ESAC workshop on

Non-traditional (big) data sources and data science for official statistics

New methods, challenges for data quality, data governance and participation models

Organisers: ESAC, EUROSTAT, ECB, STATEC

Programme

Wednesday, 20 October 2021
Virtual event - Link to registration

Purpose

The workshop will explore challenges and opportunities relating to diverse non-traditional data sources for official statistics, including crowd-sourced, remotely sensed and big data for [Official Statistics](#).

The event aims to spotlight the topic of digital transformation in the context of the Conference on the Future of Europe. Data form an important part of the EU's digital transformation. Big data refers to collected datasets, which are so large and complex that they require whole new approaches, such as innovative statistical methods, artificial intelligence and data science methods that straddle statistics and computer science. These techniques generally collect data automatically and use new methodologies to analyse data and discover patterns for knowledge extraction facilitating rapid decision-making based on statistical and machine learning tools. The timeliness of analysis is an additional feature of these methods.

In areas like industry, the environment, healthcare, agriculture, finance, services, the public sector and transport, these data and data science methods enable: companies to innovate by analysing people's wants and needs; nations to control emissions and waste; medical professionals to develop better diagnostics, treatments and drugs; farmers to make better use of natural resources; public services to increase efficiency and effectiveness; road operators to mitigate traffic jams; and providers to tailor financial services to their customers and distributors.

The workshop will focus mainly on new participatory approaches to collecting data, applying data science techniques and producing statistical information with the help of a large group of people (crowd-sourced data). It will also look at methods for collecting georeferenced data using IoT sensors and earth observation (satellite imagery and remote sensing data) and techniques based on the datafication of human activities including computer-mediated communication (text mining and natural language analysis) and financial transactions (credit cards). These methods are potentially useful in the production of official statistics, but also for developments in the sciences (citizen science) and for the sustainable development of the information society. These techniques will be examined from a methodological angle to identify opportunities and pinpoint issues, with a focus on data quality and production governance models from the point of view of experts in statistics.

Aspects of data governance regarding the need for new principles covering data accessibility, quality assessment, and participatory roles for partners in these non-traditional data sources and official statistics production will also be discussed.

Topics of interest

- Methodological issues arising from working with non-traditional data sources and diverse evidence landscapes, including citizen science, IoT, earth observation, big data and data science methods straddling statistics and computer science.
- Collaboration among universities and public and private institutions for the re-use of data and production of reliable statistics.
- Access rights, tools and skills to help users stay in full control of their data and better understand the world with statistics.
- Representations of data and quality criteria for different purposes and target audiences (e.g. EU and national policy making and action by local volunteers in municipalities or enterprises).
- Data science methods for non-traditional data source analysis and decision-making.
- New models of data governance in the production process from non-traditional sources to produce official statistics and the consequences for the role of statistical offices in society and government at national and international level.
- Concepts and strategies for integrating data from non-traditional sources into statistical production process and consequences for the usability of both traditional and non-traditional sources.
- Literacy and education aspects for traditional and new actors in the integration and use of non-traditional sources.

Target audience

The event is intended primarily for statistics stakeholders: producers of statistics, academic methodologists and experts, and representatives of different statistics user groups.

Outputs, outcomes, impacts

- Publication of papers in the Journal of Official Statistics.
- Recommendations for statistical offices regarding methodological and quality aspects when integrating non-traditional data sources.
- Recommendations regarding governance and literacy, as well as possible new roles for statistical offices in society and government.
- Identification of new areas for research and recommendations for future activities.

PROGRAMME

Wednesday, 20 October 2021

8:40   Virtual event room opens

8:50   **Opening addresses**

8:50   Maurizio Vichi, Acting ESAC Chair, Sapienza University of Rome
9:00   Mariana Kotzeva, Director General, DG ESTAT, European Commission
9:10   Silke Stapel-Weber, Director General for Statistics, European Central Bank

9:20   **European Statistics Day 2021 celebration**
Maria João Valente Rosa, ESAC Deputy Chair, NOVA University Lisbon

9:30   **Panel 1: Agenda for shaping Europe's digital future using statistics, big data and artificial intelligence**

Moderator: Emanuele Baldacci, Director, DG ESTAT, European Commission

Panellists:
Maria Coduti, Data and Innovation Policy Unit, DG CONNECT, European Commission
Francesca Kay, Chief Information Officer, Central Statistics Office Ireland
Mirko Lorenz, Innovation Manager, Co-Founder and chairman Datawrapper Reasona
Caroline Willeke, Deputy Director General Statistics, European Central Bank

10:45   *Coffee break*

11:00   **Advances in Data Analysis and Classification discussion paper:**
Is there a role for statistics in artificial intelligence?
(https://link.springer.com/content/pdf/10.1007/s11634-021-00455-6.pdf)

Moderator and presenter: Hans A. Kestler, Institute of Medical Systems Biology, Ulm University, Adalbert Wilhelm, Commerzbank Chair of Information Management, Jacobs University Bremen

Discussants**:**
David Dreyer Lassen, Prorector for Research, Copenhagen Centre for Social Data Science, University of Copenhagen
Chiara Osbat, Adviser, Directorate General Economics, European Central Bank
Vincenzo Spiezia, Senior Economist, Measurement and Analysis of the Digital Economy, OECD

11:50   **Discussion**

12:00   **Session 1: Recent developments in non-traditional data sources and new methods of data collection**

Chair: Serge Allegrezza, Director General STATEC

Speakers:
Dominik Rozkrut, President of Statistics Poland
Access to privately-held data and other principles for non-traditional data sources
Ariane König, ESAC member, University of Luxembourg, Jacek Stankiewicz, University of Luxembourg; *Can citizen science complement official data sources?*

Steve McFeely, Director of Data and Analytics at WHO, Geneva, University College Cork –
"You say you want a [data] revolution": A proposal to use unofficial statistics for the SDG Global Indicator Framework

13:05   **Discussion**

13:15   ***Lunch break***

14:15   **Session 2: Methodologies for the analysis of non-traditional data sources: challenges, opportunities and risks**

Chair: Corrado Crocetta, President of the Italian Statistical Society

Speakers:
Kimmo Soramaki, Founder and CEO, FNA (tbc)
Network analysis using big data
Diego Kuonen, Statoo Consulting & Geneva School of Economics and Management, Geneva University
How to manage artificial intelligence and data science for decision making as a process of continuous improvement in the context of official statistics?
Sophia Ananiadou University of Manchester & National Centre for Text Mining
Natural Language Processing Methods for Search and Discovery

15:20   **Discussion**

15:30   ***Coffee break***

15:45   **Panel 2: New models of data governance and principles (accessibility, cooperation, proactivity, quality framework, modernisation,) for official statistics**

Moderator: Aurel Schubert, Honorary Professor, Vienna University of Economics and Business Chair European Statistical Governance Advisory Board

Panellists:
Wieteke Dupain, Head of Knowledge, Research & Development, The European Social Enterprise Network Euclid Network
Eric Rancourt, Director General, Modern Statistical Methods and Data Science, Statistics Canada
Barry Schouten, Senior Methodologist, Statistics Netherlands; Faculty of Social Sciences and Behavioural Sciences, Utrecht University
Monica Scannapieco, Directorate for Methodology and Statistical Process Design, Italian National Institute of Statistics
Stefaan Verhulst, Co-Founder and Chief Research and Development Officer, The GovLab, New York University

17:00   **Closing session**
Maurizio Vichi, Acting ESAC Chair, Sapienza University of Rome

General information

Dates	*Wednesday, 20 October 2021*
Venue	*Online*
Language	*English*
Participation	*Please register under this link*
Organising committee	*Maurizio Vichi, ESAC* *Emanuele Baldacci, Eurostat* *Caroline Willeke, European Central Bank* *Per Nymand-Andersen, European Central Bank* *Cristina Pereira de Sá, Eurostat* *Albrecht Wirthmann, Eurostat* *Joanna Lindberg, Eurostat, ESAC Secretariat* *Annemarie Jung, STATEC* *Sylvia Mann, STATEC*
Contacts	*ESAC Secretariat* *Email: ESTAT-ESAC@ec.europa.eu*

Please note that this programme may be subject to change without notice.