata (e.g. variables, classifications, relations between variables). CSPro generates a data dictionary file for each questionnaire.
* **Cspro2sql:** thissoftware component has a key role in the integrated architecture. More specifically Cspro2sql, parsing the content of the data dictionary, generates the scripts to create the microdata database and to load microdata collected using CSPro. Further, Cspro2sql offers functionalities to generate the several types of reports (e.g. field work monitoring, age distribution, sex distribution), displayed by the Dashboard web application.
* **Dashboard:** this component is a web application implemented using open source Java frameworks. The current Dashboard displays the reports generated by Cspro2sql. Moreover, the dashboard offers the possibility to integrate GIS maps.

Figure 2 displays a simplified model of the implemented architecture.

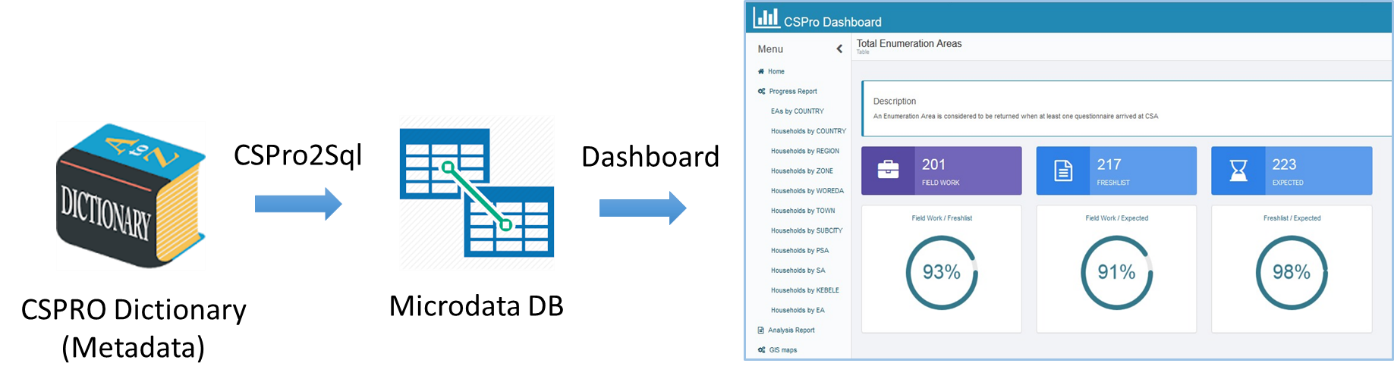


Figure 2 - Data collection architecture

Cspro2sql and the Dashboard are publicly available and released under the EUPL license.

# Conclusions

The proposed architecture provides a simple solution for monitoring electronic data collection operations, particularly in cases where the technical and financial resources to implement such a system from the ground up are lacking.

The integrated architecture described in this paper meets following requirements: (i) it is generalised, i.e. applicable to different cases without the need to develop ad hoc code; (ii) it does not require financial resources to be acquired.

The implemented software has been used in Malawi, to monitor census field work. Istat also plans to support the census data collection phase in Vietnam, where CSPro will be used as a software application for electronic questionnaires.

The requirements above underpin the sustainability of the solutions that donor countries and their implementing agencies, such as Istat and the US Census Bureau, design to support statistical agencies of developing countries.

# References

1. US Census Bureau, Select Topics in International Censuses, *New Technologies in Census Data Collection, Part 1: Planning for Mobile Data Capture*, October 2016, <https://www.census.gov/library/working-papers.html#.html>
2. CSPro Software, available at: <https://www.census.gov/data/software/cspro.html>