MODULE 12

Non-Conventional Data Sources (Big Data for gender equality)

TRAINING SYLLABUS

Curriculum on Gender Statistics Training

This product was developed under the guidance of the Subgroup on Gender Statistics Training, within the Asia-Pacific Network of Statistical Training Institutes.

Introduction

This syllabus has been designed to guide trainers on how to conduct related training. The syllabus can also be used by learners who wish to know more about this topic and people who are generally interested in understanding the range of sources of gender data.

This syllabus is part of a wider module on this area of gender statistics. Other materials within this module might include exercises, PowerPoint presentations and example quizzes. Please refer to the additional set of materials for a comprehensive and effective learning experience.

Who is this module for?

- Statisticians and other experts who are looking to understand how non-traditional data sources can be used to fill data gaps
- Policymakers and decision-makers in general, who are looking to understand how non-traditional data sources can be used for evidence-based decision-making
- Academics who wish to focus or inform their research and analysis through the use of gender data from different sources and want to enhance their knowledge of non-conventional data sources of gender data
- Civil society organizations who wish to enhance their understanding of gender data beyond the traditional data sources, and use for advocacy or communication purposes
- Anyone who wishes to acquire a thorough understanding of the range of sources of gender data

What do I need to know before going through this module?

This is a module containing introductory information on non-traditional data sources for gender statistics. It is intended to complement other Modules 1, 2, 5 and 9 of the training curriculum that introduced the different sources of gender data. It can be used by both experts and non-experts. No advanced knowledge of statistics is necessary. However, it would be good for the learner to have an idea of what the Sustainable Development Goals[1] are, including their targets and indicators[2]. It is advisable for the learner to have previously completed Modules 1 and 2, 5 and 9.

[1] For additional information on the SDGs see: https://www.un.org/sustainabledevelopment/sustainable-development-goals/

[2] See: https://unstats.un.org/sdgs/indicators/indicators-list/

Scope of this module

This module aims to equip learners with the knowledge and skills necessary to effectively understand and use Big Data for advancing gender equality and contributing to the achievement of Sustainable Development Goals. It explores the characteristics of Big Data and how it can be used as an alternative (non-traditional or non-conventional) source of data for producing gender statistics. It delves into the main features/characteristics and sources of Big Data, then highlights the opportunities that Big Data provides for leveraging gender statistics while also identifying ways to the associated challenges and pitfalls. It also includes practical applications or real-world examples that demonstrate the use of Big Data in addressing specific gender-related issues.

Learning objectives

After going through this Module, the learner is expected to:

- Understand how non-traditional sources (such as Big Data) can be used to fill gender data gaps and leverage gender statistics.
- Explain/ describe the main characteristics of Big Data and distinguish it from other traditional data sources
- Identify the range of Big Data sources and understand both the opportunities/ strengths and challenges/ limitations of each source in capturing gender-related information.
- Discover lessons and insights from real-world applications of Big Data in gender-related initiatives and appreciate the potential of Big Data in identifying and understanding gender dynamics and disparities
- Analyze case studies and best practices where Big Data has been successfully leveraged to advance gender equality and contribute to SDGs.

Note to trainer: Depending on the pace of trainer and trainees, it is expected that training for this module can be delivered in 1 hour.

Table of Contents

1. Sources of official statistics

As presented in Module 2[3], official statistics serve a crucial role in providing accurate and reliable information about various aspects of a country's economy, society, and environment. Official statistics are typically produced for the purpose of informing public policy, governance, decision-making, and research. Those statistics are based almost exclusively on data collected, compiled, and published by government agencies (usually national statistics office) or other authorized organizations. Typically, national statistics offices (NSOs) or other authorized data producers generate official statistics from traditional (conventional) sources which include censuses, household surveys (such as the Demographic and Health Survey, Multiple Indicator Cluster Survey or Labour Force Survey) and administrative records (including registers, school and hospital records). Those traditional data sources are key sources of gender data for measuring inequalities in women and men's lives and monitoring the Sustainable Development Goals (SDGs).

With the onset of the data revolution, which was characterized by the proliferation of new technologies and consequently an explosion of available data from multiple sources, Big Data has emerged as a potential data source of official statistics. Conceptually, Big Data refers to data that are different from traditional census or survey data. It is data that are generated from a variety of sources, are loosely structured, large in volume and available in real time. Owing to those key features, Big Data has been recognized as having the potential to produce more relevant and timely statistics than traditional sources of official statistics. Big Data also offers the potential for monitoring the sustainable development goals. In the field of gender statistics, Big Data offers viable new ways to gather, analyze, and interpret data related to gender disparities and inequalities.

2. What is Big Data[4]?

Big Data is used to describe the exponential growth in data that accompanied the information and communication technology (ICT) revolution as well as the frequency with which it is produced. It encompasses structured and unstructured data from various sources, including social media, sensor networks, online platforms, and administrative records. Big Data has many distinct characteristics or features that distinguish it from data produced from traditional sources such as censuses, surveys and even administrative data.

The United Nations Statistics Division[5] defines Big Data as: "data sources that can be described as; high volume, velocity and variety of data that demand cost-effective, innovative forms of processing for enhanced insight and decision making". This definition captures the 3V's that are typically used to describe Big Data namely: volume, velocity and variety.

- **Volume:** Big Data is high-volume i.e. large quantities (datasets at least terabytes in size) are continuously generated from the pervasive use of electronic devices and other digital sources.

[3] See Module 2: Gender Data Literacy and Avoiding Common Mistakes

[4] Introductory e-learning course on Big Data on the UNSD Big Data Learning hub https://learning.officialstatistics.org/course/view.php?id=46

[5] https://unstats.un.org/unsd/statcom/doc14/2014-11-BigData-E.pdf

- **Velocity:** Big Data is high velocity (volume produced in time) i.e. generated rapidly in near real-time, can be processed but also changes very quickly.
- **Variety:** Big Data is high variety i.e. produced from diverse information assets/ sources in different types. It can contain structured data and unstructured data.
-

Box 1: Characteristics of Big Data

Characteristic	Description	Attribute	Driver
Volume	*The sheer amount of data generated or data intensity that must be ingested, analyzed, and managed to make decisions based on complete data analysis*	*According to IDC's Digital Universe Study, the world's "digital universe" is in the process of generating 1.8 Zettabytes of information – with continuing exponential growth – projecting to 35 Zettabytes in 2020*	*Increase in data sources, higher resolution sensors*
Velocity	*How fast data is being produced and changed and the speed with which data must be received, understood, and processed*	• *Accessibility: Information when, where, and how the user wants it, at the point of impact* • *Applicable: Relevant valuable information for an enterprise at a torrential pace becomes a real-time phenomenon* • *Time value: real-time analysis yields improved data-driven decisions*	• *Increase in data sources* • *Improved thru-put connectivity* • *Enhanced computing power of data generating devices*
Variety	*The rise of information coming from new sources both inside and outside the walls of the enterprise or organization creates integration, management, governance, and architectural pressures on IT*	• *Structured – 15% of data today is structured, row, columns* • *Unstructured – 85% is unstructured or human generated information* • *Semistructured – The combination of structured and unstructured data is becoming paramount.* • *Complexity – where data sources are moving and residing*	• *Mobile* • *Social Media* • *Videos* • *Chat* • *Genomics* • *Sensors*

Source: Extracted from TechAmerica Foundation (2012)[6].

[6] https://www.kdnuggets.com/2012/10/techamerica-demystifying-big-data-report.html

According to the United Nations Economic Commission for Europe (2014), "Big Data cuts across all statistical activities and could be relevant for all statistical domains"[7]. Hence Big Data is relevant as a non-conventional source of data for gender statistics. As highlighted throughout this curriculum, gender statistics are essential for understanding and addressing various social and economic issues, such as gender-based discrimination, women's empowerment, and gender equality. Big Data can significantly enhance the quality and granularity of these statistics, providing valuable insights into gender dynamics.

3. Sources of Big Data

Big Data is generated from a wide range of different digital activities.

Figure 1: Big data sources

Source: Siu-Ming Tama and Gemma Van Halderen (2020)

The classification system for Big Data developed by the UN Economic Commission for Europe[8] identifies 3 main kinds of Big Data sources that are distinguished based on their characteristics:

- **Social Networks (human-sourced information):** this information is the record of human experiences, previously recorded in books and works of art, and later in photographs, audio and video. Human-sourced information is now almost entirely digitized and stored everywhere from personal computers to social networks. Data are loosely structured and often ungoverned.

- **Traditional Business systems (process-mediated data):** these processes record and monitor business events of interest, such as registering a customer, manufacturing a product, taking an

[7] https://unece.org/DAM/stats/documents/ece/ces/2014/7-In-depth_review_of_big_data.pdf

[8] https://unstats.un.org/unsd/trade/events/2015/abudhabi/gwg/GWG%202015%20-%20item%202%20(iv)%20-%20Big%20Data%20Classification.pdf

.

order, etc. The process-mediated data thus collected is highly structured and includes transactions, reference tables and relationships, as well as the metadata that sets its context. Traditional business data is the vast majority of what IT managed and processed, in both operational and BI systems. Usually structured and stored in relational database systems. (Some sources belonging to this class may fall into the category of "Administrative data")

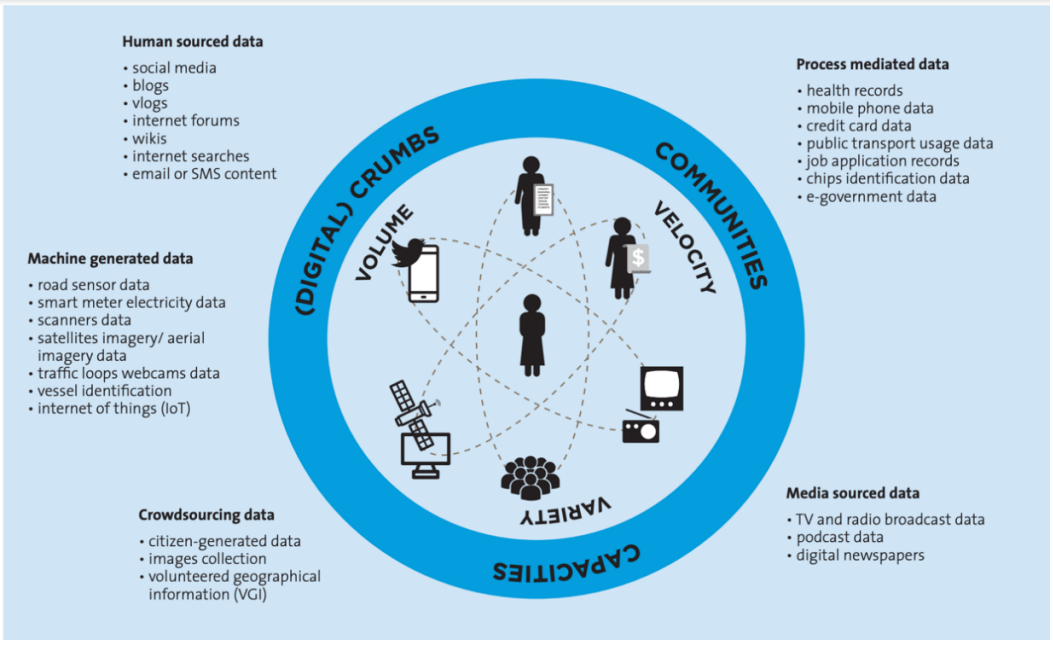
- **Internet of Things (machine-generated data):** derived from the phenomenal growth in the number of sensors and machines used to measure and record the events and situations in the physical world. The output of these sensors is machine-generated data, and from simple sensor records to complex computer logs, it is well structured. As sensors proliferate and data volumes grow, it is becoming an increasingly important component of the information stored and processed by many businesses. Its well-structured nature is suitable for computer processing, but its size and speed is beyond traditional approaches.

In addition to the 3 kinds of data sources captured in the UNECE classification system, two additional types of Big Data sources that are relevant for the production of gender statistics are: crowdsourced data and media generated data. Those sources along with relevant examples are captured in Figure 1.

- **Crowdsourced data:** is a form of secondary data that is obtained largely through the internet/ online using participatory methods that involve large groups of people. Examples include citizen-generated data or volunteered geographic information (GVI) such as OpenStreetMap project that allows individuals to contribute geographic data to create a free and editable map of the world.

- **Media generated:** refers to the information generated by media sources, including traditional media (TV, radio, newspapers) and digital media (websites, social media and streaming platforms).

Figure 2: Types, physical characteristics and social dimensions of Big Data



Source: UNWomen (2018).

Big Data also encompasses many different types of data:

- **Structured data:** refers to any data that can be processed, accessed and stored as a fixed format. Features of structured data are that such data is easier to work with; organized and easily understood by machine language and easily read by traditional analytics tools. Examples scanner data, bank transaction information, cell phone numbers or financial data.

- **Unstructured data**: refers to data that lacks any predefined format or model. Unlike structured data, unstructured data is more difficult to work with, harder for computers and humans and interpret and typically produced in large volumes. Examples include:

    - Text-only documents that lack a predefined format and can contain a mix of languages, abbreviations, slang, and grammatical variations such as emails, articles, reports, social media posts, and chat logs);
    - Social media feeds such as posts comments, likes, and shares;
    - Video and audio data.

- **Semi structured data**: refers to data that contains both structured and unstructured components. This type of data tends to be commonly available, easier to use than unstructured data but requires tools to facilitate its exploration and analysis. Examples include web-scrapped data, Hypertext Markup Language (HTML) files; and Extensible Markup Language (XML) files.

Box 2: Data sources

Traditional and non-traditional sources of data

According to Principle 5 (*Sources of Official Statistics*) of the Fundamental Principles of Official Statistics "*Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records*". For that reason, statistical surveys and administrative records have been long regarded as "**traditional**" sources of official statistics. Statistical surveys encompass both household sample surveys (Multiple Indicator Cluster Survey, Demographic Health Survey) as well as censuses (economic census, population and housing census and agriculture census) that are all invaluable sources of gender data and statistics.

Both surveys and censuses are intended specifically for producing official statistics and, as such, they follow a structured format - common definitions, standards, classifications and international recommendations. Unlike censuses that are intended to capture information on all subjects/ units in the population, surveys are administered to a small sample from a larger population and can therefore examine topics in much greater detail. In addition to surveys and censuses, a third data source is administrative records. Administrative data refers to data collected by organizations for administrative purposes (not statistical purposes). Examples include school data collected by the ministry of education or civil registration data from civil registration and vital statistics systems.

Non-traditional data refer to information collected for other (not statistical) purposes. Big data has the characteristics of being available in greater volume, higher velocity and wider variety. Some examples of Big Data include mobile phone data, social media and transaction data.

Both types of data sources have strengths and limitations. Traditional data sources provide structured and reliable information but tend to be time-consuming and expensive to collect, verify and disseminate. Big data on the other hand have the advantage of capturing dynamic and real time-data that can be generated, processed and stored at a much lower cost than traditional data.

4. Value of Big Data for enhancing gender statistics

Big Data can enhance gender statistics in several ways:

- **Granularity:** Big Data offers both the possibility of providing data from new sources as well as new types of data. The availability of larger volumes of data allows for more detailed disaggregation of some statistics as well as granular analysis of gender-related trends and patterns. It can reveal nuances that traditional data sources may miss.

- **More timely Insights:** Policy makers require up-to-date, real-time data, particularly during times of crisis or disasters – such as COVID-19 pandemic, natural disasters or other man-made or natural crises. Big Data is often generated in real-time. It therefore has the potential to provide more timely data than conventional data sources and provide timelier insights into gender dynamics/ issues and societal changes that can inform decision-making by governments, organizations etc.

- **Myriad of data sources:** Big Data brings together many different forms of data in a wide range of data sources including social media, crowdsourcing, and mobile applications, which can capture diverse perspectives on gender issues.

5. Big Data: Challenges

While Big Data offers great potential for improving gender statistics, it also comes with challenges and concerns:

I. Privacy and confidentiality

Big Data is a by-product of digital behaviour produced as users access different services and use digital devices. When Big Data, which was not initially produced for statistical purposes, is used as official statistics, the individuals and devices generating that data may be unaware how it is being used. Protecting individuals' privacy when using Big Data is a major concern, particularly in the context of gender-related data.

Maintaining the confidentiality of individual data is enshrined in principle 6 of the United Nations Fundamental Principles of Official Statistics[9] – "*Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes*". When data is collected for statistical purposes through the traditional census or household survey, there is a guarantee from the NSO to protect the identities and information of the persons or entities supplying individual data. Safeguarding confidentiality is covered in the national statistical legislation that guides the activities of the national statistics office. With Big Data, there is the potential to easily match data and identify the people from whom the data is generated. There is an additional challenge for the NSO to apply relevant data suppression techniques to safeguard confidentiality.

II. Access:

One of the major barriers to using Big Data is the lack of access. Most of the data generated from the sources described above, are commercially or privately owned and therefore not publicly accessible. In addition to the challenge of the proprietary status of some data, factors such as the prohibitive cost of the data or existing data protection legislation can act as further barriers for the NSOs. In other cases, where the data are not proprietary and are publicly, there are concerns about repurposing this data to compile official statistics.

III. Data Legislation

Statistical legislation is an important prerequisite for an effective statistical system, as it provides the legal basis for functioning of the statistical office. Key features of statistical legislation include the coverage of official statistics, statistical standards, access to official data, reporting by respondents, the protection of the confidentiality, professional independence and coordination – all of which are typically aligned with the *Fundamental Principles of Official Statistics*.

Statistical legislation on the right to access to data from government or non-government agencies and private sector varies widely across countries. For many countries the legislation may not include provisions on new

[9] https://unstats.un.org/unsd/dnss/hb/E-fundamental%20principles_A4-WEB.pdf

data sources such as Big Data, the right to access data from the private sector or the role of private data producers of statistics for that matter. The lack of such provision put restrictions on access to certain types of Big Data.

To address this, agreements or legislation can be put in place to: (i) continuously enable NSOs to receive Big Data sources: (ii) protect the privacy and confidentiality of individuals or businesses. Notwithstanding, even if such provisions may allow for access to various types of Big Data, they may not be adequate to allow the use of data from those sources for statistical purposes.

IV. Bias and impact on inclusivity:

One major shortfall of Big Data is that it can be that it is both influenced by and serves to influence existing gender inequality and biases in a society. Incomplete or skewed representation of gender-related experiences in datasets can lead to inaccurate or incomplete gender statistics.

The use of algorithms for processing Big Data has the potential to introduce biases perpetuate or amplify societal biases, discrimination and gender inequalities that disproportionately affect segments of the population.

Access to technology and digital platforms is not equal across all societies. Those without access to digital platforms may be excluded from the datasets, leading to a skewed representation of gender dynamics. This digital divide and disparities in access to technology can lead to underrepresentation of certain population groups, thereby affecting the inclusivity of Big Data in gender statistics.

V. Data Quality:

Big Data is generated from diverse sources. Having access to relevant metadata is critical in gaining an understanding of the data and assessing its quality. As explained in module 1, metadata refers to the methodological descriptions of how the data are collected and processed. On the one hand, metadata describes the quality of data and statistics. On the other hand, metadata is itself a quality component that improves the availability and accessibility of data and facilitates the interpretation and use of data over its lifetime.

There are several indicators and criteria used for determining data quality, some of which include: relevance of statistical concepts, accuracy, timeliness, accessibility and clarity of information, comparability of statistics, coherence, completeness. Ensuring those different dimensions of quality of Big Data for gender statistics can be challenging. Furthermore, inconsistencies in data quality across different sources can impact the overall reliability of official statistics derived from Big Data.

VI. Management

A shift to non-conventional data sources such as Big Data for official statistics would mean drastic increases in the volume of data that the NSO would need to process and manage. Adequate training and investment in

developing the technical capacity and expertise would be essential to harness the full potential of Big Data in generating gender statistics.

Summary of challenges:

In summary, Big Data has the potential to transform the production of official statistics (and by extension gender statistics) by providing more detailed, real-time, and varied insights into gender disparities and inequalities. However, it also presents several challenges such as privacy, legislation, data protection and safeguarding confidentiality. All those concerns need to be addressed to ensure the responsible and effective use of Big Data for official statistics, in general, and gender-related research and policy initiatives, in particular.

6. Big Data Applications for Gender Equality

This subsection highlights some applications of Big Data in producing gender statistics for research, measuring and monitoring various phenomena in the lives of women, men, boys and girls.

Big data application examples	Objective	Big data source	Methodology	Findings and implications
1. UN Women's research utilizing online searches and social media data	The study explores the impact of disasters and crises on violence against women (VAW) in the Pacific, acknowledging a gap in data within this context. It addresses challenges in traditional data collection during crises, using big data as a novel approach to understand VAW trends.	Utilizes big data from online searches, social media posts, and service provider websites to understand the impact of crises on VAW.	Analyzes trends in online behavior related to VAW, focusing on search and social media data. A web scraping and analysis tool was used to analyze the Search Engine Results Page (SERP). Further, social media (Facebook, Twitter, YouTube and Tik Tok) discourse analysis was conducted to understand in more detail the ways in which women and men engage online with respect to VAW.	Identifies an increase in online searches and discussions about VAW during and after crises, indicating a higher risk of VAW in such periods. Highlights the significance of big data in informing policies and support services for VAW victims during crises, underscoring its potential in crisis response and gender equality advocacy.
2. UN Women's research utilizing geospatial data	The study emphasizes the importance of understanding the gender-climate change nexus, especially given the lack of comprehensive data in the Sustainable Development Goals (SDG) indicator framework. It underlines the need for	The study integrates geospatial data with individual-level survey data, harnessing big data to explore connections between climate change and gender inequality in select climate change-prone Asian countries.	The analysis employs binary logistic regression and random forest machine learning techniques. These methodologies facilitate the identification of statistical associations between climate-	The findings reveal statistically significant associations between climate-related factors like droughts, aridity, and temperature with gender inequality outcomes such as child marriage, adolescent births, and intimate partner

	evidence-based analysis to inform inclusive strategies mitigating the consequences of climate change, particularly on women.		related factors and gender inequality outcomes.	violence. The study faced challenges like data limitations and the localized nature of environmental events. Solutions involved using sophisticated statistical models to capture the nuanced effects of climate variables on gender outcomes. The research contributes to a deeper understanding of how climate change exacerbates gender inequalities. It suggests the need for further research and highlights the potential of big data in shaping policies for gender equality in the context of climate change.
3. UN Women's research content and sentiment analyzing online speeches	This study delves into the representation of gender perspectives within the climate change dialogue at the Conference of Parties (COP) meetings. Acknowledging a historical absence of gender discussions in early COP agendas, it utilizes big data to bridge this gap, offering a novel approach to understanding the evolution of gender	Digital speeches from the most recent COP meetings (COP24-COP26), using online resources and publicly available transcripts.	The methodology involves a sophisticated examination of these speeches, focusing on the frequency and context of gender-related terms.	This big data analysis highlights the slow yet growing momentum in incorporating gender perspectives into climate change discussions. The findings are significant for shaping future COP agendas and for informing policy decisions at platforms like the Commission on the Status of Women

	inclusion in climate policies.			(CSW), thereby contributing to the advancement of gender equality in the context of climate action.
Analyzing gender disparities with mobile phone metadata	This analysis explored how mobile phone data can be used to predict net enrollment rates of children in Pakistan.	Call Detail Records (CDR) from a major operator in Pakistan, captured over a period of 7 days and covering 93 of 128 districts in Pakistan. The CDR included transactions (voice and text messages), as well as the timing of the activity and location of cell tower through which the call was made. The CDR also included the sex and age of each subscriber.	The approach included analysis of CDR data to extract insights on the association between the sex of the phone users in activity such as number of calls made, size of network and network metrics. Prediction algorithms were used to assess gender disparities in education.	There are statistically significant differences in the social networks of men and women in Pakistan. Women tend to make more calls within a given period; have a smaller calling network, have more clustered friend groups and are more limited in their mobility. The analysis also revealed network features that can help in predicting educational disparities: (i) gender diversity of male calling networks; (ii) clustering of friend groups; and (iii) geographical reach across all networks.

7. Key takeaways

- Official statistics provide accurate and reliable information about various aspects of a country's economy, society, and environment. They are essential for governance, research and evidence-based policy- decision-making.

- Traditional or conventional data for gender statistics include data collected through censuses, surveys and administrative records. Data from those sources have also long been the main source of data for producing official statistics.

- Non-traditional data sources encompass the vast array of newer forms of digital data generated from the pervasive use of electronic devices such as social media, mobile phones, and satellite imagery.

- There has been increased recognition and acceptance of Big Data as a source of official statistics, including gender statistics.

- Big Data sources have the potential to provide significant insights of gender-related trends and issues, support the formulation of evidence-based interventions and monitor progress towards gender equality, as well as the SDGs. Big Data for monitoring gender equality can be generated from a range of sources (human-sourced, process-mediated, machine generated) and be classified as 3 key types (structured, semi-structured and unstructured)..

- Big Data has the potential to revolutionize gender statistics by providing more detailed, real-time, and varied insights into gender disparities and inequalities. However, it also presents a myriad of technical challenges and pitfalls that need to be addressed to ensure the responsible and effective use of Big Data for gender-related research and policy initiatives.

- Real world applications and use cases of big data in diverse research projects demonstrate the range of ways in which different types of Big Data (social media data, mobile phone data) can be used to shed light on gender disparities and inequalities.

References

Data2x (2019) Big Data and Gender Brief Series
https://data2x.org/resource-center/big-data-and-gender-brief-series/

International Finance Cooperation (2020) Leveraging Big Data to advance Gender Equality
https://www.ifc.org/en/insights-reports/2020/big-data-and-gender

Rana, Ahmed Nauraiz (2020). Leveraging Big Data to advance gender equality. Retrieved:
https://www.ifc.org/content/dam/ifc/doc/mgrt/emcompass-note-86-big-data-and-gender-v2.pdf

Tam, Siu-Ming and Gemma Van Halderen (2020). The five V's, seven virtues and ten rules of big data engagement for official statistics. https://content.iospress.com/download/statistical-journal-of-the-iaos/sji190595?id=statistical-journal-of-the-iaos%2Fsji190595

TechAmerica Foundation Big Data Commission (2012). Demystifying Big Data: A Practical Guide to Transforming the Business of Government. https://www.kdnuggets.com/2012/10/techamerica-demystifying-big-data-report.html

United Nations Economic Commission for Europe. (2014). In-depth Review of Big Data. ECE/CES/2014/7 of the Conference of European Statisticians 9-11 April 2014. Retrieved from:
https://unece.org/DAM/stats/documents/ece/ces/2014/7-In-depth_review_of_big_data.pdf

United Nations Economic Commission for Europe. (2014). *What does "Big Data" mean for official statistics?*. ECE/CES/2013/20 of the Conference of European Statisticians 10-12 June 2013. Retrieved from:
https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2013/20-HLG_big_data.pdf

United Nations Global Working Group on Big Data for Official Statistics Task Team on Cross-Cutting Issues. (2015). Revision and Further Development of the Classification of Big Data. Retrieved from:
https://unstats.un.org/unsd/trade/events/2015/abudhabi/gwg/GWG%202015%20-%20item%202%20(iv)%20-%20Big%20Data%20Classification.pdf

United Nations Statistical Commission. (2014). *Big Data and modernization of statistical systems; Report of the Secretary-General*. E/CN.3.2014/11 of the forty-fifth session of UNSC 4-7 March 2014. Retrieved from:
https://unstats.un.org/unsd/statcom/doc14/2014-11-BigData-E.pdf

UNWomen (2018). Gender Equality and Big Data. Retrieved from:
https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/Library/Publications/2018/Gender-equality-and-big-data-en.pdf