*Symposium on Big Data in Finance, Retails and Commerce:*

*Statistical and Computational Challenges*

Lisbon, 2-3 November 2017, Portugal

**Education's Big Data: Management and Statistical Challenges**

**Luísa Canto e Castro (Speaker)**

*CEAUL, Faculty of Sciences, University of Lisbon, and Directorate-General for Education and Science Statistics, Portugal,* [*luisa.castro@dgeec.mec.pt*](mailto:luisa.castro@dgeec.mec.pt)

**Abstract**

The presentation will report on the challenges the Directorate-General for Education and Science Statistics (DGEEC) is facing to manage the multiple databases of education and training and the efforts it has made to make them, as a whole, an integrated and coherent information system that allows useful analyzes and readings to the political decision, transparency in the publicity to the public and agile procedures in the availability to the researchers. Examples will be given in any of DGEEC's areas of intervention - initial education (pre-school to upper secondary), higher education and science - and the main potentialities of the integration of the different data sources strategy will be explained, namely in the benchmarking of schools, courses, training areas and research areas, in the study of employability and continuation of studies, in the discontinuation of surveys with the consequent reduction of costs and burden on the respondents. In addition, the main impacts of non-matching of records during the cross-database processes will be mentioned, especially in relation to the calculation of statistical indicators where total matching is especially important, such as abandonment rates, courses completion rates in the envisaged time, continuing education rates and employability rates.

**Keywords:** Data Base Management, Raw Data, Data integration, Data exploration, Cross-databases, Non Matching problems.