Enrico di Bella (University of Genoa)

A research project proposal: “Data for qolexity: big data, nowcasting and the construction of wellbeing indicators“

Topic 3 – More rapid statistics and indicators on new phenomena

Keywords: Big data, complexity, nowcasting, social indicators, quality of life

Introduction

Big data are a top subject in international research papers and a vast debate is taking place on their actual capability of being used to complement or even substitute official statistics surveys and social indicators in particular. Connected with big data is “nowcasting“ that is a set of techniques that combine data from multiple sources to provide relevant (mainly economic and monetary) statistics using all the possible information given at a certain time and updating these estimates as long as new information becomes available. The main goal of nowcasting is to provide estimates of these statistics much earlier than the final values are officially released, so that decision making can be made in “real time“ and not ex-post. Despite the intensively growing literature on big data and nowcasting, only a small share of academic papers directly deals with the complexity of social phenomena and in particular with the measurement of the quality of life and its complexity (quolexity).

Methods / Problem statement

A straight analysis of the metadata of the Scopus database of academic papers on big data outlines that most of the existing literature is focused on software and computational issues whilst papers that are specifically focused on statistical issues and on the procedures to build social indicators from big data are a much smaller share of this vast production. Nevertheless the works that focus on these topics show promising results because in developed countries big data seem to be a good information base to create reliable proxies of social indicators, whereas in developing countries their use (for instance using satellite images) may be a viable alternative to traditional surveys. However big data based social indicators deeply suffer of a number of open issues that affect their actual use: they do not correspond to any sampling scheme and they are often representative of particular segments of the population they generally are private process-produced data whose access by national statistical offices is rarely possible while the intrinsic value of the information contained in big data have a social importance that should be shared with the whole community who is the owner of the single "big data record"? big data lack the socio-economic background on which social indicators have been founded and their help to policy makers in their decision process is a fully open point.

Results / Proposed solution

Big data may be a big opportunity for the definition of traditional or new social indicators but their statistical reliability should be further investigated and their availability and use should be internationally coordinated.

Conclusions

All these issues are of primary importance for official statistics and the final proposal of this presentation is to set up a European research group on big data to face not only the technical of software issues but the fundamentals of big data statistics.