The potential of Web2.0 communities for statistics

Donatella Fazio (dofazio@istat.it)[1], Katherine Scrivens (Katherine.SCRIVENS@oecd.org)[2], Maria Grazia Calza(calza@istat.it)[3]

1. INTRODUCTION

The Internet revolution has opened huge new opportunities for the construction and reshape of better statistics. The *liquid data* available in the Net - Open Data, Big Data and crowd sourced data - are being explored with the aim to exploit these new sources of data to integrate and complement the official statistics. The usage of new sources of data brings the National Statistical Institutes (NSIs) and the various stakeholders to face challenges moving from traditional to new ways of data collection and production, considering the opportunities given by Web2.0 tools, such as social networks and on line platforms. It is undeniable that the usage of these new sources drive relevant implications for issues such as validation, metadata, methods and techniques, including IT tools to allow combining user-generated data with the data produced by official statistics. The era of "Statistics 2.0" [1] arises the trade-off between having more and real-time information and the quality of the information produced. The exploitation of new sources of data implies a big investment in producers and users' skills to combine information coming from different sources and brings to consider crucial issues such as privacy aspects, acceptance of data re-use and the management and protection of the data. The European Statistical System (ESS), jointly with relevant stakeholders, together with producers and users of statistics, is increasingly involved in projects and research activities to set up methodologies and procedures to add information and knowledge in a systematic integrated way exploiting new sources of information and data.

Among the multiple initiatives recently carried out at European level, it stands the on-going project Web-COSI *Web Communities for Statistics for Social Innovation*, led by the Italian National Institute of Statistics (Istat) - funded by the EU FP7 ICT Work Programme 2013 - to explore the potential of Web2.0 communities focusing on statistics beyond GDP.

2. WEB-COSI PROJECT

Web-COSI was designed to respond to the call launched by the Collective Awareness Platforms for Sustainability and Social Innovation (CAPS), based on the idea that collaborate through crowd sourced platforms can produce solutions for a wide range of social needs [2]. Web-COSI (www.webcosi.eu), is a co-ordination action (2014-2015) with the general objective to foster the engagement of citizens and society at large in the area of new measures of societal progress and well-being using the opportunities given by Web 2.0. Specific objective is to implement tools for collecting, producing and visualizing information and data towards a better integration of official and non-official statistics. The release of a Wiki of progress statistics, at mid-term of the project, is envisaged with the aim to foster the use of locally generated data to bridge top-down and bottom-up approaches. Web-COSI capitalizes on the last 15 years' experience characterized by two big revolutions for the world of research and society. First, the

---

"GDP & Beyond" debate that has dominated the scene of statistical and economic research. The ESS [3], together with OECD [4] and other relevant actors, has been deeply involved to outline a strategy to develop statistical information to meet the Stiglitz's Commission recommendations [5]. Second, the Internet explosion that has radically changed the way in which information is produced and shared. Interactivity is contributing to change the roles of producers and users of data, increasing awareness and bringing to consider a bottom-up approach for the construction of statistical information.

Web-COSI is based on a consortium that sees the collaboration among two relevant Institutions (Istat and OECD), a civil society organisations (www.lunaria.org) and a social entrepreneurs' community (www.i-genius.org), representing society at large. The consortium is well-balanced and multidisciplinary creating synergies for the integration of the different approaches.

Specifically, Web-COSI work plan aims at: a) mapping and distilling best practices of existing digital initiatives for communities' involvement - a specific survey is envisaged to take stock of Web2.0 initiatives carried out or planned by NSIs; b) create a critical mass through: target campaigns, data visualization competitions, setup of a European Wikiprogress University Programme; c) facilitate the communities' access to statistics empowering the collection of civil society grass root locally generated data with the development of a Wiki of progress statistics. Moreover, various open events are organised to involve the greatest number of audience: 5 workshops, 4 focus groups and a final conference.

3. RESULTS SO FAR

To map and distil the best practices of existing digital initiatives for communities' involvement Web-COSI has carried out in 2014: i) two on-line discussions organized by OECD– "*Engaging citizens in well-being and progress statistics*" and "*Making data more accessible for society at large*"; one webinar managed by Lunaria - "*Civil society engagement in well-being statistics: good practices from Italy*"; iii) a workshop on *Using Technology to Engage Citizens with Well-being Statistics in the Perspectives from Civil Society* held at OECD. These activities have involved a large number of participants from different sectors of society, including NSIs, government, research organizations, social enterprise and civil society, generating an impressive debate on the potential of Web2.0 communities. The discussions helped to identify an initial map of the different types of initiatives set up, using collective platforms, to engage citizens with well-being and societal progress statistics. The different types of initiatives can be grouped as follows: (i) public consultation; (ii) communication; (iii) citizen-generated data, (iv) open data.

3.1. Public consultation

Consulting with the public is now widely seen as an essential step in elaboration of indicators of well-being and progress [6]. While public consultation is the type of citizen engagement most closely associated with the new measures of well-being and progress, it is an area where the potential of interactive technology is still to be fully realised. For many in the discussion, especially those working on smaller-scale community projects, citizen engagement through consultation is more suited to face-to-face events such as focus groups or community meetings, while online methods were seen as more appropriate for the collection and communication of the data. However, face-to-face events are necessary limited in terms of representativeness, which is an especially important problem for well-being measurement projects at the national level. The UK's 'Measuring National Well-being' programme, managed by the Office for National Statistics, Italy's 'Equitable and Sustainable Well-being' (BES) project, led by Istat and 'Measures of Australia's Progress' (MAP), run by the Australian Bureau of Statistics, all

used a mix of offline events and surveys, with online consultation tools such as online surveys and social media to reach as wide an audience as possible.

3.2. Communication

Finding innovative ways to communicate the underlying meaning of data by telling a story around the data (or by enabling users to play with the data and find their own stories) is a powerful way of making statistics more accessible to a broader audience. This can be done by the data producers themselves (such as government or statistical agencies) or by intermediaries such as data journalists, civil society organisations or anyone with an interest in finding the best way to communicate the key messages of datasets. Stories can be told in the traditional way, through narrative text, or they can be conveyed in a more visual manner - through infographics and charts that organise the data in such a way that the meaning is immediately apparent. Data visualisations can be very appealing, but their importance goes beyond aesthetics: they provide a unique means of highlighting new patterns in statistics and looking at the world in a different way. The mapping exercise highlighted many innovative examples of visualisation, particularly from civil society and international organisations. NSIs also recognise the need to make their data more accessible through visualisation and most of them have made some provision of interactive data content on their websites, be it in the form of tables, charts, maps or dashboards. However, the quality and amount of data made available in this way is extremely variable [7].

3.3. Citizen-generated data

Digital technologies allow members of the public to participating themselves as data producers and the prevalence of accessible yet sophisticated mapping technology through mobile platforms provides a means to crowd-source data from members of the public at minimal cost. Geographic Information Systems (GIS) allow for users to provide data in the form of Tweets, reports, photos, comments, or other types of Volunteered Geographic Information (VGI), that allow for the monitoring of outcomes related to well-being in close to real time. There are a number of different ways that platforms for citizen-generated data can function, including public reporting (e.g. of problems in their local area), monitoring of social media (to gauge public reaction to events and policies), and through the use of citizen scientists (e.g. individuals with specialised skills, hobbies, or interests can be recruited to act as data sensors to help populate scientific research databases). Given the limitations in coverage and timeliness of official statistics in many developing countries, encouraging the development and use of new forms of data collection has been seen as a core element of the data revolution needed to monitor progress towards the Sustainable Development Goals [8]. However, while citizen-generated data have a lot of potential for providing useful information and filling data gaps, they also have significant limitations when compared to official statistics related to self-selecting samples, reliability of data and comparability between areas.

3.4. Open data

For data to be truly open, not only must it be freely available online, but it should also be presented in a format that maximises its potential for re-use, with semantically tagged information, open formats, and fully downloadable information, the latter through APIs (Application Programming Interfaces), machine-readable data structures and rich metadata. For many organisations, to whom open data is a new concept, this is likely to be a gradual process, requiring significant resources. The role of advocacy organisations

will be important in order to educate government and civil society of the need to engage more with citizens through open data.

4.   CONCLUSIONS

So far Web-COSI activities have pointed out that Web2.0 technologies are exploited – at different extent - by NSIs, government, research organizations, social enterprise and civil society to foster the interaction between data producers and data users of statistics. The discussion has identified that the usage of crowd sourced data to complement and integrate official statistics, is an opportunity to evaluate in multiple terms: having data close to real time; narrowing the distance between what official statistics say and what people perceive; using new data not included in the official surveys; optimising the costs. The usage of crowd sourced data has to be considered at different levels giving them different weights: local; national; international; global. The discussion pointed out the necessity of organizing the progress statistics (generated by official and non-official data) in an integrated framework to represent a "real tool" to make statistics accessible and understandable.

Web-COSI will continue its work to conclude its activities by the end of 2015. Next steps envisage: i) the conduction of a specific survey addressed to NSIs on Web2.0 initiatives carried out and/or planned to empower statistics using new sourced of data; ii) the organization of target citizens campaigns, data visualisation competitions, youth initiatives, and the set-up of a European Wikiprogress University Programme; iii) the organisation of 3 workshops, 4 focus groups of social entrepreneurs and a final conference. At last but not at least, the development of a new data sharing portal - Wiki of progress stat on www.wikiprogress.org - designed to be a key reference for progress and wellbeing data and statistical resources such as reports, visualisations and interactive tools, able to allow external data providers to upload their own data.

Web-COSI experience is demonstrating that the integration of traditional official statistics with new sources of data is an inexorable process which requires new skills, culture and a radical change of mind set. In Web 2.0 era as the power of online communities grows ever stronger institutions of diverse type and scope cannot ignore their centrality for the "definition" of better statistics, for better policies, for a better quality of life.

REFERENCES

[1] Giovannini E. (2010). "Statistica 2.0: the next level", Introductory speech at the 10th National Conference of Statistics, Rome. Available at: http://www3.istat.it/dati/catalogo/20120621_00/atti_decima_conferenza_nazionale_statistica.pdf

[2] Sestini F. (2012). Collective Awareness Platforms: Engines for Sustainability and Ethics. IEEE Technol. Soc. Mag. 31(4) pp. 54-62 Available at: http://caps2020.eu/wp-content/uploads/2013/11/CollectiveAwarenessPlatformsEngineforSustainabilityandEthics-1.pdf

[3] ESS (2011). Final Report on the Sponsorship Group on Measuring Progress, Well-being and Sustainable Development. Available at : http://epp.eurostat.ec.europa.eu/portal/page/portal/pgp_ess/0_DOCS/estat/SpG_Final_report_Progress_wellbeing_and_sustainable_deve.pdf

[4] OECD (2012). Better Life Initiative: Measuring Well-Being and Progress. Available at: http://www.oecd.org/statistics/betterlifeinitiativemeasuringwell-beingandprogress.htm

[5]   Stiglitz et al. (2009). Final report by the Commission on the Measurement of Economic Performance and Social Progress. Available at: http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf.

[6]   Hall, J. and L. Rickard (2013). People, Progress and Participation: How Initiatives Measuring Social Progress Yield Benefits Beyond Metrics, Global Choices 1: 2013, http://www.bertelsmann-stiftung.de/cps/rde/xbcr/SID-CFCC0351-B1427148/bst_engl/xcms_bst_dms_37947_37948_2.pdf

[7]   Leib, S. and Z. Hartland (2013). ICT Delivering Tools: Catalogue of user tools and discussion relating to implementation, e-Frame Deliverable 8.1, www.eframeproject.eu

[8]   The Engine Room (2014). CIVICUS Support to People-powered Accountability and the Data Revolution: A scoping Study by the Engine Room, https://www.theengineroom.org/wp-content/uploads/CIVICUS.Scoping.Study.WEB.pdf