

# New data sources to indicate levels of active citizenship

Cristina Rosales Sánchez<sup>1</sup>, Massimo Craglia<sup>2</sup>, Arnold Bregt<sup>3</sup>

<sup>1</sup> *Joint Research Centre, Ispra, Italy; cristina.rosales-sanchez@jrc.ec.europa.eu*

<sup>2</sup> *Joint Research Centre, Ispra, Italy; massimo.craglia@jrc.ec.europa.eu*

<sup>3</sup> *Wageningen University, Wageningen, The Netherlands; arnold.bregt@wur.nl*

## Abstract

Data availability is a general constraint in the generation of indicators for decision-making processes. Web 2.0 technologies offer new potentials of data sources (also considered as big data sources) that needs to be investigated. In our research we focus on indicators of active citizenship as a relevant dimension of social capital that needs to be measured and monitored. Citizens' participation empowers communities, increasing social capital and reducing possibilities of social exclusion, thus beyond-GDP initiatives consider citizen engagement an indicator of societies' quality-of-life.

Classical measurements of active citizenship are mainly based on surveys, which are costly and time consuming, and on voter turnout, which temporal availability depends on election periodicity.

Considering that the information publicly available by online social networks may offer a sense of how much citizens are engaged in their communities, our research analyses the potential of the microblogging online site Twitter as unofficial data source to measure active citizenship. As a result, we define a method to use Twitter data to obtain indications of active citizenship that complement the results from official data sources offering a more timely, less costly information, and with higher spatial and temporal resolution.

**Keywords:** Twitter, active citizenship, social media, indicators, data mining

## 1. Web2.0: new data sources, challenges and opportunities

### 1.1. New data sources for official statistics.

Recent developments in ICTs and widespread Internet access have changed the way individuals communicate, empowering citizens with fast and easy connections. In addition, the possibility of accessing Internet via mobile devices, allows instantaneous communication, while generating real-time geolocated data (Boulos et al. 2011; Elwood et al. 2012; Goodchild

& Glennon 2010). This evolution of the Internet towards a more participatory web is conceptualized by the term Web 2.0 (Murugesan 2007; O'Reilly 2007). The so-called Web 2.0 technologies enable citizens collect the perceptions of their surroundings, and collectively share the information in their community, in the form of text, pictures, videos or audios. Thus, citizens play a new role acting as sensors and analysts of their environments (Goodchild 2007).

The potential of this vast amount of data generated by the citizens and offered publicly online has attracted the attention in wide range of domains, also including international statistical organizations (Struijs et al. 2014). The United Nations Global Working Group on Big data for official statistics, the High-Level Group for the Modernization of Statistical Production and Services, from UNECE , and the Scheveningen Memorandum (DGINS 2013) adopted by the European Statistical System Committee, are examples of initiatives promoting research on exploitation of new data sources for official statistics. Similarly, the European Framework for Measuring Progress (the “e-Frame Project”) expresses this need for new data sources and suggests the use of non-official data to measure progress in a more comprehensive way.

Despite interest in the potential of new data sources, scholars express the need of further research for a more effective use of the user-generated content (UGC) data. According to the global assessment on the use of big data for official statistics developed by UN Statistics Division (United Nations 2014), major challenges remain, particularly related to methodological issues such as data quality, representativeness, volatility and dimensionality. The study pointed to technology-related limitations as well, as the singular characteristics of big data imply that new methods of data processing and analysis are required.

### *1.1. Active citizenship: indicator of quality of life as a case study.*

Given the capacities that Web2.0 technologies offer to citizens to express their opinion, exchange information and support initiatives of their interest, in our research we explored citizens' digital trace to offer a measure of how engaged they are within their communities.

According to the most recent Eurobarometer on Media Use (DG COMM European Commission 2014), the number of social network users in Europe keeps growing. Indeed, almost half of Europeans (47%) uses social networks frequently (at least once a week). The Eurobarometer report furthermore indicates that majorities in all EU member states consider social networks a way to keep abreast of and get people interested in political affairs, and to “have your say”.

Particularly, the Twitter platform facilitates a new online political sphere where users can publicly discuss on political issues, attracting politicians, journalists, political strategists and citizens (Ausserhofer & Maireder 2013). This social networking service allows instant communication by means of short messages with maximum length of 140 characters. These messages, known as “tweets” enable linking messages to users (@-mentions), with internet content (URL links) and topics (#hashtags). Since the service was launched in 2006, the Twitter’s speed and ease of communication have been catching followers, reaching 316 million active users that produce around 500 million tweets per day, according to the company. The relevant role of Twitter in political debate, in recent uprisings, such as the Arab Spring, or mass movements as “indignados” in Spain, suggests that Twitter activity may provide indications of citizens’ engagement in their community.

## **2. Measuring active citizenship**

Active citizenship, social cohesion and community building are key factors of social development widely recognized as in the Europe’s Lisbon Strategy (European Council 2000), and the principles for social development of the World Bank. Civic engagement influences on social capital, increasing social trust, providing opportunities for cooperation between citizens and discouraging anti-social conduct (Putnam 2000). Actually, even recent alternative initiatives to measure quality-of-life integrate civic engagement as one relevant dimension of social well-being. These new initiatives for quality-of-life, also known as Beyond-GDP initiatives (Stiglitz et al. 2009) propose alternative ways to understand citizens’ well-being not only considering economic aspects, as it is the case of the GDP approach, but also including

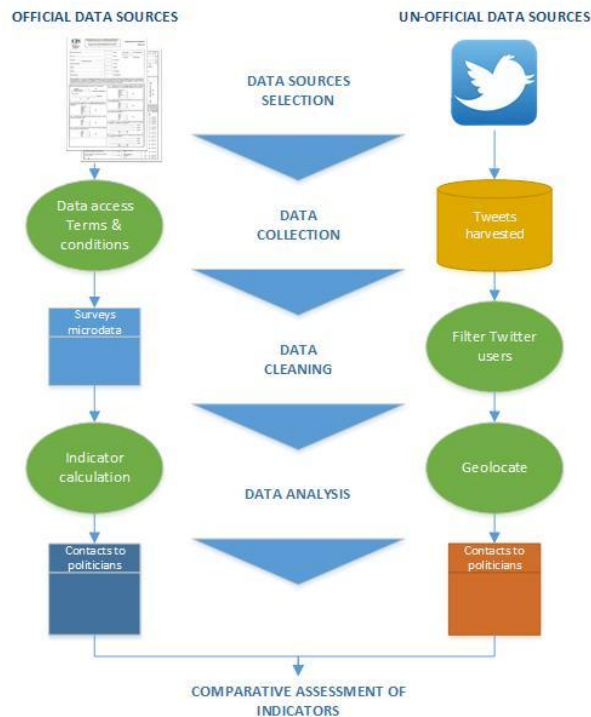
environmental and social aspects of the living environments (European Statistical System Committee 2011; Scrivens & Smith 2013). This recognition of the quality-of-life multidimensionality, however, poses substantial data challenges to official statistical organizations, as traditional data sources are mainly based on statistics from censuses and individual surveys, which are costly and time-consuming.

In the same way as other quality-of-life indicators, data availability constraints are a challenging obstacle to measure active citizenship. Official statistics, such as Eurostat's (2007) "Urban Audit" programme, classically use voter turnout and number of eligible political representatives as proxies for active citizenship. These have the advantage of being periodically available and internationally comparable, but they neglect many other facets of citizen participation.

Instead, the framework for our research is the Active Citizenship Composite Indicator (ACCI) developed by the Centre for Research on Lifelong Learning (CRELL), based on a holistic perspective of the term (Hoskins & Mascherini 2009). The ACCI incorporates four dimensions: participation in political life, participation in civil society, participation in community life and values supportive of active citizenship. Again, the ACCI measurement is hampered by data accessibility, limited due to insufficient spatial and temporal data coverage.

Under the consideration that digital trace data from social networks could offer indications of active citizenship, we aim at assessing the potential of Twitter data for measuring active citizenship in urban areas, in contrast to official measurements of the indicator. Accordingly, we developed research focusing on one variable included in the ACCI, i.e. "contacts to politicians to express an opinion". Official data sources for this indicator are traditionally based on survey data, in which questionnaires ask individuals if they contacted a politician during the past 12 months.

The method was firstly developed in 2015 using Spain as case study<sup>1</sup>. Three cities of different sizes were selected, to also explore the effect of city size in the analysis. Figure 1 shows the steps of the method proposed in our research, from data collection to data cleaning and analysis.



### 2.1. Official data source: specifications and applied methodology

The official data source used were the periodic public opinion “barometers” of the Centre for Sociological Research (CIS, from Spanish acronym), Spain’s most comprehensive data source on indicators of citizen engagement. Then, to obtain values of the indicator on a local level, we carried out the following steps:

- Data collection. The anonymized microdata from CIS barometers are freely available from the CIS website ([www.analisis.cis.es/cisdb.jsp](http://www.analisis.cis.es/cisdb.jsp)), after registration and agreement of specific terms and conditions for data protection.
- Data preparation. We extract the data from the anonymized microdata that is related to the selected cities for the study.
- Data analysis. We can then calculate the percentage of respondents who indicated having contacted a politician in a certain location as follows:

---

<sup>1</sup> Extended information on the research, method and results of the study are included in a scientific paper recently under review.

$$\text{Contacted a politician (city) (\%)} = \frac{\text{nr of interviewees that contacted a politician (city)} * 100}{\text{nr of interviewees (city)}} \quad (1)$$

## 2.2. Unofficial data source: specifications and applied methodology

Twitter users can post messages to specific users by including “mentions” in their tweets. Our method considers that tweets generated by citizens in a certain region and containing mentions of politicians and municipal representatives in that region, can be considered as direct contacts to politicians to express an opinion.

With that aim, the method to obtain a measure of active citizenship out of Twitter data, includes:

- Data collection. Tweets including mentions to selected politicians and municipal representatives were real-time harvested by using the public streaming Twitter API.
- Data processing. The method includes two steps: Twitter users’ categorization, to filter only those tweets generated by individual citizens from the selected cities; and tweets’ geographic location, geoparsing techniques are included to increase the number of tweets with geolocation information.
- Data analysis. Then, the results from the data processing are extrapolated to offer a yearly-based measurement. In such a way, we can assess the potential of Twitter data to provide an indicator of active citizenship, in comparison with the results from official sources.

## 3. Some conclusions from the comparison between official and unofficial data sources

In general, each data source offers different advantages and challenges, which complementary could offer a better understanding on active citizenship for a certain region.

Comparing the Twitter results with the survey results, Twitter data indicated a larger number of contacts for the three Spanish cities studied. However, the values were similar in magnitude

to those obtained from the official data source. The largest differences were found for the city with the smallest population size, where the data from official surveys was less representative due to the small number of observations.

Twitter as a data source presents unknown biases, in relation to population representativeness, data truthfulness and Twitter's streaming API representativeness. Though far from providing an accurate absolute measure, the Twitter data did offer a creditable indicator of active citizenship, independent of the scale and the population density of the area targeted. Additionally, the use of Twitter data presents benefits in the temporal domain, allowing analysis in almost real-time, offering a higher level of temporal resolution. Other benefits from Twitter data are related to the easy data accessibility and the possibilities of adding content-based information to the analysis.

As such, our research found that results from Twitter data present favorable opportunity to obtain indicators on active citizenship, complementing the information obtained from the surveys, more importantly in cities with low population density which are likely underrepresented in the surveys. Further analyses need to be done addressing the scalability of the method, and determining the influence of location and temporal aspects.

#### **4. References**

Ausserhofer, J. & Maireder, A., 2013. National Politics on Twitter. *Information, Communication & Society*, 16(3), pp.291–314. Available at:  
<http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.756050>.

Boulos, K. et al., 2011. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *International journal of health geographics*, 10, p.67. Available at:  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3271966&tool=pmcentrez&rendertype=abstract>.

DG COMM European Commission, 2014. *Standard Eurobarometer 82 “Media use in the European Union,”* Available at:  
[http://ec.europa.eu/COMMFrontOffice/PublicOpinion/index.cfm/Survey/getSurveyDetail/search/MEDIA USE/surveyKy/2041](http://ec.europa.eu/COMMFrontOffice/PublicOpinion/index.cfm/Survey/getSurveyDetail/search/MEDIA%20USE/surveyKy/2041).

DGINS, 2013. *Scheveningen Memorandum - Big Data and Official Statistics*, Available at:  
<http://ec.europa.eu/eurostat/documents/42577/43315/Scheveningen-memorandum-27-09-13>.

Elwood, S., Goodchild, M.F. & Sui, D.Z., 2012. Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*, 102(3), pp.571–590. Available at:  
<http://dx.doi.org/10.1080/00045608.2011.595657> [Accessed July 14, 2014].

European Council, 2000. *Lisbon Strategy. Presidency Conclusions of the Lisbon European Council, 23-24 March 2000*, Available at:  
[http://www.consilium.europa.eu/uedocs/cms\\_data/docs/pressdata/en/ec/00100-r1.en0.htm](http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/00100-r1.en0.htm).

European Statistical System Committee, 2011. *Sponsorship Group on Measuring Progress , Well-being and Sustainable Development - Final Report*, Available at:  
[http://epp.eurostat.ec.europa.eu/portal/page/portal/pgp\\_ess/about\\_ess/measuring\\_progress](http://epp.eurostat.ec.europa.eu/portal/page/portal/pgp_ess/about_ess/measuring_progress)

Eurostat, 2007. *Urban Audit Reference Guide - Data 2003-2004*, Available at:  
[http://epp.eurostat.ec.europa.eu/portal/page/portal/product\\_details/publication?p\\_product\\_code=KS-RA-07-016](http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-07-016) [Accessed August 8, 2014].

Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), pp.211–221. Available at: <http://www.springerlink.com/index/10.1007/s10708-007-9111-y> [Accessed March 4, 2013].



- Goodchild, M.F. & Glennon, J.A., 2010. Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, 3(3), pp.231–241. Available at: <http://www.tandfonline.com/doi/abs/10.1080/17538941003759255> [Accessed May 21, 2013].
- Hoskins, B.L. & Mascherini, M., 2009. Measuring Active Citizenship through the Development of a Composite Indicator. *Social Indicators Research*, 90(3), pp.459–488. Available at: <http://link.springer.com/10.1007/s11205-008-9271-2>.
- Murugesan, S., 2007. Understanding Web 2.0. *IT Professional*, 9(4), pp.34–41.
- O'Reilly, T., 2007. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, (65), pp.17–37.
- Putnam, R.D., 2000. *Bowling Alone: The Collapse and Revival of American Community*, Simon & Schuster. Available at: <http://books.google.it/books?id=rd2ibodep7UC>.
- Scrivens, K. (OECD) & Smith, C. (OECD), 2013. *Four Interpretations of Social Capital: An Agenda for Measurement*, Available at: <http://www.oecd-ilibrary.org/docserver/download/5jzbcx010wmt.pdf?expires=1389718107&id=id&accname=guest&checksum=82B9AD334FDF7E54F24260A696B46623>.
- Stiglitz, J., Sen, A. & Fitoussi, J., 2009. *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](http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf) [Accessed February 25, 2014].
- Struijs, P., Braaksma, B. & Daas, P.J.H., 2014. Official statistics and Big Data. *Big Data & Society*, 1(1). Available at: <http://bds.sagepub.com/content/1/1/2053951714538417.abstract>.
- United Nations, 2014. *Big data and modernization of statistical systems*, Available at: <http://unstats.un.org/unsd/statcom/doc14/2014-11-BigData-E.pdf>.

