

*Renza Campagni (Università degli Studi di Firenze)*

*Lorenzo Gabrielli (Institute of the National Research Council of Italy)*

*Lorenzo Gabrielli (Università degli Studi di Firenze)*

*Fosca Giannotti (ISTI-CNR Pisa)*

*Filomena Maggino (Università degli Studi di Firenze)*

## Measuring Wellbeing by extracting Social Indicators from Big Data

Topic 3 – More rapid statistics and indicators on new phenomena

Keywords: Social indicators, Big Data, Clustering analysis

### Introduction

The construction of social indicators is based upon the availability of data collected on purpose (official statistics); it is a common view that constructing social indicators, in the perspective of measuring wellbeing of countries, can benefit from the availability of new sources of data (e.g. big data). One of the big challenges in dealing with new data sources is related to possibility to describe complex human and social phenomena from different perspectives; this introduces new issues to be faced in constructing indicators.

We explore how the classical methodological approach to social indicators construction should be re-considered in light of using data collected not on the purpose of constructing indicators but for other aims (e.g., administrative). In this sense, the individual sales receipts, collected during the period 2007-2014 and made available to our group by a big Italian chain of stores (food but not only), allow us to explore not only a particular social phenomenon but also the methodological implications in dealing with big data.

We extend our previous analysis done by using data mining techniques (e.g., clustering analysis) showing some buying behavior in the period of crisis that suggested us some insights regarding the various categories of products that customers purchased, in order to identify some typical purchase behavior but also to test if and how this information allow other information to be estimated (e.g., structure of the household).

### Methods / Problem statement

We are trying to

- (i) understand what kind of information can be extracted from these data, important and informative in view of constructing social indicators;
- (ii) study different families' behavior in a crucial period, by detecting possible changes in people's lifestyle and eventually the role of the crisis of last recent years in these changes;
- (iii) elaborate and test a model aimed at extracting social indicators from big data. The study is enriched by the possibility not only to observe across different areas the groups behavior (by referring to the territorial distribution of the stores) but also to trace the individual spending behavior over time, while ensuring the anonymity of the sensitive information eventually present in data.

In our previous analysis we identified some groups of customers; one of them corresponds to customers for whom, over the years under investigation, were recorded annual amounts of constant spending, that is, customers in the years of the crisis have not changed their behavior, at least in relation to the amount spent. But analyzing the amount of products purchased by these customers, our analysis highlighted two different sub-groups of customers: the first relating to customers who have purchased a larger number of products (that is, customers who bought lower quality products?), the second for customers who instead bought less products with the same amount spent (prices increased or customers who chose the higher quality products?).

## Results / Proposed solution

In order to answer these questions we deepened the analysis by studying the categories of products that over time these customers have purchased; our work shows that, as starting from receipts of expenditure and from customers' behavior (selection of one product instead of another), it is possible to identify some typical purchase behavior but also to test if and how from this information is possible to estimate other structural information (e.g., dimension and structure of the household).