Analysis-Sensitive Conversion of Administrative Data into Statistical Information Systems

Mariagrazia Fugini, Politecnico di Milano, Italy
Mirko Cesarini, Università degli Studi di Milano-Bicocca, Italy
Mario Mezzanzanica, Università degli Studi di Milano-Bicocca, Italy

EXECUTIVE SUMMARY

Public Administrations (PAs) are facing profound institutional and organizational changes requiring managers, stakeholders, and responsible statesmen to increase their ability in quick decision making processes. In this context, information related to individual and collective needs plays a large and relevant role, and therefore a deep knowledge of the observed phenomena and an increased degree of timeliness, quality, and efficiency is required. This article presents a case study concerning the development of a Statistical Information System (SIS) out of data coming from administrative archives of the PAs. Such archives are a rich source of up to date information, but an attempt to use them as sources for statistical analysis reveals errors and incompatibilities among each other that do not permit their usage as a statistical and decision support basis. These errors and incompatibilities are usually undetected during administrative use, since they do not affect their day-by-day use in the PAs; however they need to be fixed before performing any further aggregate analysis. In the sequel of the article, the reader is engaged with the basic aspects involved in building a SIS out of administrative data, such as design of an integration model for different and heterogeneous data sources, improvement of the overall data quality, removal of errors that might impact on the correctness of statistical analysis, design of a data warehouse for statistical analysis, and design of a multidimensional database to develop indicators for decision support. Finally, some examples are presented concerning the information that can be obtained by making use of a SIS constructed out of Registry and Income Office archives.

Keywords: administrative data; data integration; data quality; eGovernment; statistical information systems

ORGANIZATION BACKGROUND

The case study described in this article concerns the development of a SIS out of data coming from administrative archives managed by some PAs, namely the Registry Office of a large municipality and the Income Office of the country (for privacy reasons we refer to the municipality with the M tag and to the country with the C tag). The project described in this case study aims at integrating
the information contents of the archives to provide to the decision makers a set of accurate and up-to-date information describing some characteristics of the population living in $M$.

Here we summarize some descriptive data about the PAs and the statistical research center involved:

- Registry of $M$
  - Type of business: PA services
  - Products/services provided: Registry activities
  - City dimensions: about 1,300,000 resident inhabitants
- Income office of country $C$
  - Type of business: PA services
  - Products/services provided: tax collection, tax related documentation collection
- Statistical Research Center that developed the SIS
  - Type of business: non profit organization
  - Products/services provided: statistical analysis, statistical information system design, and management

A huge amount of changes occurred in the economical and social context in the last year in $C$, and, consequently, traditional and well known static scenarios in PAs have been replaced by more complex and dynamic ones. The PA, as well as the political decision makers, are currently faced with various management and organizational issues, and hence they need accurate and up-to-date information on some specific topics and sectors affected by the decision they are called to make.

**SETTING THE STAGE**

The development of a SIS can provide answers to the information needs, by providing support to analysis, monitoring, and control activities. Still, the only available data in the context of National-level statistics come from National Observatories, and are mainly constructed on the basis of sample surveys. The significance of these data decreases relevantly when small territorial environments are investigated (and this is the case of the $M$ object of our study, with respect to the whole country $C$). On the contrary, PA administrative sources are very good representative data sets of the whole population and contain “fresh” data. Statistical analysis applied to them allows obtaining up-to-date information on the whole population. Moreover, the introduction of ICT has empowered the availability and fruition of administrative databases, making information accessible to organizations and institutions for further surveys (Sundgren, 1996). Some international experiences show that the integrated use of tax-related databases together with Registry databases enables to obtain rich information regarding both income trends and citizens (StatisticsDenmark, 2000). However, this possibility has not been fully exploited yet, because administrative data are often incorrect and unsuitable, in their format and contents, to be used for statistical and decision making. Hence, data need to be cleaned up from errors and pre-processed in several ways before being reversed into statistical databases.

Within an organization, a SIS is loaded and continuously fed using data sources derived from the administrative and management systems. Its role is the management of organizational and administrative data to extract knowledge, and to support domain exploration, verification, and control. An official definition of a SIS is given in UNECE (2000), where a SIS is described as “an informative system aimed at collecting storing, managing, and distributing statistical information.”