

# **A DECLARATIVE APPROACH TO DATA VALIDATION OF STATISTICAL DATA SETS, BASED ON METADATA<sub>1</sub>**

*George A. Petrakos and Gregory E. Farmakis*

*Liaison Systems S.A. – Research and Development of Information Systems  
Academias 77, Athens, 106 78 GR, (www.liaison.gr)*

*E-Mail: George.Petrakos@liaison.gr, Gregory.Farmakis@liaison.gr*

## **Abstract**

Consistent validation of Statistical Data is an essential prerequisite of any process aiming at ensuring the quality and homogeneity of Statistical Information, especially when data are acquired from distributed heterogeneous sources. The ad-hoc, “filter”-like solutions currently used are not sufficient to ensure a systematic and consistent validation process. The present paper provides a framework for the formal definition and classification of validation rules, which in turn is used as the theoretical basis for the definition of abstract data structures, able to hold the rules’ declarations. Thus, the construction of an abstract information model that stores and handles validation rules as metadata becomes feasible. This leads to a declarative formalisation of rules, as opposed to the usual procedural (algorithmic) approach. Furthermore, based on these concepts, the paper describes the implementation of this data model in the form of distributed globally accessible repositories (i.e. databases) of validation rules. Suitable validation engines can then access these repositories to consistently validate data sets, even in ad-hoc cases.

*Keywords:* Data Validation, Metadata, Outliers, Data Quality, Editing, Distributed Metadata Repositories.

*<sub>1</sub> This paper was partially supported by the Statistical Office of the European Communities through the research project SUPCOM 98-lot 23, 98/S 119-76952*