

UN Global Working Group on Big Data for Official Statistics Briefing Social Media Data Task Team

1. Achievements since the last GWG meeting

The Social Media Task Team was formed leveraging the knowledge, skills and experience of institutions like Statistics Netherland, Global Pulse, IRENA, INEGI, DANE, ITU, EUROSTAT and UNSD, in social media POCs and projects, using Twitter, Facebook, web-scraping from web sites, news online search and Wikipedia for statistical purposes, to determine for example CPI, consumer confidence, sentiment analysis, text analysis, perception of statistical offices, or to detect terrorist plans & activities, mobility and tourism.

The main deliverable of the Task Team is going to be a handbook that explains how to use social media data with Big Data techniques for producing official statistics.

The structure of the handbook is based on the UNSD guidelines for Task Teams. Each chapter has a leader institution that has volunteered on it. The leader is responsible of defining the layout of the chapter, creating teams to work on, organizing work sessions and ensuring coherence and consistency among the different chapters.

The content of the handbook will include:

Chapter 1.- **Introduction** that describe the importance of big data to NSOs, specifically Social Media Data and how does it fit into the re-engineering and modernization of how data is being collected. Key benefits and challenges of using Social Media Data, opportunities and risks, privacy issues, etc.

Chapter 2.- **Data sources** and how to get started with Big Data. It describes some of the most popular Social Media Networks, and a proposed method to collect and store data from a social media network, with step by step examples.

Chapter 3.- **Methodology for analysis.** With an introduction, Social media methods and techniques including Natural Language Processing (NLP) and Extracting information from non-message part of social media, with Technical aspects and computational resources as well as with visualization methods. And finally in this chapter, analytical approaches illustrated with examples.

Chapter 4.- **Applications.** Description of some pilot projects.

Chapter 5.- **Concluding Remarks** and further work to be done

2. Topics being presented at the Conference

The task team will make a presentation with the title “Sandbox Project On Social Media Data for Sentiment Analysis/Mobility”. The presentation will address topics like: The Sandbox environment, the access to the Twitter API, ICT tools used to make social media analysis, the use of tweets to make analysis of sentiment and mobility.

3. Future work plan for the Task Team

To finish a draft of the document for this year, circulate it, receive feedback and consolidate a final version.

4. Acknowledgements

The Task Team chair, Dr Juan Muñoz Lopez from INEGI Mexico, would like to thank all the members of the Task Team for their collaboration and contributions, and specially recognizes those members who have volunteered to lead each chapter, including: Ronald Jansen, Markie Muryawan and Nino Magallanes from UNSD for chapter 1; Silvia Fraustro and Abel Coronado from INEGI as well as Syed Tajammul Hussain from IRENA for chapter 2; Piet Daas, Ali Hürriyetoglu and

Marco Puts from Statistics Netherlands for chapter 3; Orlando Saavedra from DANE Colombia for chapter 4.