DRAFT CIRCULATED FOR COMMENT

BIG DATA

AND

THE 2030 AGENDA FOR SUSTAINABLE DEVELOPMENT

FINAL DRAFT REPORT-OPEN FOR COMMENT

BY ABBAS MAAROOF, PhD

Table of Contents

List of Tables, Figures and Boxes

Tables

Figures

Boxes

Annexes

Acknowledgements

The author would like to gratefully acknowledge advice from the UN-ESCAP Environment and Development Policy Section, especially Hala Razian. This report is a result of collaboration between UNESCAP, Environment and Development Policy Section, and the UN Global Pulse Lab Jakarta office. The author benefitted from inputs from the Pulse Lab Jakarta team. I would also thank the UN Global Pulse Lab Jakarta office for providing space and resources during the desk research. The author is also grateful for inputs and comments into the draft report by Katinka Weinberger, Rusyan Jill Mamiit and Marko Javorsek.

The findings, interpretations, and conclusions expressed in this report solely reflect the author's views.

Glossary

Algorithm - A formula or step-by-step procedure for solving a problem.

Anonymization - The process of removing specific identifiers (often personal information) from a dataset.

Big data - A term for a large data set.

Big data analytics - A type of quantitative research that examines large amounts of data to uncover hidden patterns, unknown correlations and other useful information.

Big data for development - A concept that refers to the identification of sources of Big Data relevant to policy and planning of development programs.

Citizen reporting or crowd-sourced data - Information actively produced or submitted by citizens through mobile phone-based surveys, hotlines, user-generated maps, etc; while not passively produced, this is a key information source for verification and feedback

Data exhaust - Passively collected transactional data from people's use of digital services like mobile phones, purchases, web searches, etc., these digital services create networked sensors of human behavior.

Data mining - A term refers to the activity of going through big data sets to look for relevant or pertinent information.

Data philanthropy - A term that describes a new form of partnership in which private sector companies share data for public benefit.

Data cleaning/cleansing - The detection and removal, or correction, of inaccurate records in a dataset.

Data migration - The transition of data from one format or system to another.

Data science - The gleaning of knowledge from data as a discipline that includes elements of programming, mathematics, modeling, engineering and visualization.

Data silos - Fixed or isolated data repositories that do not interact dynamically with other systems

Exabyte - A large unit of computer data storage, two to the sixtieth power bytes. The prefix exa means one billion, or one quintillion. In decimal terms, an exabyte is a billion gigabytes.

Geospatial analysis - A form of data visualization that overlays data on maps to facilitate better understanding of the data.

Real time data - A data that covers/is relevant to a relatively short and recent period of time-such as the average price of a commodity over a few days rather than a few weeks, and is made available within timeframe that allows action to be taken that may affect the conditions reflected in the data.

Infographic- A graphic visual representations of information, data or knowledge intended to present information quickly and clearly

Mashup - The use of data from more than one source to generate new insight.

Status quo - A term refers to the existing state of affairs, particularly with regards to social or political issues.

Open data - A term refers to data that is free from copyright and can be shared in the public domain.

Open web data - Web content such as news media and social media interactions (e.g. blogs, Twitter), news articles obituaries, e-commerce, job postings; sensor of human intent, sentiments, perceptions.

Online information - Web content such as news media and social media interactions (e.g. blogs, Twitter), news articles obituaries, e-commerce, job postings; this approach considers web usage and content as a sensor of human intent, sentiments, perceptions, and want.

Petabyte - A measure of memory or storage capacity and is 2 to the 50th power bytes or, in decimal, approximately a thousand terabytes.

Predictive analytics/modeling - The analysis of contemporary and historic trends using data and modeling to predict future occurrences.

Physical sensors - Satellite or infrared imagery of changing landscapes, traffic patterns, light emissions, urban development and topographic changes, etc.; remote sensing of changes in human activity.

Quantitative data analysis - The use of complex mathematical or statistical modeling to explain, or predict, financial and business behavior.

Sentiment analysis (opinion mining) - The use of text analysis and natural language processing to assess the attitudes of a speaker or author, or a group.

Structured data - Data arranged in an organized data model, like a spreadsheet or relational database.

Semantics - A term refers to the study of meaning. It focuses on the relation between signifiers, like words, phrases, signs, and symbols, and what they stand for; their denotation.

Tweet - A post via the Twitter social networking site restricted to a string up to 140 characters

Unstructured data - Data that cannot be stored in a relational database and can be more challenging to analyze from documents and tweets to photos and videos.

Executive Summary

This stocktaking report attempts to provide an overview of big data, its use in the policy-making context, the stakeholders and their roles and provides some suggested actionable steps as a discussion stimulus for the "Big Data and the 2030 Agenda for Sustainable Development: Achieving the Development Goals in the Asia and the Pacific Region" meeting in Bangkok on 14 - 15 December 2015.

Critical data for global, regional and national development policymaking are still lacking. Many governments still do not have access to adequate data on their entire populations. This is particularly true for the poorest and most marginalized, the very people that leaders will need to focus on if they are to achieve zero extreme poverty and zero emissions, and to 'leave no one behind' in the next 15 years. This is true, too, for the international community, who will not be able to support the most vulnerable and marginalized people without an overhaul of the current ways of gathering data.

While most data is technically "public", accessing it is not always easy, and mining it for relevant insights can require technical expertise and training that organizations and governments with limited resources can't always afford. Making good use of big data will require collaboration of various actors including data scientists and practitioners, leveraging their strengths to understand the technical possibilities as well as the context within which insights can be practically implemented.

Recent discussion suggests to move away from seeing Big Data in isolation, but to rather focus on the "ecosystem" of Big Data. According to this concept, Big Data is not just data—no matter how big or different it is considered to be; big data is first and foremost 'about' the analytics, the tools and methods that are used to yield insights, the frameworks, standards, stakeholders involved and then, knowledge.

Effective application of Big Data would also require changes in the decision-making process, which customarily relies on traditional statistics. Given the high frequency of Big Data, a more responsive mechanism will need to be put in place that allows the government to process the information and act quickly in response. However, this stock take finds that big data is not (yet) playing a crucial role in policy making. If at all, it is used at the agenda setting stage and/or evaluation stage of policy making. One of the reasons might be because the ecosystem is not yet functioning and crucial elements, such as standards and frameworks are still missing. National governments and other policy makers are just starting to systematically engage with big data for policy making.

The proposed steps are based on the recommendations of the UN Independent Advisory Group, and are meant to help building and maintaining the Big Data ecosystem for better development policy making:

- Establish and manage a coordination mechanism with the key UN stakeholders and other international partners;

- Develop a consensus on principles and standards among the UNESCAP member countries;
- Kick-off and institutionalize a Regional Multi-Stakeholder Mechanism to share innovations;
- Mobilize regional resources for capacity development for the less advanced UNESCAP member countries;
- Enhance in-house big data analytics capacity.

Depending on the discussions during the workshop and agreements between stakeholders, certain recommended actions could be prioritized and elaborated further.

1.Introduction

Big data applications may offer the ability to collect and analyze 'real time' information from across ESCAP's 62 member States for policies that relate to the 2030 Agenda's 17 goals and their 169 targets. The scope of this information is vast, and big data applications can facilitate policy making in the region that would otherwise require dedicated intensive and continuous human and financial resources.

This stocktaking report, commissioned by ESCAP, attempts to provide an overview of big data, its use in the policy-making context, the stakeholders and their roles in making the most out of the opportunities that big data presents. For illustrative purposes, the report then presents a selection of best practices using big data in the policy making process.

The report then will, built on existing work in this field, provide some practical ideas on how to further progress the 2030 Agenda and policy making around it using big data. The recommendations of this report also shall inform ESCAP's strategic planning for the development of targeted capacity building program activities, and the Asia Pacific Sustainable Development Roadmap.

The discussion of big data is quite complex, ranging from practical or technical challenges to legal and regulatory limitations. The below figure (Figure 1) illustrates the 3 different dimensions of big data and policies. While this report touches on the policy for data in the gaps and constraints section, the focus of this report is mainly on the inner circle: data for policy. The case studies complement the center piece - evidence informed policy-making.

The purpose of this report is to support ESCAP's work of providing rigorous analysis and peer learning; and translating these findings into policy dialogues and recommendations. It focuses on big data in the policy context and in the context of the 2030 Agenda.

Despite improvements, critical data for global, regional and national development policymaking are still lacking. Large data gaps remain in several development areas. Poor data quality, lack of timely data and unavailability of disaggregated data on important dimensions are among the major challenges. As many as 350 million people worldwide are not covered by household surveys. There could be as many as a quarter more people living on less than $1.25 a day than current estimates suggest, because they have been missed out of official surveys [1].

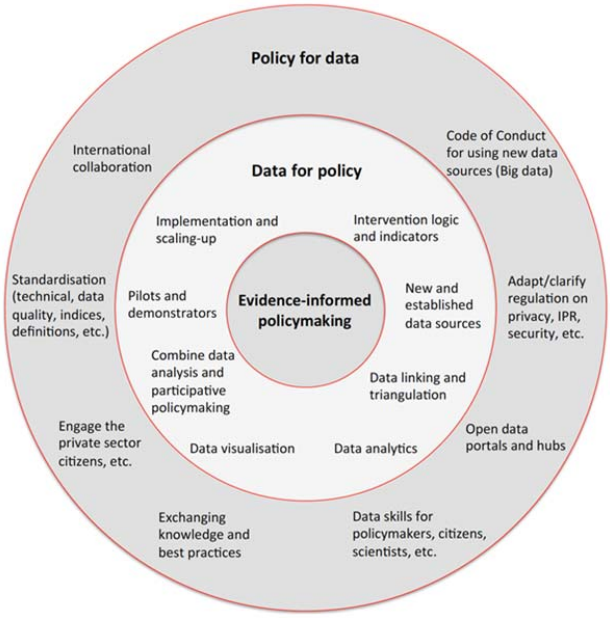


Figure 1: Big Data and Policy

As a result, many national and local governments continue to rely on outdated data or data of insufficient quality to make planning and decisions. Good quality, relevant, accessible and timely data enables governments to extend targeted services into communities, and to implement policies more efficiently. Many governments still do not have access to adequate data on their entire populations, and particularly true for the poorest and most marginalized, the very people that leaders will need to focus on if they are to 'leave no one behind' in the next 15 years [2]. This is true, too, for the international community, who will not be able to support the most vulnerable and marginalized people without an overhaul of the current ways of gathering data.

Box 1: Real-world policy constraints: the ODI survey

To confirm some of the anecdotal evidence about the lack of good data in developing country ministries, the Overseas Development Institute (ODI) interviewed a series of policy-makers based in line ministries to understand how they viewed capacity constraints in their respective countries.

Findings highlighted the problems with stability and continuity of data collection, particularly in countries in conflict where often data and institutional memory are lost during the war, impacting time-series analysis. A further challenge was more political in nature, especially around a limited understanding of how the public sector

and civil servants can work with data and how data serves them, which may cause resistance to utilization of data effectively. Political issues are sometimes misconstrued by development actors as capacity issues [3].

Data are not just about measuring changes; they also facilitate and catalyze that change. Of course, good quality numbers will not change people's lives in themselves. But to target the poorest systematically, to lift and keep them out of poverty, even the most willing governments cannot efficiently deliver services if they do not know who those people are, where they live and what they need. Nor do they know where their resources will have the greatest impact.

Policy-making takes place in an increasingly rich data environment, which poses both promises and challenges to policy-makers. Data offers a chance for policy-making and implementation to be more citizen-focused, taking account of citizens' needs, preferences and actual experience of public services, as recorded on social media platforms. As citizens express policy opinions on social networking sites such as Twitter and Facebook; rate or rank services or agencies on government applications; or enter discussions on a range of social enterprise and NGO sites, they generate a whole range of data that government agencies might harvest to good use. Policy-makers also have access to a huge range of data on citizens' actual behaviour, as recorded digitally whenever citizens interact with government administration or undertake some act of civic engagement, such as signing a petition.

Data mined from social media or administrative operations in this way also provide a range of new data, which can enable government agencies to monitor-and improve-their own performance, for example through log usage data of their own electronic presence or transactions recorded on internal information systems, which are increasingly interlinked. Governments can use data from social media for self-improvement, by understanding what people are saying about government, and which policies, services or providers are attracting negative opinions and complaints, enabling identification of a failing school, hospital or contractor, for example. They can solicit such data via their own sites, or those of social enterprises. And they can find out what people are concerned about or looking for, from the Google Search API or Google trends, which record the search patterns of a huge proportion of Internet users [4].

The recent report of the UN Secretary General's Independent Expert Advisory Group (IEAG) [5] "defines the data revolution for sustainable development as the integration of data coming from new technologies with traditional data in order to produce relevant high-quality information with more details and at higher frequencies to foster and monitor sustainable development. This revolution also entails the increase in accessibility to data through much more openness and transparency, and ultimately more empowered people for better policies, better decisions and greater participation and accountability, leading to better outcomes for the people and the planet".

2. Big Data and the Data Revolution

Big Data is not a single 'thing' - it is a collection of data sources, technologies and methodologies that have emerged from, and to, exploit the exponential growth in data creation over the past decade [6].

Big data is a buzzword; used to describe a massive volume of both structured and unstructured data that is so large it is difficult to process using traditional database and software techniques. Data is a growing element of our lives. More and more data is being produced and becoming known in the popular literature as “big data”, its usage is becoming more pervasive, and its potential for policy making and international development is just beginning to be explored [7].

2.1. From the 3 Vs to the 3 Cs of Big Data

Big data can be defined as large volumes of high velocity, complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of the information. Big data can be characterized by 3Vs: the extreme *volume* of data, the wide *variety* of types of data and the *velocity* at which the data can be processed [8,9,10]. Although big data doesn't refer to any specific quantity, the term is often used when speaking about petabytes and exabytes of data, much of which cannot be integrated easily. It is worth to mention that most recently, some data scientists and researchers have introduced a fourth characteristic, veracity, or ‘data assurance’. That is, the big data analytics and outcomes are error-free and credible. However, veracity is still a goal and not (yet) a reality [8]. Annex 1 describes the most common types of big data.

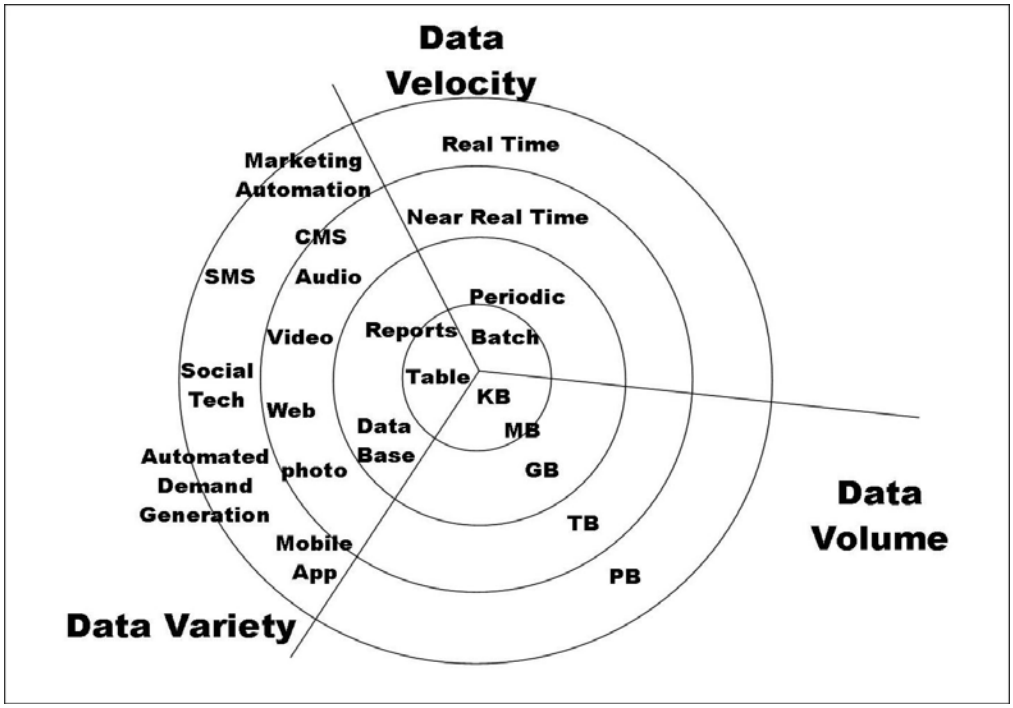


Figure 2: The 3 V's [11]

Data sets grow in size in part because they are increasingly being gathered by inexpensive and numerous information-sensing, mobile devices, remote sensing, software logs, cameras, microphones, radio-frequency identification (RFID) readers, and wireless sensor networks [12], [13],[14]. The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s [15]; as of 2012, every day 2.5 exabytes (2.5×10^{18}) of data were created. As of 2014, every day 2.3 zettabytes (2.3×10^{21}) of data were created by Super-power high-tech Corporation worldwide [16].

Letouzé, one of the Big Data for Development pioneers, has developed "the 3 Cs" of Big Data- presenting another perspective. The 3Cs stand for Big Data 'Crumbs', Big Data 'Capacities' and Big Data 'Community'; it fundamentally frames Big Data as an ecosystem, a complex system actually, not as data sources, sets or streams. And it is both in reference and opposition to the 3 Vs of Big Data [17].

According to his concept, Big Data is not just data—no matter how big or different it is considered to be; this is why and where Big Data as a field—an ecosystem. Gary King's Harvard presentation on "Big Data is not about the data" also and perhaps highlights that big data is first and foremost 'about' the analytics, the tools and methods that are used to yield insights, turn the data into information, then, perhaps, knowledge [18].

The 2nd 'C' of Big Data, for Capacities, is largely about that—the tools and methods, the hardware and software requirements and developments, and the human skills. There is

a need to both consider and develop capacities, without which crumbs are irrelevant. But it's not just about skills and chips; it's also about how the whole question is framed. This is of course related to the concept of 'Data Literacy', and the need to become sophisticated users and commentators.

The 3rd C of community refers to the set of actors—both producers and users of these crumbs and capacities; it's really the human element—potentially it's the whole world.

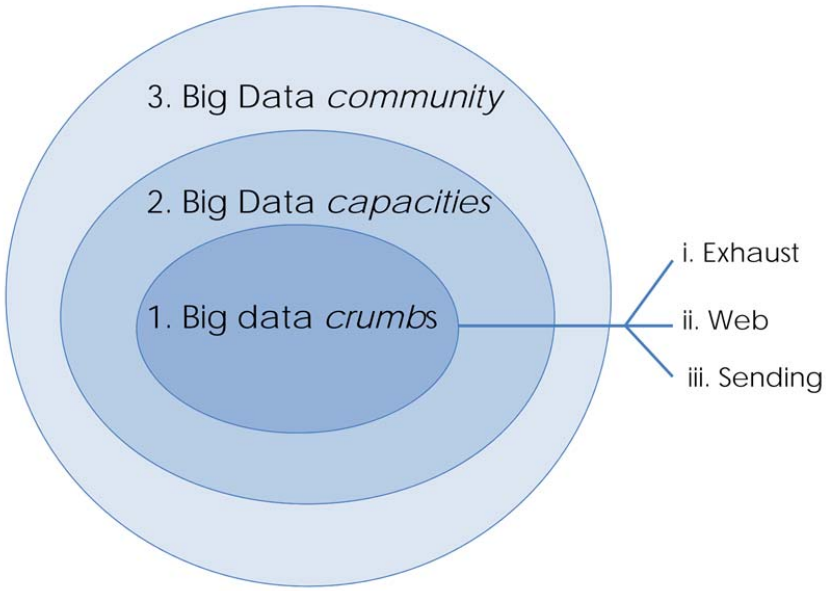


Figure 3: The 3 C's" Diagram [17]

And the resulting concentric circles with community as the larger set are a complex ecosystem—with feedback loops between them. For example new tools and algorithms produce new kinds of data, which may in turn lead to the creation of new startups and capacity needs. Letouzé and others [17,18] argue that the basic point is that Big Data is not big data; and that questions like "how can national statistical office use Big Data" don't mean much or rather they miss the point. The real important question is why and how an NSO (National Statistical Office) should *engage* with Big Data as an ecosystem, partner with some of its actors, become one of its actors, and help shape the future of this ecosystem, including its ethical, legal, technical and political frameworks. This question can then be expanded to the sustainable development actors interested in using big data and to become part of the Big Data Ecosystem. This would also involve the role of development actors as facilitators, knowledge brokers and convening powers.

This report is structured in a similar way: from a narrow focus on Big Data to promoting the establishment of a systems approach to Big Data. The focus of this report will thus focus (i) on the actors and their role in the ecosystem; (ii) the potential role Big Data can play in the Policy Cycle, and (iii) Steps towards the Ecosystem's approach and UNESCAP's potential role.

2.2. Use of Big Data

The sheer volume of data generated, stored, and mined for insights has become economically relevant to businesses, government, and consumers. In the context of policy making, big data can be used to enhance awareness (e.g. capturing population sentiments), understanding (e.g. explaining changes in food prices), and/or forecasting (e.g. predicting human migration patterns). In most countries, public sector bodies also gather enormous amounts of data from censuses, tax returns, and public health surveys, for example. Much of this data is technically "public," but accessing it is not always easy, and mining it for relevant insights can require technical expertise and training that organizations and governments with limited resources can't always afford. Making good use of big data will require collaboration of various actors including data scientists and practitioners, leveraging their strengths to understand the technical possibilities as well as the context within which insights can be practically implemented.

Box 2: Twitter Example: Use of Mobile Technology for Perception Assessment

Since 2010, Indonesia has witnessed substantial increases in food prices: the price of rice increased 51% between December 2009 and February 2012. With more than 20 million Twitter user accounts in Jakarta, a wealth of data is being produced daily. Pulse Lab Jakarta analyzed Twitter conversations discussing food price increases between March 2011 and April 2013. Taxonomies, that are groups of words and phrases with related meanings, were developed in the Bahasa Indonesia language to identify relevant content. A classification algorithm was trained to categorize the extracted tweets as positive, negative, confused, or neutral to analyze their sentiment. Using simple time-series analysis, the researchers quantified the correlation between the volume of food-related Twitter conversations and official food inflation statistics. A relationship was found between retrospective official food inflation statistics and the number of tweets speaking about food price increases. Moreover, upon analyzing fuel price tweets, it was found that perceptions of food and fuel prices were related.

This big data example was created by the Global Pulse to demonstrate the relevance within the policy context to Government of Indonesia. [19].

The public sector cannot fully exploit Big Data without leadership from the private sector [20]. The conversation around Data Philanthropy - a term which describes a new form of partnership in which private sector companies share data for public benefit - has

advanced since its emergence at the World Economic Forum in Davos in 2011. Discussions about the concept of Data Philanthropy, or private sector data sharing, have gained momentum and moved forward, reaching a broader audience. In an article about the issue, Fast Company's Co. Exist, summarized: '(t)he next movement in charitable giving and corporate citizenship may be for corporations and governments to donate data, which could be used to help track diseases, avert economic crises, relieve traffic congestion, and aid development. The public sector isn't, however, the only one to gain from Data Philanthropy: companies donating data can get advantage from it too, especially those companies interested in the sustainable economy. These companies could enhance their role in corporate social responsibilities thus shaping their branding. Also, their role as stakeholders might change as they will get to influence policies and public opinion in a broader way than related to their very own business [21].

Big data is showing promise to improve, and perhaps substantively change, public sector and the international development sector in novel ways. Of general interest is the fact that big data often is produced at a much more disaggregated level, e.g. individual instead of a country level. Whereas aggregated data glosses over the often wide-ranging disparities within a population, disaggregated data allows decision makers more objectively to consider those portions of the population who were previously neglected.

2.3. Big Data and Open Data

In the context of policy making, it is worth to elaborate on the interface between big data and the new phenomenon of "open data"- they are closely related but are not the same. Open data brings a perspective that can make big data more useful, more democratic, and less threatening. While big data is defined by size, open data is defined by its use. But those judgments are subjective and dependent on technology: today's big data may not seem so big in a few years when data analysis and computing technology improve. All definitions of open data include two basic features: the data must be publicly available for anyone to use, and it must be licensed in a way that allows for its reuse. Open data should also be relatively easy to use, although there are gradations of "openness".

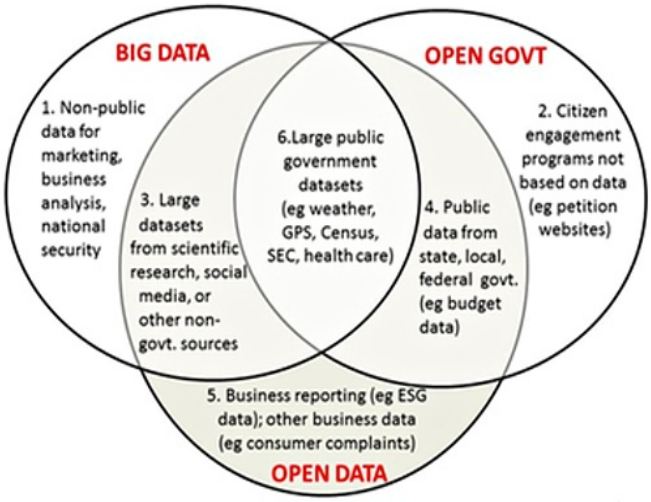


Figure 3: The Interface of Big Data and Open Data [22]

The diagram in Figure 3 maps the relationship between big data and open data, and how they relate to the broad concept of open government. There are a few important points to note:

a) Big data that is not open is not democratic: Section one of the diagram includes all kinds of big data that is kept from the public – like the data that large retailers hold on their customers, or national security data. This kind of big data gives an advantage to the people who control it.
b) Open data does not have to be big data to matter: Modest amounts of data, as shown in section four, can have a big impact when it is made public. Data from local governments, for example, can help citizens participate in local budgeting, choose healthcare, analyze the quality of local services, or build apps that help people navigate public transport.
c) Big, open data doesn't have to come from government: This is shown in section three. More and more scientists are sharing their research in a new, collaborative research model. Other researchers are using big data collected from social media – most of which is open to the public – to analyze public opinion and market trends.

But, when governments turn big data into open data, it is especially powerful: Government agencies have the capacity and funds to gather very large amounts of data, and opening up those datasets can have major social and economic benefits. Both big data and open data can transform business, government, and society – and a combination of the two is especially potent. Big data gives unprecedented power to understand, analyze, and ultimately change the world we live in. Open data ensures that power will be shared bearing huge potential to transform the way policies are made.

3. The Big Data Ecosystem

Unlike in other areas, the stakeholders in the Big Data sphere are not yet well connected and some processes need to be in place to bring them together. Making good use of big data will require collaboration of various actors including data scientists and practitioners, leveraging their strengths to understand the technical possibilities as well as the context within which insights can be practically implemented [22].

Policy stakeholders act at the international, regional, national and local level. When looking at the government actors, no single type of responsible authority emerges as a clear leader in the implementation of innovative data for policy initiatives, with the clear implication that there are opportunities for many different stakeholders and actors.

3.1. Types of Stakeholders

The EU data for policy report [23] distinguishes between the following types of stakeholders: global and European policy makers; national policy makers; regional policy makers; statistical offices; science and R&D organisations; data brokers; private providers of data analytics and visualisation tools; civil society and the policy analysis/evaluation community. For the purpose of outlining the relevant stakeholders, this report adopts the EU stakeholder categories. In the EU for example, Big Data is stimulated to promote jobs and economic growth, to promote industrial leadership and an open society (open data). It is connected to the many societal challenges that the European Commission has defined, among which are 'health, demographic change and wellbeing', 'smart, green and integrated transport' and 'climate action, environment, resource efficiency and raw materials'. However, no projects could be found in which the European Commission uses big data itself for direct use in its own policy cycle.

On a national policy-making scale, big data is often used in the areas of transport, where innovative sensor-data provides relevant information. Moreover, it is useful in detecting fraud, reducing crime and improving national security, both via defence and intelligence. National policy makers potentially possess a lot of data that could be used for informed policymaking using big data analyses. Opening up these data could be a first step (open data). Furthermore, the organisations of these policy makers have significant financial means to set up projects and improve big data for policy.

At the regional level big data could address policy issues concerning traffic, road safety, critical infrastructure, waste management, safety and security and public health. In contrast to national policymaking, data for regional policy focuses more on the policy implementation instead of agenda setting.

The statistical offices use big data to acquire better official statistics for policy means. These may concern all sorts of policy areas. Societal challenges that could be addressed are, for example, energy efficiency, infrastructure, smart transport and demographic change. The most relevant resources that the statistical offices have are knowledge and

skills related to statistically analysing large sets of data. They may also have the needed technological infrastructure to store and process big data. They have financial means to acquire and analyse data for official policy. Still, they may have to expand the experience and IT knowledge and equipment needed for big data. Most pilots are performed in cooperation with external institutes. The main benefits of big data for the statistical offices is improving the accuracy, timeliness and relevance of their statistics and reducing costs. For example, using social media data and having access to data about offline and offline retail revenues is less expensive than large-scale surveys (re-using and matching data versus collecting data).

The science community supports policy makers in all policy areas, on all governmental levels and in all steps of the policy cycles. Concerning science policy, the main policy questions have included how to promote an environment, which protects intellectual property and supports the most effective organization of disciplines and teams and resources. Steering the large resources devoted to research into the most useful and beneficial channels can be of great benefit to society, and this area has been one where there is great sophistication in the analysis and much data available. The science community has knowledge and skills related to statistically analysing large sets of data and using an evidence-based approach when researching the data-driven approaches. They often also have the needed technological infrastructure to store and process big data. They have financial means to conduct research. Moreover they have the possibility to connect multiple disciplines in their research (as the data centres demonstrate). Lastly, they possess or have access to a vast amount of large data sets (e.g. climate data, civil engineering data, social and behavioural data) and can thus more easily connect different data sets.

Data brokers could provide their data for all kinds of societal challenges and/or policy areas. Those are usually companies that collect information, including personal information about users, from a wide variety of sources for the purpose of reselling such information. An example is healthcare, in which Google Flue Trends is active. Data brokers often do not analyse or actually use the data; they often only provide it for the other actors. Data brokers have as their main resource data sets on specific groups or on societies as large. Furthermore, they have knowledge of and skills in data collection and analysis, for which they have dedicated tools. As most of the data is commercially traded, they have the financial means and incentives to invest in the improvement of data collection, storage and analysis.

3.2. Roles of Stakeholders in the Ecosystem

Table 1: Roles of Stakeholders in Data Ecosystem

	Governments	Multi-National Organizations	Statistical Bodies	R&D Bodies	Civil Society	Private Providers
Data	x		x	x	x	x
Financial Resources	x	x	x	x		x
Standards and Regulatory Frameworks	x	x				
Skills and Knowledge		x	(x)	x		x
Brokering, Facilitation, Capacity Strengthening	x	x	x	x		
IT Infrastructure			x	x		x

- **Governments** should empower public institutions to respond to the data revolution and put in place regulatory frameworks that ensure robust data privacy and data protection, and promote the release of data as open data by data producers, and strengthen capacity for continuous data innovation.
- **Multinational organizations**, **donors, governments and semi-public institutions** should invest in data, providing resources to countries and regions where statistical and technical capacity is weak. They should develop infrastructures and implement standards to continuously improve and maintain data quality and usability; keep data open and useable by all. They should also finance analytical research in forward-looking and experimental subjects.
- **International and regional organizations** should work with other stakeholders to set and enforce common standards for data collection, production, anonymization, sharing and use to ensure that new data flows are safely and ethically transformed into global public goods, and maintain a system of quality control and audit for all systems and all data producers and users. They also should support countries in their capacity-building efforts.
- **Statistical systems** should be empowered, resourced and independent, to quickly adapt to the new world of data to collect, process, disseminate and use high-quality, open, disaggregated and geo-coded data, both quantitative and qualitative.
- **All public, private and civil society** data producers should share data and the methods used to process them, according to globally, regionally, or nationally brokered agreements and norms. They should publish data, geospatial

information and statistics in open formats and with open terms of use, following global common principles and technical standards, to maintain quality and openness and protect privacy.

- **Governments, civil society, academia and the philanthropic sector** should work together to raise awareness of publicly available data, to strengthen the data and statistical literacy ("numeracy") of citizens, the media, and other "infomediaries", ensuring that all people have capacity to input into and evaluate the quality of data and use them for their own decisions, as well as to fully participate in initiatives to foster citizenship in the information age.
- **The private sector** should report on its activities using common global standards for integrating data on its economic, environmental and human-rights activities and impacts, building on and strengthening the collaboration already established among institutions that set standards for business reporting.
- **Civil society organizations** and individuals should hold governments and companies accountable using evidence on the impact of their actions, provide feedback to data producers, develop data literacy and help communities and individuals to generate and use data, to ensure accountability and make better decisions for themselves.
- **Academics and scientists** should carry out analyses based on data coming from multiple sources providing long-term perspectives, knowledge and data resources to guide sustainable development at global, regional, national, and local scales. They should make demographic and scientific data as open as possible for public and private use in sustainable development; provide feedback and independent advice and expertise to support accountability and more effective decision-making, and provide leadership in education, outreach, and capacity building efforts.

Therefore, the different stakeholders for big data, which includes owners and users, should ideally emerge into a "global data system", or big data ecosystem, to support policy making. However, the challenge will be in how to bring these different stakeholders and systems together to make the data revolution happen. These stakeholders are operating within their systems and procedures and it is important that fora and platforms are being established and managed effectively to make the big data system work.

Effective application of Big Data for Development would also require changes in the decision-making process, which customarily relies on traditional statistics. Given the high frequency of Big Data, a more responsive mechanism will need to be put in place that allows the government to process the information and act quickly in response. Also, since Big Data is often unstructured and relatively imprecise (compared to official statistics), government officials also have to learn how to effectively interpret and make use of the information provided by Big Data. This requires capacity building to turn decision makers into more sophisticated data users.

4. Big Data and Policy Making

Big data strategies for development can be important tools to formulate policies that also help successfully implementing the SDGs. However, many emerging economies or developing countries are still struggling with collecting and managing much smaller data sets and statistics. While a lot of "smaller" data exists [24], it is often not integrated, patchy and of low quality. Also, these statistics are often top-down and are missing a feedback loop to communities. The big data discussion might overlook the very fact that capacity constraints are one challenge that needs to be systematically addressed as part of the big data discussion.

4.1. Best Practices

The discussion of data-driven approaches to support policy making commonly distinguishes between two main types and uses of data. The first is the use of public data sets (administrative (open) data and statistics about populations, economic indicators, education etc.) that typically contain descriptive statistics, which are now used on a larger scale, used more intensively, and linked. The second is data from social media, sensors and mobile phones, which are typically new sources for policy making.

Best practices are still evolving where innovative approaches complement existing uses of data for policy. According to a study for the EU, the most common uses of big data in policy making include pilots where new sources of data are being used for agenda-setting and policy implementation; use of open data for transparency, accountability and participation and using administrative and statistical data for monitoring the outputs and impact of policies. Below (Box 4) an example of a state of the art tool (APPA) that is revolutionizing elements of policy making.

Countries in the Asia and the Pacific region, including among others Singapore, Indonesia, Republic of Korea, and the Philippines, as well as the US and Japan are already successfully innovating with and opening up data to solve complex policy problems, increase allocative efficiency and improve democratic processes [25].

Data analysis in the process component of the Policy Circle is more complex than in problem identification because policymakers weigh their decisions on a number of criteria. Data analysis expands from the technical aspects of an issue and focuses on the political costs and benefits of policy reform [3] to posit that policymakers tend to make their decisions based on a number of criteria, including: 1) the technical merits of the issue; 2) the potential affects of the policy on political relationships within the bureaucracy and between groups in government and their beneficiaries; 3) the potential impact of the policy change on the regime's stability and support; 4) the perceived severity of the problem and whether or not the government is in crisis; and 5) pressure, support, or opposition from international aid agencies [26].

Rather, big data is an additional means that has huge potential to improve policies. Interestingly, an EU study [27] finds that mostly big data is used at the early stage of the policy cycle, by making use of data and foresight, agenda setting, problem analysis and for identification and design of policy options. According to the study, less than a third of initiatives have a focus on the middle-stage policy cycles for the implementation of policies and interim evaluation.

Also, this stock take finds that big data is not (yet) playing a crucial role in policy making. If at all, it is used at the agenda setting stage and/or evaluation stage of policy making. One of the reasons might be that because the ecosystem is not yet functioning and crucial elements, such as standards and frameworks are still missing. National governments and other policy makers are just starting to systematically engage with big data for policy making.

4.2 The Policy Cycle

There are opportunities for full-scale implementation of data-driven approaches across all stages of the policy cycle, including evaluation and impact assessment. The following section identifies some data driven approaches in each step of the policy cycle:

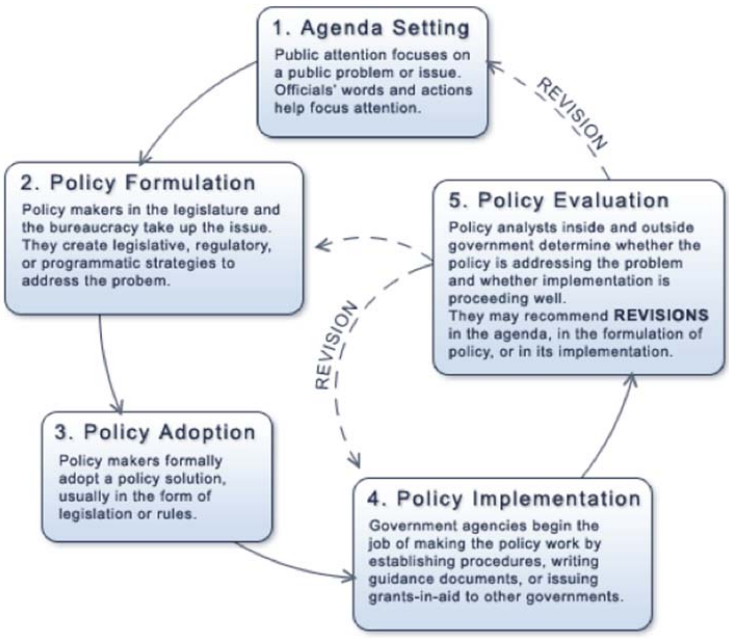


Figure 5: Policy-Making Process [28]

Policy Cycle Step 1-Agenda Setting: The agenda setting stage is one of the major steps in the policy making cycle. Once a problem requiring a policy solution has been identified, the process of policy development includes how the problem is framed by various stakeholders (issues framing), which problems make it onto the policymaking agenda, and how the policy (or law) is formulated. Together, these steps, determine whether a problem or policy proposal is acted on. Activities in policy development include advocacy and policy dialogue by stakeholders and data analysis to support each step of the process. Issue framing influences stakeholders' ability of getting the issue on the policymakers' agenda so that a problem is recognized and policy response is debated. Issue framing often sets the terms for policy debate. Agenda setting refers to actually getting the "problem" on the formal policy agenda of issues to be addressed by presidents, cabinet members, Parliament, Congress, or ministers of health, finance, education, or other relevant ministries.

Stakeholders outside of government can suggest issues to be addressed by policymakers, but government policymakers must become engaged in the process for a problem to be formally addressed through policy. Government policymaking bodies "can only do so much in its available time period, such as the calendar day, the term of office, or the legislative session. The items, which make it to the agenda pass through a competitive selection process, and not all problems will be addressed. Inevitably, some will be neglected, which means that some constituency will be denied. Among the potential agenda items are holdovers from the last time period or a reexamination of policies already implemented which may be failing" [29].

At any given time, policymakers are paying serious attention to relatively few of all possible issues or problems facing them as national or subnational policymakers. In decentralized systems, sometimes issues are placed on the agenda of various levels of government simultaneously to coordinate policymaking.

In order not to make things overwhelming, it is key to begin with questions that need to be answered in the policy making process, not with data. Once the setting for the analysis is defined, the focus of the research can move to the behaviors of interest and the consequent data generation process. Key exemplary strategies described in the boxes potentially can move the policy arena forward in a productive way. They are by no means exhaustive. Also, literature is actually missing on *how exactly* big data has influenced policy making vs. traditional data.

Policy Cycle Step 2-Policy Formulation: Policy formulation is the part of the process by which proposed actions are articulated, debated, and drafted into language for a law or policy. Written policies and laws go through many drafts before they are final. Wording that is not acceptable to policymakers key to passing laws or policies is revised. Policy formulation includes setting goals and outcomes of the policy or policies [30]. The goals and objectives may be general or narrow but should articulate the relevant activities and indicators by which they will be achieved and measured. The goals of a policy could include, for example, the creation of greater employment opportunities, improved

health status, or increased access to reproductive health services. Policy outcomes could include for example ensuring access to ARV treatment for HIV in the workplace or access to emergency obstetric care for pregnant women. Goals and outcomes can be assessed through a number of lenses, including gender and equity considerations. Activities Related to the Process—Advocacy, Policy Dialogue, and Data Analysis. While issues framing, agenda setting, and policy formulation are stages that policies go through, each of these stages can include a number of activities, namely advocacy, policy dialogue, and analysis of evidence related to the problem and policy responses.

The interpretation of this information will include various policy stakeholders- these include the legislature, CSO's and other relevant stakeholders. The executive will have to produce actionable insights with the possible objective of influencing the behaviors of interest considered. This also includes mapping the landscape – understanding the policy arena's issues and current challenges. Key players and stakeholders in the policy arena and their relationships to each other need to be identified and mobilized. Big data now allow creating multiple scenarios to understand how the policy landscape may evolve. Also, community participation can be enhanced with mobile technology.

Policy Cycle Step 3- Policy Adoption: The policy adoption process is typically still applying the conventional policy institutionalization methods- drafting laws and regulations. However, the dissemination of new policies can be faster and wider with the Internet, apps etc. The potential to the compliance and take-up of new policies can increase dramatically.

Of course, all this information is useless unless it is used to generate insights that leaders can act on. Fortunately, advances in analysis and visualisation tools (interactive charts, infographics, deep zooming applications, etc.) mean it is now feasible to bring granular and up-to-date evidence to bear on leadership challenges. This applies across the board – from analysing and optimising the impact of policies, through to gathering and acting on feedback from citizens on certain policies. In many instances, important sources of big data for learning live outside traditional organisational boundaries [31].

Policy Cycle Step 4- Policy Implementation: Procedures, guidelines and resources need to be made available for policy implementation. SIM Government (Box 3) is one of the few examples available where big data is used for policy implementation.

Box 3: SIMGovernment

Like the popular computer game *SimCity*, APPA creates a SimGovernment for policy makers to build possible policies and then test the effects of those policies in a realistic environment. As the amount of data grows and the analytic techniques become more sophisticated, it is possible to measure the impact of policies on other issue landscapes. For example, policy makers could model how a new health policy will affect environmental and educational issues, along with health issues.

A major advantage of APPA is that it will also help in identifying the undesired results of policies. With current policy making, it takes time to collect the data and observe the results of a policy. This delay often worsens the undesired effects of a policy- sometimes for years. With APPA, policy makers can spot and prevent the undesired effects of their policies before implementation. [32]

Policy Cycle Step 5- Policy Evaluation: Policies can be evaluated in a variety of informal and formal methods and this can be initiated and driven by a whole range of different actors, such as the legislature, CSO's, the executive, academia or other relevant stakeholders. However, formal methods tend to be difficult to carry out and informal methods can be riddled with bias. Policies can be evaluated while they are being implemented or after they have been implemented. They are difficult to evaluate when they aim to accomplish broad conceptual goals, have competing objectives, or possess multiple objectives. Most policies fail to be evaluated due to assessment difficulties and the tendency of the policy process to favor the status quo. Also, policy evaluations can be expensive to do. Public administrations are not necessarily well equipped to design evaluations- scope, sequencing, etc. External parties might be available, but this choice has not been widely applied [33].

In the policy evaluation and policy revisions elements, big data can potentially play a big role as it can provide feedback loops and information that was previously not available.

Box 4: Agile Predictive Policy Analysis (APPA)

Agile predictive policy analysis (APPA) is built upon the concepts behind other data-based policy making functions such as the Obama administration's PortfolioStat IT [34]. Data is blended from various sources to create a dashboard that displays key performance indicators where decision makers can create and monitor policies. This gives policy makers near-real-time feedback on the performance of policies and governance decisions.

The goal of APPA is to not only use data to accurately report on the current state of the agencies and policies but to create accurate models of the landscape, protagonists, and the policy struggle to create the most likely scenarios. This is accomplished by using transactional data sources from agency operations and using data science techniques such as machine learning and predictive analytics to better model agency decisions. The relationships between the agencies and other policy stakeholders are modeled along with any relevant environmental factors in the policy landscape. This all goes toward creating a simulation of the policy landscape, which gives both the current status and future scenarios.

There is nothing new in using research techniques developed in academia to analyze data by public policy practitioners. One can see a cycle where public agencies create the data and analytical challenges that lead to academic research in more effective policy making techniques, which in turn leads to even more data collection and more complex analytical challenges. APPA is the latest iteration in this cycle where complexity theory and data science will lead to more sophisticated policy making, which anticipates policy events rather than just reacts to them.

Big data technologies alone are not, however, a silver bullet for transforming the public sector. Underlying data issues like quality, standards and bias still need to be recognized and addressed. And governments must have the capability to conduct, interpret and consume the outputs of data and analytics work intelligently. This is only partly about cutting-edge data science skills. Just as important - if not more so -is ensuring that public sector leaders and policymakers are literate in the scientific method and confident combining big data with sound judgment. Governments will also need the courage to pursue this agenda with strong ethics and integrity. The same technology that holds so much potential also makes it possible to put intense pressure on civil liberties.

5. Challenges and Opportunities

Several challenges and considerations with big data must be kept in mind. This report touches on some of them and does not pretend to provide answers and solutions but rather to promote discussion.

A World Bank study [35] shows that about half of the 155 countries lack adequate data to monitor poverty and, as a result, the poorest people in these countries often remain invisible. During the 10-year period between 2002 and 2011, as many as 57 countries (37 per cent) had none or only one poverty rate estimate. Lack of well-functioning civil registration systems with national coverage also results in serious data gaps.

5.1. Challenges

Institutional Frameworks

Institutional frameworks, meaning the institutions that are required to protect pillars of democracy such as privacy, are often not in place when it comes to big data. This is a key challenge that needs to be addressed in order to scale up the use and useability of big data for sustainable development. Privacy, defined as the right of individuals to control what information related to them may be disclosed, is a pillar of democracy, and protections must be put in place to avoid compromising this basic human right in the digital age. Privacy is an overarching concern for anyone wishing to explore Big Data for

development, since it has implications for all areas of work, from data acquisition and storage to retention, use and presentation. In many cases, the production of data itself raises concerns, as people may be unaware of the sheer quantity or types of data they are generating on a daily basis, as well as that data they unknowingly consent to the collection and usage of without understanding how it may be used [36]. In this context, it is important to note that suitable legal frameworks, ethical guidelines and technological solutions for protected data sharing are at the center of efforts to leverage Big Data for development.

Digital Divide

Although the data revolution is unfolding around the world in different ways and at different speeds, the digital divide is closing faster than many had anticipated. The availability and types of digital data, however, differ from country to country. For instance, countries with high mobile phone and Internet penetration rates will produce more data directly generated by citizens, while nations with large aid communities will produce more program-related data. Data also varies between age groups, economic income brackets, gender and geographic location. These types of biases must be addressed in the way Big Data can influence policies, and particular attention must be given to the countries that are producing less data and/or have less capacity in data analytics to avoid adding new facets to digital divide [37]. It is important to also highlight that it is not only a digital divide between countries, but also *within* countries. Are the poorest of the poor able to access any of the technologies or services that would collect their data-are they then represented in big data statistics and figures? Analysis of big data results has to take this into consideration.

Access and Partnerships

Although much of the publicly available online data has potential utility for development purposes, private sector corporations hold a great deal more data that is valuable for development. Companies may be reluctant to share data due to concerns about competitiveness and their customers' privacy. Working with big data requires a new form of partnership between data makers, data users (see data system section above), and data storage stakeholders/institutions to ensure that the potential of big data is realized. It is a new way of working, and the challenge of bridging the worlds together is a big one.

Analytical and Capacity Challenges

The process of mining Big Data (using Big Data analytics techniques to extract relevant information) contains certain analytical risks that may reduce the accuracy of the results. Analyzing Big Data for policy inputs and evaluations poses different challenges that are in part methodological, or related to interpretation accuracy, methods of analysis, and detection of anomalies [38], which will be not further, elaborated in this report.

The capacity to effectively utilize all the potential that big data brings along is still very limited. The institutional frameworks missing also impede the strengthening of capacities of different stakeholders and their roles.

5.2. Opportunities

The use for big data for policy making is about turning imperfect, complex and often unstructured data into actionable information. Despite the many challenges that big data analysis presents, understanding the growing amount of digital information human communities' produce can be invaluable in providing them with support and protection.

Citizen-Focus and Participation

Big data offers a chance for policy-making and implementation to be more citizen-focused, taking account of citizens' needs, preferences and actual experience of public services, as recorded on social media and other platforms [39]. As citizens express policy opinions on social networking sites such as Twitter and Facebook or rate or rank services or agencies on government applications, policy makers also have access to a huge range of data on citizens' actual behavior, as recorded digitally whenever citizens interact with government administration or undertake some act of civic engagement, such as signing a petition. Data mined from social media or administrative operations in this way also provide a range of new data which can enable government agencies to monitor – and improve – their own performance, for example through log usage data of their own electronic presence or transactions recorded on internal information systems, which are increasingly interlinked. And they can use data from social media for self-improvement, by understanding what people are saying about government, and which policies, services or providers are attracting negative opinions and complaints, enabling identification of a failing school, hospital or contractor, for example. They can solicit such data via their own sites, or those of social enterprises. And they can find out what people are concerned about or looking for. Efficient procedures to draw links between large-scale data-processing technologies and existing expert knowledge in major policy domains would potentially offer chances to make policy development processes more citizen-focused, taking into account public needs and preferences supported with actual experiences of public services.

Big data can contribute to the transformation of citizen-state relations. Data can be used to track service provision, enable citizens to reallocate local budgets, make changes in their communities, hold their governments to account and to participate better in democratic processes to ensure their needs and concerns count-often for the first time.

Evidence/ More and Better Analytics

The notion that policy decision should be based on sound evidence has become widely adopted by many public administrations. Strengthening science-policy interface is also highlighted in the Rio+20 outcome document "The Future We Want [40] as well as the 2030 Agenda for Sustainable development. Data technologies are amongst the valuable tools that policymakers have at hand for informing the policy process, from identifying issues, to designing their intervention and monitoring results. More data often means we can do more with analytics, especially advanced analytics.

Big data and new forms of data collection will give citizens new information they need to live better lives and earn more secure livelihoods. They can tell people the best time to avoid traffic, when best to plant crops, and which waterholes are free from arsenic, fluoride, iron, and parasites.

Variety

Validation is a key success factor to benefit from analytical insights. The variety of data available nowadays makes it easier to determine if certain insights are consistent with data from multiple sources (triangulation). Given the low cost of attaining and the size of available data, replication is now often easier, and anything online can be easily tested.

Real-Time Information

Real-timeliness refers to data being available much faster and sometimes in real-time. Internal data can be available in a week; clickstream data could probably be obtained an hour after it is captured-provided the initial setup and coding has been done — and social media comments can be watched in real-time. It is widely believed that the use of information technology can reduce the cost of public services while improving its quality. Data can be routinely captured and created in the day-to-day business of government. It is important to note that, for the purpose of global development, "real time" does not always mean occurring immediately, but rather refers to information that is produced and made available in a relatively short and relevant period of time and within a timeframe that allows action to be taken in response, creating a feedback loop.

Early Warning System

Data collected through new technologies can act as an early-warning system. Even if we do not know at the macro level the precise number of clinics or pharmacies that stock vital medicines, if people can alert their government via SMS to stock outs, this signals problems in a certain area, meaning that action can be taken before a full dataset is available.

Economic Value

Good quality data yield not only social benefits, but also real economic returns, such that, in the medium term, a data revolution could pay for itself. First, if governments

invest in better economic data, this can improve investor confidence. The IMF has found that, if countries invest in better-quality data, it is cheaper for them to borrow internationally. It investigated the effect of its data standards on sovereign borrowing costs in 26 emerging market and developing countries and estimated that countries that sign up to its more stringent data standard reduce borrowing spreads (that is, the cost of borrowing) by an average of 20% [24].

Another important aspect is the cost reduction in policy making- replacing or substituting traditional data collection and evaluation methods. The large amount of data readily available will enable more timely analysis of policy interventions.

6. Big Data and Policy for the 2030 Agenda for Sustainable Development

Therefore, the section below is suggesting *actionable* steps that can be taken by policy stakeholders. Certainly, it is not suggested to dismiss other methods of gaining inputs for policy making. Also, big data is not necessarily for policy making in every sector. Data driven methods are especially beneficial for policy areas with large volumes of data- such as health, macroeconomics, transport, migration and the environment.

The availability of big data provides a unique opportunity to support the achievements of the SDG's like never before. As the post-2015 development agenda has now been established, strengthening data production and the use of better data in policymaking and monitoring are becoming increasingly recognized as fundamental means for development. The MDG monitoring experience has clearly demonstrated that effective use of data can help to galvanize development efforts, implement successful targeted interventions, track performance and improve accountability. Thus, the sustainable development demands a data revolution to improve the availability, quality, timeliness and disaggregation of data to support the implementation of the new development agenda at all levels in all regions. Big Data is an essential part of the data revolution, and chapter six below identifies potential areas where UN-ESCAP can play a vital role to support policy making using big data for sustainable development. Localizing the SDG's based on local priorities will be key to make them tangible and relevant targets.

6.1. A Vision for Big Data and the 2030 Agenda

As described above in chapter three on ecosystems, the very nature of big data requires new forms of inter-institutional relationships in order to leverage data resources, human talent, and decision-making capacity. The necessary capabilities enable the integration of big data into ongoing policy processes rather than one-time policy decisions, thereby enabling its value to be continually released and refined. Spaces will be needed in which technical, cultural, and institutional capabilities can commensurately develop. Given the variety and pervasiveness of the necessary capabilities to utilize big data to address big problems, collaborative spaces are needed

to enhance the capacity of individuals, organizations, businesses and institutions to elucidate challenges and solutions in an interactive manner, strengthening a global culture of learning.

Some elements of this new ecosystem are already emerging. The UN Statistical Commission established a global working group (GWG) mandated to provide strategic vision, direction and coordination of a global program on Big Data for official statistics [23]. The group found that nontraditional sources of data, especially big data that thus far have been underutilized in producing official statistics. Big Data sources need to be leveraged and considered for adequacy to enrich the sources of official statistics so that the data needs in new development areas can be satisfied and timely, detailed and spatially disaggregated data can be produced and made available to decision makers. This implies that the innovative and transformative power of information technology may be harnessed: from the collection stage (through, for example, the use of computer-assisted collections through mobile devices), to the dissemination stage (through advanced visualization tools, such as data on maps).

6.2. Possible Action

The UN's Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG) is calling for action to mobilize the data revolution for sustainable development [4]. The recommendations out of this group relevant for better policy making have been taken as a basis for a possible action steps described below.

UNESCAP is in a unique position to support emerging groups and networks, leveraging existing knowledge and resources. It can facilitate dialogue and bring technical expertise into the consultations- agreeing on concrete actions and shared responsibilities among all stakeholders.

In order to gain the maximum benefit that big data offers to the policy making process and sustainable development, it is important that various aspects are addressed in a holistic way. These aspects include high level stakeholder agreements as well as access to innovations and capacity strengthening. Based on the recommendations of the UN Independent Advisory Group on a Data Revolution of Sustainable Development [4] and UN's Global Pulse [41], the following steps are recommended as an initial suggested regional implementation roadmap:

1. **Establish and manage a coordination mechanism with the key UN stakeholders,** (Global Working Group (GWG) on Big Data for Official Statistics, Global Pulse and its regional offices) and other international partners. Resources are limited, and an effective coordination can enhance knowledge sharing, faster replication of innovations and advance progress in a novel way.
2. **Develop a consensus on principles and standards among the UNESCAP member countries.** This would include a participatory and inclusive series of stakeholder meetings, bringing together the public, private and civil society to build trust and

confidence among data users. This can then feed into the "Global Consensus on Data" to be facilitated by the UN.

In addition, UNESCAP governments should be brought together as a subgroup of stakeholders-focusing specifically on the *use and availability* of data for policy making.

3. **Kick off and Institutionalize a Regional Multi-Stakeholder Mechanism**
    - **to share innovations.** The Pulse Lab Jakarta could support this effort. The ultimate mechanism can be a digital network, equivalent to a UNESCAP Community of Practice (CoP) on Big Data based Policy Innovations. This CoP could then also lead the identification of specific areas of innovation focus (e.g. incentives, research etc.)
    - **to define Local Tangible Benefits-** Big Data should not be an end in itself. While the Big Data is an interesting area to explore in itself, it is important to bear in mind that the application of big data in the policy making discourse should ultimately benefit the people of the Asia Pacific Region by achieving the Sustainable Development Goals. It is not likely that projects using data will get this right the first time. It will be a matter of testing, re-testing, adjusting and learning. The point here is not to experiment all day in boutique labs with little regard to impact, but rather to integrate experimentation and adaptation at the heart of how we implement at scale.
    - **to initiate the Localization of SDG's indicators.** The big data discourse in this context can be used to engage Partner Governments in the drafting of targets and to include all relevant stakeholders.
4. **Mobilize regional resources for capacity development for the less advanced UNESCAP member countries**

    While big data is available in almost unlimited amounts, the capacity to actually use the data and to feed it into policy making, is limited, especially in many of the UNESCAP member countries. While some governments are quite advanced in the utilization of big data, such as South Korea and Singapore- other countries barely benefit from the data revolution.

    A capacity development approach based on peer to peer learning, *connected* to the above mentioned CoP's should be further discussed. Also, available resources should be mapped. In addition, UNESCAP should work with its networks to mobilize additional resources.
5. **Enhance in-house big data analytics capacity.** The geographic coverage of UNESCAP is huge and a support need among its members varies significantly. Additional technical resources are needed to maintain the momentum and make the big data revolution happen. UNESCAP to establish and manage a Regional Sustainable Development Big Data Policy Secretariat in collaboration with the Global Pulse and Pulse Lab Jakarta. This secretariat could lead the above mentioned stakeholders mechanism, provide analytical capacity strengthening, mobilize the stakeholders and coordinate the proposed actions.

These initial recommendations are supposed to stimulate a discussion during the Big Data and the 2030 Agenda for Sustainable Development: Achieving the Development Goals in the Asia and the Pacific Region Workshop in Bangkok on 14 - 15 December 2015.

References

[1] Carr-Hill, R. (2013) 'Missing millions and measuring development progress', World Development 46: 30-44.
[2] Granoff, I. et al. (2014) 'Targeting zero: achieving zero extreme poverty on the path to zero net emissions.' London: Overseas Development Institute.
[3] Elizabeth, Stuart, and Others (2015). The data revolution. Finding the missing millions. ODI Development progress.
[4] Helen Margetts (2013); in http://blogs.oii.ox.ac.uk/policy/promises-threats-big-data-for-public-policy making/
[5] United Nations (2014), p.6. A world that Counts: Mobilizing a Data Revolution for Sustainable Development by the Independent Expert Advisory Group on a Data Revolution for Sustainable Development. New York.
[6] UNECE Statistics Wikis.How Big is Big Data.
http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=99484307.
[7] World Bank (2014). Big Data in Action for Development. Central America.
[8] Dinesh Mavaluru, et al. (2014). Big Data Analytics in Information Retrieval: Promise and Potential. Proceedings of 08th IRF International Conference. Bengaluru, India.
[9] K, Arun; L. Jabasheela(2014).Big Data: Review, Classification and Analysis Survey. International Journal of Innovative Research in Information Security (IJIRIS) ISSN: 2349-7017(O). Volume 1 Issue 3.
[10] Emmanuel Letouzé (2015). Big Data and Development: An Overview.Data-Pop Alliance Primers Series.
[11] Emmanuel Letouzé (2012). Big Data for Development: What May Determine Success or failure? OECD Technology Foresight, Paris.
[12] Christopher, Surdak (2014).Data Crush: How the Information Tidal Wave is Driving New Business Opportunities.Amacom
[13] Hellerstein, Joe (9 November 2008).Parallel Programming in the Age of Big Data. Gigaom Blog.
[14] Segaran, Toby; Hammerbacher, Jeff (2009). Beautiful Data: The Stories Behind Elegant Data Solutions. O'Reilly Media. P257.
[15] Hilbert, Martin; López, Priscila (2011). The World technological capacity to Store, Communicate, and Compute Information. Science 332(6025). P60-65.
[16] IBM (2013). What is Big Data? Bringing Big Data to the Enterprise. www.ibm.com.
[17] http://www.kdnuggets.com/2015/04/interview-emmanuel-letouze-democratizing-benefits-big-data.html.
[18] Gary King (2013). https://gking.harvard.edu/files/gking/files/evbase-gs.pdf. Institute for Quantitative Social Science, Harvard University. Talk at the Golden Seeds Innovation Summit, New York City.
[19] United Nations Global Pulse (2014). Mining Indonesia Tweets to Understand Food Price Crises. New York.
[20] United Nations Global Pulse (2015). Blog Data Philanthropy: Where Are We Now? New York

[21] United Nations Global Pulse (2013). http://www.unglobalpulse.org/data-philanthropy-where-are-we-now. New York.
[22] Joel Gurin; in: http://www.theguardian.com/public-leaders-network/2014/apr/ 15/ big-data-open-data-transform-government.
[23] United Nations (2014), p.6. A world that Counts: Mobilizing a Data Revolution for Sustainable Development by the Independent Expert Advisory Group on a Data Revolution for Sustainable Development. New York.
[24] Martijn, Poel, and others (2015). Data for Policy: A study of big data and other innovative data-driven approaches for evidence-informed policymaking. Brussels.
[25] Joel Gurin; in: http://www.theguardian.com/public-leaders-network/2014/apr/ 15/ big-data-open-data-transform-government.
[26] Thomas, J., and M. Grindle. (1994). "Political Leadership and Policy Characteristics in Population Policy Reform." Population and Development Review 20 Supp: 51–70.
[27] http://www.policyproject.com/policycircle/content.cfm?a0=4
[28] Luis Crouch (2015). A Relevant Data Revolution for Development. RTI Press International. New York.
[29]https://texaspolitics.utexas.edu/archive/html/bur/features/0303_01/policy.htmlTh e Texas Politics Project at the University of Texas at Austin, USA.
[30] Hayes, M.T. (2001) The Limits of Policy Change: Incrementalism, Worldview, and the Rule of Law (Washington D.C.: Georgetown University Press).
[31] Isaacs, S. and Irvin, A. 1991. Population Policy: A manual for policymakers and planners, Second Edition. New York: The development Law and Policy Program, Center for Population and Family Health, Columbia University, and Futures Group.
[32] Chris Yiu (2013). The Big Data Opportunity Making government faster, smarter and more personal. Policy Exchange.
[33] William. A. Brantley (2012). Agile Policy Making: How Complexity Theory, Big Data and Data Science Research is Changing The Practice of Policy Making. American Society for Public Administration.
[34] Steven, Vanroekel (2013). PortfolioStat 2.0: Driving Better Management and Efficiency in Federal IT. White House, USA.
[35] Joel, Gurin, and Laura, Manley (2015). Open Data for Sustainable Development. World Bank.
[36] Kenneth, Neil, Cukier, and Viktor, Mayer-Schoenberger (2013). The Rise of Big Data. Foreign Affairs.
[37] Jackie, Hoi-Wai, Cheng, National Economist of UNDP China UNDP (2014). Big Data for Development in China. UNDP China.
[38] Data For Policy a study of big data and other innovative data-driven approaches for evidence-informed policy making. http://www.data4policy.eu/#!appendixb/c1kb8
(2015). The workshop on Data-Driven Innovations for Better Policies, Brussels.
[39] Rockefeller Foundation Bellagio Centre conference, (2014).Big data and positive social change in the developing world: A white paper for practitioners and researchers Oxford: Oxford Internet Institute.
[40] United Nations (2012). The Future We Want. Outcome Document of the United Nations Conference on Sustainable Development.Rio de Janeiro, Brazil.

[41] Global Pulse (2012). Big Data for Development: Challenges & Opportunities. New York.
[42] Shailendra Kumar (2014). The Data within Big Data and the myth around Unstructured Data.
[43] UNECE Statistics Wikis (2013). http://www1.unece.org/stat/platform/display/bigdata/Classification+of+Types+of+Big+Data.
[44] Judith. Hurwitz,et al (2013). Big Data For Dummies .John Wiley & Sons, Inc.

Annex 1: Big Data Types

Variety is one of the principles of Big Data as described previously. The Big Data can be divided into three types [42,43,44]: Structured Data, Semi-Structured Data, and Unstructured Data. Definitions and examples of each can be described as follows:

Structured Data

Structured data generally refers to data that has a defined length and format. Most organizations are storing large amounts of structured data in various divisions, in normalised/ deformalised formats in a database: Data warehouses, relational database management system (RDMSs), and various other environments. The data can be queried using a language like structured query language (SQL) in which the datasets can be updated with new data, and deleted, read or any other activity.

The evolution of technology provides newer sources of structured data being produced - often in real time and in large volumes. The sources of data are divided into three categories:

(a) Computer- or Machine-Generated Structured Data

Machine-generated data generally refers to data that is created by a machine without human intervention. They can include the following:

- **Sensor data:** Examples include radio frequency ID (RFID) tags, smart meters, medical devices, and Global Positioning System (GPS) data. Another example of sensor data is smartphones that contain sensors like GPS that can be used to understand customer behavior in new ways. For example, RFID is rapidly becoming a popular technology. It uses tiny computer chips to track items at a distance. An example of this is tracking containers of produce from one location to another. When information is transmitted from the receiver, it can go into a server and then be analyzed. Companies, for example, are interested in this for supply chain management and inventory control.

- **Web log data**: When servers, applications, networks, etc operate, they capture all kinds of data about their activity. This can amount to huge volumes of data that can be useful, for example, to deal with service-level agreements or to predict security breaches.

- **Point-of-sale data:** When the cashier swipes the bar code of any product that you are purchasing, all that data associated with the product is generated. Just

think of all the products across all the people who purchase them, and you can understand how big this data set can be.

- **Financial data:** Lots of financial systems are now programmatic; they are operated based on predefined rules that automate processes. Stocktrading data is a good example of this. It contains structured data such as the company symbol and dollar value. Some of this data is machine generated, and some is human generated.

(b) Human-Generated Data:

This is data that humans, in interaction with computers, supply.

- **Input data:** This is any piece of data that a human might input into a computer, such as name, age, income, non-free-form survey responses, and so on. This data can be useful to understand basic customer behavior.

- **Click-stream data:** Data is generated every time you click a link on a website. This data can be analyzed to determine customer behavior and buying patterns.

- **Gaming-related data**: Every move you make in a game can be recorded. This can be useful in understanding how end users move through a gaming portfolio.

The way data is structured is a vital element. If the structures aren't coherent and understandable, data is liable to be misused (misunderstood) and will fail to facilitate "bringing together" data from disparate sources to produce new knowledge/evidence. This is a metadata schema related issue – or brought down to a simple example what headings/terms are being used for columns of data in a spread sheet and how can the person using the spread sheet understand the context.

Semi-Structured Data

Semi-structured data is a kind of data that falls between structured and unstructured data. This type of data became a talking point. Mostly data coming from Facebook, Twitter, Blogs, publically available websites, etc. makes the basis of semi-structured data. These data sources usually have defined structures and mostly contain text information.

The free flow text generated through the social media is the only unstructured component whilst the remaining data is structured. Most of the times, the social data is mistaken with unstructured data. The social data is NOT unstructured data, it is semi-structured and in fact, some of the social data contains industry standard structures.

- **Social media data:** This data is generated from the social media platforms such as YouTube, Facebook, Twitter, LinkedIn, and Flickr.

Unstructured Data

DataUnstructured data does not have any defined, consistent fields and it may even do not have any numbers and text. Unstructured data can be divided also into either machine generated or human generated and described as flows:

(a) Machine-Generated Unstructured Data Examples

- **Satellite images:** This includes weather data or the data that the government captures in its satellite surveillance imagery. Just think about Google Earth, and you get the picture (pun intended).

- **Scientific data:** This includes seismic imagery, atmospheric data, and high energy physics.

- **Photographs and video:** This includes security, surveillance, and traffic video.

- **Radar or sonar data:** This includes vehicular, meteorological, and oceanographic seismic profiles.

(b) Human-generated Unstructured Data Examples

- **Mobile and Voice data:** This includes data such as text messages and location information. Human voice contains a lot of information and it needs access and mined. The spectrogram of the human voice reveals its rich harmonic content including pitch, tone, emotion, bass, etc.

- **Web behavior and content:** This comes from any site delivering unstructured content, like YouTube, Flickr, or Instagram. The scope of web behavior is huge. There are nearly five billion indexed web pages on the Internet and for each page there are traffic statistics ranging from the number and duration of visits to far richer information on user behavior on a large proportion of websites. Big Data also encompasses the content of those web pages and the changes that occur on them. Also included in this category is the vast amount of search engine data constantly being generated.

- **Image and Video Data:** Total number of pictures taken in last 5 years is more than double the pictures taken in 1900 -2000. This gives us an opportunity to use patterns within the pictures and mine the information available to us. Various techniques like pixilation, pattern matching, image processing, feature extracting, etc. allows us covert the pictures into data and further mine it using classification algorithms. Examples of image data use cases: One of the most

common use case is the thumb print recognition which is now available in our phones and one large bank is using the image mining technique and predicting likelihood of a customer to be fraud while in case of the video One very large security agency uses the video data to identify trouble making candidates in the premise by using predictive analysis based on the sequence of actions performed by the individual.

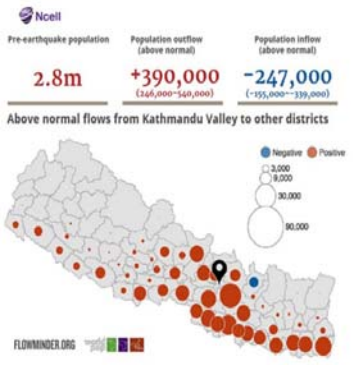
- **Machine Data:** As the size of computer chips is reducing, there is potential of having a computer chip in almost all the machines, e.g. cars, the mobile phones, ships, etc. The data residing in these machines is unstructured and is not of a standard format to be available for mining. This unstructured data is being extracted by large organisations and then used to understand the hidden patterns to drive efficiency. Example of machine data use case: A large Telco in the US is using mobile app data to advertise and promote retail offers by understand customer behaviour and A large car company is collecting data from the cars to understand the reason behind engine failure to optimise the performance and reduce engine failure possibilities.

Annex 2: Big Data Case Studies

1. Case Study I

USING FLOWMINDER TO FOLLOW POPULATION DISPLACEMENT AFTER THE NEPAL EARTHQUAKE IN APRIL 2015

Flowminder.org developed a tool to provide key information on large scale displacement taking place after the Nepal disaster. Through the use of anonymous mobile operator data they were able to measure and visualize population movements and this resulted in more equitable support to people struck by the earthquake regardless of their location.



PROBLEM: Asia Pacific is the most disaster prone region of the world. Annually, millions of people remain at risk to earthquakes, tsunamis, tropical cyclones, typhoons, floods and storm surges[1]. The poor are more impacted by natural disasters because they are more vulnerable and usually their livelihoods depend on climate and land based subsistence. They are also less likely to have social protections, insurance, or capacity to recover after a disaster. Thus, disaster and risk reduction policies and measures should be incorporated in poverty reduction, development and environmental strategies to create more disaster resilient societies and communities, facing decreased level of risk and vulnerability. Following major disasters there is a pattern of population movement, and twenty to thirty million people are displaced due to natural disasters every year. In most cases, traditional tools used in disaster response and preparedness - including eyewitness accounts, manual counting of people, registration in camps or satellite or aerial images of shelters or changes in vegetation[2] - are not able to document in a timely and accurate manner. Predicting and monitoring population displacement can reduce the population's vulnerability and help provide targeted relief assistance and prevent diseases.

USING BIG DATA TO UNDERSTAND AFFECTED POPULATION MOVEMENTS DURING A DISASTER: As opposed to traditional disaster response and preparedness tools, utilizing Big Data and new technologies can offer an excellent alternative to map affected people and their movements. Flowminder[3] works with large mobile operator's databases. The underlying technology that Flowminder is using refers to geographic positions of SIM cards which are determined by the location of the mobile phone tower through which each SIM card connects when calling. Through analysis of these data sets, Flowminder

---

1 ESCAP Trust Fund for Tsunami, Disaster and Climate Preparedness Brochure

2 Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J (2011) Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. PLoS Med 8(8): e1001083. doi:10.1371/journal.pmed.1001083

3 Flowminder Foundation is a non-profit organization with a mission to provide global public goods through the collection, analysis and integration of anonymous mobile operator, satellite and household survey data

can map the distributions and characteristics of vulnerable populations in low and middle income countries.

Following the devastating 7.8 magnitude earthquake on 25 April 2015 in Nepal in the Ghorka district which killed more than 9,000 people and caused injuries to more than 23,000, Flowminder supported the Nepali Government, United Nations entities and other relief agencies with displacement analyses. Flowminder entered into a partnership with Ncell, the largest mobile operator in Nepal to have access to the anonymized data of 12 million phones. As shown in the <Figure 1>, the pre-earthquake population was 2.8 million with abnormal outflows from the Kathmandu Valley to other districts of 390,000 people[4,5].

KEY PLAYERS: Nepali Government, Flow minder, UN relief agencies, NCell

OUTCOMES: The data that Flowminder gathered and analyzed with the contribution of Ncell were shared with different UN and non-UN relief actors such as the UN Office for Coordination of Humanitarian Affairs (OCHA), UN World Food Program, and the International Office for Migration. The information was used to plan aid distribution and estimate the number of people affected. Organizations can use the real-time data to understand displacement mechanisms and develop targeted systems for provision of relief response[6.] In the case of Nepal Flowminder , through their analysis, found that after adjusting for normal movement patterns, which would have taken place in the absence of the earthquake, an estimated additional 500,000 people had left the Kathmandu Valley two weeks after the earthquake. The majority of these went to the surrounding districts and the Terai areas in the South and Southeast of Nepal (Flowminder Nepal Case study). Though analysis of the Nepal research results is ongoing, a previous study conducted by the Flowminder team after the Haiti earthquake in 2010 showed that there was a correlation of displaced people's destination to where they had significant social bonds[7] . Big Data offer unprecedented insight into human behavior that is unparalleled to the previous methods enlisting surveys and static methods to collect self-reported indications of action.

CONCLUSIONS: Natural Disasters are a major threat to Sustainable Development. Some countries, especially those countries with special needs, do not yet have the mechanisms in place to provide effective disaster response and preparedness. As a country highly prone to natural disasters, a priority for Nepal is development of policies and practices that emphasize disaster resilience and preparedness to minimize the impact on poverty eradication and sustainable development efforts. Big Data offers an opportunity to enhance early warning systems, strengthen resilience and ensure

---

4 Nepal Earthquake 2015, Flowminder Case Study (2015)

5 Ncell Picture :Accessible at http://i.imgur.com/xnGbX92.jpg?1

6 Nepal Earthquake 2015, Flowminder Case Study (2015)

7 Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J (2011) Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. PLoS Med 8(8): e1001083. doi:10.1371/journal.pmed.1001083

efficient and effective action after a disaster has occurred to limit damage. More specifically, the utilization of technology can massively increase the efficiency of provision of aid and better structures for relief response for displaced populations. The real time predictive mechanism that Flowminder is using through the analysis of large data sets can give an insight in population displacement immediately following a natural disaster. The type of insights that we can draw from large data sets shed light to human behavior regarding 1) mobility, 2) social interaction and 3) economic activity[8]. These data can help policymakers identify the appropriate policy strategies associated with disaster response. Knowing where the displaced populations are amassed can lead agencies to better target poverty-alleviation policies such as food and nutrition, unemployment assistance and microfinance[9].It is clear that there is a role for multi-stakeholder partnerships to deliver on the potential of the use of big data in disaster preparedness and response. In the case of Nepal, the private sector (NCell) got involved and after the devastating earthquake they utilized data-sharing for social good. Data sharing or Data Philanthropy is essential to ensure free access to large data sets which can be used to improve public policies.

Prepared by Erifyli Nomikou, Consultant, EDD/ESCAP

References:

[1] ESCAP Trust Fund for Tsunami, Disaster and Climate Preparedness Brochure

[2,7] Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J (2011) Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. PLoS Med 8(8): e1001083. doi:10.1371/journal.pmed.1001083

[3] Flowminder Foundation. About us page Accessible : http://www.flowminder.org/about

[4,6] Nepal Earthquake 2015, Flowminder Case Study (2015)

[5] Picture Credits Ncell :Accessible at : http://i.imgur.com/xnGbX92.jpg?1

[5] Shakya.A, 29 May - 5 June 2015 #760 " Where are we: Ncell partners with Flowminder to track movement of Nepalis post-earthquake" Accessible at : http://nepalitimes.com/article/nation/Ncell-Flowminder-track-movement-of-nepalis-post-earthquake,2278

[8] United Nations Global Pulse (October 2013) Mobile Phone Network Data for Development.

[9] OECD Policy Development Center, Natural Disaster and Vulnerability , Policy briefing 29 Accessible: http://www.oecd.org/dev/37860801.pdf

8 United Nations Global Pulse (October 2013) Mobile Phone Network Data for Development.

9 OECD Policy Development Center, Natural Disaster and Vulnerability , Policy briefing 29

2. Case Study II

USING BIG DATA TO SUPPORT E-WASTE MANAGEMENT IN CHINA

Baidu Recycle, a web based application launched by UNDP and Baidu, helps to properly dispose e-waste in China. China is the second biggest e-waste producer and biggest e-waste importer. The application has successfully been used to collect and recycle 11,429 electronic items since its inception in August 2014. In November 2015, the Baidu Recycle Green Service Alliance was established by Baidu and UNDP to further help the App scale up and promote and internet –based nationwide e-waste management ecosystem.

PROBLEM: Asia-Pacific is among the world's top regions generating and importing high levels of electrical and electronic equipment waste or e-waste for short. E-waste covers items of all types of electrical and electronic equipment (EEE) and its parts that have been discarded by the owner as waste without the intention of re-use[10]. China is the second biggest e-waste producer and biggest e-waste importer[11]. Chinese national e-waste grew from 2009 to 2013 at an annual average of 21.6%, out of the 3.6 million tons of e-waste being generated domestically, only about 40% were processed by formal channels[12]. The informal sector plays an important role in the collection and disposal of e-waste in China and other emerging countries.

Ensuring sustainable consumption and production patterns (SDG 12) is important for proper e-waste management and resource efficiency. A commitment from member states to international regulation and technical standards would enhance environmental sustainability, ensure that precious and scarce resources are not lost and lead to healthier environments which promote the human well-being (SDG 3).

USING BIG DATA TO ENSURE PROPER E-WASTE MANAGMENT: Technology can fuel innovative ways to manage e-waste and create responsible recycling behavior. Baidu Inc., a leading Chinese company in web services, and UNDP entered into a strategic partnership to co-create a Big Data joint laboratory. Beyond its primary focus on environmental issues, the laboratory will explore the use of Big Data technologies to solve other global problems such as health, education and disasters. The first product launched by the UNDP-Baidu Joint Big Data Lab was a web based application aiming to improve monitoring of e-waste disposal and recycling behavior, and raise awareness about environmentally appropriate approaches to e-waste disposal through procedures that do not fall in the informal market. The users do not need to download an app but rather use a picture of their electronic device on Baidu's Recycle search app.

The result of that research yields name, type and estimated value of the electronic item. Then users can arrange for a door-to door e-waste pick up. The success of the first version of this web based app led to the expansion of the apps service[13.] Initially the coverage of the research databases included only TVs, washing machines, refrigerators and digital products which expanded to include cell-phones and laptops.

---

10 Solving the E-Waste Problem (Step) Initiative White Paper. "One Global Definition of E-waste". (2014)

11 Cheng, J. UNDP China working Paper. "Big Data for Development in China" (2014)

12 UNDP. "Harnessing the Power of Big Data" (2014)

13 UNDP. "China: Turning E-Trash into Cash". (2015)

KEY PLAYERS: Government of the People's Republic of China, UNDP, Baidu Inc.

INSIGHTS & OUTCOMES: Baidu has been vastly successful in helping to develop intelligent solutions for e-waste recycling. Using photographs to match electronic equipment across different types with data sets is an innovative way to allow customers to share their disposal needs while creating an efficient management of e-waste. In August 2015, according to UNDP, 11,429 electronic items have been successfully recycled and treated, 370,000 page views of Baidu Recycle App had been reached and the total daily searches for the app numbered 50,000[14]. These data show the tremendous potential for the Recycle app to scale up and reach other cities in the world's most populated country. The use of the Baidu Recycle app actively renders citizens to develop greener recycle conscious and contribute to the cut down of the informal recycle stations. In a continuous effort to support this initiative, in November 2015, Baidu and UNDP launched the Baidu Recycle Green Service Alliance to further enhance the use of Baidu Recycle app and attract more stakeholders. The Alliance aspires in the collaboration with electronic manufacturers in order to build an internet-based nationwide e-waste management ecosystem[15].

CONCLUSION: Unsafe e-waste management is posing a threat to Sustainable Development. Policymakers need to assess the available opportunities in order to mitigate environmental threats deriving from improper e-waste disposal. Good policies will include proper recycling infrastructure, shifting e-waste collection from the informal sector to the formal, the creation of green jobs and a shift in people's behavior toward a green approach to e-waste disposal. Ensuring that these policies are in place and more Big Data initiatives in the form of Public- Private Partnerships are formed can support the region to achieve a sustainable development future which is inclusive for all and does not lessen the environmental and health standards.

Prepared by Erifyli Nomikou, Consultant, EDD/ESCAP

References:

[10] Solving the E-Waste Problem (Step) Initiative White Paper. "One Global Definition of E-waste". (2014). Accessible at: http://www.step-initiative.org/files/step/_documents/StEP_WP_One%20Global%20Definition%20of%20E-waste_20140603_amended.pdf

[11] Cheng, J. UNDP China working Paper. "Big Data for Development in China" (2014)
[12] UNDP. "Harnessing the Power of Big Data" (2014). Accessible at : http://www.cn.undp.org/content/china/en/home/presscenter/pressreleases/2014/08/harnessing-the-power-of-big-data.html

14 UNDP. "China: Turning E-Trash into Cash". (2015)

15 UNDP. "UNDP and Baidu Launched Green Alliance to Step up E-waste Recycling Service" (2015)

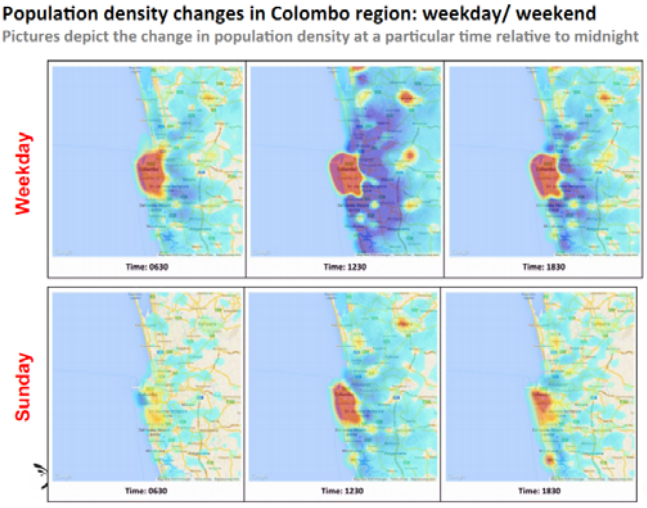
[13,14] UNDP. "China: Turning E-Trash into Cash". (2015). Accessible at: http://www.asia-pacific.undp.org/content/rbap/en/home/ourwork/development-impact/innovation/projects/china-ewaste.html
[15] UNDP. "UNDP and Baidu Launched Green Alliance to Step up E-waste Recycling Service" (2015). Accessible at : http://www.cn.undp.org/content/china/en/home/presscenter/pressreleases/2015/11/04/undp-and-baidu-launched-green-alliance-to-step-up-e-waste-recycling-service0.html

3. Case Study III

USING MOBILE META-DATA FOR URBAN AND TRANSPORTATION PLANNING IN PURPOSES IN SRI LANKA

Sri Lanka's major city Colombo is facing a tremendous population congestion challenge. LIRNEasia has gained access to historical and anonymized mobile data to better understand population movements in Colombo City and make informed and timely urban planning and urban transportation policy recommendations.

PROBLEM: "In 2010, the Asia-Pacific region's urban population was 754 million people, and it is expected that the urbanization rate in the region will reach 50 per cent in 2026[16]. Making cities and human settlements inclusive, safe, resilient and sustainable (SDG 11) is a cross-cutting issue across the integrated 2030 Agenda. Thirteen out of the total twenty-two Mega-cities are located in the region. Population density and road congestion are among the difficulties that urban populations are facing as mega-cities are expanding. Additionally, the impacts of poverty in cities are exacerbated by inadequate accommodation, slum dwellings, and unsanitary and unsafe living conditions. Existing infrastructure is often unable to accommodate the impacts of the growth rate of Asian cities, and growth patterns are leading to unstainable consumption and production patterns[17]. A focus on urban planning; urban transportation and urban infrastructure is the springboard of sustainable urbanization. In Colombo, Sri Lanka population congestion is a major challenge for public policy. 47% of the city's daytime population comes from outside the



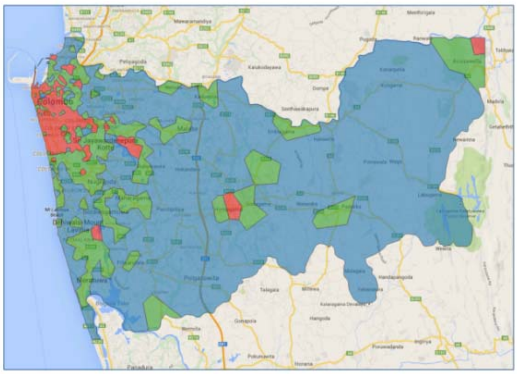
16 ESCAP. "Urbanization Trends in Asia Pacific" (2013) Accessible at : http://www.unescapsdd.org/files/documents/SPPS-Factsheet-urbanization-v5.pdf
17 ESCAP. "Urbanization Trends in Asia Pacific" (2013) Accessible at : http://www.unescapsdd.org/files/documents/SPPS-Factsheet-urbanization-v5.pdf

city[18]. The population density level, as observed in the heat map <figure 1>, reaches its peak during weekdays when people remain in the inner city for work or entertainment. The issues that arise from the population congestion pose threats to the livability of the city and by extension to many cities in the Asia- Pacific that face similar population congestion issues.

USING BIG DATA FOR URBAN AND TRANSPORTATION PLANNING: The use of emerging new technologies and more specifically, Big Data technologies are creating a new smart profile for cities and a new type of citizenship which promotes social activism and citizens engagement for more participatory governance in the urban system. Digital urbanism or the Internet of things is changing the urban landscape through the utilization of information and communication technologies (ICTs) to tackle urban challenges. In Sri Lanka, LIRNEasia[19], a pro-poor, pro-market think tank, partnered with multiple telecom operators to gain access to historical and anonymized telecom network big.

Those operators offered access to Call Detail records including Calls, SMS and Internet and Airtime Recharge Records. Through the use of SIM-movements data, new insights can be drawn regarding location and timeline of the population congestion, origin/home location and destination/work location and frequency and quantity of mobile interaction of users within the administrative boundaries of the city. Big data offers a cheaper and more effective alternative to traditionally costly census and household surveys to gather information which will be valuable for urban and transportation planning. At the same time the opportunity to leverage Big Data is huge due to the tremendous high coverage of the population by mobile phones in developing economies, gaining insight on mobility frequency and geography.

INSIGHTS & OUTCOMES: Due to the increased use of mobile phones and the wide coverage that the operators are offering it is easy to resort to mobile data information for understanding geographic locations, mobility patterns and the frequency of movements of populations. This information is extremely useful for policymakers, especially when it is timely, efficient and not costly. One of the insights in the Colombo City was that municipal boundaries are no longer valid. In terms of transportation policy, the focus should turn to the creation of high volume



[18] LIRNEasia, Mobile network big data for urban and transportation planning in Colombo, Sri Lanka" Data for Policy conference, Presenters: Samarajiva,R. Lokanathan, (2015). Accessible at: http://lirneasia.net/wp-content/uploads/2013/09/Samarajiva_Cambridge_June15.pdf

[19] LIRNEasia mission is to catalyze policy change through research to improve people's lives in the emerging Asia Pacific by facilitating their use of hard and soft infrastructures through the use of knowledge, information and technology (LIRNEasia).

transportation corridors for mass transit[20]. Furthermore, the Colombo District was mapped in three spatial clusters <Figure 2> From this almost real time monitoring of urban land use, it concludes that the central business district in Colombo has expanded[21].

CONCLUSION: Big Data can play a key role in achieving Sustainable Development through the valuable insights that large groups of data can generate especially for improving urban and transportation planning in cities. The research findings and recommendation of LIRNEasia, can provide insight in understanding changes in the urban population density and mobility. These findings can help urban planners and policy makers to create more sustainable cities and benefit from the cost savings associated with the new technologies instead of traditional less effective mechanisms to gather these data. The private sector offered access to historical and anonymized mobile data to LIRNEasia. This was an opportunity to leverage Big Data using private sector's Data Philanthropy for public policy insights. It is also an opportunity for mobile and other companies to draw insight concerning the population they are servicing for commercial and profit making uses.

Prepared by Erifyli Nomikou, Consultant, EDD/ESCAP

References:

[16,17] ESCAP. "Urbanization Trends in Asia Pacific" (2013) Accessible at : http://www.unescapsdd.org/files/documents/SPPS-Factsheet-urbanization-v5.pdf
[19] LIRNEasia. About us. Accessible at : http://lirneasia.net/about/
[18, 20, 21] LIRNEasia, Mobile network big data for urban and transportation planning in Colombo, Sri Lanka" Data for Policy conference, Presenters: Samarajiva,R. Lokanathan, (2015). Accessible at: http://lirneasia.net/wp-content/uploads/2013/09/Samarajiva_Cambridge_June15.pdf
[21] Samarajiva,R . "Using mobile-network big data for urban and transportation planning in Colombo" LIRNEasia (2015). Accessible at : http://www.iesl.lk/Resources/Documents/My%20Docs/Event%20PDF/PL%20L%201601 2015.pdf

[Figure 1, Figure 2} Samarajiva,R . "Using mobile-network big data for urban and transportation planning in Colombo" LIRNEasia (2015). Accessible at: http://www.iesl.lk/Resources/Documents/My%20Docs/Event%20PDF/PL%20L%201601 2015.pdf

[20] LIRNEasia, Mobile network big data for urban and transportation planning in Colombo, Sri Lanka" Data for Policy conference, Presenters: Samarajiva,R. Lokanathan, (2015). Accessible at: http://lirneasia.net/wp-content/uploads/2013/09/Samarajiva_Cambridge_June15.pdf

[21] LIRNEasia, Mobile network big data for urban and transportation planning in Colombo, Sri Lanka" Data for Policy conference, Presenters: Samarajiva,R. Lokanathan, (2015). Accessible at: http://lirneasia.net/wp-content/uploads/2013/09/Samarajiva_Cambridge_June15.pdf

4. Case Study IV

USING SOCIAL MEDIA TO TRACK WORKPLACE DISCRIMINATION AGAINST WOMEN IN INDONESIA

Gender-based discrimination is prevalent in the Asia-Pacific region. Women are presented with less employment opportunities, wage gaps and are most frequently victims of sexual harassment at work. The ILO in collaboration with Pulse Lab Jakarta used social media to explore whether online data can act as a source for drawing real time information for discrimination against women in the workplace.

PROBLEM: A cross-cutting issue which needs urgent action is achieving gender equality and empowering all women and girls (SDG 5). Over the last two decades, employment rates have increased for women in the region. So has the level of discrimination, not only based on gender and ethnic origin, but also due to sexual harassment. Women in Indonesia, experience limited access to employment opportunities and training, and unequal terms of employment – both in terms of wages, with a wage gap of 35 per cent[22], as well as in terms of professional responsibilities. Over the past decade, women participation rates in the labor force have been between 50-53 per cent while for men it is between 80-83 per cent. Historically, gender based workplace discriminations are very difficult to monitor as incidents usually remain unreported[23].

USING BIG DATA TO UNDERSTAND DISCRIMINATION IN THE WORKPLACE: Big Data provides an innovative way to gain useful insights on population behavior in real time. In Indonesia, social media data mining and more specifically, leveraging tweets, can be a good alternative to costly traditional ways of collecting data through lengthy surveys to gain new sources of information for workplace discrimination. In partnership with the government of Indonesia and the ILO, the UN Global Pulse Lab in Jakarta tested whether social media monitoring can provide signals for real- time workplace discrimination against women. They filtered tweets and extracted online conversations in the Bahasa Indonesia language from 2010 to 2013. Tweets falling in one of the 8 topics[24] were filtered and then analyzed for volume and content using a social data analytics platform called Crimson Hexagon to detect whether they provided sufficient volume to analyze further as potential signals of perceptions, opinions and incidents of discrimination[25.]

KEY PLAYERS: The Government of Indonesia, United Nations Global Pulse Lab Jakarta, International Labor Organization, Twitter.

---

22 UN Global Pulse, 'Feasibility Study: Identifying trends in Discrimination against Women in the Workplace in Social Media", Global Pulse Project Series no 11, 2014'.

23 UN Global Pulse, 'Feasibility Study: Identifying trends in Discrimination against Women in the Workplace in Social Media", Global Pulse Project Series no 11, 2014'.

24 The categories were the following: 1) Permission to work, 2) Appropriateness of work, 3) the burdens of working women 4) Discrimination in job requirements 5) Lack of skills or education 6) cost to access employment 7) Home-based workers and 8) sexual harassment in the workplace. UN Global Pulse, 'Feasibility Study: Identifying trends in Discrimination against Women in the Workplace in Social Media", Global Pulse Project Series no 11, 2014'.

25 UN Global Pulse, 'Feasibility Study: Identifying trends in Discrimination against Women in the Workplace in Social Media", Global Pulse Project Series no 11, 2014'.

INSIGHTS & OUTCOMES: From 2010 to 2013, social media inputs and online conversations were analyzed demonstrating that only four topics had sufficient tweets to look into. Those were: Permission to work (3,000 tweets); appropriateness of work (5,000 tweets); burdens of working women (21,000 tweets) and discrimination in job requirements (78,000 tweets)[26]. Given that the volume of relevant online conversations is increasing it was concluded that further research is needed and that existing monitoring mechanisms could be supplemented by digital tools to create a decent work environment. The private sector is an important player in offering large data sets for the analysis. More specifically, Twitter is extremely popular in Indonesia, and in particular in Jakarta, and people are very likely to share experiences using their twitter handles or other social media outlets. The understanding of the power of data sharing, tools and expertise from private sector is instrumental to the completion of data projects.

Harnessing digital data for social good relies heavily in the contribution of the private sector. There are two underlying motives behind data sharing for private companies. Firstly, Data Philanthropy which is the understanding of the importance of data for social good and complies with data- driven Corporate Social Responsibility and secondly, ensuring that developing countries population will not return to poverty levels which do not allow for viable consumption patterns. Last but not least, the pattern of privacy concerns is the norm in almost any use of large data sets of user's information. Ensuring that the identity of the user cannot be identified due to historical and location data and that the data sets remain anonymized will assist in leveraging Big Data to prevent any type of discrimination at workplace and remaining on path to achieve Sustainable Development.

CONCLUSIONS: To mark the development of the next 15 years, achieving gender equality and empowerment of all women and girls is the way forward for the region. Gender based discrimination results in female workers being demoted and dismissed. Discrimination at workplace further exacerbates alienation and violates human and labor rights. Drawing information from real time data can assist governments and the international community to understand further drivers of discrimination in the workplace and prevent incidents from occurring, ensuring a decent and equitable work environment.

Prepared by Erifyli Nomikou, Consultant, EDD/ESCAP

References:

[22,23,24,25,26] UN Global Pulse, 'Feasibility Study: Identifying trends in Discrimination against Women in the Workplace in Social Media", Global Pulse Project Series no 11, 2014'.

26 UN Global Pulse, 'Feasibility Study: Identifying trends in Discrimination against Women in the Workplace in Social Media", Global Pulse Project Series no 11, 2014'

[27] How the UN lab in Indonesia uses twitter Accessible at: http://www.fastcolabs.com/3007178/open-company/how-uns-new-data-lab-indonesia-uses-twitter-preempt-disaster
[28] ILO. Discrimination at Work in Asia. Accessible at: http://www.ilo.org/wcmsp5/groups/public/---ed_norm/---declaration/documents/publication/wcms_decl_fs_89_en.pdf

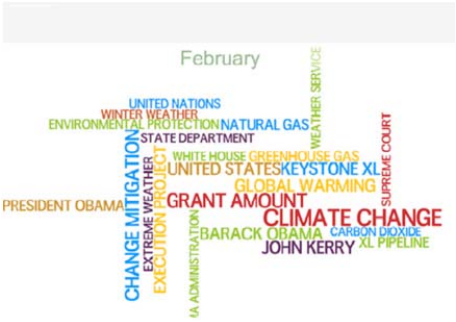
5. Case Study V

USING SOCIAL MEDIA TO MEASURE PUBLIC AWARENESS FOR CLIMATE CHANGE

The 2014 Climate Summit and the upcoming COP21 have offered a unique opportunity to explore and monitor real-time social media conversations about climate change. Since April 2014, UN Global Pulse has been measuring the total volume of tweets, links and hashtags about different climate change topics providing insight on public awareness and engagement.

PROBLEM: Asia-Pacific is one of regions most prone to climate change. The impacts of climate change are projected to intensify in the future and world leaders have pledged to combat climate change focusing on adaption and mitigation to achieve the 2030 Agenda for Sustainable Development (SDG 13). Climate change is not only a regional priority but for the region to be successful in adaptation and mitigation, leaders must mobilize people and enhance public interest around climate change issues. There are not enough data available through traditional data collection tools to provide us with insight on the public's awareness and engagement in tackling the climate change challenge[27].

USING BIG DATA TO ENACT POLICIES FOR CLIMATE CHANGE: The increasing number of mobile users in the developing world and the new technologies which offer an unprecedented opportunity for interconnectivity and civic engagement can be a valuable source of digital data for social good. Social media has revolutionized the way citizens respond to cross-cutting issues. Digital data is an innovative way to gain insight in citizen's behavior and promote participatory and inclusive policy making. Leveraging tweets allow citizens to undertake an active role through social media outlets, creating high levels of civic engagement and social activism among them. Citizen's inclusion in climate change policies through their online presence and active engagement can be a game changer to the formulation of respective regionals priorities moving forward the transition towards sustainable development in the post- 2020 climate regime.



[27] Picture: Tweets hashtags that trended in relation to climate change in February 2015, UN Global Pulse http://unglobalpulse.net/climate/google/

KEY PLAYERS: Policy Makers, Citizens, UN Global Pulse, Twitter

INSIGHTS & OUTCOMES: Leveraging tweets through the monitoring of the volume and content can inform decision and policy makers on what citizens are mostly concerned about and to develop communications to target priority regions. Global Pulse and the Secretary General's Climate Change Support Team created a tool to monitor real-time social media engagement prior to and after the Climate Summit in 2014. On a daily bases tweets in English, French and Spanish were monitored across different topics related to climate change. Measuring and visualizing tweets overtime created a baseline of engagement; increased engagement around Climate Summit[28]. Hashtags, links, and tweets were an innovative and unprecedented tool to measure public engagement, reflect public opinion and enact data-driven policy making for climate change. The methodology used was the development of a taxonomy of 1,000 words and phrases which filtered over 15 million tweets since April 2014 in English, French and Spanish. Out of eight topics the "economy" and "politics" showed the highest number of public conversations about climate change. The baseline volume remained at 140,000 English language tweets per day with that number increasing at over 400,000 on the day of events such as the Climate Summit or the People's Climate March[29]. Following the summit the baseline increased between 10 and 15[30] percent indicating that climate change was not a temporary engagement but people sustained their interest in climate change issues.

CONCLUSIONS: The rapid growing world economy and population once threatened to collide with the planet's finite resources and fragile ecosystems[31] . Today this threat is a global crisis. Climate change is a cross-cutting issue and action is needed immediately. The year 2015 is critical for setting the agenda of the next 15 years and it cannot be achieved without active citizen's engagement. The ability to monitor real-time conversations in social media and draw insights can be a driver to measure and increase public awareness and help climate policy makers to make informed decisions relevant to the climate change policy priorities identified by the people in each region. People can be game changers is building climate change solutions for adaption and mitigation.

Prepared by Erifyli Nomikou, Consultant, EDD/ESCAP

References:

[27] UN Global Pulse, Picture: Tweets hashtags that trended in relation to climate change in February 2015. Accessible at: http://unglobalpulse.net/climate/google/
[28,29,30] UN Global Pulse, "Using Twitter to Measure Global Engagement on Climate Change", Global Pulse Project Series no 7, 2015.
[31] Sachs, J. " The Age of Sustainable Development". Columbia Press. New York (2015)

28 UN Global Pulse, "Using Twitter to Measure Global Engagement on Climate Change", Global Pulse Project Series no 7, 2015.

29 UN Global Pulse, "Using Twitter to Measure Global Engagement on Climate Change", Global Pulse Project Series no 7, 2015.

30 UN Global Pulse, "Using Twitter to Measure Global Engagement on Climate Change", Global Pulse Project Series no 7, 2015.

31 Sachs, J. "The Age of Sustainable Development". Columbia Press. New York (2015)