

New data sources to measure social variables: Twitter to contact politicians

27 – Big Data & Web Scraping

3rd June 2016

Cristina Rosales^{1,2}, Max Craglia¹, Arnold Bregt²

¹ Digital Earth and Data Reference Unit – IES -

Joint Research Centre (EC), ² Wageningen University

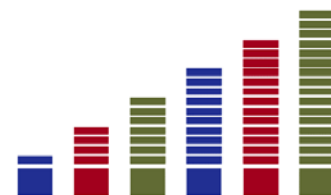
cristina.rosales-sanchez@jrc.ec.europa.eu

New data sources to measure social variables: Twitter to contact politicians

Introduction

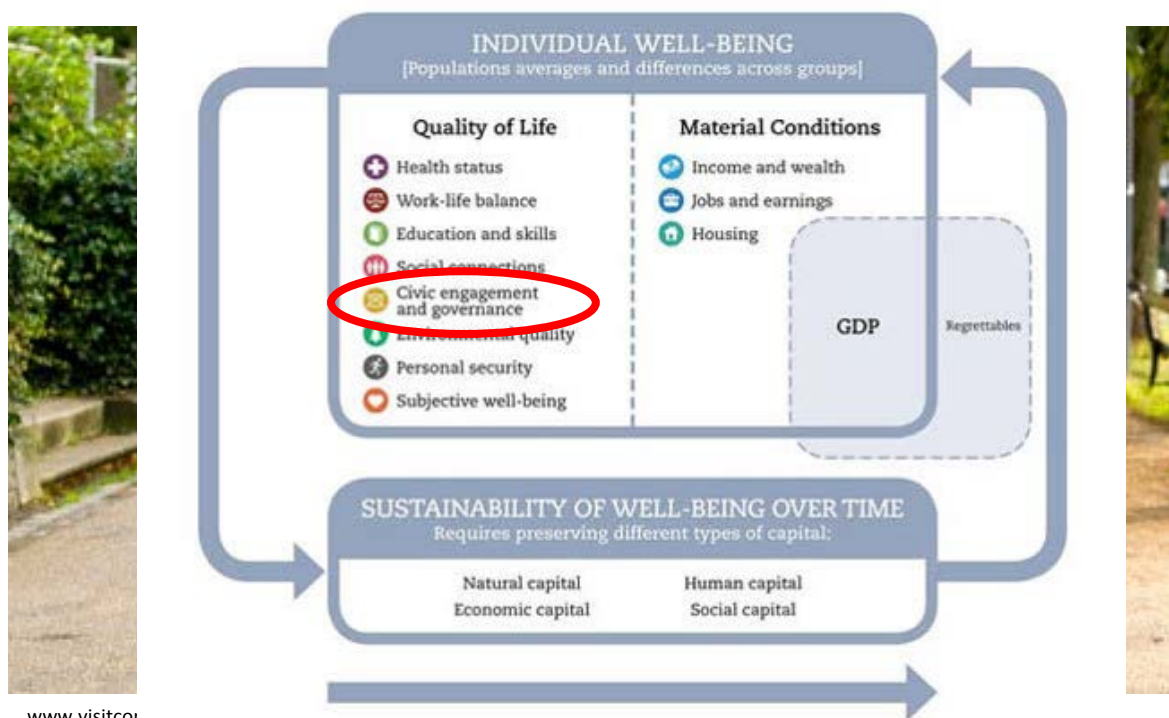


"The Data Deluge", Brett Ryder, The Economist (Feb 2010)



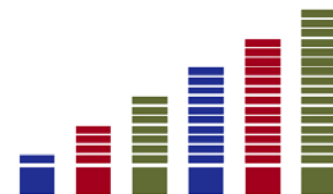
New data sources to measure social variables: Twitter to contact politicians

Introduction



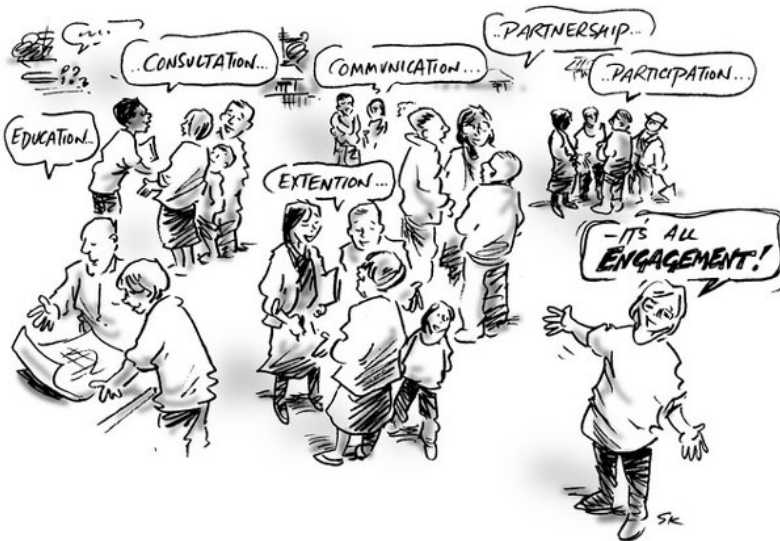
www.visitco...

OECD, 2013



New data sources to measure social variables: Twitter to contact politicians

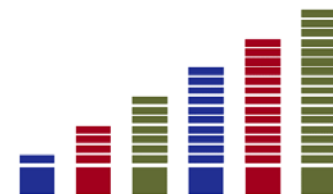
Introduction



www.dse.vic.gov.au › Effective Engagement › Introduction to Engagement

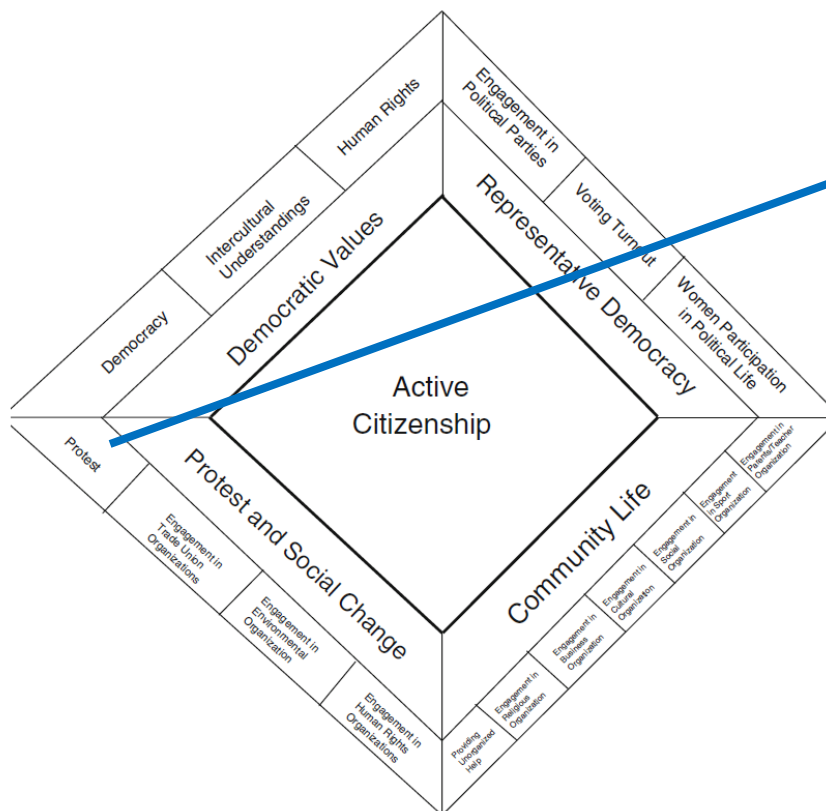


<https://www.aegge.org/yvote2014/voting-guide/>



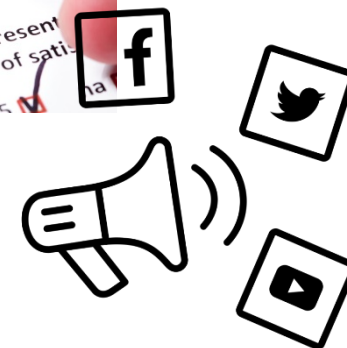
New data sources to measure social variables: Twitter to contact politicians

Introduction



number of citizens who contacted
a politician or governmental representative
to express an opinion

of 1 to 5 where 1 represents
and you rate your level of satisfaction
3 ☐ 4 ☒ 5 ☐ na ☐
ale of 1 to 5 where 1 represents
ould you rate your level of satisfi
2 ☐ 3 ☐ 4 ☐ 5 ☒ na ☐
a scale of 1 to 5 where 1 represent
you rate your level of satisfi
5 ☒ na ☐



Measuring Active Citizenship through the Development
of a Composite Indicator

Bryony L. Hoskins · Massimiliano Mascherini

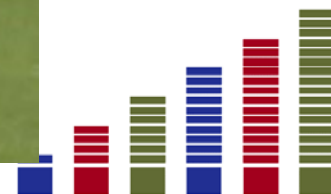


New data sources to measure social variables: Twitter to contact politicians

Research Objectives



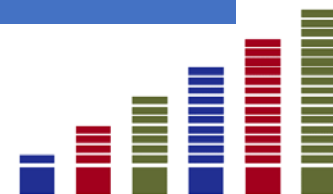
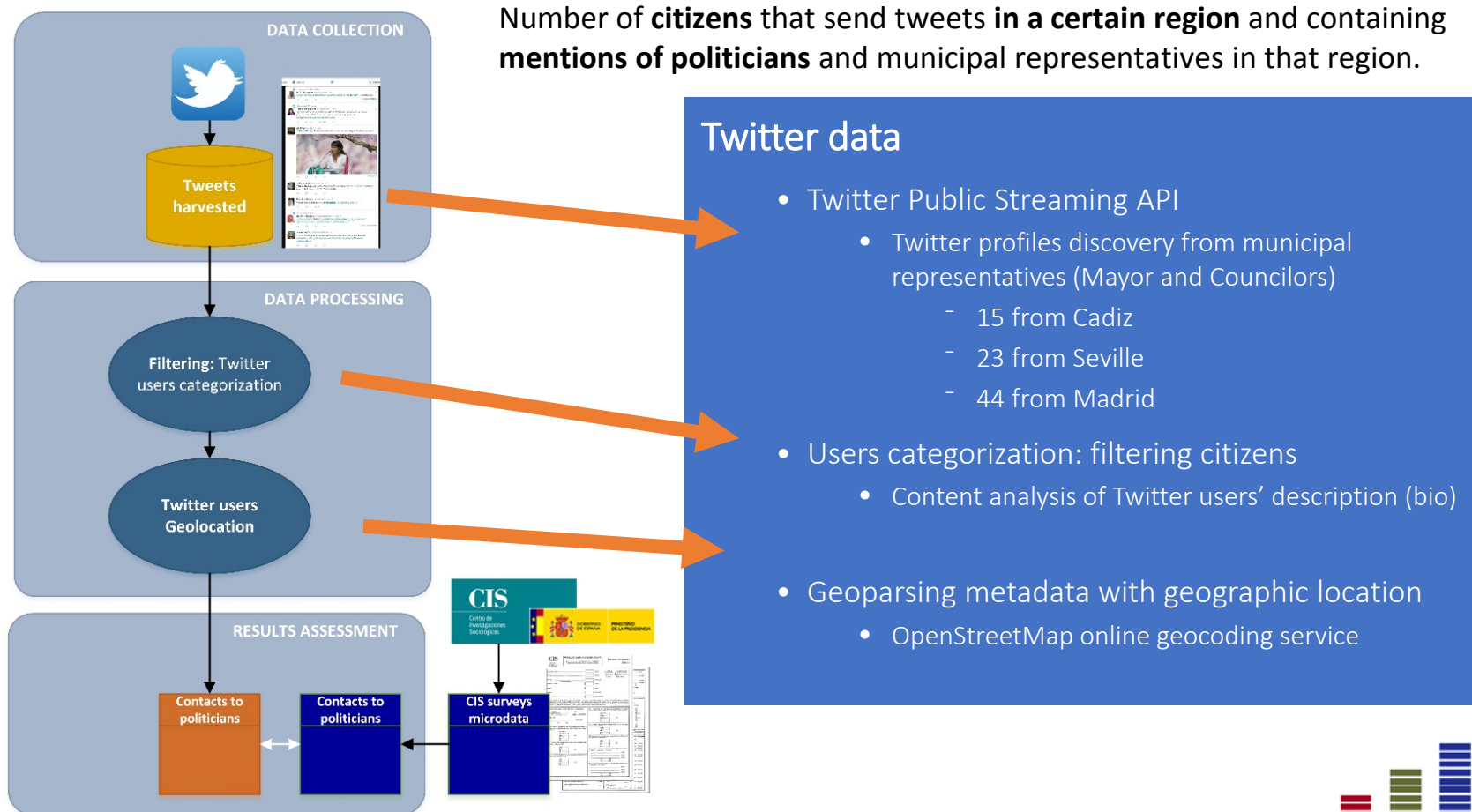
- to explore the potential of new data sources to complement traditional methods of measuring active citizenship
- to design a method for extracting data from Twitter on contacts with politicians





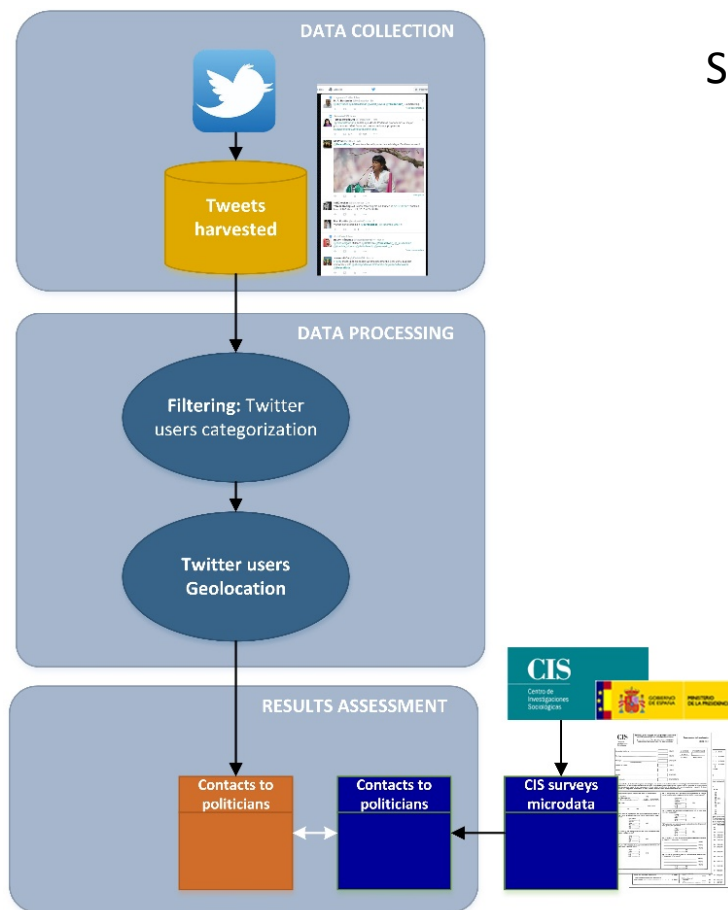
New data sources to measure social variables: Twitter to contact politicians

Method



New data sources to measure social variables: Twitter to contact politicians

Method



Survey data:

- Selection of microdata from Madrid, Seville and Cadiz
- Inferences from sample:

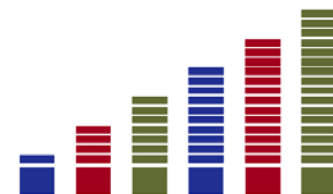
Contacts to politicians (city) = $\text{Contacts} * \text{Inhab} / \text{Interviewees}$

Where:

Contacts = # interviewees that contacted a politician (city)

Interviewees = # interviewees (city)

Inhab = # inhabitants (city)

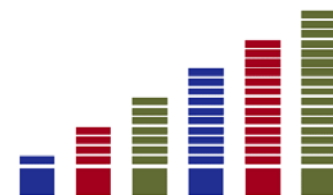
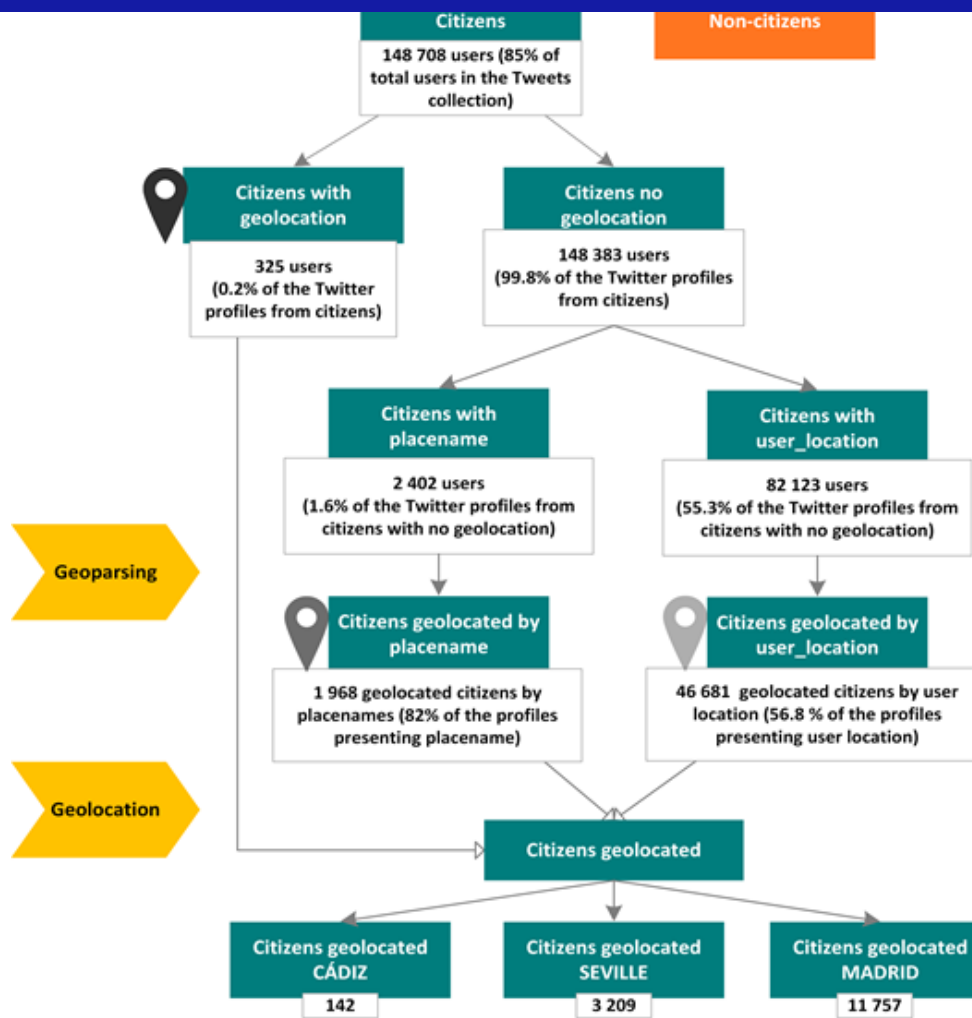


New data sources to measure social variables: Twitter to contact politicians

Filtering Users



Results





New data sources to measure social variables: Twitter to contact politicians

Conclusions

EVALUATION CRITERIA	OFFICIAL DATA SOURCE	UNOFFICIAL DATA SOURCE
	(CIS survey)	(Twitter)
Accuracy	High	Very low
Scope and coverage	Defined target universe	Demographics of Twitter users are unknown
Spatial dimension	Country level	Municipal level
Timeliness	Yearly	Almost real-time
Cost	High	Very low
Flexibility	Fixed questionnaire	Flexible method
Method feasibility	Data is accesible after user registration and signature of stricts terms and conditions	Twitter data is accessible by Oauth authentication, and depends on continuous electricity supply, adequate internet connection and storage capacity

The developed method:

needs to address scalability, and determine the impact of the number of politicians considered, and the influence of location and temporal aspects.

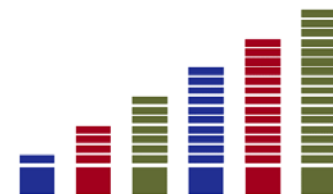
Each data source presents its own strengths and weakness -> best results may be achieved by the combination of both.

Results from Twitter data:

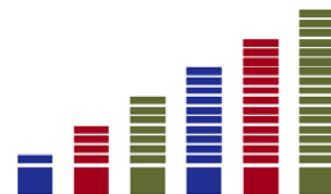
- offers more timely and less costly information, with higher spatial and temporal resolution
- are independent of the scale and the population density of the area targeted
- provides added insights on the main concerns expressed by users over time
- BUT, presents several biases

Data from official surveys:

- are statistically robust and representative of a total population, usually just at a national/regional level
- BUT, are costly and time-consuming



Thanks for your attention!



Stay in touch

Cristina Rosales Sánchez
Digital Earth and Reference Data Unit
Joint Research Centre - European Commission
cristina.rosales-sanchez@jrc.ec.europa.eu



JRC Digital Earth: <https://ec.europa.eu/jrc/en/research-topic/digital-earth>



LinkedIn: [Cristina Rosales Sanchez](#)

